Interactome-based framework to translate disease genetic data into biological and clinical insights

Kalyani Bindu Karunakaran

Thesis submitted for the Degree of Doctor of Philosophy

School of Chemistry, Food and Pharmacy
Department of Pharmacology

Declaration of original authorship

I certify that the research presented in this thesis is my own work. All resources and contributions have been properly and fully acknowledged.

Kalyani Bindu Karunakaran

The interactome-based framework proposed in this thesis has been discussed in the research publications and conference presentations* listed below:

*included in the thesis

Peer-reviewed publications

1. ***Karunakaran, Kalyani B.**, Satoko Amemori, N. Balakrishnan, Madhavi K. Ganapathiraju, and Ken-ichi Amemori. Generalized and social anxiety disorder interactomes show distinctive overlaps with striosome and matrix interactomes. Scientific reports 11, no. 1 (2021): 1-25.

2. ***Karunakaran, Kalyani B.**, Srilakshmi Chaparala, Cecilia W. Lo, and Madhavi K. Ganapathiraju. Cilia interactome with predicted protein–protein interactions reveals connections to Alzheimer's disease, aging and other neuropsychiatric processes. Scientific reports 10, no. 1 (2020): 1-16.

3. ***Karunakaran, Kalyani B.**, George C. Gabriel, Narayanaswamy Balakrishnan, Cecilia W. Lo, and Madhavi K. Ganapathiraju. Novel protein–protein interactions highlighting the crosstalk between hypoplastic left heart syndrome, ciliopathies and neurodevelopmental delays. Genes 13, no. 4 (2022): 627.

4. ***Karunakaran, Kalyani B.**, Naveena Yanamala, Gregory Boyce, Michael J. Becich, and Madhavi K. Ganapathiraju. Malignant pleural mesothelioma interactome with 364 novel protein-protein interactions. Cancers 13, no. 7 (2021): 1660.

5. ***Karunakaran, Kalyani B.**, N. Balakrishnan, and Madhavi K. Ganapathiraju. Interactome of SARS-CoV-2 modulated host proteins with computationally predicted PPIs: Insights from illustrative translational systems biology studies. Frontiers in Systems Biology (2022): 2.

6. ***Karunakaran, Kalyani B.**, Srilakshmi Chaparala, and Madhavi K. Ganapathiraju. Potentially repurposable drugs for schizophrenia identified from its interactome. Scientific reports 9, no. 1 (2019): 1-14.

7. ***Karunakaran, Kalyani B.**, Anand Thiyagaraj, and Kirankumar Santhakumar. Novel insights on acetylcholinesterase inhibition by Convolvulus pluricaulis, scopolamine and their combination in zebrafish. Natural Products and Bioprospecting 12, no. 1 (2022): 1-15.

8. Ganapathiraju, Madhavi K., **Karunakaran, Kalyani B.**, and Josefina Correa-Menéndez. Predicted protein interactions of IFITMs may shed light on mechanisms of Zika virus-induced microcephaly and host invasion. F1000Research 5 (2016).

9. Xu X, Jin K, Bais AS, Zhu W, Yagi H, Feinstein TN, Nguyen PK, Criscione JD, Liu X,

Beutner G, **Karunakaran KB**, Rao KS, He H, Adams P, Kuo CK, Kostka D, Pryhuber GS, Shiva S, Ganapathiraju MK, Porter GA Jr, Lin JI, Aronow B, Lo CW. Uncompensated mitochondrial oxidative stress underlies heart failure in an iPSC-derived model of congenital heart disease. Cell Stem Cell. (2022) May 5;29(5):840-855.e7.

10. Faubel, Regina J., Veronica S. Santos Canellas, Jenna Gaesser, Nancy H. Beluk, Tim N. Feinstein, Yong Wang, Maya Yankova, **Karunakaran, Kalyani B.**, King, Stephen M, Ganapathiraju. Madhavi K. and Lo Cecilia W. Flow blockage disrupts cilia-driven fluid transport in the epileptic brain. Acta Neuropathologica (2022): 1-16.

Manuscripts accepted for publication

1. *****Karunakaran, Kalyani B.**, Madhavi Ganapathiraju, Sanjeev Jain, Samir Brahmachari, and Narayanaswamy Balakrishnan. Drug contraindications in comorbid diseases: A protein interactome perspective. Network Modeling Analysis in Health Informatics and Bioinformatics (2023).

Manuscripts under revision

1. *****Karunakaran, Kalyani B.**, and Madhavi K. Ganapathiraju. Malignant peritoneal mesothelioma interactome with 417 novel protein-protein interactions. Research Square (2021).
Current status: Under revision at BJC Reports

Preprints

1. *****Karunakaran, Kalyani B.**, N. Balakrishnan, and Madhavi K. Ganapathiraju. GPX4-associated Sedaghatian type spondylometaphyseal dysplasia: A protein interactome perspective. bioRxiv (2022).

Conference presentations including published abstracts

1. **Karunakaran, Kalyani B.**, and Madhavi K. Ganapathiraju. Alzheimer's disease, aging and the sonic hedgehog pathway connections highlighted through protein interactome. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 15.7 (2019): P1012-P1013.
Presented at: Alzheimer's Association International Conference (AAIC) 2019, Los Angeles, USA

2. **Karunakaran, Kalyani B.**, and Ganapathiraju, Madhavi K. Malignant pleural mesothelioma interactome and web-platform for biological hypotheses on novel protein- protein interactions. Science 2017, University of Pittsburgh, US

3. **Karunakaran, Kalyani B.**, and Ganapathiraju, Madhavi K. Malignant pleural mesothelioma

interactome. International Mesothelioma Interest Group 2018, Ottawa, Canada

4. **Karunakaran, Kalyani B.**, and Ganapathiraju, Madhavi K. Malignant pleural mesothelioma interactome. UPMC Hillman Cancer Centre Scientific Retreat 2018

5. **Karunakaran, Kalyani B.**, and Ganapathiraju, Madhavi K. Malignant pleural mesothelioma interactome. UPMC Hillman Cancer Centre Scientific Retreat 2019

6. **Karunakaran, Kalyani B.**, Thiyagaraj, Anand, Santhakumar, Kirankumar. Analysis of acetylcholinesterase inhibitory activity of selected plant extracts in zebrafish. Alzheimer's Association International Conference (AAIC) Satellite Symposium 2018, Bengaluru, India

Abstract

The study of human diseases has evolved from exploring non-causal pathophenotypes to discovering aetiological genetic factors through genome-wide association (GWA) studies. While GWA studies identify disease-associated genetic variants, they fail to reveal their functional implications. The emergence of network biology and molecular interaction mapping helped conceptualise disease as a breakdown of the protein-protein interaction (PPI) network or the 'interactome' due to these genetic variants. The interactome drives cellular processes and responds to genetic and environmental changes by leveraging its inherent interdependencies. Simultaneously, the focus has shifted from single-target drugs to the polypharmacological effects of drugs within the PPI network, inducing therapeutic and non-therapeutic effects. Despite these conceptual advances, several factors hinder the widespread application of interactome analysis in disease genetics and drug discovery. These include the lack of an integrated conceptual framework to derive biological and clinical insights from genetic data, the lack of context-sensitive interactomes, the failure to integrate computationally predicted PPIs to circumvent interactome sparsity, and the absence of methods to study correlations across multiple disease interactomes and drug target networks. This thesis proposes a two-pronged interactome-based framework to address these limitations. The first arm focuses on constructing the disease interactomes of complex and Mendelian disorders using both experimentally validated and computationally predicted PPIs, refining them using multi-omics datasets, and deriving insights into disease mechanisms using functional enrichment analyses, identifying repurposable drugs targeting the interactome using comparative analysis of drug-induced and disease-associated transcriptomes, and studying their activity in animal models. The second arm employs multivariate data analysis to explore relationships of multiple interactomes, revealing biological and clinical themes in cross-disorder relationships. This framework has demonstrated its potential by providing insights into eight disorders, identifying disease subgroups, and refining disease categorization based on genetic structures. The methodology yields clinically actionable results, including repurposable drugs and insights into drug activity that can inform safety and efficacy evaluations in clinical trials. This thesis proposes a comprehensive interactome-based framework to uncover hidden patterns in emerging multi-omics disease data and enhance our understanding of disease biology and therapeutics.

Acknowledgements

Table of Contents

Abbreviations

APEX	Ascorbate peroxidase
AP-MS	Affinity-purification mass spectrometry
ASD	Autism spectrum disorder
BioGRID	Biological general repository for interaction datasets
BioID	Proximity-dependent biotin identification
CDD	CDKL5 deficiency disorder
CMAP	Connectivity map
Co-IP	Co-immunoprecipitation
COVID-19	Coronavirus disease
DES	Differential expression signature
DTN	Drug target network
DULIP	Dual luminescence-based co-immunoprecipitation
FDA	Food and drug administration
GAD	Generalised anxiety disorder
GEO	Gene expression omnibus
GWA	Genome-wide association
HBB	Haemoglobin subunit beta
HIPPIE	Human integrated protein--protein interaction reference
HiPPIP	High-precision protein-protein interaction prediction
HLHS	Hypoplastic left heart syndrome
HPRD	Human protein reference database
IP	Immunoprecipitation
LUMIER	Luminescence-based mammalian interactome
MPeM	Malignant peritoneal mesothelioma
MPM	Malignant pleural mesothelioma
OCD	Obsessive-compulsive disorder
PCA	Principal component analysis
PD	Panic disorder
PPI	Protein-protein interactions
PTSD	Post-traumatic stress disorder
SAD	Social anxiety disorder
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SCZ	Schizophrenia
SMDS	Spondylometaphyseal dysplasia, Sedaghatian type
SNP	Single nucleotide polymorphism
STRING	Search tool for the retrieval of interacting genes/proteins
SVD	Singular value decomposition
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

Summary

The conceptualisation of human disease and its underlying aetiological factors have evolved with advancing technology from the late 1880s to the 2000s. Initially, the points of inquiry into aetiological mechanisms were observable non-causal and intermediate cell and tissue-level pathophenotypes, but it later shifted towards unobservable preclinical and causal genetic factors. In the 1900s, as gene discovery techniques improved, the theories regarding single gene perturbations underlying human diseases transformed into a polygenic perspective. This new perspective conceptualised diseases as the cumulative result of multiple gene perturbations influenced by developmental and environmental factors. The genome-wide association (GWA) framework, which emerged in the early 2000s, played the most significant role in championing this concept.

The GWA framework helped identify disease-associated genetic variants. However, it failed to explain the functional implications of these genetic variants on disease causation and accurately model the heritability of human diseases as the additive effect of these variants. Network biology concepts and molecular interaction mapping techniques that emerged simultaneously with the GWA framework helped address these limitations. Network biology concepts proposed that higher-level interactions among the genes carrying the variants could explain the missing component of disease heritability. Moreover, it could provide a mechanistic framework for understanding disease causation by characterising disease as a network phenomenon resulting from perturbations in the interactions of the proteins encoded by disease-associated genes – i.e., protein-protein interactions (PPIs) – that regulate cellular functions.

Conceptual advances in drug discovery progressed parallel to advancements in disease aetiology investigations. It transitioned from the magic bullet era of the early 1900s – where treating disease symptoms involved discovering one drug interacting with a single target – to the systems therapeutics era. Throughout much of the 20th century, the prevailing magic bullet hypothesis resulted in high drug failure rates during clinical trials and led to adverse events in marketed drugs due to incomplete assessments of drug activity. The newer systems therapeutics approach integrated a more biologically realistic understanding of drugs as having multiple effects (the polypharmacological model) with network biology to study their multimodal impacts on disease alleviation, i.e., drug-induced perturbations of the PPI network elicited either beneficial therapeutic responses or adverse events.

Network biology has significantly advanced our understanding of disease mechanisms and drug discovery, leading to numerous studies exploring the organisational principles of proteins encoded by disease-associated genes and drug targets in the interactome (the complex network of interacting proteins). However, five factors continue to restrict the scope and widespread application of network biology principles to examine disease genes and drug targets in their network neighbourhoods. These include the lack of an integrated conceptual framework incorporating network biology principles to analyse disease interactomes and drug target networks (DTNs), their inadequate exploration leading to a scarcity of biologically insightful and clinically actionable findings from disease-associated genetic and

pharmacological data, underutilisation of computational PPI predictions to complement the sparse discovery of the human interactome through experimental methods, limited integration of disease-associated multi-omics data at organ, tissue, and cell levels resulting in context-insensitive interactomes, and the lack of exploration of the higher-level relationships between disease interactomes and DTNs, which could potentially elucidate complex patterns of cross-disorder relationships and drug interactions.

This thesis proposes a two-pronged interactome-based framework to translate disease-associated genetic data into biological and clinical insights and address these limitations. The first arm of the framework introduces a methodology for analysing single disease interactomes. This arm involves the compilation of disease-associated genetic data from various sources and disease interactome construction using both experimentally verified and computationally predicted PPIs, disease interactome validation and refinement using disease-associated multi-omics datasets, and interactome characterisation using functional and phenotypic enrichment and sub-network analyses, leading to the generation of testable biological hypotheses. The steps to generate clinical insights include the identification of drug targets within the disease interactome and comparative analysis of disease-associated and drug-induced transcriptomes to uncover repurposable drugs. Promising candidate drugs are selected and tested in animal models, leading to a prioritised list of drugs for clinical testing. This methodology produced valuable insights into complex disorders with polygenic architectures, namely, Alzheimer's disease (**Chapter 2** and **Appendix section 14.2**), schizophrenia (**Chapter 8**), hypoplastic left heart syndrome (**Chapter 3**), malignant pleural mesothelioma (**Chapter 4**), malignant peritoneal mesothelioma (**Chapter 5**), as well as COVID-19 (**Chapter 6**). Additionally, the framework generated valuable results in Mendelian disorders believed to have monogenic architectures despite their complex pathophenotypes. The thesis proposed methods to uncover the embedding of these disorders in polygenic genetic landscapes. Specifically, this approach provided insights into Sedaghatian type spondylometaphyseal dysplasia, a bone dysplasia, in **Chapter 7**.

The second arm of the framework presents a methodology for analysing the interrelationships between multiple disease interactomes and DTNs. For disease interactomes, this involves the identification of context-specific disease interactomes and comparative disease interactome analysis, leading to the discovery of biological themes underlying multiple disorders, as demonstrated for five anxiety disorder subtypes in **Chapter 9**. For DTNs, the methodology involves the identification of context-specific DTNs and comparative analysis of DTNs, leading to the elucidation of biological themes underlying clinical activity in multiple disorders, as demonstrated in the study of drug contraindications in six comorbid disease pairs in **Chapter 10**.

Overall, the thesis introduces an analytical pipeline to investigate disease interactomes and DTNs, addressing the limitations of current network biology studies and providing valuable insights into disease mechanisms and interrelationships. It covers heart diseases, cancers, viral infections, skeletal disorders, and neurological and psychiatric disorders, showcasing potential applicability to multiple disorders. The methodology extracts disease-disease relationships, uncovers subgroups and common biological themes among disorders, and refines categorisation of diseases based on their genetic structures. It yields clinically

actionable results – such as repurposable drugs and insights into comorbidity effects on drug action – that can augment drug safety and efficacy evaluation in the clinical developmental pipeline. The interatomic framework integrates diverse disease-associated genetic data and allows the extraction of context-specific sub-networks from disease interactomes and DTNs. In summary, the comprehensive approach proposed in this thesis can help uncover hidden patterns in the genetic architecture of diseases and advance understanding of disease biology and drug action.

1. Introduction

Understanding the aetiological factors underlying human disease has been a centuries-old quest, punctuated by three eras, the Oslerian, classical genetics, and post-genomics. Various conceptual and technological advances in human genetics and drug development shaped these eras and significantly altered the trajectory of disease research. Specifically, over the past 140 years, there has been a gradual shift in the conceptualisation of disease. A reductionist monogenic conceptualisation prevailed from the late 1880s to the late 1990s. In early 2000s, a holistic polygenic conception of disease arose, which recently evolved into characterising disease as an emergent property of a network of genes. The network medicine paradigm conceptualises disease as a phenomenon arising from the breakdown of multiple functional units in the cell constituted by interacting genes, which, in turn, perturbs the physiological state produced by the functional units. The field of network biology aims to systematically identify the multiple choke points and paths through which disorders emerge in this cellular landscape of interacting genes. These inquiries will inform studies on the systems-scale effects of drugs and help devise rational therapeutic strategies. To chart the emergence of the network medicine paradigm – the central theme of this thesis – it is critical to understand the evolution of the relationship between disease determinants and disease manifestation during the three eras and the incremental capacity of the concepts and technologies to capture the complexity of the disease phenotype.

1.1 A historical perspective on the conceptualisation of human disease and the underlying aetiological factors: from Oslerian tradition to network biology

Oslerian, classical genetics, and post-genomic eras transformed our understanding of human disease and its aetiology. Sections 1.1.1 to 1.1.4 describe the chronological sequence of events in genetics, drug development and network biology that shaped these three major eras (**Fig. 1**). The shortcomings of the conceptual and technological advances and the evolution of the disease determinant – phenotype relationships (**Fig. 2**) in each of these eras led to the development of the network medicine paradigm, described in sections 1.1.5 and 1.1.6.

1.1.1 The age of Oslerian formalism: biological reductionism and its effects on the understanding of disease processes and therapeutics

Oslerian formalism – named after William Osler, the father of modern medicine – has dominated the study of human disease and therapeutics since the late 19th century (Osler and McCrae, 1892, Loscalzo and Barabasi, 2011). In this traditional paradigm, clinicians correlated the clinical presentation of the disease with the gross anatomic pathology and the histopathology of the organ system in which the symptoms manifested. Disease diagnosis relied on deducing syndromic patterns based on structured clinical observation and
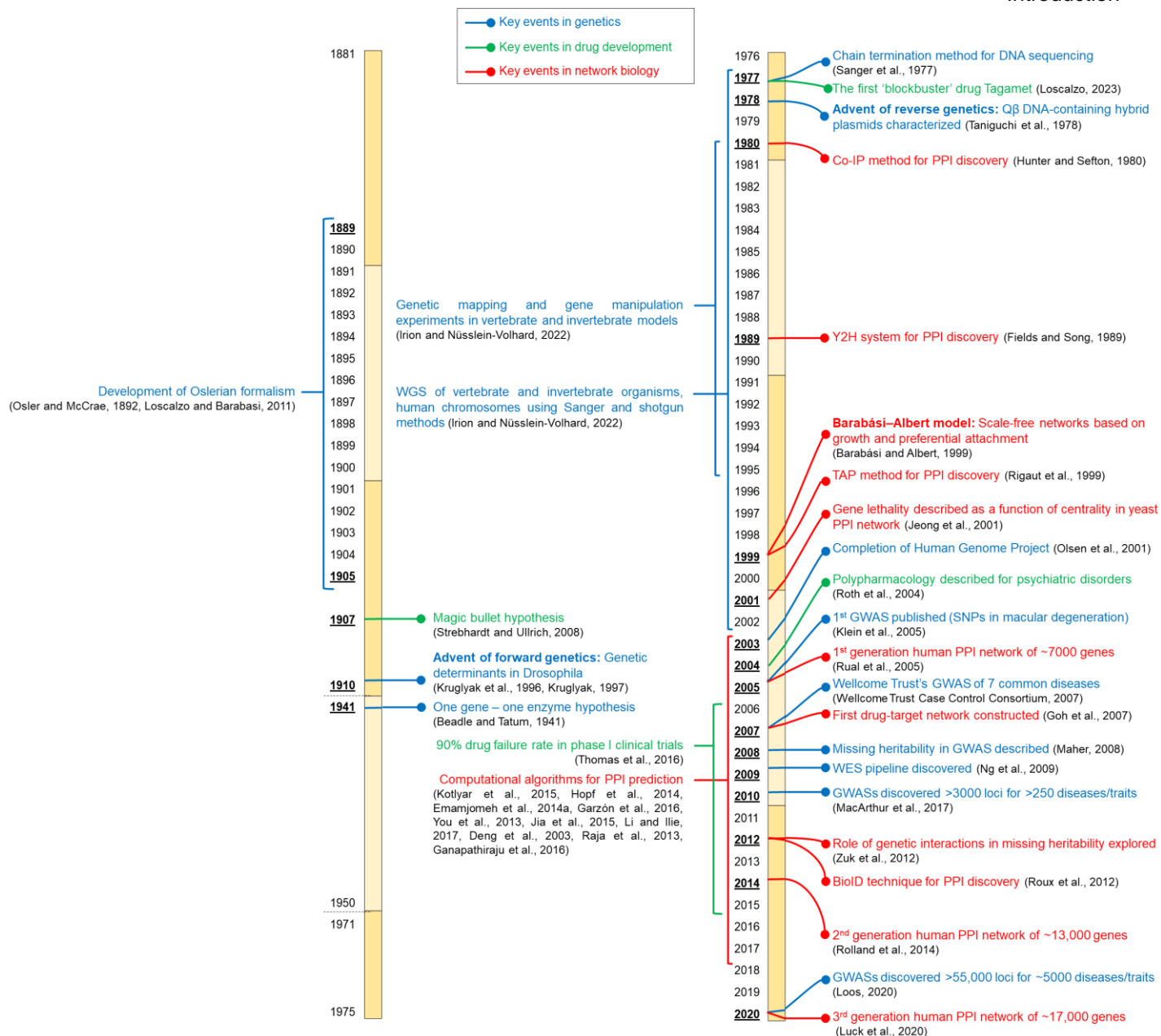
Figure 1: A timeline of the key events that influenced the conceptualisation of human disease and therapeutics. The events in the fields of genetics, drug development and network biology have been marked using different colours, as shown in the legend. Abbreviations: BioID: Proximity-dependent biotin identification; Co-IP: Co-immunoprecipitation; GWAS: Genome-wide association study; PPI: Protein-protein interactions; SNP: Single nucleotide polymorphism; TAP: Tandem affinity purification; WES: Whole-exome sequencing; WGS: Whole-genome sequencing; Y2H: Yeast two-hybrid system. The figure was created in Microsoft PowerPoint.

formulating treatment regimens focused singularly on reversing these presumed pathophenotypes. The Oslerian paradigm conceptualised a phenomenon as complex as human disease in terms of a mechanistic clinicopathological relationship. 'Higher-level' observed clinical behaviours were thought to be precipitated by specific abnormalities in 'lower-level' tissues and cells (**Fig. 2**). The reductionist bias in scientific thinking that pervaded various disciplines at the time – a tendency to explain the attributes of a complex system as the sum of its constituent parts – magnified the influence of this paradigm (Regenmortel, 2004, Loscalzo and Barabasi, 2011). This powerful model influenced both the structural organization of clinical medicine as a discipline and the guiding philosophy underlying investigations of disease mechanisms at the molecular level. It helped formalise and streamline the services of the healthcare system and led to the discovery and refinement of cutting-edge medical
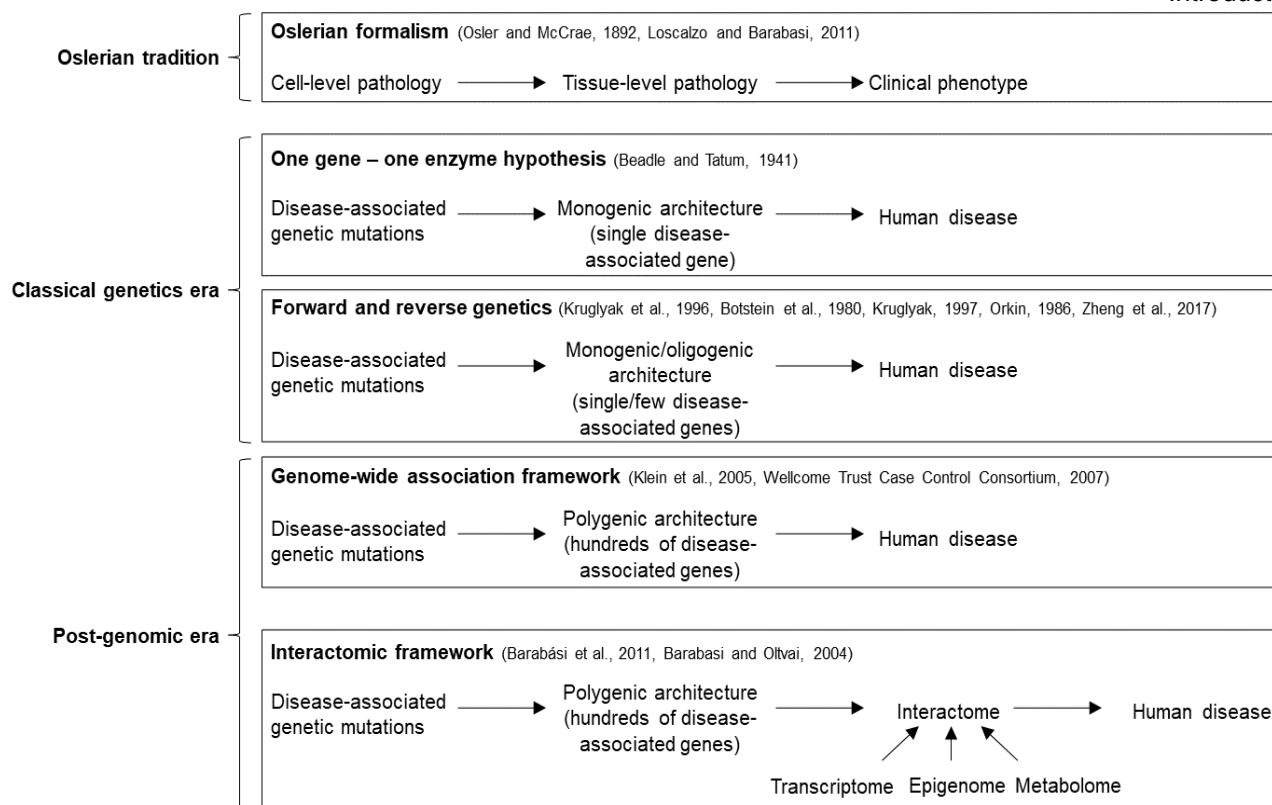
Figure 2: The evolution of the relationship between disease determinants and disease phenotypes. In the Oslerian era, cell- and tissue-level pathophenotypes (often end processes of the actual etiological factors) were believed to produce clinical presentations of the disease. In the classical genetics era, the focus shifted to a single gene or a group of genes that produced the disease phenotype. In the post-genomic era, GWA studies revealed the multifactorial and polygenic architectures of human diseases. The network medicine paradigm showed that this polygenic architecture was embedded within the human interactome, which when integrated with biological evidence from other levels (transcriptome, epigenome and metabolome), resulted in disease phenotypes. The figure was created in Microsoft PowerPoint.

diagnostic procedures and tools (Geyman, 1983). However, the approach had several limitations.

Formal medical education and healthcare became categorised into clinical silos specializing in ailments affecting roughly ten organ systems (Geyman, 1983). Scientific inquiries into disease causation focused on dissecting the pathological processes underlying broad sets of intermediate phenotypes, such as fibrosis, thrombosis, inflammation, and necrosis. These phenotypes possibly reflected later-stage manifestations unrelated to the actual etiological factors (Loscalzo and Barabasi, 2011, Osler and McCrae, 1892). The therapeutic targeting of the processes underlying these stages merely led to symptom alleviation (Loscalzo and Barabasi, 2011). When shared between etiologically unrelated disorders, these pathophenotypes impeded differential diagnosis (Pearn, 2011). Treating observable intermediate phenotypes as points of disease inquiry shifted the focus away from critical preclinical, albeit unobservable, states possibly regulated by genetic predisposition and environmental factors (Loscalzo and Barabasi, 2011). The interaction of genetic and environmental factors leads to variable disease susceptibility, onset ages, phenotype expressivity and penetrance, disease progression patterns, and prognosis (Loscalzo and Barabasi, 2011). These phenomena naturally lead to patient stratification, and the resultant patient groups remained routinely unrecognised by the Oslerian healthcare system and did not receive personalised diagnosis and treatment (Loscalzo and Barabasi, 2011). Altogether,

the scope of Oslerian formalism was too limited to encapsulate the complexity and diversity of human disease processes and suggest disease-specific targeted therapies. It did not allow the identification of critical factors, such as unique etiological and stochastic environmental factors, which influence the susceptibility and preclinical stages of disease development (Loscalzo and Barabasi, 2011).

1.1.2 The classical genetics era: the monogenic view of diseases and the advent of forward and reverse genetic approaches for interrogating genetic mechanisms

From the 1940s, etiological genetic factors began to be identified with the emergence of modern genetic theory and technologies. However, the notion propounded by Beadle and Tatum in 1941 that a simple correlational relationship exists between the genotype and the phenotype (Beadle and Tatum, 1941) continued the reductionist trend and characterised disorders as having a monogenic aetiology (**Fig. 2**). Specifically, a perturbation in a single gene was thought to produce disease phenotypes. It was only in the 1980-1990s that identifying the causative gene was made possible through forward and reverse genetic approaches (**Fig. 2**). Linkage mapping and positional cloning were part of the forward genetics approach in which the knowledge of a specific disease phenotype inherited within a family was sufficient to elucidate the disease-associated gene (Kruglyak et al., 1996). In parametric linkage studies, the linkage intervals within the human genome were tagged using restriction fragment length polymorphic markers or single sequence repeats (Kruglyak et al., 1996, Botstein et al., 1980, Kruglyak, 1997). These markers helped identify the chromosomal regions co-segregating with the inheritance pattern of the disease phenotypes (Kruglyak et al., 1996, Botstein et al., 1980, Kruglyak, 1997). Positional cloning and confirmatory mutational studies in cell lines and animal models further helped identify the causative genes (Kruglyak et al., 1996, Botstein et al., 1980, Kruglyak, 1997). Huntington's disease was associated with dynamic mutations in *HTT* gene located on chromosome 4 using forward genetics (Gusella et al., 1983, MacDonald et al., 1993). Genes were selected based on prior knowledge in the reverse genetic approach (Orkin, 1986, Zheng et al., 2017), e.g., cellular function, differential expression in disease conditions, or the suspected involvement of genes from the same gene family in the same disorder. Gene silencing and targeted mutagenesis were used to abrogate gene function and functional studies to characterise the phenotypic effects (Orkin, 1986). For example, animal models carrying mutations in the *FBN* gene – involved in extracellular microfibril formation – recapitulated the elastic fibre abnormalities seen in patients with Marfan's syndrome (Magenis et al., 1991, Hayward and Brock, 1997, Dietz et al., 1991).

The classical genetic approaches helped discover ~1200 disease-associated genes underlying monogenic diseases and traits by the 2000s (Botstein and Risch, 2003). The discovery was propelled by the development of gene sequencing methods such as Sanger sequencing (Sanger et al., 1977) and the completion of the draft human genome sequence (Olsen et al., 2001), both of which allowed finer mapping of the disease-causing alleles in genetic loci linked to disease. Meanwhile, burgeoning evidence hinted at a complex interplay between an organism's genotype and phenotype. This interplay was evident even in studies before the classical molecular genetics era. Genetically identical makeup in identical environments produced variable phenotypes. For example, phenotypic variability was seen in the pod sizes of isogenic bean plants bred in controlled environments (Vidal et al., 2011).

Additionally, selective transcriptomic and proteomic expression was seen in proximally grown and genetically identical bacterial and yeast cells (Vidal et al., 2011). Lastly, remarkable differences were noted between identical twins with comparable life histories and exposed to the same environment (Vidal et al., 2011).

1.1.3 The post-genomic era: the evolution towards a polygenic view of disease with the advent of the genome-wide association approach

Evidence from forward and reverse approaches and evolutionary models of disease-associated genes suggested that a polygenic architecture underlay the phenotypic heterogeneity seen not only in complex disorders, but also in Mendelian disorders. Mendelian disorders such as sickle cell disease with well-characterised monogenic causes (single point mutation in the gene coding for haemoglobin subunit beta), presented multiple pathophenotypes, outside of its primary manifestation (i.e., polymerization of haemoglobin driven by its altered oxygen affinity), namely, haemolytic events, stroke, and acute chest syndrome (Vidal et al., 2011). The polygenic architecture was likely made of a large number of genes with no prior suspected link to the disease and harbouring rare and common alleles (Altshuler et al., 2008). In the case of Mendelian subtypes of common diseases, the genes seemed to harbour numerous rare alleles that accounted for only a few disease cases in the population and in the more prevalent subtypes of these diseases, relatives of affected individuals showed lower disease risk and no definitive underlying causal gene (Altshuler et al., 2008). Hence, multiple genes seemed to modulate disease phenotypes in complex ways. Purifying selection could have limited the segregation of rare alleles that produce deleterious phenotypes (Altshuler et al., 2008). On the other hand, common alleles could have persisted at high frequency due to multiple reasons. First, the later onset of some associated phenotypes and their lower impact on reproductive fitness (Altshuler et al., 2008). Second, the recent expansion of *Homo sapiens* enabling mildly deleterious alleles to achieve moderate frequency (Altshuler et al., 2008). Third, modern living conditions changing the effect of a previously advantageous or neutral mutation (Altshuler et al., 2008). Fourth, an allele associated with disease risk also conferring a beneficial phenotype (Altshuler et al., 2008).

Breakthroughs in the identification of disease-associated gene variants came in the form of new statistical frameworks powered by population genomics, namely, genome-wide association (GWA) studies (Klein et al., 2005, Wellcome Trust Case Control Consortium, 2007) and high-throughput sequencing technologies, including whole-genome sequencing (WGS) (Sanger et al., 1977) and whole-exome sequencing (WES) (Ng et al., 2009). In GWA studies, the genotypes of individuals with a specific disease in millions of genomic locations containing single nucleotide polymorphisms (or SNPs compiled using DNA microarrays) were compared with the genotypes of individuals without the disease (Klein et al., 2005, Wellcome Trust Case Control Consortium, 2007). Disease-associated allelic variations – likely including causal and risk-conferring SNP loci – occur more frequently in the disease cohort than in the control cohort. WES and WGS allowed finer SNP mapping than the precompiled DNA microarrays. In the WES protocol, the protein-coding regions in the genome, i.e., exomes, were sequenced to enable the targeted capture of highly penetrant rare variants (Ng et al., 2009). In the WGS protocol, all genomic regions, including coding, non-coding and mitochondrial regions, are sequenced, increasing the likelihood of identifying disease-

associated rare variants and common variants not in linkage disequilibrium with the SNPs catalogued in DNA microarrays (Sanger et al., 1977).

GWA studies unravelled the polygenic tapestry of complex disorders with common genetic variants of reduced effect size and highly penetrant rare variants contributing additively to genetic liability (**Fig. 2**). They provided critical insights into the multifactorial aetiology of several disorders, most remarkably demonstrated by gene discovery in neuropsychiatric disorders (Demontis et al., 2019, Grove et al., 2019, Stahl et al., 2019, Wray et al., 2018, Pardiñas et al., 2018, Mullins et al., 2021, Ripke et al., 2014, Levey et al., 2020, Consortium, 2013, Lee et al., 2019) and comorbid disorders affecting other organ systems, e.g., congenital heart disease (Zaidi and Brueckner, 2017, Egbe et al., 2014, Marino et al., 2012, Homsy et al., 2015, Jin et al., 2017, Willsey et al., 2018).

1.1.4 Shortcomings of the genome-wide association framework: roadblocks on the path to elucidating the biological themes underlying the polygenic architecture of diseases

GWA studies helped identify a large number of disease-associated variants. However, they had several limitations. Firstly, the statistical framework of GWA studies did not help elucidate the functional implications and pathological impact of the genetic variants. GWA studies only tagged disease-associated SNPs in specific genomic regions. They did not directly implicate causal alleles in disease aetiology (Breen et al., 2016, Gallagher and Chen-Plotkin, 2018) or provide a framework to deduce the biological context of the variants, e.g., by integrating expression, protein, and methylation data. Attempts were made to characterise functional variants localising to regulatory regions and believed to impact disease pathogenesis. These were largely unsuccessful due to the moderate influence of these variants on the disease phenotype and their complex spatiotemporal dynamics, which were intractable by the molecular technologies at the time (Gallagher and Chen-Plotkin, 2018). Secondly, the contribution of the common variants to disease heritability – the phenotypic variance in a population explained by additive genetic factors – was small (Manolio et al., 2009). It was possible that many more common and rare variants remained to be captured by microarrays and identified as disease-associated in GWA frameworks (Manolio et al., 2009). However, a competing possibility was that the current estimates of 'missing heritability' were inaccurate, since they considered only the simple additive genetic contributions of the variants and excluded potentially higher-level interactions between the genes (Maher, 2008), reflecting the functional architecture produced by the genome. This plausible hypothesis was arguably the first milestone in the road that led to the conceptualisation of disease as a network phenomenon resulting from a breakdown of a web of molecular interactions underlying cellular functions. Unfortunately, molecular interaction mapping techniques remained insufficiently developed at the time. Moreover, sequencing efforts to identify and statistically prioritise genetic variants outpaced the efforts to devise integrative mechanistic frameworks to elucidate the functional impact of genetic variants (Woodsmith and Stelzl, 2017).

1.1.5 Emergence of network biology: re-conceptualisation of disease as a state resulting from perturbed molecular interdependencies within the polygenic architecture of diseases

In 1999, the Barabási–Albert model proposed that real-world networks resulted from a steady process of growth in the number of its components or 'nodes' (N) and preferential attachment based on the number of pre-existing connections or 'edges' (known as degree and denoted by k) that each node has with others (Barabási and Albert, 1999). This process gave rise to networks in which the number of nodes of a given degree, i.e., the degree distribution, followed a power law. The probability that a particular node had k links was given by $P(k) \sim k^{-\gamma}$, where γ is the degree exponent, and $2 < \gamma < 3$. This degree distribution produced a remarkable network structure, in which a large number of small degree nodes had very few edges and a small number of high degree nodes known as 'hubs' had many edges (Barabási and Albert, 1999). In contrast, most of the nodes in the Erdös–Rényi model of a random network had a similar number of edges, comparable to the average degree of the network. Since the networks produced by the Barabási–Albert model had a heterogeneous structure consisting of nodes with a wide range of degrees, they were called scale-free networks (Barabási and Albert, 1999). An average degree (or scale), as seen for random networks, cannot be used to characterise them. Several real-world cellular, technological, and social networks showed this scale-free structure, including the cellular protein-protein interaction (PPI) network, the World Wide Web and the internet, and the author citation network (Barabási and Albert, 1999, Albert and Barabási, 2002). Therefore, it became clear that the interdependencies of the nodes in these networks and their organizing principles were not products of random processes.
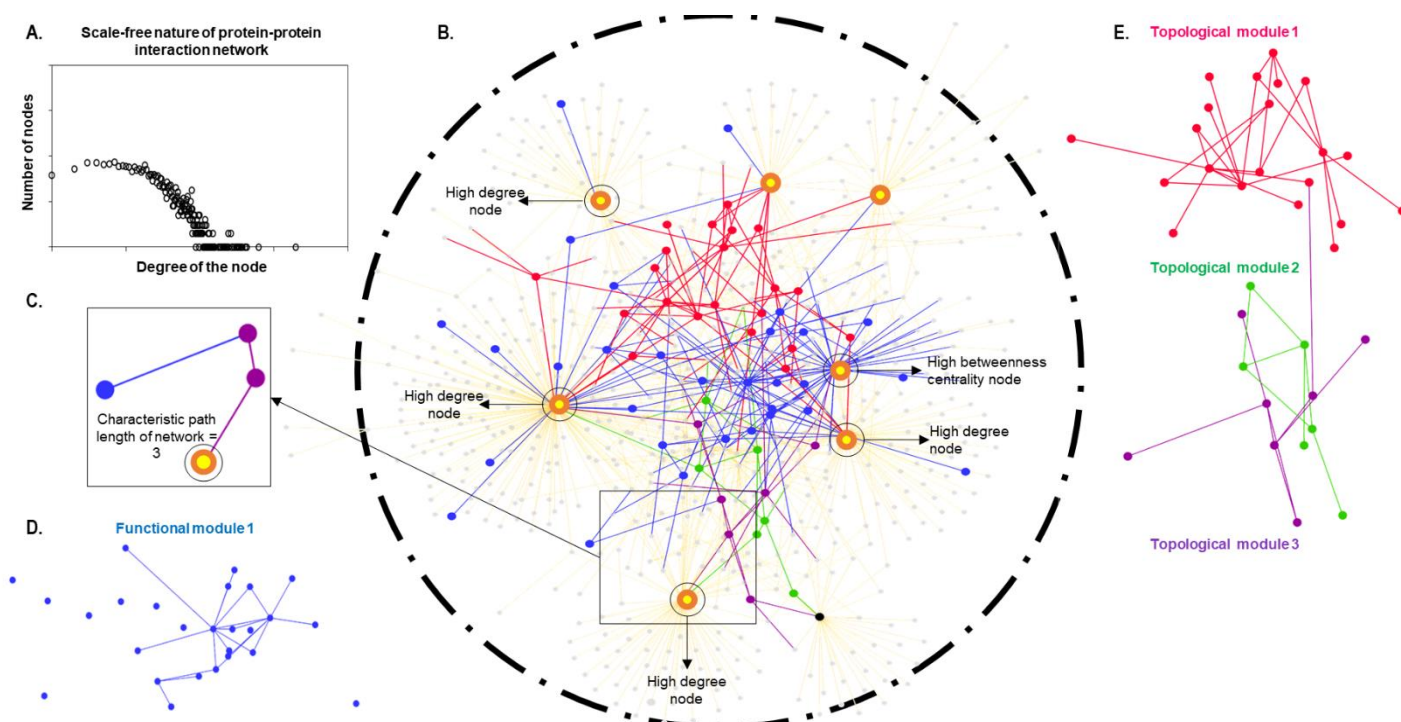
Figure 3: Characteristics of the protein-protein interaction (PPI) network. A. PPI networks are scale-free networks with a large number of small degree nodes and small number of high degree nodes. **B.** The few high degree nodes in the network called 'hubs' hold the network together. 'Bottlenecks' are nodes with high betweenness centrality that direct the flow of a significant amount of information in the network. Betweenness centrality measures the fraction of shortest paths passing through a node. Note that the nodes marked in this network as 'high degree nodes' also have high betweenness centrality and vice versa. **C.** Path length is the number of edges between any two nodes. The characteristic path length of a network is the average number of edges between all pairs of nodes in the networks, and can be used to assess the navigability of the network. **D.** Groups of nodes involved in the same cellular function are called functional modules. **E.** Groups of nodes in the network that are highly interconnected with one another compared to the rest of the nodes in the network, form topological modules. The network diagrams were created using Cytoscape. The degree distribution graph was created using the Cytoscape core app NetworkAnalyzer. The figures were assembled in Microsoft PowerPoint.

21

Instead, they were higher-level topological manifestations of systematically coordinated microscopic processes. These higher-level relationships could provide critical insights into the behaviour of the system. **Fig. 3** summarises the key characteristics of the scale-free networks, illustrated in the context of the PPI network.

The most notable characteristic of real-world networks that emerged from the scale-free network structure was that they were highly tolerant of random node failures and yet intolerant of (or vulnerable to) the removal of high-degree nodes (Albert and Barabási, 2002). Random node failures were more likely to affect the more abundantly present low-degree nodes, which do not dictate network topology. On the other hand, removing the high-degree nodes resulted in a breakdown of the topological structure, since they acted as scaffolds providing links to multiple nodes and shaping the network. Strikingly, the PPI network of yeast proteins interconnected via direct physical interactions demonstrated this same feature, i.e., differential response to node failure based on topological parameters (Jeong et al., 2001). In mutagenesis experiments, yeast cells showed high tolerance to the deletion of a vast number of proteins. In line with this, the topological structure of the yeast PPI network remained unaffected when randomly selected yeast proteins were removed using computational simulations (Jeong et al., 2001). On the other hand, single-gene deletion of ~62% of the proteins, each having more than 15 interactions (i.e., hub proteins), was lethal to yeast cells, although these high-degree nodes constituted only ~0.7% of the yeast proteins (Jeong et al., 2001). This finding that hub proteins were essential for survival led to several other studies (Barabási et al., 2011), which collectively showed that such proteins evolved more slowly and regulated more phenotypes than lower-degree nodes.

Slowly, it emerged that the PPI network could explain the non-linear relationship between the polygenic architecture of diseases revealed by genetic advances leading up to GWA studies and the clinicopathological correlations and phenotypic complexity seen during the Oslerian, Mendelian, and classical genetic eras. Now, human diseases were conceptualised as emergent properties of the cellular PPI network, which uses its inherent interdependences, i.e., PPIs, to respond to genetic and environmental perturbations (**Fig. 2**) (Barabási et al., 2011, Barabasi and Oltvai, 2004). Since PPIs drive the cellular machinery by facilitating biological processes, including signal transduction, formation of cellular structures, and enzymatic complexes, they are central to elucidating disease mechanisms. Individually, the effect of a disease-associated variant – uncovered through a GWA study – can be studied in terms of how it affects the protein structure and function. However, as shown in the study of the yeast PPI network (Jeong et al., 2001), proteins do not function in isolation within the cellular landscape. Instead, they interact with one another and give rise to complex networks with emergent properties. The network medicine paradigm was thus proposed as an integrative framework for examining the mechanistic effects of disease-associated genes within the context of the human PPI network (Barabási et al., 2011, Barabasi and Oltvai, 2004).

1.1.6 Discovering the human interactome: experimental methods for protein-protein interaction mapping and computational algorithms for protein-protein interaction prediction

Fortunately, in the early 2000s, alongside the re-conceptualisation of diseases as a network science problem, molecular interaction mapping techniques such as the yeast two-hybrid system, co-immunoprecipitation (Co-IP), several variations of Co-IP such as luminescence-based mammalian interactome (LUMIER) and dual luminescence-based co-immunoprecipitation (DULIP), proximity-dependent biotin identification (BioID) and affinity-purification mass spectrometry (AP–MS) (Blasche and Koegl, 2013, Trepte et al., 2015, Le Sage et al., 2016, Dunham et al., 2012) were developed. Co-IP captures interactors (or 'prey' proteins) based on their biochemical affinity for the protein of interest (i.e., the 'bait'). BioID, on the other hand, is based on addition of biotin tags to prey proteins (Le Sage et al., 2016). Biotinylated proteins are then bound by streptavidin, a protein that has a high affinity for biotin. The affinity of streptavidin for biotin has been reported to be the strongest non-covalent biological interaction known with a dissociation constant (K_d) in the femtomolar range (Sano et al., 1998, Stayton et al., 1999, Laitinen et al., 2006). The advantages of biotinylation over immunoprecipitation (IP) include the capture of transient interactors (fewer false negatives), higher capture specificity (fewer false positives) and greater enrichment of true interactors (Branon et al., 2017). In a study that mapped interaction networks at the centrosome−cilium interface, for the same 10 baits, >40% of the interactors detected by IP were also detected using BioID, with these shared interactions representing only 21% of the interactors identified by BioID (Gupta et al., 2015). False-positive PPIs are eliminated, since BioID is performed in living cells (before lysis) unlike IP and stringent washes may be used to separate biotinylated proteins from non-specific proteins (after lysis). Non-specific interactors detected using IP such as tubulin and HSPA9 were not detected in BioID (Branon et al., 2017). Greater enrichment of true interactors is observed since the biotin−streptavidin interaction used for capture of interactors is extremely strong and stringent washes may be used during purification to eliminate non-specific proteins. Interactors of the microprotein modulator of retroviral infection (MRI) were enriched 10-fold with ascorbate peroxidase (APEX)-mediated proximity biotinylation and only ~2-fold with IP (APEX is a peroxidase that catalyses proximity-dependent biotinylation in living cells) (Chu et al., 2017). LUMIER is a Co-IP assay which detects PPIs in mammalian cells by quantifying luminescence activity after immunoprecipitation (Blasche and Koegl, 2013). DULIP is a second-generation LUMIER assay (Trepte et al., 2015). Study with positive PPI reference set composed of high-confidence PPIs from the Human Integrated Protein Interaction Reference (HIPPIE) database (HIPPIE score ≥0.99) revealed that DULIP can detect known human PPIs with a sensitivity of ~ 79.5% (Trepte et al., 2015). Positive PPIs were recovered from hsPRS-v1 (Homo sapiens positive reference set) with a success rate of ~ 35% (Trepte et al., 2015). DULIP has been reported to detect human PPIs with high sensitivity and specificity, detecting both low and high-affinity interactions and the effects of point mutations on interaction strength (Trepte et al., 2015). AP–MS is a pull-down assay widely used to detect PPIs (Dunham et al., 2012). To counter the problem of distinguishing true interactors from non-specific contaminants in AP-MS, improved methods have been devised in which samples are quantitatively compared with control assays using rigorous statistical controls (Choi et al., 2011, Keilhauer et al., 2015).

Since PPI discovery through experimental techniques is prohibitively expensive, laborious and time-consuming, PPI prediction through computational algorithms (Kotlyar et al., 2015, Hopf et al., 2014, Emamjomeh et al., 2014a, Garzón et al., 2016, You et al., 2013, Jia et al., 2015, Li and Ilie, 2017, Deng et al., 2003, Raja et al., 2013) has been undertaken. Several methods

that emerged over the past ten years attempted to identify feature sets co-occurring in PPIs in a context-independent manner. In FpClass, the likelihood of an interaction was estimated from features that appeared to be unrelated but co-occurred in interacting partners, such as the protein domain in the first partner and subcellular localization of the second partner (Kotlyar et al., 2015). EV complex, PPIccc and DWIN used new features functionally relevant to PPIs such as the co-variance of amino acids, codon usage and variation of gene expression over time (Hopf et al., 2014),(Emamjomeh et al., 2014a),(Deng et al., 2003). The algorithm called PPInterFinder extracted relational keywords from MEDLINE abstracts to predict PPIs (Raja et al., 2013). However, overall, the number of text mining approaches for PPI prediction has dwindled over the years. Some of the most commonly used features for PPI prediction include information on domains, gene expression, Gene Ontology terms, orthology, evolutionary conservation of sequences and network topology. While some methods used an ensemble of these evidences such as Pre-PPI, in which a Bayesian approach was adopted to predict PPIs from genomic and structural features (Garzón et al., 2016), other methods used only one or two features such as iLOOPs, which is purely based on protein architecture (loops and motifs) (Planas-Iglesias et al., 2013). In a method based on personalised-recommendation system widely used in e-commerce, the problem of the incompleteness or 'sparsity' of the experimentally determined human interactome was treated in a manner similar to known user-item pairs being fewer in number than the unknown ones (Luo et al., 2015). Another method based on new measures of link prediction such as node graphlets attempted to capture the topologically similar deep neighbourhoods of interacting partners (Hulovatyy et al., 2014). PPI prediction based on sequence information were also widely used. PCA-EELM and iPPI-Esml used pseudo amino acid composition and distribution patterns of amino acids to predict PPIs using ensembles of learning machines (You et al., 2013),(Jia et al., 2015). PRISM considered sequence conservation in addition to structural similarity to interacting proteins in a template-based method (Baspinar et al., 2014). Evolutionary profiles were derived from sequences represented as kernels in a method called profile-interaction kernel (Hamp and Rost, 2015). Another emerging trend is the use of ensembles of learning machines to provide the final PPI prediction. Locfuse used an ensemble of learners such as random forest, naïve Bayes, multilayer perceptron and radial basis function to obtain the final prediction, also incorporating the concept of 'classifier fusion' in which every feature space has its own base classifier (Zahiri et al., 2014). An ensemble learning method that employed random forest, naïve Bayes, support vector machine and multilayer perceptron as base learners was used to predict PPIs between human and hepatitis C virus proteins (Emamjomeh et al., 2014b). The entire interactomes of *Bacillus licheniformis* and the human-microbial oral interactome were predicted based on the transfer of orthologous data and domain-domain interactions (Coelho et al., 2014, Han et al., 2016). An algorithm called SPRINT predicted the human interactome using a sequence-based algorithm that assessed the contribution of similar sub-sequences to the likelihood of interaction (Li and Ilie, 2017). Only a few of these studies reported experimental validation of the predicted PPIs. For example, 137 of the PPIs predicted by FpClass were validated using Co-IP, Glutathione S-Transferase pull-down assay and Mammalian Membrane Two-hybrid assay (Kotlyar et al., 2015).

Altogether, PPI mapping efforts using experimental methods – including three large-scale efforts that employed high-throughput yeast two-hybrid system (Rual et al., 2005, Rolland et al., 2014, Luck et al., 2020) – led to the discovery of ~150,000 PPIs in human cells out of the

estimated ~600,000 PPIs (data from Human Protein Reference Database or HPRD (Keshava Prasad et al., 2009) and Biological General Repository for Interaction Datasets or BioGRID (Stark et al., 2006)). The PPIs predicted using computational algorithms remain unintegrated into the global PPI map, due to the lack of experimental validation. Systematic efforts to curate PPIs validated in small-scale functional studies and large-scale efforts for interactome discovery led to the formation of at least 23 PPI databases, which additionally contained data on at least 26 modalities affecting PPIs (see **Appendix – Table 1**).

1.2 The disease interactome: an integrative framework to examine the disease state from multiple angles

The PPI network had a modular structure, containing several subunits of nodes (or 'modules') that were highly intra-connected and less inter-connected with other subunits (**Fig. 3**) (Barabási et al., 2011, Barabasi and Oltvai, 2004). The nodes in these topological modules often participated in the same cellular function and formed functional modules(Barabási et al., 2011, Barabasi and Oltvai, 2004). In addition to modules, the network also exhibited the small-world phenomenon, which meant that any two nodes could be connected through a relatively small number of edges (Barabási et al., 2011, Barabasi and Oltvai, 2004). From these, it became clear that the effects of perturbations – such as gene mutation, abnormal gene or protein expression, and treatment-related and environmental perturbations – could propagate in this network, produce disease and influence multiple aspects of disease expression, including shared symptomatology, comorbidity, and response to drugs (Barabási et al., 2011, Barabasi and Oltvai, 2004). In line with this, studies showed that ~67% of disease-associated variants either alter the binding affinities between proteins or abolish/establish PPIs (Laddach et al., 2018, Sahni et al., 2015). Other studies showed that genetic variants localizing to the protein core may disrupt its tertiary structure and abolish all chances of the protein interacting with any of its interaction partners (node removal in the interactome) (Barabási et al., 2011), and variants localizing to interaction interfaces may perturb specific interactions (edge perturbation in the interactome) (Barabási et al., 2011).

The PPI network or the 'interactome' started to be used as a skeletal framework to elucidate biologically relevant relationships existing at a higher level among genes harbouring disease-associated variants, which may not be apparent by examining individual genes (Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a). The most important use of the interactome model was in creating integrated landscapes for specific human diseases. Heterogeneous disease genetic data spread across independent studies, which hitherto remained unconnected, could now be interconnected, as was demonstrated for ataxias and autism spectrum disorders (ASDs). Purkinje cell dysfunction and degeneration emerged as the common themes underlying the ataxia interactome constructed using previously unrelated genes linked to 23 inherited ataxias as starting points (Lim et al., 2006). Similarly, despite the clinical heterogeneity of ASDs, the ASD interactome seeded using genes harbouring mutations producing (or associated with) syndromic ASD or severe language delay showed tight interconnectivity and enrichment for copy number variants seen in ASD patients (Sakai et al., 2011). The interactome also converged on common subcellular locations and biological functions.

Five other factors make the interactome framework better poised to address disease aetiology (**Fig. 4**). Firstly, unlike the GWA framework, the PPI network could be used as a mechanistic framework to integrate disease-associated cellular data from disparate sources, such as transcriptomic and methylation data. The effect of DNA variation is most proximally reflected by the RNA, as per the central dogma of biology (Schadt, 2009). Studies showed that SNP genotypes were responsible for more than 50% of the variation in RNA abundance levels, supporting this notion (Schadt, 2009). Hybridization-based microarray techniques and RNA-sequencing technologies developed alongside genome sequencing technologies for disease gene discovery (Schadt, 2009). Hence, there was ample opportunity to integrate transcriptomic datasets derived from patients, cell lines, and animal models into the PPI network. The interactome also served as an integration platform for cross-species data. In this manner, vital information on human diseases learned from animal models could be incorporated into the disease model. Specifically, interactions experimentally validated or predicted based on co-expression patterns and genomic information (as seen in the Search Tool for the Retrieval of Interacting Genes/Proteins or the STRING database (Szklarczyk et al., 2021)) in other species can be transferred to the human interactome and used to construct networks of disease-associated genes. A breast cancer interactome study elegantly demonstrated the power of the interactome as an integration platform for various kinds of biological datasets and cross-species data (Pujana et al., 2007). The interactome was seeded by four tumour suppressor genes, expanded by including genes that showed correlated expression in several transcriptional datasets of human tissues and cell lines, and augmented using differential expression profiles of breast tumours, cross-species genetic interactions, and PPIs. This interactome showed higher connectivity than random networks containing the same
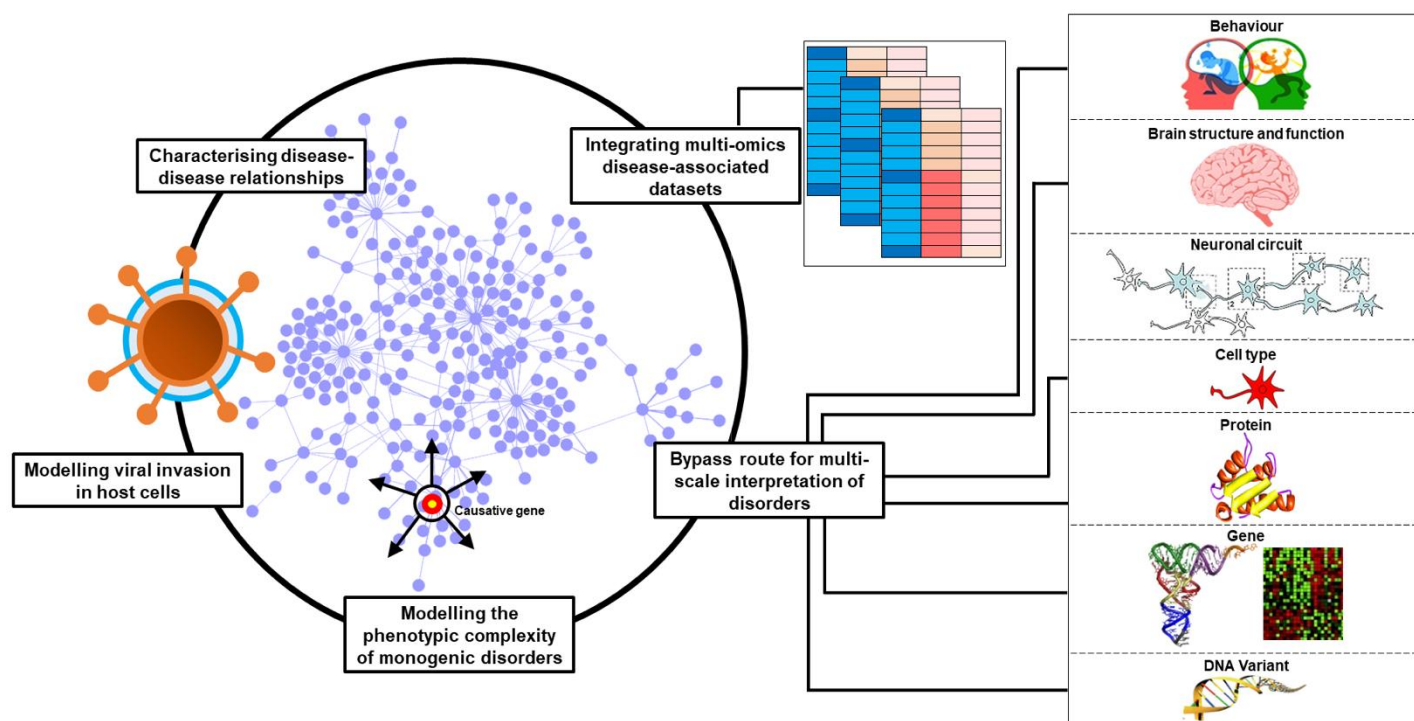
Figure 4: Major applications of the interactome framework. The five applications of the interactomic framework that will help elucidate disease mechanisms are (a) integrating multi-omics datasets, (b) acting as a bypass route for various biological levels, (c) explaining the phenotypic complexity of monogenic disorders by charting the neighbourhood networks of the causative genes, (d) re-formulating the problem of viral hijack of the host cellular machinery as that of host protein interactome perturbation by viral proteins and (e) examining the neighbourhood networks of multiple diseases to elucidate higher-order etiological relationships. The figure was created in Microsoft PowerPoint. The network diagram was created using Cytoscape.

number of nodes and sourced from the same transcriptional datasets, indicating its biological cohesion. The interactome also helped identify previously unknown functional links of known tumour suppressor genes, such as *BRCA1*, with genes carrying disease-associated variants in independent studies. In another study, a large expression dataset of cancer tissues and cell lines was used to identify the regulatory interactions of genes (Ergün et al., 2007). The regulatory network was then used to filter the differential expression datasets of two prostate cancer states (non-recurrent primary and metastatic states) and identify the genetic mediators regulating these states.

Secondly, it was becoming increasingly clear that complex disorders need to be examined through the lens of the multiscale networks spanning various levels of biological organization and interacting with one another to produce structure, function, and phenotype. Therefore, the traditional gene-centric approach was rapidly converging with the systems approach, the effect of which most strongly impacted the field of neuroscience, as shown by the linear paradigm of biological discovery in psychiatry (Totah, 2016). According to this paradigm, it was essential to examine multiple intervening levels for elucidating the effects of variants on behavioural phenotypes, i.e., gene expression, protein function, protein activity in multiple cellular landscapes converging in local neural circuits, and their interactions in global neuronal wiring. This multi-level investigation is often impossible in a single study. However, in small-scale PPI studies, interactome analysis was used to bypass various levels of biological discovery, normally traversed to elucidate the effect of a variant on a neuropsychiatric phenotype. For example, discovering the PPIs of the proteins encoded by the chromosomal region 16p11.2 (which harbours recurrent aberrations associated with ASD and schizophrenia (SCZ)) led to several findings on their involvement in the KCTD13-Cul3-RhoA pathway and the late mid-foetal brain development seen in the psychiatric disorders (Sharma et al., 2015b). Similarly, a SCZ-predisposing mutation was shown to lead to abnormal dopaminergic modulation in a circuit of prefrontal cortical interneurons (Choi et al., 2017). PPIs between the immunoglobulin superfamily proteins DIP-α and Dpr6/10 have been shown to regulate layer-specific synaptic circuitry (Xu et al., 2018a). Altogether, contrary to previous models such as Oslerian formalism, the ability of the interactome to serve as an integration model for multiscale disease data made it better poised to address real-world phenotypic heterogeneity. An example of a multiscale relationship in neuroscience is the correlation between the spatial distribution of brain gene expression and functional connectivity between various brain regions (Anderson et al., 2018, Richiardi et al., 2015, Wang et al., 2015, Krienen et al., 2016, Patania et al., 2019, Mills et al., 2018). Specific brain regions or neural circuits strongly drive psychiatric morbidities (as seen for anxiety disorders (Whalen et al., 2008, Hattingh et al., 2013)). The correlation between gene expression and functional connectivity suggests that these disorders may exhibit abnormalities in region-specific transcriptional signatures. However, gene expression linked to psychiatric disorders identified in post-mortem human brain tissues and pharmacogenomic animal models are from disparate sources and have limited overlap. The interactome model can reveal the hidden interdependencies of genes identified in disparate studies and the genes expressed in specific brain regions. An interesting example is an ASD study that described an interactome module showing statistically significant enrichment for ASD-associated variants (Li et al., 2014). This module showed dichotomised expression in the corpus callosum and the hippocampus.

Thirdly, the interactome model also provides an opportunity to construct a complex interactome space for Mendelian disorders and interpret their phenotypic complexity in terms of the interdependencies of the causative genes in the interactome. A seminal study introduced the random walk with restarts algorithm to help prioritise 'candidate' genes associated with Mendelian diseases for experimental studies based on their proximity to known disease-associated genes in the human interactome (Köhler et al., 2008). Another seminal study showed that missense mutations in Mendelian diseases lead to widespread perturbations in the interactome, and that different mutations in the same genes could alter interaction profiles and produce distinct disease phenotypes (Sahni et al., 2015).

Another attractive utility of the interactome model is in elucidating the host invasion and response mechanisms in viral epidemics. When viral proteins bind to proteins in the host cell, this effect may spread in the interactome through regulatory and biophysical interactions, perturbing other proteins in the PPI network, and affecting viral infection, host immunity, and the response to therapeutics. The host could use these complex interactions to restrict viral replication in host cells, or the virus could hijack them to allow its perpetuation, as has been shown by their preferential targeting of hubs in the human interactome (Calderwood et al., 2007, Shapira et al., 2009).

Two competing approaches exist to define and categorise mental disorders (Thaxton et al., 2022). In the first ('lumping') approach, disorders are broadly categorised based on shared symptom dimensions, as seen in Research Domain Criteria, which suggests a framework to investigate mental disorders. In the second ('splitting') approach – as seen in the Diagnostic and Statistical Manual of Mental Disorders – the symptomatic distinctions are highlighted to allow differential diagnosis of disorders and their subtypes. Through the interactome model, these approaches can be used as complementary indicators to define the interactomic space of a disease based on genetic components underlying the shared and unique aspects of its symptoms. Disease-disease relationships have been characterised in a seminal study that proposed the concept of the human 'diseasome', a web of pathophenotypes linked based on shared genes harbouring disease-associated mutations (Goh et al., 2007). Distinct disease modules could be delineated from the diseasome. The genes associated with similar disorders were more likely to interact with one another and show correlated expression patterns. As in the yeast interactome (Jeong et al., 2001), essential genes in the human genome were highly likely to encode hub proteins and show expression in a large number of tissues (Goh et al., 2007). On the other hand, majority of the disease genes were non-essential, and they were highly likely to encode non-hub proteins (Goh et al., 2007). Their expression was uncoupled from, i.e., not correlated with, the expression of the other genes. Purifying selection will likely weed out mutations in essential hub genes critical to survival and development to prevent spontaneous abortions and postnatal lethality (Goh et al., 2007). Hence, it is highly likely that disease-associated mutations localise to topologically neutral non-hub proteins, ensuring that the organism survives up to reproductive years, allowing the mutations to persist in the population (Goh et al., 2007).

In summary, for functional interpretation of genetic variants associated with complex disorders, it is imperative that we surpass the simplistic view of 'variants-disrupt-protein function' and place variants in the complex web of PPIs to unravel disease mechanisms. Mapping disease-

associated variants onto proteins and PPI networks will help pull more disease-associated genes into the network and examine communities of proteins involved in biological pathways relevant to the disease aetiology. The interactome model allows a more complex conceptualisation of human disease. As demonstrated by the multiple studies above, it provides a framework to 'translate' genetic data into discoveries on disease mechanisms.

1.3 A historical perspective on drug discovery: from the magic bullet hypothesis to the network medicine paradigm

The classical pharmacology and systems therapeutics eras shaped our understanding of disease therapeutics and sections 1.3.1 and 1.3.2 describe the chronological sequence of events spanning these two eras. Drug discovery driven by classical pharmacology principles saw limited success and led to the co-evolution of polypharmacology and the network medicine paradigm in the systems therapeutics era. **Fig. 1** provides a bird's eye view of the events in drug development and network biology described in these sections; note that the figure also depicts the events in disease genetics, described earlier in section 1.1. **Fig. 5** summarises the evolution of the relationship between drug-target interactions and clinical manifestations in the classical pharmacology and systems therapeutics eras.
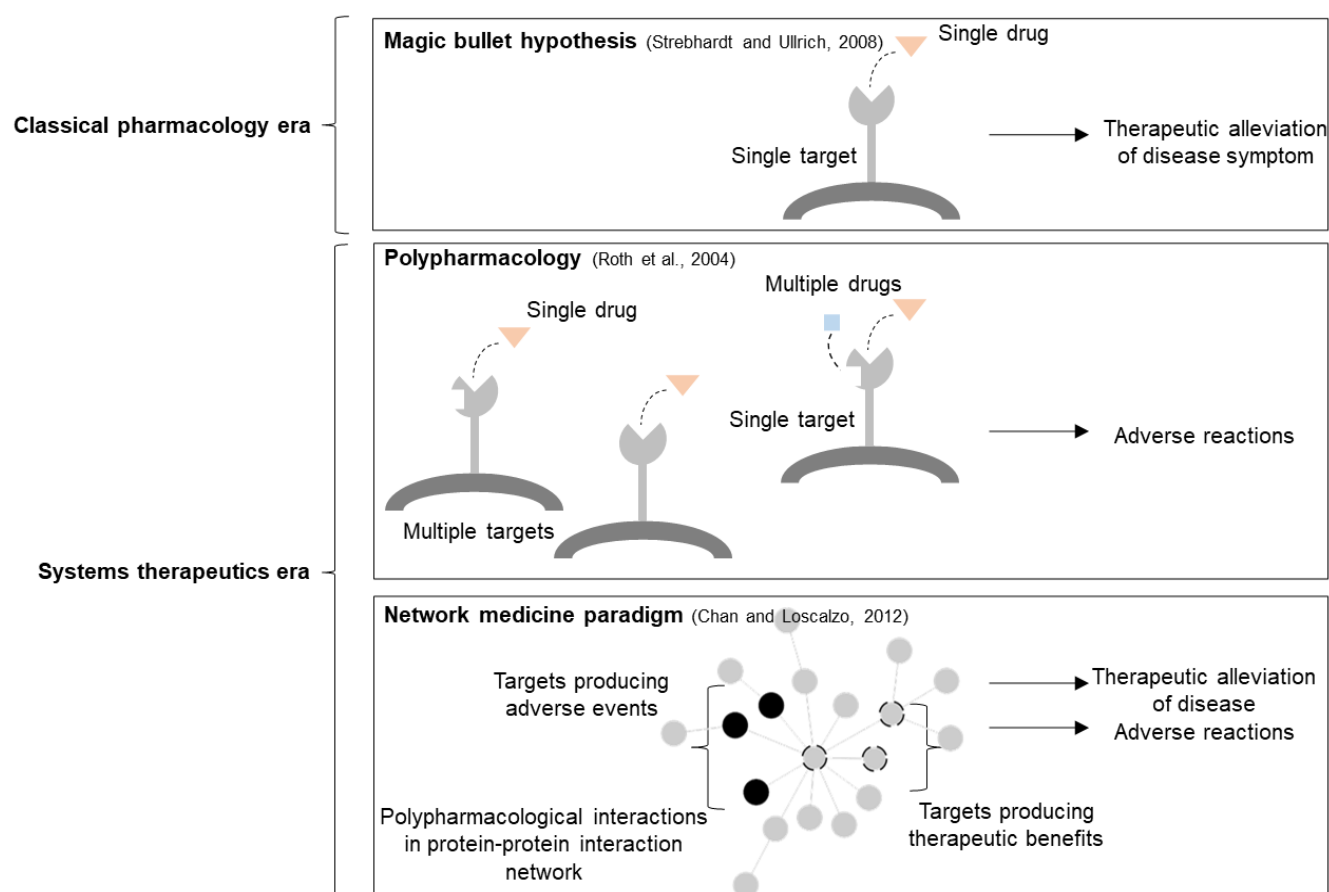
Figure 5: The evolution of the relationship between drug-target interactions and clinical manifestations. Clinical investigations in the classical pharmacology era focused on individual drugs believed to act on single cellular targets and alleviate disease symptoms. In the systems therapeutics era, a single drug was found to act on multiple targets and multiple drugs on a single target, and such 'promiscuous' and 'off-target' interactions were believed to result in adverse events. This multi-drug-multi-target interaction principle or 'polypharmacology' was later integrated into the network medicine paradigm, and shown to be responsible for both therapeutic and non-therapeutic effects of drugs. The figure was created in Microsoft PowerPoint. The network diagram was created using Cytoscape.

1.3.1 The classical pharmacology era: drug discovery driven by the magic bullet hypothesis and the resulting decline in clinical development success rates

Coupled with the problem of understanding disease aetiology is the need to devise targeted treatment strategies. Therefore, drug discovery followed close at the heels of disease biology, as the latter transitioned from the reductionist era of Oslerian formalism to the holistic age of network medicine. Traditional drug discovery revolved around the 'magic bullet' hypothesis proposed by Paul Ehrlich in 1907 (**Fig. 5**) (Strebhardt and Ullrich, 2008). Small molecules or biologics that target and modify the function of a specific protein were developed to alleviate or prevent disease symptoms. Although several beneficial drugs were developed using this approach, particularly antimicrobial and antiviral agents (Strebhardt and Ullrich, 2008), it produced unsatisfactory results in complex disorders. For example, drugs such as donepezil, rivastigmine, and galantamine, which reversibly inhibit the enzyme acetylcholinesterase by forming hydrolysable carbamylated compounds with it, are widely used for alleviation of dementia in Alzheimer's disease (Moss et al., 2017, Greig et al., 2001, Ballard, 2002, Poirier, 2002, Fukuto, 1990, Morisset et al., 1996). Only a small section of the people treated with these drugs experience significant cognitive improvement (Di Santo et al., 2013). Cognitive decline in a vast majority of people is delayed by only 6-9 months (Masters et al., 2015). Further, the top-ten grossing drugs in the U.S. benefit only 4% to 25% of the patients treated with them (Schork, 2015).

The fact that drugs benefit only small subsections of the patient population led the pharmaceutical industry to develop 'blockbuster' drugs for common diseases (Loscalzo, 2023). The reasoning was that a large section of the patient population had to benefit from the drugs, for drug development to be economically profitable (Loscalzo, 2023). However, such blockbuster drugs did not produce the intended effects. They had low efficacy rates, possibly because the patient population was naturally stratified into various groups based on their genetic predisposition and responded differently to drugs (Loscalzo, 2023). Additionally, these patient subgroups had already been exposed to existing drugs whose interactions with the blockbuster drugs remained unexplored (Loscalzo, 2023). Collectively, these failures have led to a decline in drug discovery. Despite increasing investments in research and development, drugs take longer to be developed, have high attrition rates at various phases of clinical trials, and have a much higher cost (Loscalzo, 2023).

The trends revealed by a study that examined the clinical development success rates from 2006 to 2015 provide some critical insights into the specific challenges faced by the industry (Thomas et al., 2016). Firstly, only 9.6% of the drug candidates in phase I clinical trials transitioned into phase II, with psychiatric drugs the least likely to make this transition (Thomas et al., 2016). The safety of the candidate drug is tested in phase I. Therefore, the high attrition rate in this phase could indicate that the information on drug activity examined in the preclinical trials was incomplete. The caution exercised by the regulatory bodies in phase I is warranted. Serious adverse drug reactions constitute the fourth leading cause of death in the U.S., with 100,000 deaths per year and about 2 million patients in the U.S. experiencing adverse drug reactions per year (Giacomini et al., 2007). Patient fatalities have led to the withdrawal of 19 drugs from the U.S. market during 1998-2007 (Giacomini et al., 2007). Secondly, drugs intended to treat rare diseases and targeting specific biomarkers were more likely to have high

approval rates in all the developmental phases (Thomas et al., 2016). On the contrary, drugs targeting common diseases were the least likely to advance to phase II (Thomas et al., 2016). These results suggested that developing drugs targeting specific patient subgroups identified using pharmacogenomics approaches will increase their success rates. Altogether, these results highlighted the need to revise the strategies for drug development, with a focus on reducing adverse events and increasing drug efficacy by incorporating genomic insights. To address these two problems, the conventional 'one drug for one target for one disease' approach had to be replaced by a more complex approach.

1.3.2 The systems therapeutics era: re-conceptualisation of polypharmacology by the network medicine paradigm to explain the multimodal effects of drugs

Historically, drug design was based on findings from studies that describe genetic and pharmacological modulation of specific targets and pathways, which elicit measurable changes in pathophenotypes (Chan and Loscalzo, 2012). This framework suggested that side effects of specific drugs arose from the unintended manipulation of 'off-targets' in other pathways, a phenomenon attributed to the concept of polypharmacology, i.e., a drug has multiple targets, and a target may bind several ligands (**Fig. 5**) (Chan and Loscalzo, 2012). In conventional drug development, polypharmacology was treated as an undesirable property of drugs and drug targets. However, the network medicine paradigm reconceptualised drugs as having multimodal effects in the complex cellular landscape of PPIs (Barabási et al., 2011, Barabasi and Oltvai, 2004, Yıldırım et al., 2007), thereby re-contextualising polypharmacology (as described by (Roth et al., 2004)) as a property critical to modelling the complex phenotypic responses to drugs (**Fig. 5**). Specifically, drugs that target proteins could perturb the PPI network to elicit the intended therapeutic response in a disease or an unintended adverse event or side effect (Chan and Loscalzo, 2012).

In the network medicine paradigm, the preclinical phase in the drug development pipeline will involve the following steps: identification of the disease module within the cellular interactome, identification of multiple druggable targets in this module and construction of a drug target
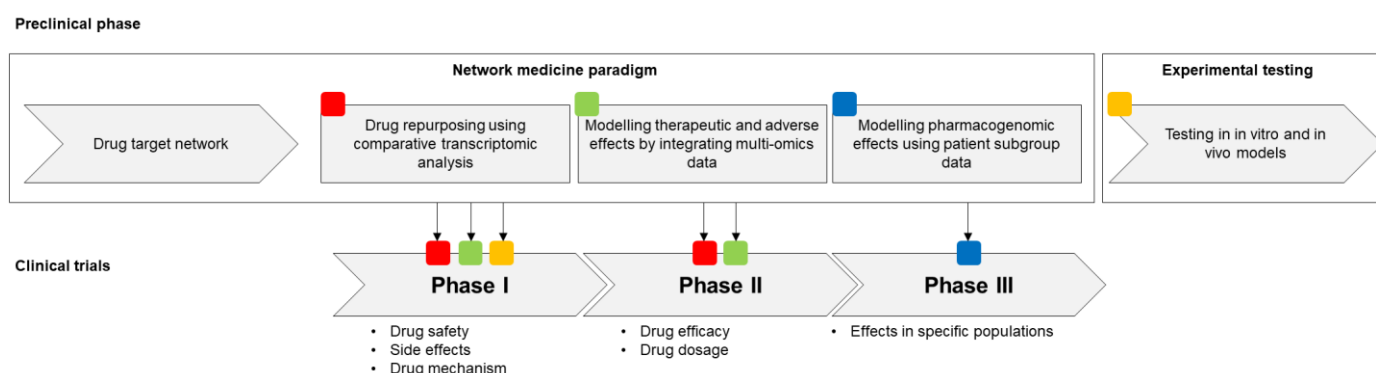
Figure 6: Integration of the network medicine paradigm into the clinical developmental pipeline. Each application of the network medicine paradigm (marked with coloured boxes) will inform phase I-III clinical trials in patients. Most repurposed drugs will have passed clinical trials and will be safe for human use. Hence, drug repurposing results will inform phase 1, which focuses on drug safety. Additionally, they will inform phase II, which focuses on drug efficacy. Analyses integrating the DTN with multi-omics datasets that help elucidate drug mechanisms will inform phase I efforts to characterise the therapeutic and non-therapeutic effects of drugs and phase II efforts to assess drug efficacy. Integrating the DTNs with patient subgroup-specific data will help identify the pharmacogenomic interactions of the drug. This, in turn, will inform phase III, which assesses the effects of drugs in multiple populations. The figure was created in Microsoft PowerPoint.

network (DTN), refining the list of drugs in this network by assessing their efficacy through comparative analysis of drug-induced and disease-associated transcriptomes (as seen in 'drug repurposing', described below), modelling therapeutic and non-therapeutic benefits of the drugs by integrating DTNs with multi-omics datasets, and experimental testing in cell lines and animal models (**Fig. 6**). This paradigm redefines adverse events as the consequence of PPIs in the network, providing opportunities to mitigate them before the clinical trials. This, in turn, will help prioritise drug combinations to reduce or neutralise adverse events and select drugs with specific biological profiles less correlated with adverse events, before the clinical trials. Network pharmacology studies have provided several key insights into the biological correlates of adverse events. Pairs of drugs used for the same disease induce adverse events when the network modules of their protein targets overlap with each other and the network of disease-associated genes ('overlapping exposure') (Cheng et al., 2019). A 2.6-fold greater risk of side effects was seen with drugs targeting genes having 5 specific genetic features, including non-specific tissue gene expression (Duffy et al., 2020). The findings from this study also suggested that side effects arise from drug delivery to multiple tissues (including those unrelated to the disease) (Duffy et al., 2020). Network medicine opens up the possibility of tailoring drugs for specific patient subgroups, whose characterization should be made possible by clinical trials using comprehensive phenotyping strategies (Loscalzo, 2023). The interactome can be used as a framework to map the heterogeneous genetic data of the subgroups and isolate targetable sub-networks after integration with multi-omics datasets that help predict drug efficacy. This step should also be integrated into the preclinical phase of drug development (**Fig. 6**).

The clinical pipeline is yet to incorporate the rational strategies offered by the network medicine paradigm to address the issues of adverse events and stratified drugs. On the other hand, a concept derived from the paradigm called 'drug repurposing' has been widely adopted in preclinical research, and has helped identify several drugs, particularly during the coronavirus disease (COVID-19) pandemic (Chen et al., 2021, Zhou et al., 2020, Gysi et al., 2020, Li and De Clercq, 2020). Drug repurposing involves identifying drugs approved by regulatory bodies such as the Food and Drug Administration (FDA) for treating a specific disease and repositioning them for a different disease. Investigational drugs (whose clinical efficacy is under investigation) can also be repurposed. Diseases may share druggable proteins, as indicated by the interconnectivity of the human interactome and the diseasome (Barabási et al., 2011, Barabasi and Oltvai, 2004, Goh et al., 2007). The drugs targeting these proteins could be profiled and refined for clinical specificity, and subsequently therapeutically switched to treat a condition other than the original indication. The repurposing route, compared to novel drug discovery, offers several advantages. Since the drugs identified in this manner are more likely to have passed safety trials in preclinical models and early-stage trials, they are less likely to fail the subsequent phases of clinical development and more likely to reduce the overall time for drug development and lead to the discovery of new pathways and targets (Dudley et al., 2011, Athauda et al., 2017, Kinnings et al., 2009, Evans et al., 2005). Multi-omics datasets measuring drug activity at single-cell resolution and the level of molecular interaction perturbations are now becoming available (Srivatsan et al., 2020, Lambert et al., 2019). Their integration with DTNs will help devise new strategies to identify drugs with low attrition rates in clinical trials.

Overall, mapping drugs and drug targets to the complex landscapes of disease-associated interactomes helped harness the potential of polypharmacology for drug identification. The network medicine approach will widen the scope of the conventional clinical development pipeline. However, the methods currently available for drug discovery, repurposing, adverse event minimization, and personalizing therapeutics, are varied. It is important to standardise these methods and test them in several diseases. Research on disease biology through the lens of the interactome will also help fine-tune these methods for different categories of disorders.

1.4 The current state of network biology: factors limiting its scope in disease mechanism elucidation and drug discovery and past attempts to address them

Subsequent to the discovery of the scale-free nature and the small world property of PPI networks, scientists active in the field of network biology primarily focused on advancing four key areas: (i) deriving the rules governing the distribution of disease-associated genes and drug targets in the human interactome (Barabási et al., 2011, Barabasi and Oltvai, 2004, Yıldırım et al., 2007, Cheng et al., 2019), (ii) elucidating the relationship of the PPI network –
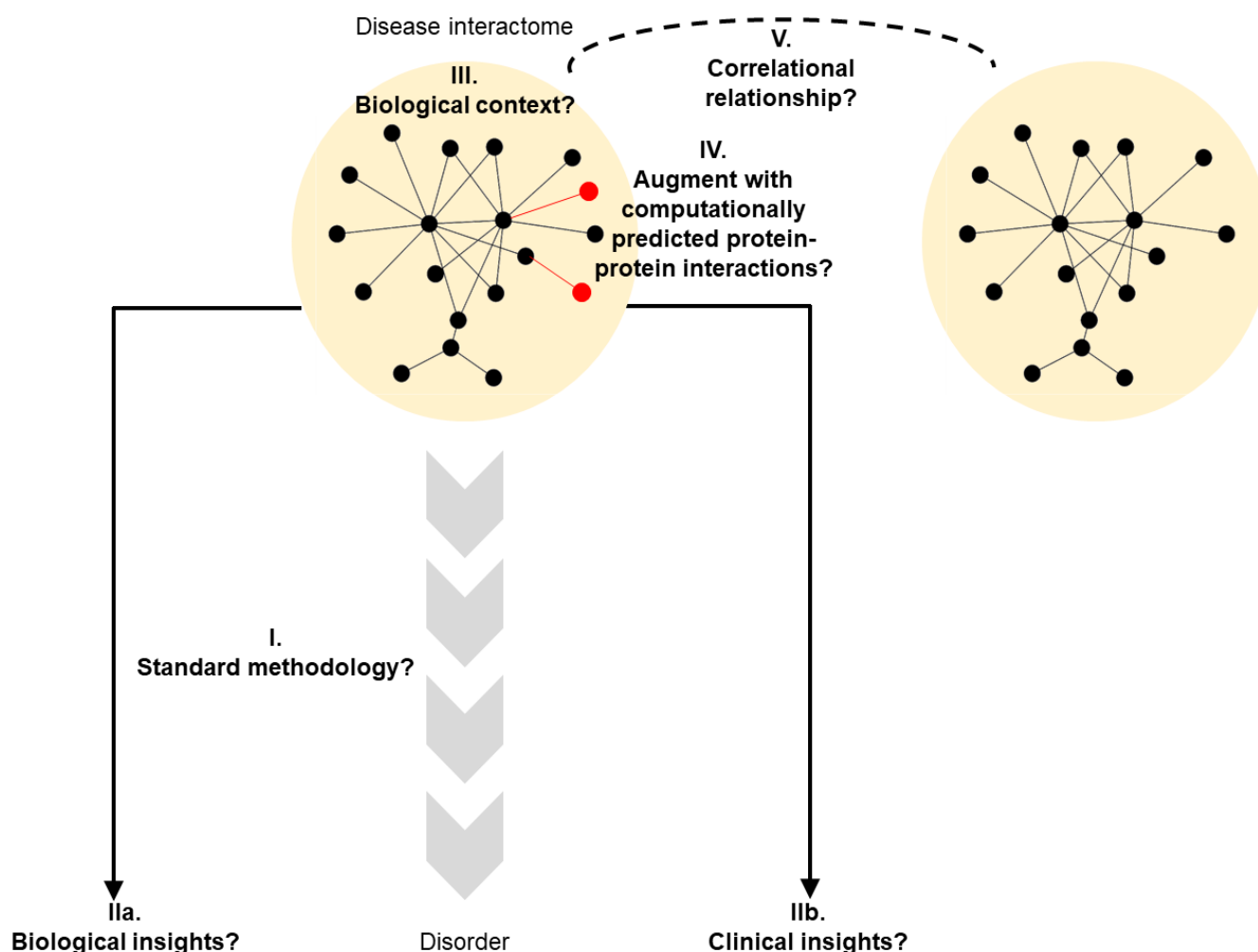
Figure 7: Factors limiting the application of network biology to disease mechanism elucidation and drug identification. The five factors limiting the scope of the network medicine paradigm in disease biology and therapeutics are: (I) the lack of a standard methodology (II) to derive biological and clinical insights into the disease, (III) based on a context-specific disease interactome (IV) augmented with computationally predicted PPIs (to circumvent the sparsity of the experimentally validated PPI map), and (V) the lack of a methodology to examine the correlation of this disease interactome with other interactomes. The figure was created in Microsoft PowerPoint. The network diagram was created using Cytoscape.

which contains biophysical interactions between proteins – with other types of cellular networks, such as epistatic genetic interactions (Ulitsky and Shamir, 2007, Kelley and Ideker, 2005) and gene co-expression networks (Paci et al., 2021), which capture biological pathways and processes, (iii) producing derivative networks such as the diseasome (Goh et al., 2007) that chart higher-level disease-disease relationships based on shared genes and (iv) investigating whether the organizational principles of disease-associated genes in the cellular networks are also reflected in the interrelationships of the human disease phenotypes (Hidalgo et al., 2009); these phenotypic relationships were inferred based on multiple indicators of disease comorbidity (such as relative risk and Matthews correlation coefficient) computed using the disease history of millions of patients. Large-scale network studies have provided crucial insights into these areas: the functional coherence and the higher interconnectivity of interactome neighbours in disease modules (Barabási et al., 2011, Barabasi and Oltvai, 2004), the various configurations of disease modules and DTNs in the human interactome that produce therapeutic benefits and adverse events (Cheng et al., 2019), the high overlap between PPIs of proteins encoded by essential genes (including protein complexes) and genetic interactions (Costanzo et al., 2016), the localization of 'switch' genes – that are inferred from gene co-expression profiles to regulate disease transitions – to disease modules in the human interactome (Paci et al., 2021), the peripheral localization of most human disease-associated genes and the separation of related disorders into distinct clusters based on shared genes (Goh et al., 2007), and the network patterns underlying the development, progression and prognosis of comorbidities (Hidalgo et al., 2009).

The conceptual advances provided by these and similar studies helped identify the characteristics of the microscopic processes that drive large-scale molecular and phenotypic networks. These processes, when perturbed, contribute to disease development, manifestation, progression, therapeutics, and prognosis. However, five factors have since limited the scope of network biology (**Fig. 7**).

First, no single unifying schema exists to examine disorders based on network biology principles. The subsets of network biology principles and methods used to analyse disease-specific genetic data vary between independent research groups across the globe (Sun et al., 2009, Sharma et al., 2015a, Sharma et al., 2018, Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a). In this scenario, it is critical to devise a specific methodology adoptable by an interdisciplinary team – consisting of biologists, bioinformaticians, clinicians, and computer scientists – and applicable across disorders.

Second, due to the lack of a standard methodology, a single disease interactome is not sufficiently explored in a single study or a series of related studies, leading to a dearth of disease-associated actionable information (Sun et al., 2009, Sharma et al., 2015a, Sharma et al., 2018, Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a). Therefore, it is critical to devise a protocol that uses disease-associated genes as starting points and generates (i) biologically insightful results, seeding further scientific investigations, and (ii) clinically actionable results, leading to clinical trials and therapeutic interventions.

Third, PPIs predicted by computational algorithms are often not integrated into network analyses (Kotlyar et al., 2015, Hopf et al., 2014, Emamjomeh et al., 2014a, Garzón et al., 2016,

You et al., 2013, Jia et al., 2015, Li and Ilie, 2017, Deng et al., 2003, Raja et al., 2013). Approximately 75% of the PPIs that are estimated to exist remain undiscovered by experimental methods, and several disease-associated genes have no known PPIs (Rual et al., 2005, Rolland et al., 2014, Luck et al., 2020). Therefore, augmenting the networks of experimentally determined PPIs with computationally predicted PPIs could help construct a more populated version of the interactome, allow the characterization of under-studied proteins through functional associations of their predicted interactors, and facilitate the discovery of previously unknown disease mechanisms.

Fourth, the field is witnessing far more attempts at drawing generalizable context-insensitive conclusions across all human disorders than examining disorders, disorder subtypes, or related subtypes, individually or jointly, in a context-sensitive manner (Goh et al., 2007, Hidalgo et al., 2009). Gene specificity at various physiological levels – organs, tissues, and cell types – is crucial in determining the pathological consequences of disease-associated variants. Therefore, integrating the interactome with multi-omics datasets generated at different physiological levels and for specific diseases will help refine and contextualise the wiring diagram of the molecular interactions of disease-associated genes.

Fifth, standard methodologies to study the higher-level correlational relationships between the interactomes of related disorders or drug targets, although proposed (Goh et al., 2007, Cheng et al., 2019), have not been widely adopted. In the former case, it is critical to devise methods that help understand the converging points of disease development and the diverging points of etiological differentiation and reconsider the norms of disease classification. In the latter case, it is critical to devise methods that help understand the complex molecular relationships producing favourable or unfavourable drug-drug interactions embedded in the interactome.

1.5 An interactome-based framework to translate disease-associated genetic data into biological and clinical insights

To address these issues, an interactome-based framework to 'translate' disease-associated genetic data into biological and clinical insights is proposed in this thesis. Detailed workflows for this framework can be found in **Appendix section 14.2**. Its utility is demonstrated in six disorder classes, specifically, a neurological disorder (Alzheimer's disease in **Chapter 2** and **Appendix section 14.2**), two neuropsychiatric disorders (anxiety disorders in **Chapter 9**, and schizophrenia or SCZ in **Chapter 8**), a cardiovascular disorder (hypoplastic left heart syndrome in **Chapter 3**), two cancers (malignant pleural mesothelioma in **Chapter 4** and malignant peritoneal mesothelioma in **Chapter 5**), a viral disease (COVID-19 in **Chapter 6**) and one bone dysplasia (sedaghatian type spondylometaphyseal dysplasia in **Chapter 7**); note that in **Chapter 10**, six comorbid disease pairs belonging to multiple disorder classes have been examined. This framework can be used to functionally characterise a single interactome (**Fig. 8**) and compare multiple disease interactomes and DTNs (**Fig. 9**), with the dual aim of elucidating disease mechanisms and identifying clinically actionable targets. Following are the three underlying components of the framework (**Fig. 8**):
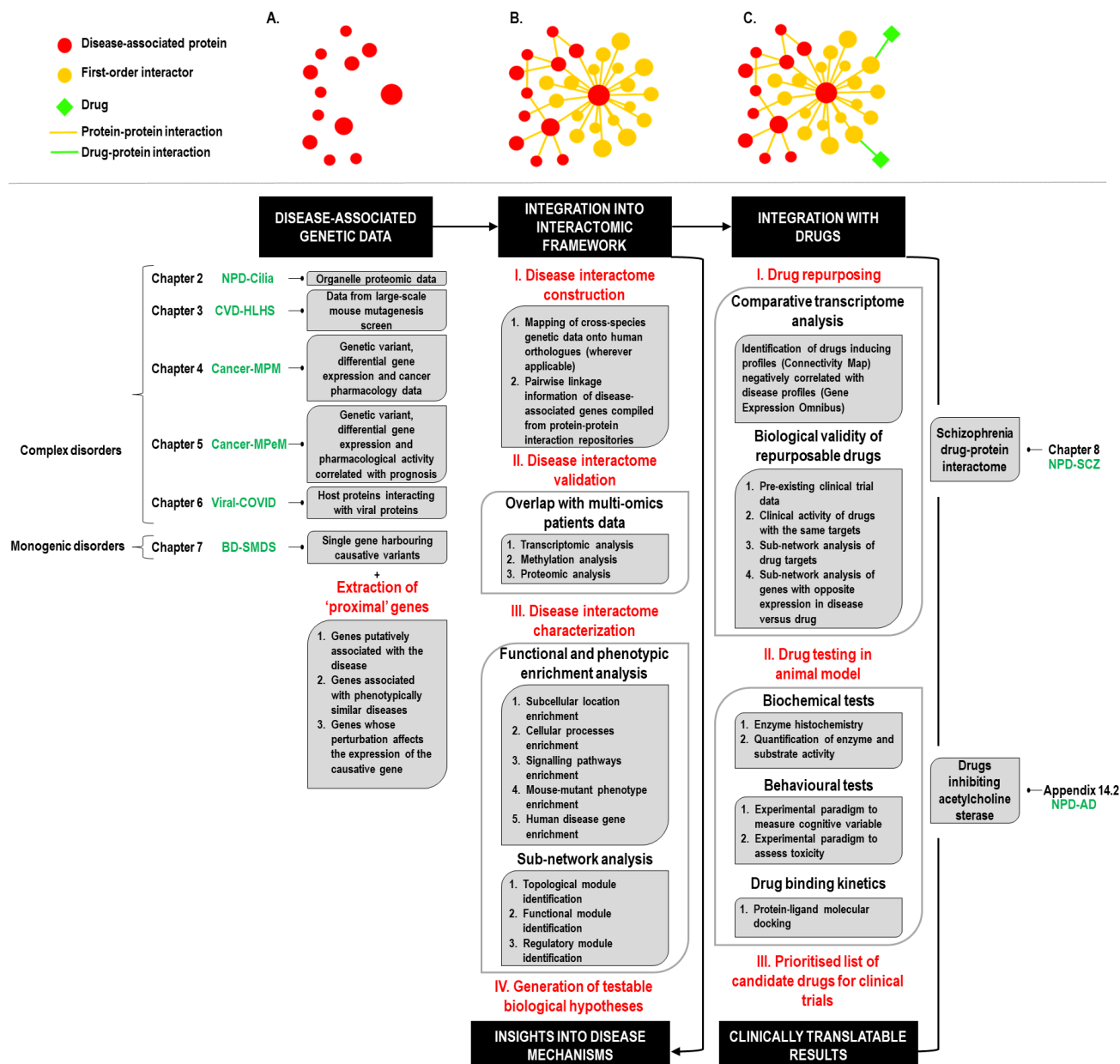
Figure 8: The framework to translate disease genetic data into biological and clinical insights using single interactome analysis. The various steps in the proposed framework to examine a single interactome seeded by disease-associated genetic data have been enumerated. Note that specific codes (shown in green colour) have been used to indicate the disorder class and the specific disorder examined in Chapters 2-9; disorder classes: CVD: Cardiovascular disorders; NPD: Neuropsychiatric/neurological disorders; Viral: Viral infections; specific disorders: AD: Alzheimer's disease; BD: Bone dysplasias; HLHS: Hypoplastic left heart syndrome; MPeM: Malignant peritoneal mesothelioma; MPM: Malignant pleural mesothelioma; SCZ: Schizophrenia; SMDS: Spondylometaphyseal dysplasia, Sedaghatian type. The figure was created in Microsoft PowerPoint.

A. Compilation of disease-associated genetic data
B. Mapping of the proteins encoded by these genes onto the human protein interactome
C. Addition of the drugs interacting with the proteins in the interactome

A series of methods used for single interactome analysis (**Fig. 8**) originating in (B) leads to the generation of insights into disease mechanisms. Another series of methods stemming from (C) yields clinically translatable results. The methods used for the generation of biological and clinical insights from multiple interactomes are different (**Fig. 9**).
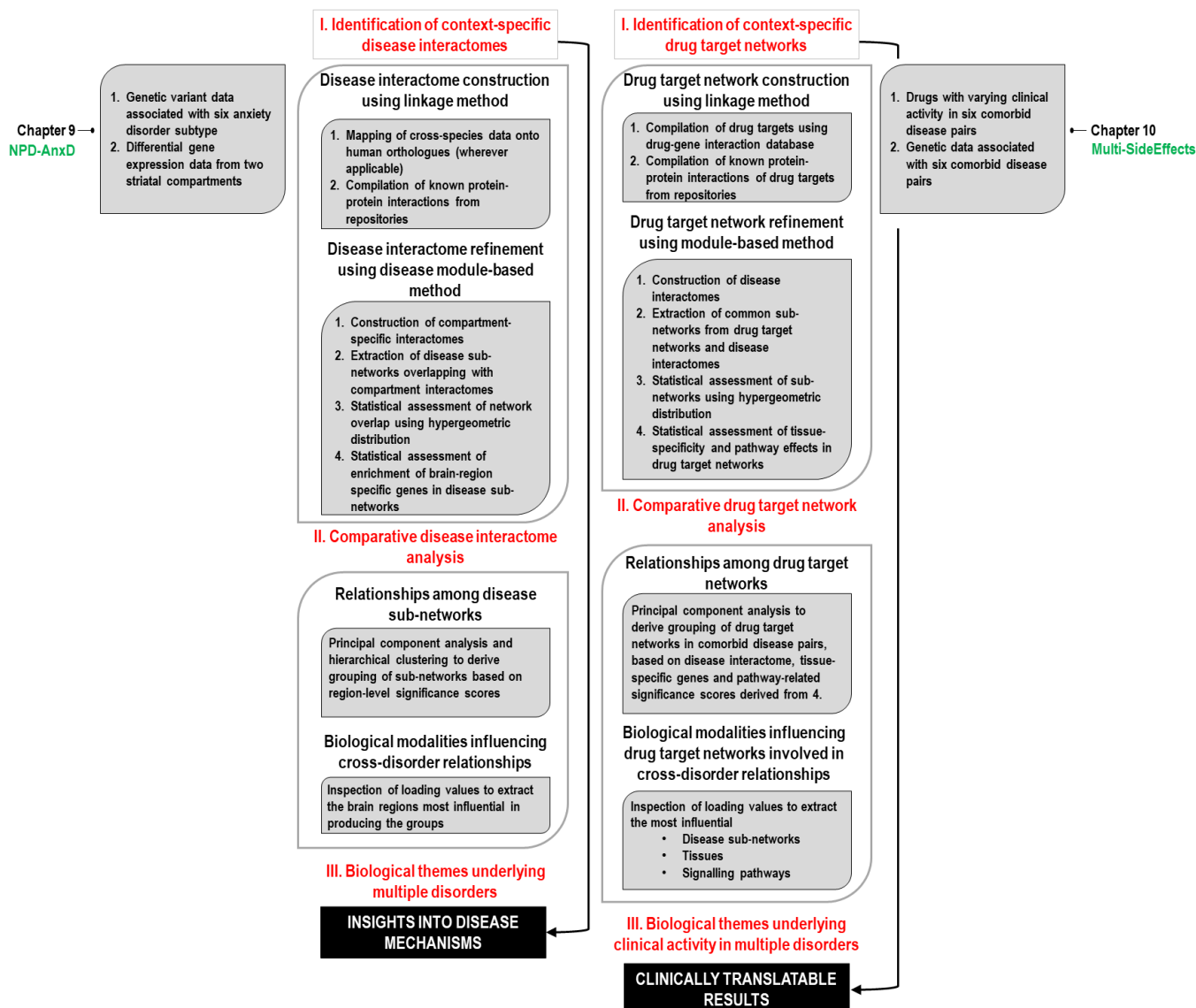
Figure 9: The framework to translate disease genetic data into biological and clinical insights using multiple interactome analysis. The various steps in the proposed framework to examine multiple interactomes seeded by disease-associated genetic data have been enumerated. Note that specific codes (shown in green colour) have been used to indicate the disorder class and the specific disorder examined in Chapter 9 and Chapter 10; disorder classes: Multi: Multiple disorders; Anx: Anxiety disorder. The figure was created in Microsoft PowerPoint.

Below, a brief description of the single interactome analysis framework (see **Fig. 8**), as implemented across **Chapters 2-9**, is provided. The sub-sections that follow describe the constituent steps of the framework in detail.

1.5.1 The network biology-based framework for examining a single interactome

Since complex disorders have polygenic architectures, multiple disease-associated genes – compiled from disparate sources – were available for disease interactome construction. On the contrary, only single (possibly causative) genes were available for monogenic disorders. Neither were these genes amenable to functional interpretation, nor could they explain the complexity of the disease phenotype. To circumvent this, genes 'proximal' to the causative gene or the disease phenotypes were extracted. The proteins encoded by the disease-associated genes were interlinked and their first-order interactors were identified using the 'linkage' method for interactome construction. The biological validity of the resultant disease

interactomes was ascertained by computing their overlap with multi-omics patient data. Testable biological hypotheses on disease mechanisms were generated by conducting functional and phenotypic enrichment analysis of the interactome and identifying critical sub-networks. Next, the drugs that targeted specific proteins in the disease interactome were identified. The drugs that induced differential gene expression profiles (in cell lines) negatively correlated with disease-associated expression profiles (in patients) were prioritised as 'repurposable drugs' after performing multiple analyses to ascertain their biological validity. It is critical to test potential drugs in an experimental preclinical study. Therefore, an animal model was used to assess the utility of the drug. Biochemical and bioinformatics analyses were used to quantify drug activity and binding kinetics. Behavioural tests involving experimental paradigms were used to measure symptom alleviation.

The methodology to derive biological and clinical insights from single interactomes, described thus far, has been demonstrated in **Chapters 2-9**.

1.5.1.1   Compilation of disease-associated genetic data for complex disorders

In **Chapters 2-6**, the aim was to examine the mechanisms underlying complex disorders under polygenetic regulation, namely, Alzheimer's disease (**Chapter 2**), hypoplastic left heart syndrome (**Chapter 3**), malignant pleural mesothelioma (**Chapter 4**), malignant peritoneal mesothelioma (**Chapter 5**) and COVID-19 (**Chapter 6**). **Fig. 10** shows the methodologies proposed for compiling the gene sets and constructing the interactomes of complex and monogenic disorders.

Disease-associated genetic data were obtained from different sources, deemed suitable for the disorders in question. These genetic data were used as seed nodes to identify active and/or conserved disease-specific network modules in the human interactome (**Fig. 10**). Transcriptomic and genomic disease-specific data capture dynamic mechanisms correlated with disease pathophysiology (Mitra et al., 2013). When mapped to the interactome, this data will help reveal 'active' network modules underlying differential molecular activity and phenotypic manifestation in disease (Mitra et al., 2013). In the SCZ, mesothelioma, Alzheimer's disease, and COVID-19 studies, active network modules were extracted. In the SCZ study, the genes historically linked to the disease in the pre-GWAS era (Farrell et al., 2015) and the variant-harbouring genes identified in GWA studies (Ripke et al., 2014) were used as seed nodes. The historical genes included genes harbouring structural variations, linked to antipsychotic pharmacology and positional candidates in linkage studies (Farrell et al., 2015). Two other the studies were on the pleural and peritoneal subtypes of mesothelioma (a rare and aggressive cancer). In the pleural mesothelioma study, genes reported by the Ingenuity Pathway Analysis suite (Cedres et al., 2012) to be related to this specific cancer through gene expression changes or genetic variants, or by being targeted by drugs clinically active against the cancer, were used. In the peritoneal mesothelioma study, a list of genes associated with the cancer was compiled from nine studies (Joseph et al., 2017, Ugurluer et al., 2016, Chirac et al., 2016, Foster et al., 2010, Hung et al., 2020, Pillai et al., 2013, Varghese
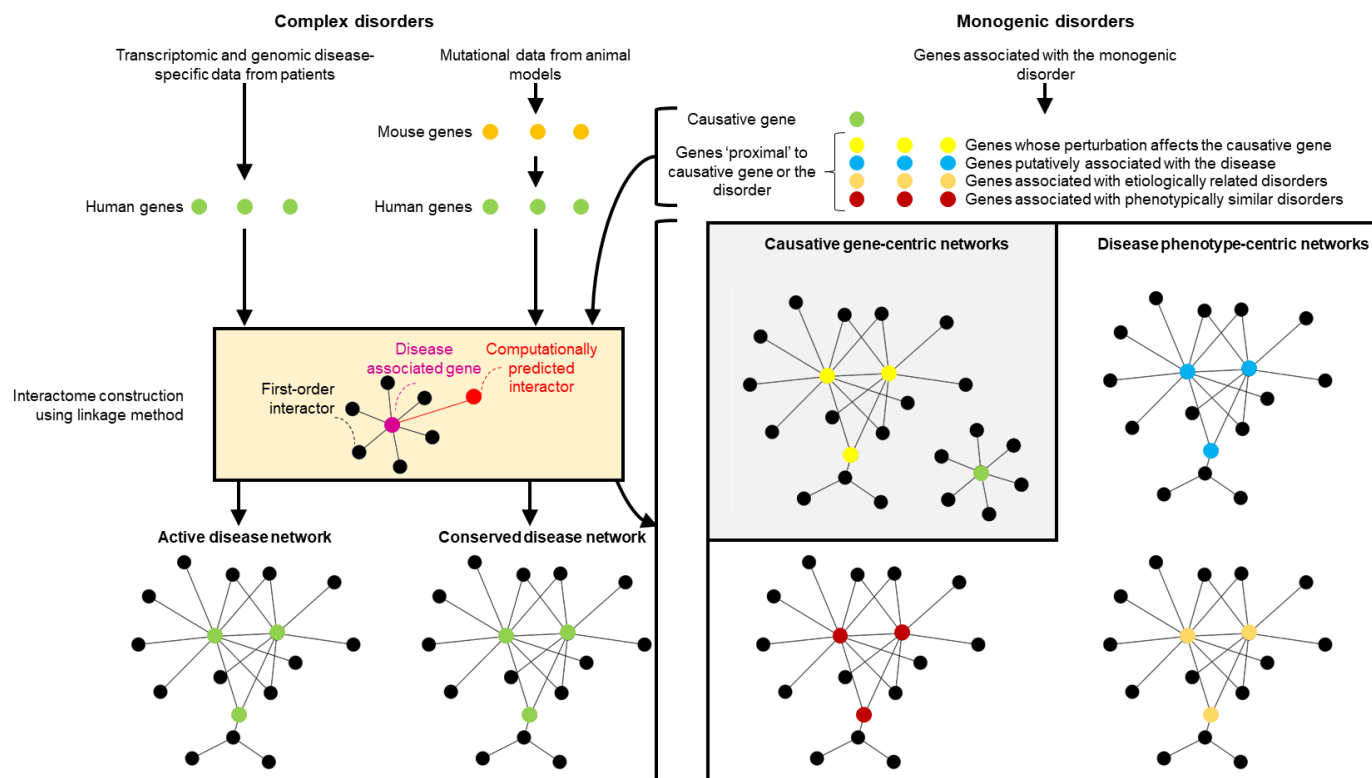
Figure 10: The proposed methodology for gene set compilation and interactome construction in complex and monogenic disorders. In the case of complex disorders, genes carrying disease-associated mutations or showing differential expression in patients can be compiled. An active disease network can be assembled by identifying the direct interactors of the proteins encoded by these genes (i.e., the 'linkage' based method for interactome construction). Alternatively, genes from animal models of human diseases can be compiled and their human orthologues can be identified. These genes will help construct conserved networks. In the case of monogenic disorders for which only causative genes will be available, one causative gene-centric gene set and three disease-centric gene sets can be compiled, and used for interactome construction. The network diagrams were created using Cytoscape. The figure was assembled in Microsoft PowerPoint.

et al., 2011, Zaffaroni et al., 2007, Hung et al., 2018). These genes harboured mutations, copy number aberrations, rearrangements, or showed expression patterns in peritoneal mesothelioma surgical specimens correlated with poor prognosis in patients or reduced cell survival/less favourable response to drugs. In the study that examined the role played by the organelle cilia in Alzheimer's disease pathology, the genes associated with Alzheimer's disease were extracted from DisGeNET (Piñero et al., 2015), with the gene-disease association score > 0.2 to include only expert-curated disease genes. In the COVID-19 study, the host proteins interacting with 27 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) viral proteins identified from the 2019-nCoV/USA-WA1/2020 strain (Gordon et al., 2020) were used to construct the network module active in host invasion and response mechanisms.

Active modules, when preserved across species, are called 'conserved' modules (Mitra et al., 2013). Mapping mutational data from animal models to the human interactome will help identify conserved network modules (Mitra et al., 2013). The disease-associated genes in the hypoplastic left heart syndrome (HLHS) study were obtained from animal models. Specifically, genes harbouring HLHS-associated mutations were recovered from a large-scale mutagenesis screen conducted with eight independent mouse lines (Liu et al., 2017, Li et al., 2015) to construct an active and conserved network module.

Altogether, it is clear that polygenic sets associated with complex disorders can be used as seed nodes in the interactomic framework. These gene sets could be curated using expert knowledge and in alignment with the specific hypotheses formulated for each disorder.

Several diseases are believed to be under monogenic regulation, and yet their complex phenotypes remain unexplained by single causative genes. The interactome framework allows one to map the relationships of the causative genes with additional lists of genes suspected to be active in the disease, based on gene-centric or phenotype-centric inferences (**Fig. 10**): (i) genes speculated to be associated with the disease by the BeFree text mining algorithm that employs a kernel-based approach based on morphosyntactic and dependency information to identify gene-disease associations (Bravo et al., 2014), (b) genes associated with etiologically related disorders, (c) genes associated with disorders sharing phenotypic similarity with the disease in question and (d) genes whose perturbation is known to cause significant overexpression or underexpression of the causative gene. **Chapter 7** presents a protein interactome perspective on the rare and lethal skeletal dysplasia called spondylometaphyseal dysplasia, Sedaghatian type (SMDS). SMDS exhibits an autosomal recessive pattern of inheritance, and has been attributed to at least 3 mutations in the gene *GPX4* (Smith et al., 2014, Aygun et al., 2012), which codes for the protein glutathione peroxidase. This study showed that the interactomic framework unravels the interdependencies of this causative gene in the neighbourhood network that help produce the complex disease phenotype.

1.5.1.2   Construction of disease interactomes using the linkage method

In **Chapters 2-7**, the disease interactomes were constructed by identifying the direct interacting partners of the proteins encoded by the disease-associated genes (i.e., seed nodes) from the experimentally validated PPIs catalogued in two PPI databases, namely, HPRD (Keshava Prasad et al., 2009) and BioGRID (Stark et al., 2006). This method of interactome construction is called linkage-based (**Fig. 10**), because it assumes that the direct interactors of a protein encoded by a gene associated with a disease phenotype are likely to produce or be associated with the same disease phenotype (Barabási et al., 2011, Barabasi and Oltvai, 2004). For example, within a disease locus identified using the forward genetics approach and containing an average of 100 genes, it is more than 10% likely that the (genes coding for the) direct interacting partners of the known disease proteins are the true disease-causing genes than the other genes in the locus (Barabási et al., 2011, Barabasi and Oltvai, 2004). The linkage-based method differs from modularity-based and diffusion-based methods (Barabási et al., 2011). The modularity-based method assumes that proteins belonging to the same topological or functional modules are highly likely to be involved in the same disease. The diffusion-based method identifies proteins closest to those encoded by disease-associated genes using random walkers. Note that in **Chapter 3** (hypoplastic left heart syndrome), the disease-associated genes identified in mouse models were mapped to their corresponding human orthologues using the Homologene database. After piecing together the neighbourhood network of the disease-associated genes using experimentally validated PPIs, we augmented the network using novel PPIs computationally predicted using the High-precision Protein-Protein Interaction Prediction (HiPPIP) algorithm. As mentioned in section 1.1.6, a large part of the human interactome remains unknown, and computational algorithms

developed to predict PPIs in human as well as model organisms (Kotlyar et al., 2015, Hopf et al., 2014, Emamjomeh et al., 2014a, Garzón et al., 2016, You et al., 2013, Jia et al., 2015, Malavia et al., 2017b, Deng et al., 2003, Raja et al., 2013) help circumvent the resulting sparsity of the interactome.

The training dataset for developing the HiPPIP model consisted of 20,000 known (i.e., experimentally verified) interactions from HPRD, combined with 80,000 non-interacting pairs (Ganapathiraju et al., 2016). The test set comprised 0.3% known interactions out of 160,000 pairs, with proteins having more than 50 known PPIs considered as hubs. The model employed a random forest algorithm, incorporating protein features such as cellular localisation, molecular function, gene location, expression, domains, and tissue membership. Trained with 30 trees, the model produced a continuous score within the [0,1] range. Evaluation on a held-out test dataset, using a threshold of 0.75, resulted in 97.5% precision and 5% recall. Additional evaluations included generating ranked lists for hub genes, treating pairs with a score > 0.5 as novel interactions. These predictions were ranked by score, and precision-recall analysis was conducted by adjusting the threshold from 1 to 0. Scanning the ranked lists from top to bottom allowed the computation of true positives versus false positives. The HiPPIP model was initially introduced for SZ genes (Ganapathiraju et al., 2016). Each historical (pre-GWAS) and GWAS SZ gene was paired with every other human gene. Pairs with a score > 0.5 were considered predicted interactions, and these, combined with known PPIs, constituted the SZ interactome. 0.5 was chosen as the threshold based on evaluations with hub proteins that showed that the pairs that received a score > 0.5 are highly likely to be interacting pairs. This observation was further validated by experimentally confirming a few novel PPIs above this score.

One of the strengths of HiPPIP is the translational impact of the predicted PPIs, as illustrated in individual studies. For example, based on HiPPIP predictions, the human OASL protein was shown to interact and co-localise with RIG-I involved in activating cellular innate immunity to virus infections (Zhu et al., 2014). OASL was found to mediate host responses to viral infections by activating RIG-I through its C-terminal ubiquitin-like domain, mimicking polyubiquitin (Zhu et al., 2014). Additionally, interactome-level analysis of disease-specific proteins including their predicted PPIs revealed the central role of cilia in congenital heart disease (Li et al., 2015) and identified mitochondrial proteins as crucial in hypoplastic left heart syndrome, a subtype of congenital heart disease called (Liu et al., 2017). Interactome analysis also revealed genetic and biological overlaps between schizophrenia and rheumatoid arthritis, potentially explaining their inverse epidemiological relationship (Malavia et al., 2017). Computational evaluations indicate that HiPPIP outperforms other state-of-the-art algorithms in PPI prediction (Dunham and Ganapathiraju, 2022). However, the algorithm has notable limitations. First, there is a scarcity of large-scale experimental confirmation for the predicted PPIs, with only 17 PPIs tested and validated experimentally (Zhu et al., 2014, Ganapathiraju et al., 2016). Second, the generalisation of the model to diverse and less-explored biological contexts may be limited by the protein sets represented in the training data. Selective discovery of PPIs introduces sampling biases into the training set (Von Mering et al., 2002). These biases stem from examining PPIs from specific cellular environments or the PPIs of proteins highly expressed in specific tissues or conserved across multiple species (Von Mering et al., 2002). Third, the training data was sourced from HPRD (Keshava Prasad et al., 2009),

which has not been updated since 2010. The relevance of the model to the current state of knowledge may be limited, given the rapid accrual and update of PPI data and gene annotations since then. Lastly, the new PPIs predicted by HiPPIP exhibit limited overlap with recently released interactome maps (Luck et al., 2020, Huttlin et al., 2020) (**Chapter 6**). The predicted PPIs included in the interactomes reported in **chapters 2-8** are all sourced from HiPPIP. Given its limitations, future studies would benefit from a more robust approach, involving the use of a consensus method, incorporating PPIs predicted by multiple algorithms, including those discussed in section 1.1.6.

The decision to use the known (i.e., experimentally verified) PPIs from the HPRD database (in conjunction with the BioGRID database) in **chapters 2-8** was influenced by the addition of HPRD PPIs in the HiPPIP

Figure 11: Parameters of the enrichment analysis based on hypergeometric distribution. The Venn diagram shows the four parameters used to compute the hypergeometric p-values for the enrichment analyses performed with disease interactomes and multi-omics datasets. Note that ∩ indicates intersection (i.e., the number of genes shared between K and n). The figure was created in Microsoft PowerPoint.

training set. However, using more comprehensive and updated datasets with context-specific information, particularly those curated through international collaborations such as the IMEx consortium, would have improved the studies. The IMEx consortium curates molecular interactions from various databases (Porras et al., 2020), including DIP (Xenarios et al., 2002), IntAct (Xenarios et al., 2002), MINT (Licata et al., 2012), MatrixDB (Chautard et al., 2010), IID (Kotlyar et al., 2019), Innate DB (Breuer et al., 2013), and UniProt (UniProt Consortium, 2019), all in the Proteomics Standards Initiative Molecular Interactions (PSI-MI) standard format. IMEx contains ~660,000 PPIs, covering 86% of the human interactome (assuming an estimated interactome size exceeding 767,000 PPIs), and these interactions are extensively annotated with biological features (Porras et al., 2020).

1.5.1.3   Validation of the disease interactomes using multi-omics disease data

The biological validity of the disease interactomes – constructed from the proteins encoded by the disease-associated genes, and their experimentally determined and computationally predicted PPIs – was ascertained by computing their overlap with multi-omics patient datasets, depending on their availability (as seen in **Chapters 2-6**). Statistically significant overlaps with the genes differentially regulated in transcriptomic, methylation and proteomic datasets of patients compared to control subjects validated the dynamical integrity of the interactomes under disease conditions. The statistical significance of the overlaps between the list of genes in the interactome and the differentially regulated genes was computed using an overrepresentation (or enrichment) analysis based on hypergeometric distribution (**Fig. 11**). In this method, p-value is computed from the probability of k successes in n draws (without replacement) from a finite population of size N containing exactly K objects with an interesting feature.

$$P\left(X=k\right)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Population size N = Total number of genes whose differential regulation under disease conditions was tested

Number of successes in the population K = Number of genes that were differentially regulated under the disease condition

Sample size n = Number of genes in the interactome

Number of successes in the sample k = K ∩ n, where ∩ indicates intersection (i.e., the number of genes shared between K and n).

The p-values derived from the enrichment analysis were corrected for multiple hypothesis using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). In this method, the hypergeometric p-values are sorted from small to large, multiplied by the total number of tests and then divided by their rank order. A p-value < 0.05 after Benjamini-Hochberg correction was considered to be statistically significant. Prior to the enrichment analyses, differentially regulated genes were identified from the multi-omics datasets using specific criteria. For example, in differential transcriptomic datasets, genes with fold change >2 or ½ were considered as significantly overexpressed and underexpressed respectively at p-value < 0.05. In differential methylation datasets, methylation values (M-values) were computed as M = $log_2[\beta\ (1-\beta)]$ for both control and test disease cases, where β is the ratio of methylated probe intensity and overall intensity. Difference between M-values of test and control cases was then computed, and genes with M-value > 1 and M-value < 1 were considered to be hypermethylated and hypomethylated respectively at p-value < 0.05.

1.5.1.4   Characterization of the disease interactome

To identify the functions overrepresented in the disease interactome as a whole, it is critical to examine the enrichment of its constituent genes in various biological modalities from various biological levels. Therefore, in **Chapters 2-7**, enrichments for biological processes, cellular components and molecular functions (Gene Ontology (Consortium, 2004)), pathways (Reactome (Croft et al., 2014)), diseases (OMIM (Hamosh et al., 2005) and DisGeNET (Piñero et al., 2015)), mutant phenotypes (Mammalian Phenotype Ontology (Smith and Eppig, 2012)) and transcription factor targets (MSigDB (Liberzon et al., 2011)) were computed using the gene set analysis toolkit called WebGestalt (Liao et al., 2019). WebGestalt computes enrichment of specific functional groups (e.g., a Reactome Pathway) in an input list (e.g., genes in the disease interactome). Statistical significance is computed using Fisher's exact test and corrected using the Benjamini-Hochberg method for multiple test adjustment. WebGestalt was chosen for its user-friendly interface, intuitive plots, large collection of functional categories from different types of functional databases and multiple enrichment analysis methods. This analysis typically yielded information from different biological levels potentially influenced by the disease state. For example, analysis with the SARS-CoV-2-modulated host protein interactome (**Chapter 6**) provided insights into the conditions comorbid with COVID-19, and subcellular locations, cellular processes and pathways potentially targeted by SARS-CoV-2.

Apart from the general functional associations of the interactome, it is critical to identify the topological, functional and regulatory modules contained within it. Some neighbourhoods in the interactome may consist of genes that have a high tendency to interact with genes in the same local neighbourhoods than with genes in other neighbourhoods (Barabási et al., 2011, Barabasi and Oltvai, 2004). Such topological modules can be identified using the Netbox algorithm (Cerami et al., 2010) that is blind to the function of the individual genes (**Chapter 3** and **Chapter 6**); the data can consist of the disease-associated genes (i.e., seed genes) and all human PPIs. It expands the set of seed genes by adding nodes from the entire human interactome whose number of links to the seed genes are statistically significant compared to its degree in the human interactome. From this sub-network, it identifies highly interconnected modules by computing a scaled modularity score, as compared with the modularity observed in 1000 random permutations of the sub-network. Scaled modularity refers to the standard deviation difference between the observed sub-network and the mean modularity of the random networks (Wang and Zhang, 2007). In contrast with the network clustering algorithms detecting topological modules, those detecting functional modules partition the interactome based on the aggregation of genes with similar or related functions in local neighbourhoods. Functional modules can be extracted using the HumanBase toolkit (**Chapter 5** and **Chapter 7**) (Krishnan et al., 2016). HumanBase uses shared k-nearest-neighbours and the Louvain community-finding algorithm to cluster the genes sharing the same network neighbourhoods and similar Gene Ontology biological processes into functional modules. The p-values of the terms enriched in the modules are calculated using Fisher's exact test and Benjamini-Hochberg method. ReactomeFIViz, a Cytoscape plugin (Wu et al., 2014), was used to extract a regulatory module involved in viral budding and interferon signalling pathways from the host protein interactome (**Chapter 6**). Unlike topological and functional modules that contain undirected (non-directional) interactions, a regulatory module contains directed interactions; an interaction in a regulatory module is usually between two proteins participating in the same biochemical reaction as components of a protein complex, or as an input, catalyst, activator or inhibitor (Barabási et al., 2011, Barabasi and Oltvai, 2004). Overall, the topological modules (that show enrichment for cellular functions), and the functional and regulatory modules detected from the disease interactomes could either correspond to protein complexes in which proteins interact within a specific location/time/condition to perform a function in a coordinated manner (e.g., RNA splicing machinery and transcription machinery), or to form dynamic, yet functionally coherent units, in which the proteins interact with one another at different times/conditions to carry out a biological process (e.g., signalling pathways and cell cycle regulation) (Spirin and Mirny, 2003, Barabási et al., 2011, Barabasi and Oltvai, 2004). The construction of disease interactomes based on disease-associated genes, their validation of using real-world clinical datasets and their functional characterization using multiple biological data modalities, led to the generation of new insights into the mechanisms underlying the diseases examined in **Chapters 2-7**.

1.5.1.5   Drug repurposing based on disease-associated drug target networks

GWA studies offer several unexplored opportunities for hypothesis-driven drug discovery and therapeutic interventions in human disorders. Although of small effect sizes, common genetic variants may help identify new drug targets, and biomarkers for improved disease screening and allow patient stratification for better matching of indications to high/low risk populations

(Okada, 2014, Breen et al., 2016). Unfortunately, there are several challenges in using genetic data to derive clinically actionable results. GWA studies tag disease-associated SNPs in specific genomic regions and do not directly implicate genes and causal alleles in disease aetiology (Breen et al., 2016). It is difficult to pinpoint causal alleles due to incomplete expression, protein and methylation data and the lack of large-scale brain sampling (from diverse spatiotemporal points) (Breen et al., 2016). Additionally, it is difficult to pinpoint specific drug targets within the polygenic architecture of complex disorders (Breen et al., 2016). However, the network medicine paradigm can be effectively used for clinical translation of genetic data. Drug targets are three times more likely to be found among the direct interactors of the proteins encoded by the genes in GWA loci than in the loci themselves (Cao and Moult, 2014). Hence, interactomes of variant-harbouring genes can be examined to identify novel drug targets or reposition existing drugs for novel therapies. A generalizable framework for drug identification in complex disorders can combine the drug information obtained from the interactome with the widely applied comparative transcriptomics approach employing differential (gene) expression signatures (DES) (Dudley et al., 2011, Athauda et al., 2017, Kinnings et al., 2009, Evans et al., 2005). Unique DES often underlie disease mechanisms (Sirota et al., 2011, Maertzdorf et al., 2012, Chaussabel et al., 2005, Kumar et al., 2014, Chiu

et al., 2013). Drugs administered to treat these diseases revert the expression of these genes to their normal levels (Duran-Frigola et al., 2017, Pushpakom et al., 2019). DES for patients with the disease (compared to healthy subjects) are quantified using gene expression analysis based on microarrays and RNA sequencing methods. Online repositories such as the Gene Expression Omnibus (GEO) contain DES datasets, and make them freely available for integrated computational analyses (Barrett et al., 2012). Similarly, DES for drug-treated versus untreated cell lines is made available through Connectivity Map (CMAP) (Lamb et al., 2006). The transcriptomics approach involves comparing the differential gene expression profiles induced by drugs targeting the proteins in a disease module with the expression profiles of the
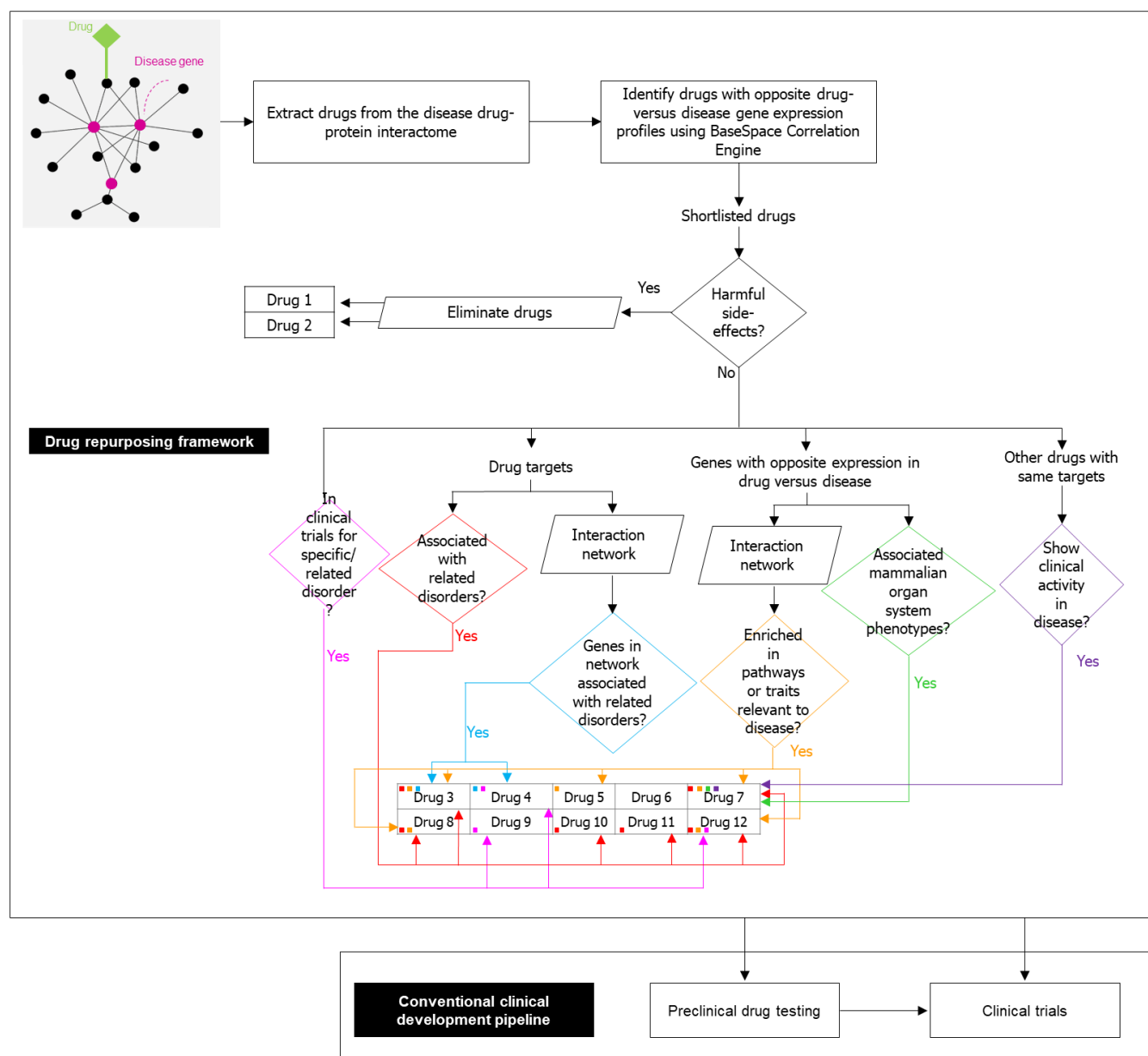
Figure 12: The proposed methodology for repurposing drugs. Drugs can be extracted from the drug-protein interactome and screened for negative correlation of drug-induced versus disease-associated gene expression profiles. Drugs shortlisted in this manner should be checked for their toxicity, and eliminated if they are found to have harmful side effects. Subsequently, the drugs can be subjected to a series of network-based analyses to assess their biological validity, as shown in the figure. Lines of different colours indicate the different pieces of evidence supporting the biological validity of the drugs. Each drug is marked by coloured boxes that correspond to specific pieces of evidence. Drugs that get tagged with several pieces of supporting information can be directly tested in clinical trials, or subjected to additional examinations in preclinical models. The figure was created in Microsoft PowerPoint.

patients (compared to controls). This procedure will help assess the real-world clinical activity of the drugs in the disease. Drug repurposing based on the negative correlation of drug-induced versus disease-associated gene expression profiles has resulted in some valuable results in the past (Dudley et al., 2011, Athauda et al., 2017, Kinnings et al., 2009, Evans et al., 2005). Topiramate, an anti-convulsant drug used to treat epilepsy was identified to be potentially repurposable for inflammatory bowel disease based on the negative correlation of drug-induced profiles extracted from CMAP (Lamb et al., 2006) and disease-associated profile from GEO (Barrett et al., 2012, Dudley et al., 2011). They further demonstrated its efficacy in a rodent model of inflammatory bowel disease (Dudley et al., 2011). In **Chapter 8**, FDA-approved repurposable drugs were identified for SCZ using comparative transcriptomic analysis of drug-induced and SCZ-associated gene expression profiles. Their biological validity was assessed using clinical trial data, network-based analyses and literature review. This work led to the testing of cromoglycate as adjunctive therapy for SCZ (Nimgaonkar, 2019) and acetazolamide for treatment-resistant SCZ (Nimgaonkar, 2022). In **Chapter 4**, **Chapter 5**, **Chapter 6** and **Chapter 7**, repurposable drugs for other disorders were identified using the approach described above. **Fig. 12** shows the proposed framework to identify repurposable drugs.

1.5.1.6   Drug testing in animal models

It is often difficult to directly translate the repurposable drugs identified using the methodology described in the previous section to clinical trials. In this case, further preclinical studies in animal models may be necessary to test the efficacy of the drug and the mechanisms underlying its pharmacological actions. **Appendix section 14.2** describes drug testing against a complex neurological disorder. Accordingly, the study describes the use of enzyme histochemistry to visualise enzyme activity, in vitro assays to quantify enzyme and substrate activity, bioinformatics methods to model drug-enzyme binding kinetics, and experimental paradigms to measure the cognitive variable (memory acquisition and retention) and toxicity (dysregulation of motor activity). Analogous variables could be examined for other complex disorders, i.e., visualising and quantifying drug target activity, modelling the binding kinetics of the drug with the target protein, and measuring quantifiable phenotypes indicative of therapeutic benefit and adverse events. Therefore, the drugs that target the interactome of a disorder, if identified as repurposable and not directly transferable to the clinical development pipeline, could be tested in preclinical animal model studies using the framework illustrated in **Appendix section 14.2**.

Altogether, the integration of the disease interactome with drugs, drug repurposing based on comparative transcriptome analysis and preclinical studies in animal models led to the clinically translatable results described in **Chapters 4-9**.

Below, a brief description of the framework for the analysis of multiple interactomes (see **Fig. 9**), as implemented in **Chapter 9** and **Chapter 10**, is provided. The sub-sections that follow describe the constituent steps of the framework in detail.

1.5.2 The network biology-based framework for examining multiple interactomes

To understand the higher-order relationships among disorders (e.g., subtypes of a psychiatric disorder), disease interactomes were constructed using the linkage method and then intersected with context-specific data (e.g., region-specificity) following the 'disease module-based method' for interactome construction. The relationships of the resultant context-specific disease sub-networks, and the influence of context-specific data on these relationships, were examined using various clustering and dimensionality reduction methods. To understand the higher-order relationships governing clinical activity in multiple disorders (e.g., presence/absence of drug contraindications in comorbid diseases), DTNs were generated, i.e., the proteins targeted by drugs (indicated for specific diseases) and their first-order interactors. The DTNs were refined by intersecting them with disease interactomes and tissue-specific and pathway data. Clustering and dimensionality reduction methods were employed to plot the relationships between the DTNs (producing varying clinical activity in multiple disorders) and assess the influence of specific disease sub-networks, tissues, and pathways on these relationships.

The methodology to elucidate the biological themes underlying the aetiology of multiple disorders and their complex clinical relationships has been demonstrated in **Chapter 9** and **Chapter 10**.

1.5.2.1 Identification of context-specific disease interactomes

The analyses described in section 1.5.1 aimed to discover and functionally characterise single disease interactomes. Hence, the interactomes were constructed using the linkage-based method, i.e., retrieving the pairwise linkage information between the proteins encoded by disease-associated genes and their first-order direct interactors in the human interactomes. However, while comparing multiple disease interactomes to elucidate common disease mechanisms and points of aetiological differentiation, it is necessary to identify the specific sub-networks of the disease interactomes that are biologically active. Therefore, disease interactome construction and refinement using specific biological modalities is the first step in investigating the higher-order relationships of any group of disorders using the interactomic framework. The disease module-based method will help refine the disease interactome into its biologically active sub-network. This method assumes that genes in the same topological or functional module are highly likely to be involved in the same disease (Barabási et al., 2011). Therefore, after disease interactomes are constructed using the linkage-based method, these interactomes could be intersected with a tissue- or cell line-specific interactome, resulting in context-specific sub-networks, whose biological validity could further be ascertained by examining their expression patterns or enrichment for specific functions (Barabási et al., 2011). **Fig. 13** shows the proposed methodology – derived based on **Chapter 9** – to examine the unifying and differentiating biological themes underlying related disorders or subtypes of the same disorder.

1.5.2.2 Comparative disease interactome analysis

For comparing the disease interactomes, their enrichment in genomic and proteomic datasets – extracted from various levels of spatiotemporal diversity and granularity (single cell, tissue, organism) – could be examined. For example, large-scale analysis incorporating disease

interactomes and tissue-specific transcriptomes may reveal previously hidden associating or disassociating patterns of regional specificities among the disorders. Hierarchical clustering and principal component analysis (PCA) will help understand the grouping patterns of the diseases. These algorithms will delineate groups based on the expression variations of the interactome genes within the tissue, e.g., at the level of histochemically segregated compartments and cell type. Component loading values in PCA will help infer the influence of specific tissue compartments and cell types on the grouping patterns. The hierarchical clustering and PCA methods can be implemented as follows.

PCA can be used to capture relationships between the disease interactomes. The negative log-transformed p-values indicating the statistical enrichment of various biological modalities in the interactomes can be assembled into a data matrix containing the modalities as rows and the disorders as columns; each cell in the matrix contained a $-\log_{10}P$ value. As seen in previous studies, $-\log_{10}$ transformed p-values can be used as inputs for PCA (Chang and Keinan, 2014, McGuirl et al., 2020). The data matrix can be pre-processed to include only those rows and columns that contained less than 70% missing values. The $-\log_{10}P$ values in the matrix can be centred using the unit variance scaling method, in which the values are
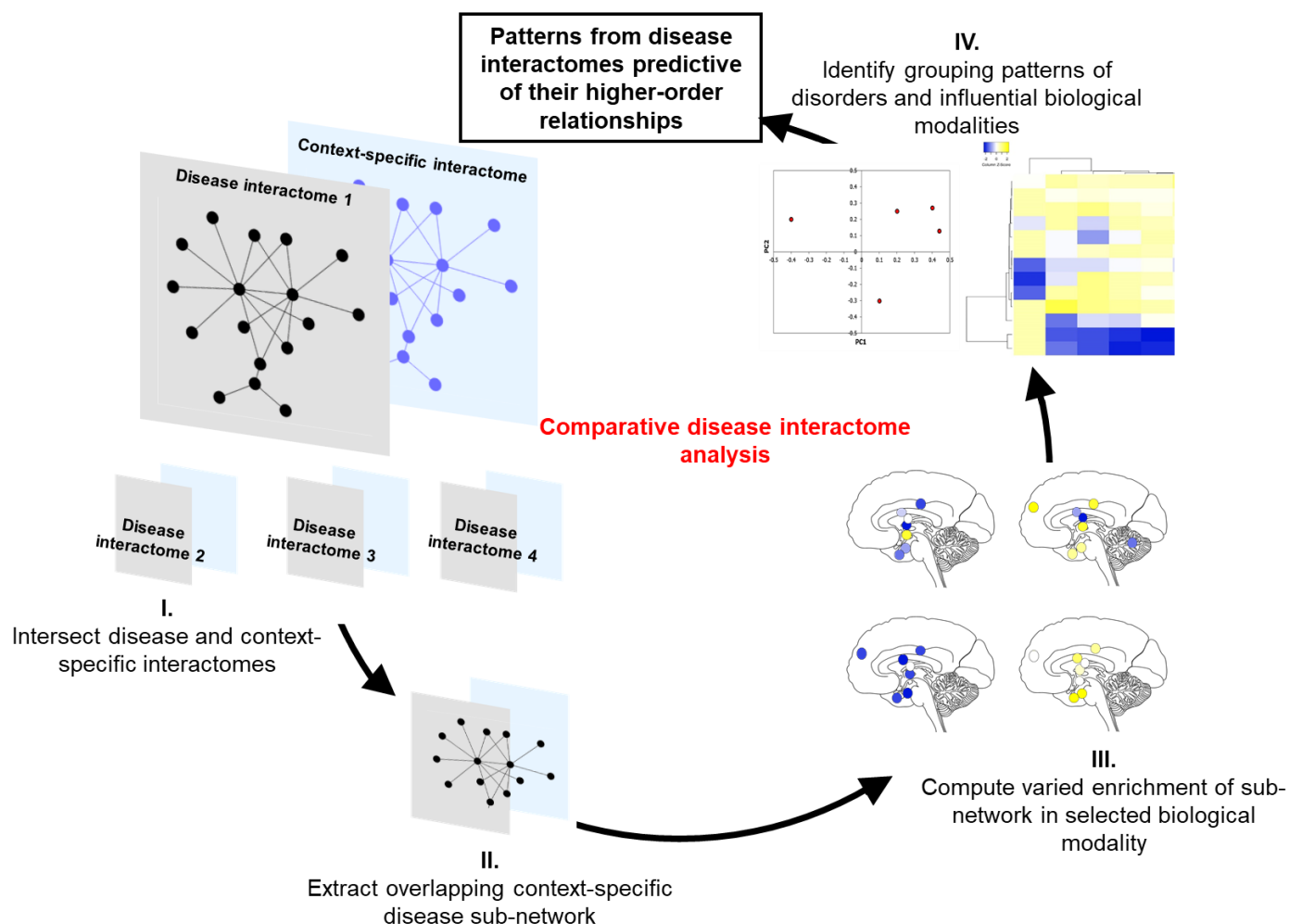
Figure 13: The proposed methodology to elucidate the higher-order relationships between disorders. First, the disease interactomes should be intersected with context-specific (such as cell line- or tissue-specific) interactomes. Then, the overlapping sub-networks can be extracted. The differential enrichment of the various disease sub-networks in a specific biological modality can be computed. The metrics from this analysis can be used to identify the grouping patterns of the diseases and pinpoint the specific biological modalities that influenced these patterns. Network diagrams were created using Cytoscape, the heatmap using the web tool Clustvis and the PC plot using Microsoft Excel. The figure was assembled in Microsoft PowerPoint.

divided by standard deviation so that each row or column has a variance of one; this ensures that they assume equal importance while finding the components. The method called singular value decomposition (SVD) with imputation can be used to extract principal components. In this method, missing values are predicted and iteratively filled using neighbouring values during SVD computation, until the estimates of missing values converge. The number of principal components computed will be equal to the number of column dimensions in the data matrix, i.e., the number of disorder interactomes. PCA essentially transforms the original variables ($-\log_{10}P$) into uncorrelated variables called principal components. These principal components can be ranked in the descending order of the percentage of total variance explained by them. The first two components that (generally) cumulatively explain the maximum percentage of variance seen in the enrichment patterns, i.e., PC1 and PC2, can be selected. The component scores of each disorder can be plotted on a 2D plane to capture the angle of highest variability and delineate grouping patterns of the disorders based on approximated distances between the scores. The positions of each observation in the PC plot are called component scores and are calculated as linear combinations of the original variables and the corresponding weights a_{ij} (also known as loading values). For example, the score for the rth sample on the kth principal component is calculated as

$$Y_{rk} = a_{1k}x_{r1} + a_{2k}x_{r2} + \cdots + a_{pk}x_{rp}$$

The importance of each biological modality is reflected by the magnitude of their corresponding loading values on the principal components (PC1 and PC2), and these values can be used to identify the modalities that were most influential in producing the grouping patterns seen in the PC plot.

The data matrix of biological modalities (rows) and disease interactomes (columns) can be subjected to hierarchical clustering to check whether the grouping patterns observed in the PC plot are valid. Pairwise distances in the data matrix can be calculated using Pearson correlation and the data points can be grouped using the average linkage method, i.e., the dendrograms can be generated by merging the data points with the smallest distance first, and those with larger distances later. In the average linkage method, the average distance of all possible pairs is considered while clustering.

The approach adopted in **Chapter 9** can be applied across related disorders or disorder subtypes to delineate underlying biological themes.

1.5.2.3   Identification of context-specific drug target networks

Adverse events precipitated by drugs in individual diseases have been investigated within the framework of the PPI network (Mizutani et al., 2012, Fliri et al., 2005, Wang et al., 2013, Campillos et al., 2008, Brouwers et al., 2011, Hase et al., 2009). However, the effects of multiple drugs and their contraindications on comorbid conditions remain largely unexplored. In **Chapter 10**, the mechanisms underlying drug contraindications in pairs of comorbid diseases were elucidated by examining DTNs. **Fig. 14** shows the proposed methodology – derived based on **Chapter 10** – to examine the target networks of drugs to elucidate patterns predictive of their clinical activity. The Drug Bank database (Wishart et al., 2008) can be used to compile the lists of drugs indicated for the diseases. The proteins targeted by the drugs can be obtained by querying the DGIdb (drug gene interaction database) web portal (Griffith et al., 2013). To construct the DTNs, the PPIs of the drug targets can be compiled from HPRD (Keshava Prasad et al., 2009) and BioGRID (Stark et al., 2006). To refine the DTNs, 3 types of data that reflect their biological profiles can be examined, namely (i) disease PPI networks, (ii) biological pathways and (iii) tissue gene expression. Specifically, gene overrepresentation analyses based on hypergeometric distribution can be conducted to check the enrichment of the DTNs among disease interactomes, genes showing high/moderate expression in 53 tissues across the human body (Consortium, 2015), and proteins involved in ~1000 biological pathways (Croft et al., 2014). Overlaps computed in this manner were considered to be
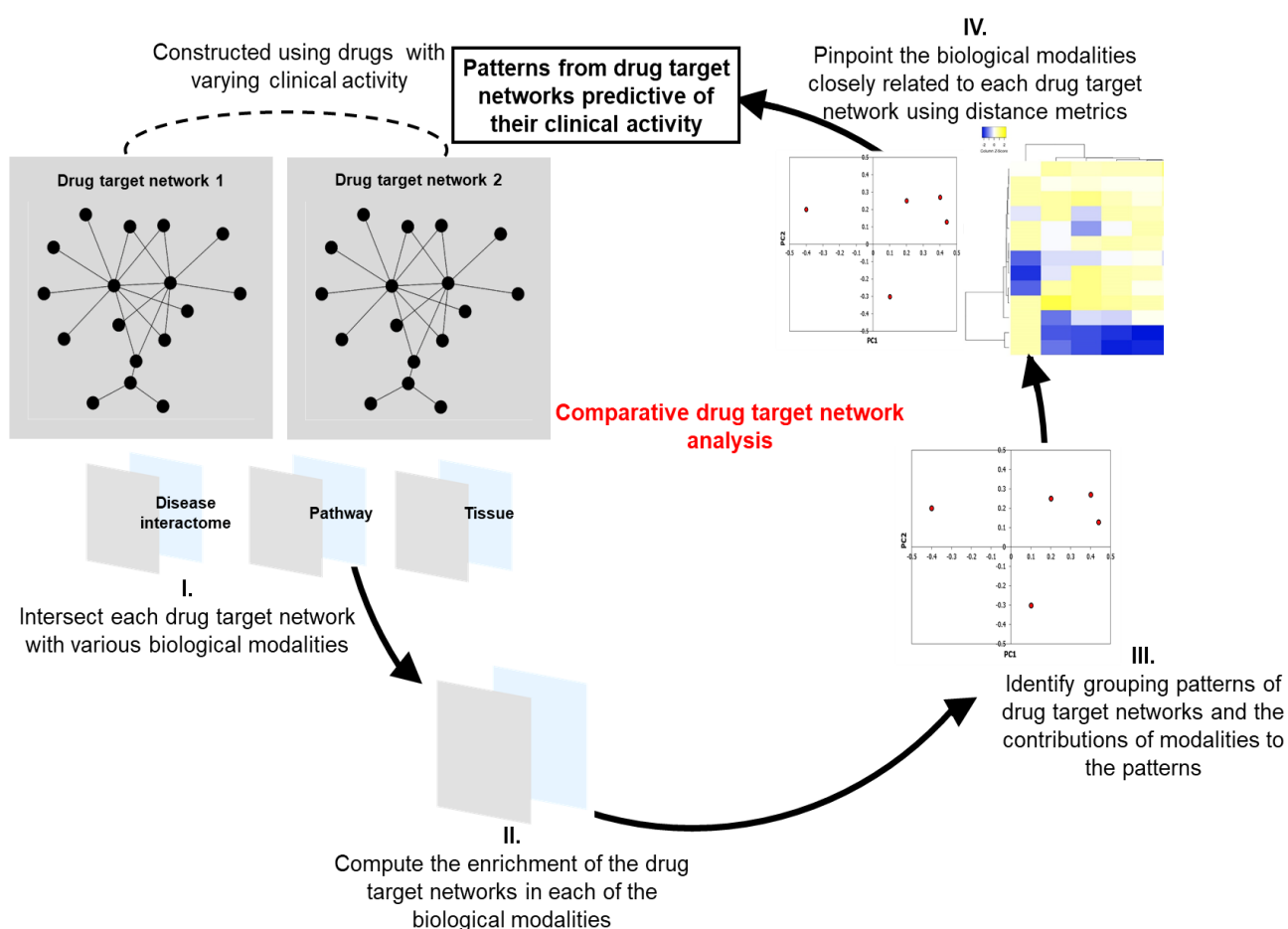
Figure 14: The proposed methodology to elucidate the higher-order relationships between drug target networks (DTNs). First, the DTNs should be intersected with various biological modalities (e.g., the disease interactome, pathways, and tissues). The enrichment of the DTNs for each of these modalities should be computed. The grouping patterns of the DTNs should be identified based on their differential enrichment metrics. The relative influence of specific biological modalities on each of the DTNs (with varying clinical activity) can be identified using an additional set of analyses involving Euclidean distance metrics. Network diagrams were created using Cytoscape, the heatmap using the web tool Clustvis and the PC plot using Microsoft Excel. The figure was assembled in Microsoft PowerPoint.

statistically significant at p-value < 0.05 after correction for multiple hypotheses using the Benjamini-Hochberg method. The sub-networks obtained by intersecting the DTNs with each of the biological modalities were used for further analysis.

1.5.2.4   Comparative drug target network analysis

PCA was used to compare the DTNs and identify their grouping patterns. PCA was performed with a data matrix containing the DTNs (columns) versus specific disease protein sets (from the disease interactome), pathways or tissues (rows). For example, for the data modality 'disease protein set', the rows would be 'common to both the networks', 'unique to disease A network' and 'unique to disease B network' and for the data modality 'tissue', the members would be 'amygdala', 'aorta', 'lungs', etc. Each cell contained $-\log_{10}$ transformed p-values, which have been used as inputs for PCA in previous studies (Chang and Keinan, 2014, McGuirl et al., 2020). All the PCs generated after this analysis were considered for our study, and the PC scores of the DTNs were used to identify their grouping patterns. Following this, the component loading values denoting the weights of each of the biological modalities on the PCs were extracted. Lastly, specific disease protein sets, pathways and tissues that were more closely related to each of the 4 DTNs were isolated in terms of Euclidean distance. Specifically, the Euclidean distance between the PC scores of each of the DTNs and the corresponding component loadings of the biological modalities was calculated. In summary, the framework helps examine the biological modalities influencing specific clinical events (e.g., adverse events as shown in **Chapter 10**), and can inform rational drug development strategies in complex disorders with aetiological and clinical associations.

Overall, the thesis proposes a comprehensive methodology for extensively exploring individual disease interactomes and drug target networks and their interactions at higher levels encompassing multiple networks. This pipeline addresses the significant limitations observed in current network biology studies. Its primary importance lies in its potential for incorporation into routine examinations of the multigenic origins of diseases and the polypharmacological aspects of drug action.

Interactome-based framework to translate disease genetic data into biological and clinical insights

2. Cilia interactome with predicted protein–protein interactions reveals connections to Alzheimer's disease, ageing and other neuropsychiatric processes

The experimental chapter is based on the following peer-reviewed publication:
Karunakaran, Kalyani B., Srilakshmi Chaparala, Cecilia W. Lo, and Madhavi K. Ganapathiraju. Cilia interactome with predicted protein–protein interactions reveals connections to Alzheimer's disease, ageing and other neuropsychiatric processes. Scientific reports 10, no. 1 (2020): 1-16.

Summary of this chapter

Neurological and neuropsychiatric phenotypes often co-occur with ciliopathies. The crucial role of the cilium in neuronal signalling, neurogenesis and neuronal migration indicated that a systems-level investigation of how ciliary proteins function together may provide insights into their contribution towards nervous system phenotypes. In this chapter, I demonstrate how the interactomic framework was used to derive biological insights into the interlinked biology of cilia and brain disorders. I constructed the cilia interactome using the experimentally identified and computationally predicted protein-protein interactions (PPIs) of primary and motile ciliary proteins. Overlap with genetic and transcriptomic datasets and functional interactions ascertained its biological validity. My analysis of the interactomes of seven neuropsychiatric and neurological disorders generated using variant-harbouring disease-associated genes revealed that they shared statistically significant overlaps with the cilia interactome, which additionally overlapped with the genes differentially expressed in patients affected with six of these disorders. The interactome was significantly associated with cellular pathways related to neuropsychiatric processes and contained proteins targeted by ~100 drugs used to treat nervous system disorders. As a case study, I identified an interactome sub-network in which Alzheimer's disease (AD)-associated, ciliary and ageing-related proteins intersected. Its integration with multi-omics data revealed the potential role played by ciliary sonic hedgehog signalling in hippocampal neurogenesis and memory deficits in AD. Finally, I generated testable hypotheses on novel PPIs involved in neuropsychiatric disorders, primary ciliary dyskinesia, hydrocephalus, ciliogenesis and ciliary membrane receptor trafficking. Altogether, the interactomic framework suggested that ciliary defects underlay neuropsychiatric processes.

Contribution to this chapter (60%)

- Conceptualised, performed and derived the conclusions for cilia interactome validation, functional enrichment and intersection analyses
- Conceptualised, performed and derived the conclusions for the study of the ciliary association of Alzheimer's disease and ageing
- Generated all the testable hypotheses for novel PPIs

Interactome-based framework to translate disease genetic data into biological and clinical insights

- Wrote and edited the manuscript and prepared Fig. 1, Fig. 3, Fig, 4, Table 1, Table 3 and Supplementary File 5

SCIENTIFIC REPORTS
natureresearch

Check for updates

OPEN

Cilia interactome with predicted protein–protein interactions reveals connections to Alzheimer's disease, aging and other neuropsychiatric processes

Kalyani B. Karunakaran[1,6], Srilakshmi Chaparala[2,3,6], Cecilia W. Lo[4] & Madhavi K. Ganapathiraju[2,5 ✉]

Cilia are dynamic microtubule-based organelles present on the surface of many eukaryotic cell types and can be motile or non-motile primary cilia. Cilia defects underlie a growing list of human disorders, collectively called ciliopathies, with overlapping phenotypes such as developmental delays and cognitive and memory deficits. Consistent with this, cilia play an important role in brain development, particularly in neurogenesis and neuronal migration. These findings suggest that a deeper systems-level understanding of how ciliary proteins function together may provide new mechanistic insights into the molecular etiologies of nervous system defects. Towards this end, we performed a protein–protein interaction (PPI) network analysis of known intraflagellar transport, BBSome, transition zone, ciliary membrane and motile cilia proteins. Known PPIs of ciliary proteins were assembled from online databases. Novel PPIs were predicted for each ciliary protein using a computational method we developed, called High-precision PPI Prediction (HiPPIP) model. The resulting cilia "interactome" consists of 165 ciliary proteins, 1,011 known PPIs, and 765 novel PPIs. The cilia interactome revealed interconnections between ciliary proteins, and their relation to several pathways related to neuropsychiatric processes, and to drug targets. Approximately 184 genes in the cilia interactome are targeted by 548 currently approved drugs, of which 103 are used to treat various diseases of nervous system origin. Taken together, the cilia interactome presented here provides novel insights into the relationship between ciliary protein dysfunction and neuropsychiatric disorders, for e.g. interconnections of Alzheimer's disease, aging and cilia genes. These results provide the framework for the rational design of new therapeutic agents for treatment of ciliopathies and neuropsychiatric disorders.

Abbreviations
PPI   Protein–protein interaction
GO   Gene ontology
HiPPIP High-confidence protein–protein interaction prediction model

Cilia are dynamic organelles projecting from the surface of many types of eukaryotic cells. They detect changes in the extracellular environment and transduce signals into the cell to regulate a wide variety of physiological and developmental processes. They can be either motile or non-motile, and exhibit a microtubule organization

[1]Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, India. [2]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA. [3]Health Sciences Library System, University of Pittsburgh, Pittsburgh, PA, USA. [4]Department of Developmental Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [5]Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA. [6]These authors contributed equally: Kalyani B. Karunakaran and Srilakshmi Chaparala. ✉email:

of 9 + 2 or 9 + 0, respectively[1]. Primary cilia are sensory organelles modulating several core signaling and cellular polarity pathways that are fundamental for tissue homeostasis and embryonic development[2]. Motile cilia drive the flow of bodily fluids including mucus and cerebrospinal fluid[3,4]. Defects involving the primary cilia are observed in various human ciliopathies such as Bardet-Biedl syndrome (BBS), Joubert syndrome and Meckel–Gruber syndrome. Motile cilia defects are seen in primary ciliary dyskinesia (PCD), male infertility and laterality defects[5]. The cilium is a complex organelle comprising over 600 proteins[1]. Underscoring their functional importance, many of these ciliary proteins are highly evolutionarily conserved including the intraflagellar transport (IFT) complexes located within the axoneme involved in bidirectional protein transport between the ciliary base and the tip, complexes localizing to the transition zone (TZ) at the ciliary base acting as a 'ciliary gate' regulating protein trafficking into and out of the cilia and BBSomes mediating cilia assembly[6–8].

Primary cilium is increasingly viewed as a hub for neuronal signalling. A large body of evidence has emerged demonstrating the role of cilia in the development and function of the central nervous system (CNS)[9–11]. Gene knockdown of BBS proteins such as BBS1, BBS4-5, BBS7, and BBS9-12 lead to cortical defects and improper neuronal migration, highlighting the significance of cilia genes in brain development[12]. Additionally, neural tube defects are observed in the brain with the disruption of cilia-transduced sonic hedgehog signaling (Shh) and Wnt signaling[13,14]. Indeed, many ciliopathies are known to be associated with neurological deficits such as developmental delays, cognitive impairment and neuropsychiatric disorders including ataxia, autism spectrum disorders and schizophrenia[10,12]. Importantly, the ciliary proteins AHI1, ARL13b, CDKL5 and EFHC1 have been implicated in autism spectrum disorder, epilepsy, and schizophrenia[15–18]. A recent study identified neuropsychiatric risk genes (NEK4, SDCCAG8, FEZ1, CEP63, PDE4B and SYNE1) to be linked to cilia assembly and function[15]. In addition, several ciliary proteins interact with proteins that are known to play a role in neuropsychiatric disorders: PCM1, BBS4 with DISC1 in schizophrenia, bipolar disorder and depression[19,20], KIF3A, PCNT with DCDC2 in dyslexia[21,22], and PCM, AHI1 with HTT in Huntington disease[12,23,24]. Hydrocephalus, a phenotype observed frequently in BBS and other ciliopathies, may reflect the role of motile cilia in the flow of cerebrospinal fluid in the brain[10]. Ciliopathies have also been associated with obesity, suggesting a role for cilia in the neural circuitry responsible for monitoring food intake and satiety[25]. The obesity-related genes MC4R and ADCY3 co-localize to primary cilia of hypothalamic neurons, and impairing this localization or blocking their signalling in primary cilia led to gain in body weight in mice[26].

Given the importance of large multi-protein complexes in its assembly and function, knowledge of the protein–protein interactions (PPIs) of ciliary proteins would help to elucidate the potential role of cilia biology in neuropsychiatric diseases. Studies based on PPI networks have significantly advanced our knowledge of specific proteins or the diseases that they are associated with, such as DISC1 in schizophrenia, or the NPHP-JBTS-MKS protein complex in ciliopathies[27]. DISC1 was a novel protein with well-characterized domains but of unknown function with no known human homolog, when it was identified as being associated with schizophrenia[28,29]. To understand the function of DISC1, its PPIs were determined using yeast 2-hybrid technology[30,31]. This led to a large number of studies, which connected DISC1 to cAMP signaling, axon elongation and neuronal migration. A study revealed that the role played by DISC1 in dopamine signaling, which is implicated in schizophrenia, may also involve primary cilia on neurons[19]. DISC1 localized to primary cilia on rat striatal neurons and was found to be involved in the formation and maintenance of cilia with certain dopamine receptors[19]. The PPI network of ciliary proteins CEP290 and RPGR revealed their connection to photoreceptors, and disruption of this network has been shown to cause blindness on rapid degeneration of photoreceptors, a finding associated with several ciliopathies[32].

Large-scale proteomic and protein interactome analyses have significantly advanced our understanding of its role in developmental biology and disease[33–37]. Multidimensional protein identification technology (MudPIT) was used to identify 195 candidate primary cilia proteins localizing to sensory cilia, or linked to known ciliopathies[33]. 850 interactors of nine NPHP/JBTS/MKS proteins (i.e. Nephronophthisis/Joubert/Meckel-Gruber syndromes) were identified using the G-LAP-Flp purification strategy, and several cilia-specific modules, namely 'NPHP1-4-8' functioning at the apical surface, 'NPHP5-6' at centrosomes and 'MKS' linked to hedgehog signaling were uncovered[34]. In another study, in vivo proximity-dependent biotinylation (BioID) was used to identify more than 7,000 interactions of 58 centriole, satellite and ciliary transition zone proteins, which revealed protein modules involved in cilia and centrosome biogenesis[35]. The interactome of CPLANE (ciliogenesis and planar polarity effector) proteins, namely that of *Inturned* (INTU), *Fuzzy* (FUZ) and *Wdpcp* (WDPCP), consisting of ~ 250 interactions, was identified using LAP-tagged immunoprecipitation, and it was shown that the CPLANE proteins govern IFT-A/B trafficking[36]. Systematic tandem affinity purifications coupled to mass spectrometry was employed to identify 4,905 interactions and 52 complexes for 217 proteins with known or suspected involvement in ciliary function or disease, and this study linked vesicle transport, the cytoskeleton and ubiquitination to ciliary signaling and proteostasis[37]. None of these experimental methods are single handedly capable of identifying all the possible interactions of ciliary genes. In fact, it is the ability of an experimental method to discover interactions not detected by another method that makes it truly valuable. Machine learning methods can computationally predict new interactions that other high throughput detection methods may fail to capture and serve as hypotheses-generation methods that may be validated by other experimental methods. Here, we applied computational method that we developed previously to discover novel PPIs of 165 ciliary proteins and analyzed the resulting ciliary PPI interactome for novel associations and potential connections to neuropsychiatric diseases.

Experimental procedures

Dataset. Compilation of Cilia Gene List: We obtained a list of 165 cilia genes that were curated from literature by prioritizing the genes based on their association with cilia from Dr. Gregory Pazour's lab building upon their prior work[38]. This list includes IFT proteins, BBS proteins, TZ proteins, ciliary membrane proteins,

and proteins restricted to motile cilia. Known PPIs were collected from Human Protein Reference Database (HPRD)[39] and Biological General Repository for Interaction Datasets (BioGRID)[40]. Gene-drug associations and ATC classifications were collected from DrugBank[41], while neuropsychiatric gene-disease associations were collected from the GWAS catalog (www.ebi.ac.uk/gwas/). Random gene sets used in shortest path comparisons were sampled from about twenty thousand human proteins listed in the Ensembl database (www.ensembl.org). Novel PPIs were predicted using the HiPPIP model that we developed[42]. Each ciliary protein (say C_1) was paired with each of the other human genes say, $(G_1, G_2, … G_n)$, and each pair was evaluated with the HiPPIP model. The predicted interactions of each of the cilia genes were extracted, which resulted in 620 newly dis- covered PPIs of cilia genes. The average shortest path distance was computed using the Networkx package in python. Pathway associations were computed using Ingenuity Pathway Analysis suite. GO term enrichment was carried out using BinGO[43]; for each C_1, a list of its known and predicted interacting partners (i.e. $B_1, B_2, … B_n$) are given as input to BinGO, which extracts the GO terms of all these genes and finds which of the GO terms are statistically enriched in comparison to the background distribution of GO terms of all human proteins. All statistically significant terms are assigned as network-based enriched GO terms of C_1.

Gene expression datasets in Gene Expression Omnibus were used to compute the overlap of the cilia interactome with genes differentially expressed in various neuropsychiatric disorders: major depressive disorder (GSE53987[44]), schizophrenia (GSE17612[45]), bipolar disorder (GSE12679[46]), autism spectrum disorder (GSE18123[47]), Alzheimer's disease (GSE29378[48] and GSE28146[49]), Parkinson's disease (GSE28894) and non-syndromic intellectual disability (GSE39326[50]). Genes with fold change > 2 or < ½ were considered as significantly overexpressed and underexpressed respectively at p value < 0.05. A gene with transcripts per million ≥ 2 was considered to be 'expressed' while analyzing the overlap of the interactome with genes expressed in the amygdala, anterior cingulate cortex, caudate, cerebellum, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord and substantia nigra extracted from GTEx[51]. Time-dependent gene expression variation in the hippocampal region was extracted from BrainSpan Atlas containing RNA-Seq data from post-conceptional weeks to middle adulthood[52]. 78 genes associated with Alzheimer's disease were extracted from DisGeNET[53] (with score > 0.2 to include only expert-curated disease-gene associations). Then, to construct the Alzheimer's disease interactome, whose overlap was to be checked with the cilia interactome, 4,742 known PPIs extracted from HPRD[54] and BioGRID[55], and 490 computationally predicted PPIs of these 78 genes were assembled. The biological validity of the interactome was shown by the fact that 676 genes out of the 3,944 genes in the AD interactome are differentially expressed in CA1 hippocampal gray matter from patients with severe Alzheimer's disease versus healthy controls (GSE28146[49]), out of which 71 were novel interactors (p value = 1.138e−20).

Results

We assembled a list of 165 genes encoding proteins known to be associated with primary and/or motile cilia, including IFT, BBS, TZ, and ciliary membrane proteins, as well as proteins restricted to motile cilia. Known PPIs of ciliary proteins were assembled from HPRD and BioGRID[40,56]. Novel PPIs were predicted for each of the cilia genes using our High-precision Protein–Protein Interaction Prediction (HiPPIP) model[42]. In this manner, a ciliary protein interactome was assembled comprising 165 ciliary proteins (red color square shaped nodes) with 1,011 known PPIs (blue edges) and 765 novel PPIs (red edges) that connect to 800 previously known interactors (light blue nodes) and 705 novel interactors (red nodes) (Fig. 1 and Table 1). We predicted 216 new interactions for 50 out of the 56 cilia genes that had no known PPIs. For example, GPR83 has 12 novel PPIs, LRRC48 has 10, PKD1L1 has 10, and SPEF has 10 novel PPIs. The number of known and novel PPIs of cilia genes are given in Supplementary File 1, and the lists of all genes and PPIs is given in Supplementary File 2.

For each of the ciliary proteins, we computed the enrichment of gene ontology (GO) terms among its interacting partners in order to aid in the discovery of its function using BinGO (Biological Networks Gene Ontology tool)[43]. This information is especially useful for those ciliary proteins that have either no known or very few known GO biological process terms. For example, there are 11 genes that have no known GO terms, and we predicted new GO terms for each of those genes, for e.g. 27 novel GO terms for ARMC4, 11 for CCDC63, and 30 for DNAAF2.

We computed the pathway associations of genes in the interactome, using the Ingenuity Pathway Analysis (IPA) suite (Ingenuity Systems, www.ingenuity.com). This showed a significant overlap of neuronal pathways with the cilia interactome (see selected pathways in Table 2). The complete list of all pathways, their p values and the genes from the interactome that are associated with these pathways, are given in Supplementary File 3. We also extracted information about drugs targeting the genes in the interactome. This analysis showed that there are several genes that are targets to drugs belonging to the Anatomic category of "nervous system", highlighting the connection between cilia and the nervous system as shown in Fig. 2 and Supplementary File 4.

Experimental Validation of novel cilia PPIs in independent studies. Four of the novels PPIs that we predicted for cilia genes were independently recovered by other groups. TMEM237-SFXN5 and DYNLL2-c17orf47 were recovered by yeast two-hybrid experiments in the recent release of the human protein interactome map[57]. We also predicted two PPIs of IFT140 that were discovered as part of the CPLANE interactome using affinity purification-mass spectrometry, but were not deposited in BioGRID or HPRD: IFT140-TELO2 and IFT140-TRAP1[36]. It is also worth noting that 8 novel interactors in the interactome appeared among the proteins isolated from the primary cilia of mouse kidney cells using a method called Mud-PIT (multidimensional protein identification technology)[33]: ABCE1, CCDC47, CCT5, G3BP1, GBF1, RAB10, RAN and USP14. 94 genes in the cilia interactome, including 44 cilia genes, 36 known and 14 novel interactors, were also recovered as regulators of the ciliary sonic hedgehog pathway in a CRISPR genetic screen (p
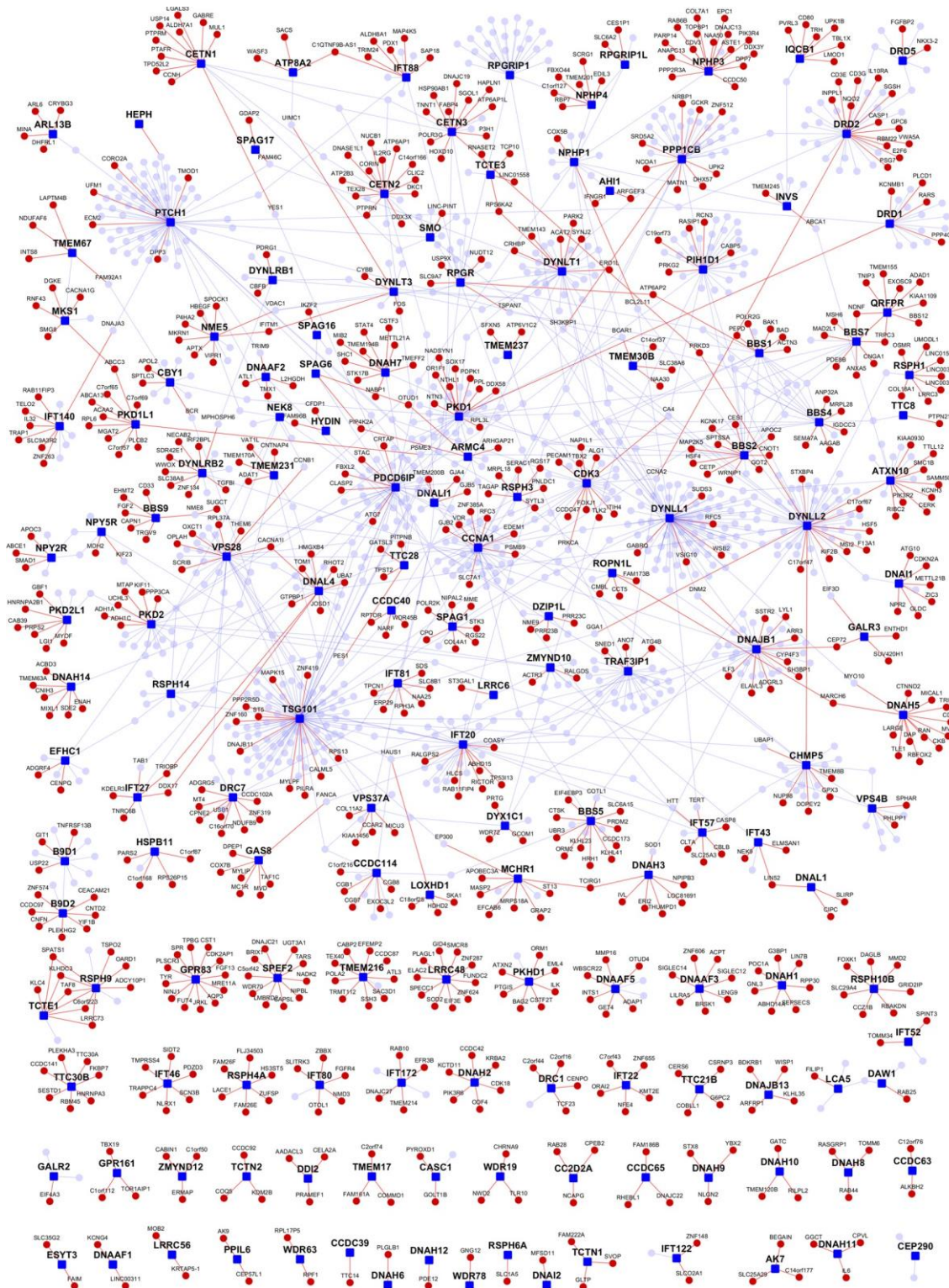
Figure 1. Cilia interactome. Cilia genes are shown as small dark-blue colored nodes and interactor genes are larger round nodes; the interactors are colored in light blue if they are previously known interactors and in red if they are found only through novel PPIs. PPIs are shown as edges, where blue color edges are known PPIs and red color edges are novel predicted PPIs. Most genes at the bottom of the figure have had zero known PPIs, and have multiple novel predicted PPIs.

value = 2.28e−19)[58]. The interactome was also significantly enriched with genes differentially expressed in bronchial biopsies of primary ciliary dyskinesia patients (p value = 2.64e−02)[59].

Cilia gene	K	N	Novel interactors
AHI1	1	2	ARFGEF3, IFNGR1
AK7	0	3	BEGAIN, C14orf177, SLC25A29
ARL13B	1	4	MINA, CRYBG3, ARL6, DHFRL1
ARMC4	3	3	ARHGAP21, OTUD1, PIP4K2A
ATP8A2	1	3	C1QTNF9B-AS1, WASF3, SACS
ATXN10	9	9	TTLL12, RIBC2, SMC1B, GGA1, KCNH3, PIK3R2, KIAA0930, SAMM50, CERK
B9D1	7	3	TNFRSF13B, USP22, GIT1
B9D2	2	7	PLEKHG2, YIF1B, CCDC97, CNFN, CEACAM21, ZNF574, CNTD2
BBIP1	0	0	None
BBS1	13	6	ACTN3, BAK1, BAD, ATP6AP2, PEPD, POLR2G
BBS2	11	10	APOC2, WRNIP1, CETP, GOT2, SPTSSA, CES1, CNOT1, HSF4, MAP2K5, KCNK17
BBS4	13	5	ANP32A, AAGAB, MRPL28, SEMA7A, IGDCC3
BBS5	5	11	SLC6A15, CCDC173, KLHL41, CTSK, COTL1, EIF4EBP3, UBR3, KLHL23, ORM2, PRDM2, HRH1
BBS7	5	8	ANXA5, PDE8B, MSH6, MAD2L1, TRPC3, NDNF, QRFPR, CNGA1
BBS9	0	6	NME8, CD33, EHMT2, TRGV9, FGF2, CAPN1
CASC1	1	2	GOLT1B, PYROXD1
CBY1	8	3	APOL2, BCR, SPTLC3
CC2D2A	0	3	CPEB2, NCAPG, RAB28
CCDC114	8	4	CGB8, EXOC3L2, CGB7, CGB1
CCDC135	6	0	None
CCDC39	0	1	TTC14
CCDC40	1	3	NARF, RPTOR, WDR45B
CCDC63	0	2	ALKBH2, C12orf76
CCDC65	0	3	FAM186B, DNAJC22, RHEBL1
CCNA1	32	7	VDR, RFC3, GJB2, ZNF385A, PSMB9, SLC7A1, EDEM1
CDK3	17	10	CCDC47, CA4, ALG1, TLK2, ITIH4, PRKCA, PECAM1, NAP1L1, FOXJ1, TBX2
CEP290	2	0	None
CETN1	4	10	ALDH7A1, CCNH, GABRE, MUL1, YES1, LGALS3, PTPRM, USP14, TPD52L2, PTAFR
CETN2	9	12	ATP6AP1, C14orf166, ATP2B3, NUCB1, DKC1, PTPRN, CLIC2, DNASE1L1, TEX28, IL2RG, DDX3X, CORIN
CETN3	6	11	ATP6AP1L, HSP90AB1, HOXD10, SGOL1, FABP4, HAPLN1, TNNT1, P3H1, DNAJC19, POLR3G, PRKD3
CHMP5	14	6	NUP98, UBAP1, DYNLL2, TMEM8B, DOPEY2, GPX3
DAW1	2	1	RAB25
DDI2	0	3	AADACL3, CELA2A, PRAMEF1
DNAAF1	0	2	LINC00311, KCNG4
DNAAF2	1	4	ATL1, TRIM9, L2HGDH, TMX1
DNAAF3	0	7	ACPT, BRSK1, SIGLEC14, SIGLEC12, LENG9, LILRA5, ZNF606
DNAH1	0	7	ABHD14A, SEPSECS, LIN7B, POC1A, RPP30, GNL3, G3BP1
DNAH10	0	3	GATC, RILPL2, TMEM120B
DNAH11	0	3	CPVL, GGCT, IL6
DNAH12	0	1	PDE12
DNAH14	1	6	ACBD3, CNIH3, ENAH, TMEM63A, SDE2, MIXL1
DNAH17	0	0	None
DNAH2	0	6	CDK18, CCDC42, KCTD11, ODF4, PIK3R6, KRBA2
DNAH3	0	7	ERI2, TCIRG1, SOD1, THUMPD1, NPIPB3, LOC81691, IVL
DNAH5	0	13	CTNND2, CKB, CDH6, DAP, LARGE, MYO10, TLE1, MVP, RAN, TRIO, MARCH6, MICAL1, RBFOX2
DNAH6	0	1	PLGLB1
DNAH7	2	9	CSTF3, SHC1, STK17B, STAT4, METTL21A, TMEFF2, NABP1, TMEM194B, MIB2
DNAH8	0	3	TOMM6, RAB44, RASGRP1
DNAH9	0	3	YBX2, NLGN2, STX8
DNAI1	2	6	ATG10, CDKN2A, ZIC3, NPR2, METTL21B, GLDC
DNAI2	0	1	MFSD11
DNAJB1	28	10	CYP4F3, ARR3, CEP72, ADGRL3, ELAVL3, DNM2, ILF3, LYL1, SH3BP1, SSTR2
DNAJB13	0	4	BDKRB1, ARFRP1, KLHL35, WISP1
DNAL1	0	3	CIPC, LIN52, SLIRP
DNAL4	9	9	DDX17, CACNA1I, UBA7, TRIOBP, TOM1, GTPBP1, RHOT2, JOSD1, HMGXB4
DNALI1	5	3	GJA4, GJB5, TMEM200B
Continued			

Cilia gene	K	N	Novel interactors
DRC1	1	4	C2orf44, CENPO, C2orf16, TCF23
DRC7	6	8	CCDC102A, ADGRG5, CPNE2, C16orf70, ZNF319, USB1, NDUFB9, MT4
DRD1	11	5	AP3M1, PLCD1, RARS, PPP4C, KCNMB1
DRD2	20	13	CASP1, CCNA2, CD3G, CD3E, GPC6, SGSH, VWA5A, INPPL1, IL10RA, PSG7, NQO2, RBM22, E2F6
DRD5	5	2	NKX3-2, FGFBP2
DYNLL1	76	5	RFC5, SUDS3, GABRQ, WSB2, VSIG10
DYNLL2	47	9	C17orf67, CHMP5, C17orf47, HSF5, STXBP4, F13A1, KIF2B, PRKD3, MSI2
DYNLRB1	6	2	CBFB, PDRG1
DYNLRB2	4	8	WWOX, SDR42E1, TGFBI, ZNF134, SLC38A8, NECAB2, IRF2BPL, MPHOSPH6
DYNLT1	28	7	CRHBP, ACAT2, SYNJ2, RPS6KA2, TMEM143, PARK2, ERO1L
DYNLT3	7	4	CYBB, IFITM1, UIMC1, FOS
DYX1C1	4	3	GCOM1, PRTG, WDR72
DZIP1L	4	3	PRR23B, NME9, PRR23C
EFHC1	6	2	CENPQ, ADGRF4
ESYT3	0	2	FAIM, SLC35G2
GALR2	1	1	EIF4A3
GALR3	0	4	EIF3D, CEP72, ENTHD1, SUV420H1
GAS8	0	7	FANCA, COX7B, DPEP1, MYLIP, MVD, MC1R, TAF1C
GPR161	0	3	C1orf112, TOR1AIP1, TBX19
GPR83	0	12	FUT4, CST1, CDK2AP1, AQP3, FGF13, PLSCR3, TYR, TPBG, JRKL, MRE11A, NINJ1, SPR
HEPH	0	1	MSN
HSPB11	2	4	C1orf168, C1orf87, RPS26P15, PARS2
HYDIN	1	2	FAM96B, CFDP1
IFT122	1	2	ZNF148, SLCO2A1
IFT140	4	7	DNAJA3, ZNF263, SLC9A3R2, RAB11FIP3, TELO2, IL32, TRAP1
IFT172	2	4	EFR3B, DNAJC27, RAB10, TMEM214
IFT20	28	7	ABHD15, COASY, HLCS, RALGPS2, RICTOR, TP53I13, RAB11FIP4
IFT22	0	5	C7orf43, NFE4, ORAI2, KMT2E, ZNF655
IFT27	2	5	DDX17, TNRC6B, KDELR3, TRIOBP, TAB1
IFT43	2	3	ELMSAN1, LIN52, NEK9
IFT46	0	6	NLRX1, TMPRSS4, SCN3B, TRAPPC4, SIDT2, PDZD3
IFT52	1	2	SPINT3, TOMM34
IFT57	4	6	CLTA, CASP8, CBLB, HTT, SLC25A3, TERT
IFT74	0	0	None
IFT80	1	5	FGFR4, OTOL1, SLITRK3, NMD3, ZBBX
IFT81	1	6	ERP29, NAA25, TPCN1, SLC8B1, SDS, RPH3A
IFT88	6	6	C1QTNF9B-AS1, ALDH8A1, MAP4K5, PDX1, TRIM24, SAP18
INVS	5	2	ABCA1, TMEM245
IQCB1	1	6	CD80, UPK1B, TRH, LMOD1, PVRL3, TBL1X
LCA5	3	1	FILIP1
LOXHD1	0	4	HDHD2, HAUS1, C18orf25, SKA1
LRRC48	0	10	ELAC2, GID4, EIF3E, FUNDC2, SMCR8, ZNF287, SPECC1, PLAGL1, ZNF624, SOD2
LRRC56	0	2	KRTAP5-1, MOB2
LRRC6	2	1	ST3GAL1
MCHR1	2	8	GRAP2, EP300, APOBEC3A, MASP2, EFCAB6, TCIRG1, MRPS18A, ST13
MKS1	2	5	ABCC3, DGKE, CACNA1G, RNF43, SMG8
NEK8	4	0	None
NME5	2	8	HBEGF, APTX, MKRN1, IFITM1, VDAC1, SPOCK1, VIPR1, P4HA2
NME8	1	2	BBS9, SUGCT
NPHP1	15	2	BCL2L11, COX5B
NPHP3	2	15	CCDC50, ANAPC13, COL7A1, ASTE1, NAA50, DPP7, EPC1, DDX3Y, CDV3, DNAJC13, PIK3R4, TOPBP1, PARP14, PPP2R3A, RAB6B
NPHP4	1	6	FBXO44, EDIL3, C1orf127, TMEM201, SCRG1, RBP7
NPY2R	4	3	APOC3, ABCE1, SMAD1
NPY5R	4	2	MDH2, KIF23
PDCD6IP	37	5	ATG7, CLASP2, FBXL2, CRTAP, STAC
PIH1D1	31	5	CABP5, C19orf73, RCN3, RASIP1, PRKG2
Continued			

60

Cilia gene	K	N	Novel interactors
PKD1	24	9	NTHL1, NADSYN1, PDPK1, OR1F1, DDX58, NTN3, RPL3L, SOX17, PPL
PKD1L1	0	10	C7orf69, ABCC3, C7orf65, ACAA2, C7orf57, MGAT2, ABCA13, RPL6, PLCB2, PSME3
PKD2	14	6	KIF11, ADH1C, MTAP, ADH1A, PPP3CA, UCHL3
PKD2L1	3	6	MYOF, HNRNPA2B1, LGI1, CAB39, GBF1, PRPS2
PKHD1	1	7	ORM1, CSTF2T, ILK, ATXN2, EML4, BAG2, PTGIS
PPIL6	0	2	AK9, CEP57L1
PPP1CB	26	9	NRBP1, DHX57, GCKR, NCOA1, MATN1, SRD5A2, UPK2, SH3KBP1, ZNF512
PTCH1	65	5	CORO2A, ECM2, DPP3, TMOD1, UFM1
QRFPR	0	9	BBS12, BBS7, NDNF, EXOSC9, ADAD1, KIAA1109, TRPC3, TNIP3, TMEM155
RABL5	0	0	None
ROPN1L	3	4	FAM173B, 42435, CMBL, CCT5
RPGR	11	5	NUDT12, ATP6AP2, USP9X, SLC9A7, TSPAN7
RPGRIP1	23	1	SLC6A2
RPGRIP1L	3	2	CES1P1,SLC6A2
RSPH1	1	7	LINC01547, OSMR, COL18A1, LINC00313, LINC00334, LRRC3, UMODL1
RSPH10B	0	7	DAGLB, CCZ1B, FOXK1, MMD2, RBAKDN, GRID2IP, SLC29A4
RSPH3	5	6	PNLDC1, MRPL18, RGS17, TAGAP, SERAC1, SYTL3
RSPH4A	0	6	LACE1, HS3ST5, FAM26F, FLJ34503, FAM26E, ZUFSP
RSPH6A	0	1	SLC1A5
RSPH9	2	9	C6orf223, LRRC73, OARD1, KLHDC3, ADCY10P1, TSPO2, TCTE1, TAF8, SPATS1
RTDR1	13	0	None
SMO	8	1	LINC-PINT
SPAG1	3	7	MME, NIPAL2, CPQ, POLR2K, RGS22, COL4A1, STK3
SPAG16	2	1	IKZF2
SPAG17	0	2	GDAP2, FAM46C
SPAG6	2	2	OTUD1, PIP4K2A
SPEF2	0	10	C5orf42, BRIX1, NIPBL, LMBRD2, DNAJC21, CAPSL, NADK2, UGT3A1, TARS, WDR70
TCTE1	1	7	KLC4, C6orf223, LRRC73, KLHDC3, SPATS1, TAF8, RSPH9
TCTE3	1	4	TCP10, LINC01558, RPS6KA2, RNASET2
TCTN1	0	3	GLTP, SVOP, FAM222A
TCTN2	0	3	COQ5, KDM2B, CCDC92
TCTN3	0	0	None
TMEM17	0	3	FAM161A, COMMD1, C2orf74
TMEM216	1	9	TEX40, SAC3D1, ATL3, EFEMP2, CABP2, CCDC87, SSH3, POLA2, TRMT112
TMEM231	3	4	TMEM170A, CNTNAP4, ADAT1, VAT1L
TMEM237	2	2	ATP6V1C2, SFXN5
TMEM30B	0	4	BCAR1, SLC38A6, C14orf37, NAA30
TMEM67	1	4	NDUFAF6, INTS8, FAM92A1, LAPTM4B
TRAF3IP1	27	3	ATG4B, ANO7, SNED1
TSG101	86	10	PILRA, RPS13, MYLPF, DNAJB11, CALML5, ST5, PPP2R5D, MAPK15, ZNF160, ZNF419
TTC21B	0	4	COBLL1, G6PC2, CSRNP3, CERS6
TTC28	3	4	GATSL3, TPST2, PES1, PITPNB
TTC30B	0	7	FKBP7, CCDC141, PLEKHA3, RBM45, HNRNPA3, SESTD1, TTC30A
TTC8	1	1	PTPN21
VPS28	27	7	THEM6, CCNB1, OXCT1, OPLAH, RPL37A, SCRIB, CACNA1I
VPS37A	7	4	MICU3, CCAR2, COL11A2, KIAA1456
VPS4B	7	2	SPHAR, PHLPP1
WDR19	0	3	NWD2, TLR10, CHRNA9
WDR35	0	0	None
WDR63	0	2	RPL17P5, RPF1
WDR78	0	1	GNG12
ZMYND10	10	2	ACTR3, RALGDS
ZMYND12	0	3	ERMAP, C1orf50, CABIN1

Table 1. Novel interactors of each of the cilia genes. The table shows the number of known and computationally predicted novel PPIs for each of the 165 cilia genes, and lists their corresponding novel interactors.

Neuronal pathways in cilia interactome	p value	Number of proteins	Proteins
Huntington's disease signaling	1.00E−13	15	PLCB2,SHC1,CASP1,GNG12,PIK3R6,POLR2G,PIK3R4,CLTA,PIK3R2,CAPN1,PDPK1,RPH3A,CASP8,POLR2K,PRKD3
Dopamine-DARPP32 feedback in cAMP Signaling	2.00E−08	8	PLCB2,PPP2R3A,PRKG2,CALML5,PLCD1,PPP2R5D,PPP3CA,PRKD3
CREB signaling in neurons	3.02E−08	11	CALML5,PLCB2,SHC1,GNG12,PIK3R6,POLR2G,PIK3R4,PIK3R2,PLCD1,POLR2K,PRKD3
nNOS signaling in neurons	8.13E−06	4	PPP3CA,CAPN1,CALML5,PRKD3
nNOS signaling in neurons	8.13E−06	4	PPP3CA,CAPN1,CALML5,PRKD3
Axonal guidance signaling	8.91E−06	15	PLCB2,GIT1,SHC1,PIK3R6,ACTR3,NTN3,GNG12,MICAL1,PIK3R4,MYLPF,PLCD1,PPP3CA,PRKD3,SEMA7A,PIK3R2
eNOS signaling	1.17E−05	13	BDKRB1,CALML5,AQP3,CNGA1,CASP8,SLC7A1,HSP90AB1,PIK3R6,PIK3R4,PIK3R2,PDPK1,CHRNA9,PRKD3
Synaptic long term potentiation	1.26E−05	5	PPP3CA,PLCB2,PRKD3,CALML5,PLCD1
Wnt/Beta-catenin signaling	1.62E−05	6	CDKN2A,TLE1,PPP2R3A,ILK,SOX17,PPP2R5D
Neuregulin signaling	3.63E−05	8	SHC1,TMEFF2,HSP90AB1,HBEGF,BAD,PIK3R2,PDPK1,PRKD3
Neuropathic pain signaling in dorsal horn neurons	0.000224	7	PLCB2,FOS,PIK3R6,PIK3R4,PIK3R2,PLCD1,PRKD3
Calcium signaling	0.000603	8	TNNT1,SLC8B1,CABIN1,CALML5,TRPC3,PPP3CA,ATP2B3,CHRNA9
Dopamine receptor signaling	0.00251	3	PPP2R5D,SPR,PPP2R3A
Glutamate receptor signaling	0.00398	1	CALML5
Synaptic long term depression	0.00676	7	PLCB2,PPP2R3A,NPR2,PLCD1,PRKG2,PPP2R5D,PRKD3
Wnt/Ca + pathway	0.0102	3	PPP3CA,PLCB2,PLCD1
Dendritic cell maturation	0.0102	12	PLCB2,STAT4,COL11A2,COL18A1,CD80,TRGV9,IL6,PIK3R6,PIK3R4,IL32,PIK3R2,PLCD1
Reelin signaling in neurons	0.0407	3	PIK3R6,PIK3R4,PIK3R2

Table 2. Overlap of neuronal pathways in cilia interactome. Neuronal pathways which were present in cilia interactome with number of novel interactors.
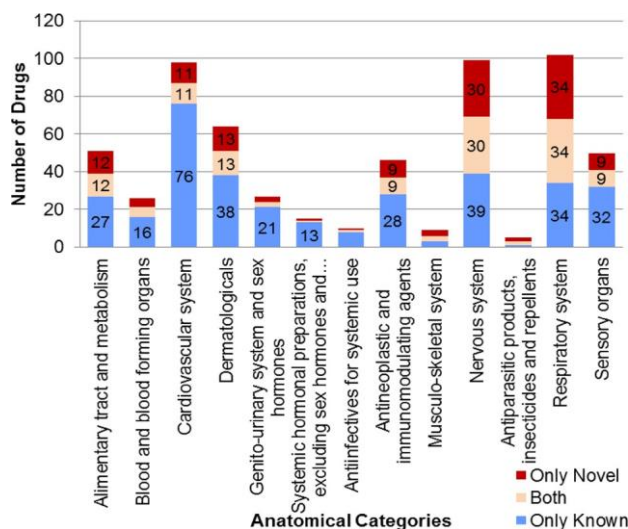
Figure 2. Number of drugs targeting genes in the cilia interactome. The numbers are shown separated by the anatomic category of the drugs (anatomic, therapeutic and chemical classification) and also separated by whether they target known interactors (blue) or novel interactors (red) or both (cream-colored).

Functional interactions of cilia genes with predicted novel interactors. We used ReactomeFIViz[60], a Cytoscape plugin, to extract known functional interactions between cilia genes and their novel interactors. Five novel PPIs had such functional interactions, namely, IFT57-CLTA, DYNLL2-KIF2B, IFT57-HTT, CHMP5-UBAP1 ('*part of the same complex*', '*bound by the same set of ligands*') and IFT57→CASP8 ('*activation*').

Discussion

We developed the interactome of ciliary proteins that included IFT, BBS, TZ, ciliary membrane proteins and proteins in motile cilia. The interactome includes novel computationally predicted PPIs for multiple proteins, including proteins with few or no previously known PPIs.

Both analysis of individual novel PPIs and the cilia interactome as a whole has the potential to highlight connections to specific neurological disorders and lead to biologically insightful and clinically translatable results. We interpreted the functions of individual novel PPIs using literature-based evidence and top pathways obtained from IPA (See Supplementary File 5 for testable hypotheses on novel PPIs involved in neuropsychiatric disorders, primary ciliary dyskinesia, hydrocephalus and in biological processes such as ciliogenesis and trafficking of membrane receptors in cilia). The following is a demonstrative example of a systems-level analysis.

Cilia, Alzheimer's disease and aging. Alzheimer's disease (AD) is a progressive neurodegenerative disease with an estimated prevalence of 10–30% in the population aged 65 years and more, characterized by memory loss (dementia), behavioral changes, impaired cognition and language[61]. Around two-thirds of dementia cases is attributed to AD[61]. Hippocampus, a region in the brain critical to memory and learning, exhibits signs of neurodegeneration in the early stages of AD[62]. It has been speculated that memory and learning deficits in AD may be associated with aging and reduced neurogenesis in the hippocampus[62–64]. It is interesting to note that primary cilia have been shown to mediate sonic hedgehog signaling (Shh) to regulate hippocampal neurogenesis[65,66]. So, we explored interconnections of AD, aging and cilia in the PPI network (the 'interactome'), while asking the following questions: Are genes associated with AD, aging and cilia closely connected in the interactome? Will such a network also include genes involved in Shh signaling and neurogenesis, and genes expressed in the hippocampus? What specific biological processes may underlie the connections of AD to aging, and will they interact with the Shh pathway?

Significant overlap was found between cilia and the AD interactomes (p value = 0.022). The AD interactome was highly significantly enriched in 'human aging-related genes' (p value = 1.77e−37), compiled from the GenAge database[67]. 51 aging genes co-occurred in AD and cilia interactomes. The subnetwork of these 51 genes and their AD and cilia interactors is shown in Fig. 3. In this subnetwork, aging genes connected cilia genes with/without Shh involvement to AD genes (Fig. 3).The next question we asked was: do any of the 51 genes directly interact with a ciliary gene involved in the Shh pathway? 15 cilia genes in the network were also recovered as regulators of the Shh pathway in a CRISPR genetic screen: ARL13B, BBS1-2, BBS4-5, BBS7, CBY1, DYNLL1, IFT140, IFT20, IFT52, IFT81, PTCH1, STUB1 and TRAF3IP1[58]. These 15 genes had direct interactions with 14 aging genes, 6 AD genes and 2 cilia genes. This included 13 novel predicted interactions connecting aging genes to cilia genes including 4 Shh genes (in *italics*): BAK1-*BBS1*, CDKN2A-DNAI1, TRAP1-*IFT140*, PDPK1-PKD1, SOD1-DNAH3, CCNA2-DRD2, TERT-*IFT57*, HTT-*IFT57*, FOS-DYNLT3, EP300-MCHR1, SHC1-DNAH7, PRKCA-CDK3 and RICTOR-*IFT20*. The network was significantly enriched in the GO term '*neurogenesis*' (p value = 5.66e−12) and in genes expressed in the hippocampus (transcripts per million ≥ 2) (p value = 2.54e−09). The cilia genes DYNLT1 and PKD1 were associated with neurogenesis, and IFT20, IFT140, PTCH1 and BBS4 were Shh regulators also associated with neurogenesis. Reduced size of hippocampus was noted in mutant mouse models of 5 cilia genes, namely BBS1, BBS2, BBS4, BBS7 and PDCD6IP (Mammalian Phenotype Ontology term: *small hippocampus*)[68–70]. We next identified the biological processes that may be specifically affected in AD in relation to its links with aging. 75 genes in the network were differentially expressed in the hippocampus of AD patients compared with non-AD subjects (GSE48350[71], GSE36980[72], GSE1297[73], GSE28146[49], GSE29378[48]). We then examined the fold change in the normal expression of these 75 genes in the hippocampus at 40 years compared with 8 post-conceptional weeks. For this, we used the 'developmental transcriptome' from the Brain-Span Atlas containing RNA-Seq data of up to 16 brain regions from post-conceptional weeks (number of weeks elapsed from the first day of the last menstrual period and the day of the delivery) to middle adulthood (up to 40 years)[52]. The genes were grouped based on the specific direction in which their expression varied in AD versus aging (i.e. fold change in same/opposite directions in *AD versus non-AD hippocampal samples* compared with *expression at 40 years versus 8 post-conceptional weeks*) (Fig. 4). 42 genes showed an expression change in the opposite direction in AD versus aging. Out of this, 18 genes were underexpressed in AD but overexpressed in aging; they were enriched in the GO term *'calcium-mediated signaling'* (p value = 8.72e−09). It has been postulated that calcium signaling pathways involved in cognition may be remodeled by an activated amyloidogenic pathway in AD, resulting in elevated levels of calcium and a constant erasure of new memories through enhancement of mechanisms involved in long term depression[74]. It is also worth noting that Shh signaling requires calcium mobilization[75]. The 18 genes included the cilia genes DYNLL1, DYNLT3, PKD1 and MCHR1, and the ciliary Shh regulator BBS7. 24 genes were overexpressed in AD but underexpressed in aging; they were enriched in '*circulatory system development*' (p value = 3.04e−07). Loss of hippocampal blood vessel density accompanied by ultrastructural changes in the blood vessels have been observed in a senescence-accelerated rat model of AD[76]. It is interesting to note that circulatory system processes were found to be upregulated in early stages of AD-like pathology in this model, while they were found to be downregulated with age, similar to our observations[76]. It is also interesting to note that neovascularization requires Shh signaling[77]. The 24 genes included the cilia genes CCDC40, SPAG6, ZMYND10, DNALI1 and SPAG1, BBS2 and CBY1 which are ciliary Shh regulators, DYNLT1, a cilia gene involved in neurogenesis and PTCH1 which is an Shh ligand also involved in neurogenesis. 25 genes showed an expression change in the same direction (either under/overexpression) in AD versus aging including the cilia genes VPS4B, CCNA1, DYNLRB2, NPHP1, DNAH7 and the ciliary Shh regulator BBS5; '*negative regulation of cell death*' was enriched in this group (p value = 1.59e−09). Shh maintains neural stem cells in the hippocampus by inhibiting cell death[78].

In summary, our analysis demonstrates that aging and AD genes directly interact with ciliary Shh regulators in the interactome. This network is enriched in genes associated with neurogenesis and expressed in the hippocampus. Genes involved in *calcium-mediated signaling* and *circulatory system development* are differentially expressed in the opposite direction in AD versus aging, whereas genes involved in *regulation of cell death* are differentially expressed in the same direction.
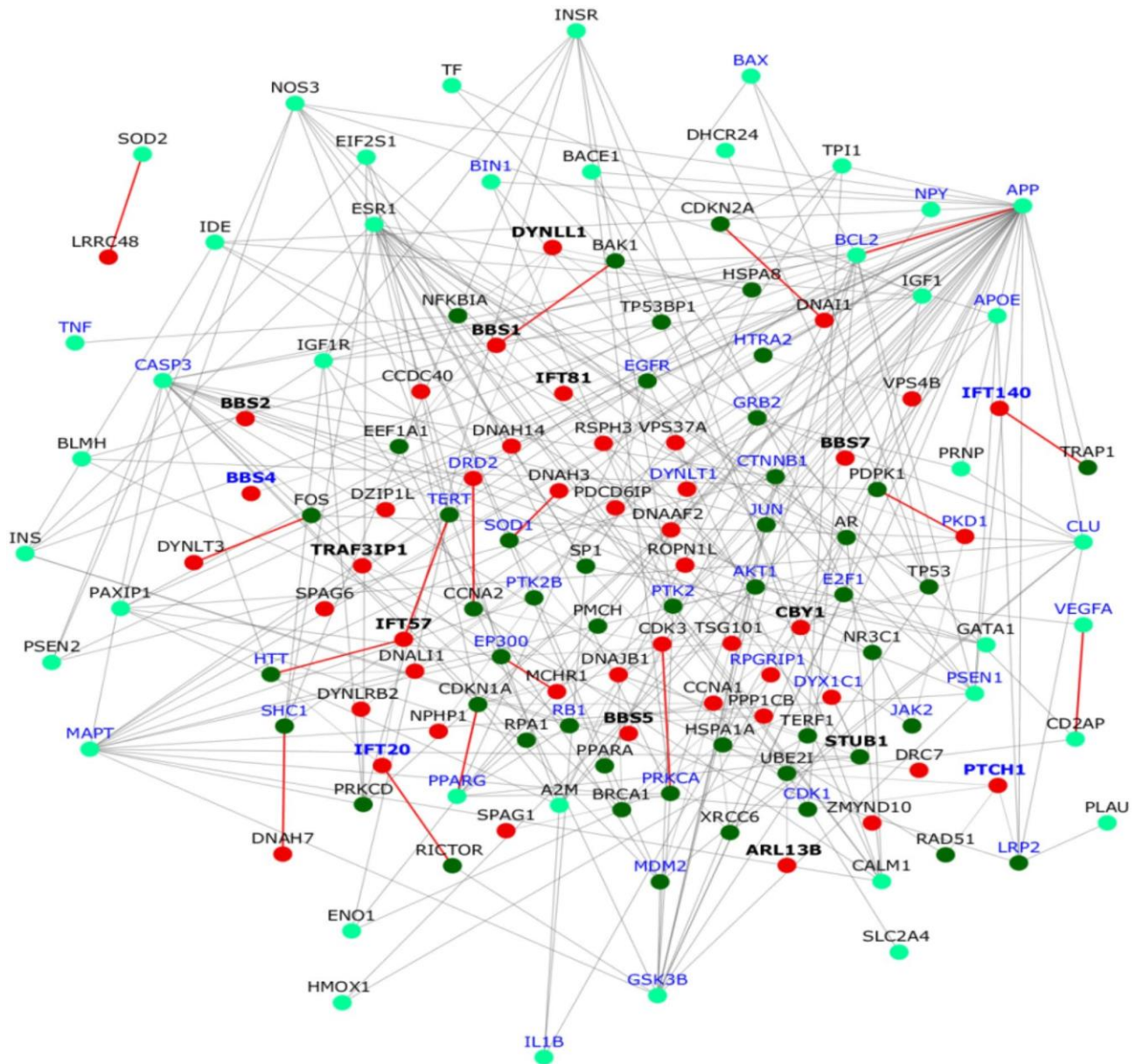
Figure 3. Interconnections between cilia, aging and Alzheimer's disease genes. Cilia genes are shown as red nodes; AD genes are colored in cyan and aging genes in green. PPIs are shown as edges, where grey edges are known PPIs and red color edges are novel predicted PPIs. Genes with bold labels are involved in the sonic hedgehog (Shh) pathway and those with blue labels are involved in neurogenesis. Note that, in this case, a bold blue-labeled gene indicates a cilia gene with Shh involvement, which is also involved in neurogenesis.

Cilia and neuronal pathways and functions. Our pathways enrichment analysis of the cilia interactome revealed several neuronal pathways with high statistical significance (see Table 2 and Supplementary File 3). This included *axonal guidance signaling pathway* with 15 novel cilia interactors, *Huntington disease signaling* with 15 novel interactors, *eNOS signaling pathway* with 13, *Wnt signaling* with 6, *DARPP32 feedback in cAMP signaling* with 8, *dopamine receptor signaling* with 3, *synaptic long-term depression* with 7, and *synaptic long-term potentiation* with 5 novel interactors. Dopamine receptors are localized in the membrane of neuronal cilia[19], suggesting that these novel cilia interacting partners may have a role in neurotransmission. *Dopamine signaling*, *eNOS signaling*, and *synaptic long term potentiation* pathways are also known to be associated with neuropsychiatric disorders such as schizophrenia[79,80]. The identification of Huntington's disease (HD) pathway in the cilia interactome is also notable given that the protein huntingtin (HTT) localizes to the centrosome and plays an important role in ciliogenesis. The HD mutant mouse model exhibits abnormal cilia motility and cerebrospinal fluid flow[23]. Recovery of Wnt signaling thought to be involved in schizophrenia etiology is also of interest[81,82].

Analysis of the known and novel PPIs and GO term associations identified a role for cilia in neuronal disease pathogenesis. While consistent with the known role of cilia in several key processes in the nervous system
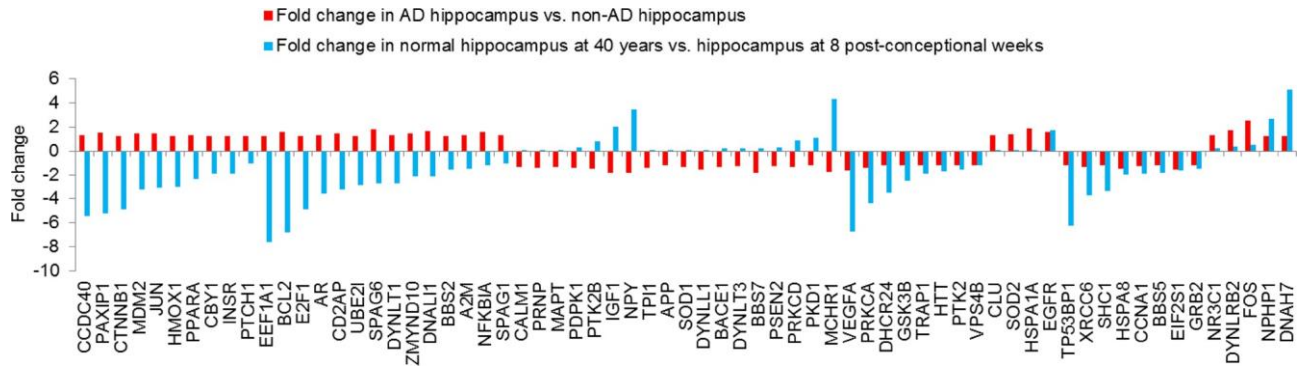
Figure 4. Direction of fold change of gene expression levels in AD vs. non-AD hippocampal samples compared with expression at 40 years vs. 8 post-conceptional weeks. The X-axis shows the 67 genes that are differentially expressed in the hippocampus of Alzheimer's disease patients compared with the hippocampus of healthy subjects. The red bars on the Y-axis show the fold change of the differential expression of these genes in AD vs. non-AD hippocampal samples. The blue bars on the Y-axis show the fold change in the expression level of these genes in normal hippocampus at 40 years after birth (middle adulthood) compared with 8 weeks after conception (fetal life) in healthy humans.

such as the neuronal signaling and development, these findings reveal novel connections between cilia and these functional modules. The defects in neuronal migration and differentiation are the underlying cause of abnormal neural circuitry in psychiatric disorders[12]. This is further supported by the reported linkage of neuropsychiatric risk genes to cilia[15,19] and the finding of neuropsychiatric phenotypes and brain abnormalities in ciliopathies[5,12]. Our interactome analysis shows that TCTN2, cilia gene with known role in neuronal development and migration[12] has 3 novel interactors and neuronal GO terms such as *initiation of neural tube closure*, *midbrain morphogenesis* and *mid brain development* are enriched among the interacting partners. The GO terms that are enriched for interacting partners of ARMC4 include *sympathetic neuron projection guidance*, *axonogenesis*, *axon extension*, and *axon fasciculation*. Dynein gene, DNAAF2 has only one known but 4 predicted interactions. Two of those novel interactors, ATL1 and TRIM9 are shown to be associated with cognitive performance and psychosis respectively through GWAS. The GO terms such as *axonogenesis*, *neuron maturation*, *synaptic growth at neuromuscular junction* are enriched among the interacting partners. Ciliary membrane genes DRD1 and DRD2 that are implicated in neurotransmission and linked to mental illnesses such as schizophrenia[83] were identified with 4 and 12 novel interactors, respectively; the associated GO terms were *neuronal action potential* and *synaptic plasticity regulation*. We also observed 4 novel interactors for the cilia protein TMEM67, including two proteins associated with cilia assembly, LAPTM4B and NDUFAF6, with NDUFAF6 also known to be associated with Alzheimer's disease[84]. Both ATG7, a novel interactor of the ciliary protein PDCD6IP, and SPR, a novel interactor of GPR83 have been associated with Parkinson's disease[85,86]. GIT1, a novel interactor of B9D1 is associated with attention deficit hyperactivity disorder and MME, a novel interactor of SPAG1with Alzheimer's disease[87]. On inspecting mammalian phenotype ontology (MPO) terms (www.informatics.jax.org/), 42 novel interactors were found to be associated with various morphological or physiological aspects of brain in mice. For example, the novel interactor ITSN1 was associated with *decreased brain size*, *abnormal corpus callosum*, *hippocampal fimbria*, *hippocampal fornix*, *brain white matter* and *anterior commissure morphology*. These findings support the role of these novel interactions and the GO terms in understanding the crucial role played by cilia biology in neuropsychiatric disorders.

Overlap of cilia and neuropsychiatric disorder interactomes. To examine the connection between cilia and neuropsychiatric disorders, we computed the overlap between their interactomes. We considered 7 neuropsychiatric disorders (NPDs), namely Attention Deficit Hyperactivity Disorder (ADHD), Major Depressive Disorder (MDD), schizophrenia, bipolar disorder, autism spectrum disorder, Alzheimer's disease and Parkinson's disease. We extracted the genes associated with each disorder from the GWAS catalog (www.ebi.ac.uk/gwas/) and then assembled disorder-specific interactomes with known PPIs from HPRD and BioGRID. We then computed how closely connected the cilia genes are to NPD genes by computing how many genes or interactors were shared between the cilia interactome and each NPD interactome. This analysis showed the overlap to be statistically significant (Table 3). For example, cilia interactome has an overlap of 88 genes with ADHD interactome (*p* value = 1.2E−16) of which 17 are novel interactors of cilia. Similar comparisons with other NPDs also showed overlaps as shown in Table 3.

Overlap of cilia interactome with genes differentially expressed in neuropsychiatric disorders. 965 genes in the cilia interactome were found to be expressed (transcripts per million ≥ 2) in several brain regions including amygdala, anterior cingulate cortex, caudate, cerebellum, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord and substantia nigra, from GTEx RNA-Seq data[51] (*p* value = 3.93E−58). Novel interactors of cilia genes were found to be highly statistically enriched among these genes expressed in the human brain (*p* value = 8.14E−09).

NPD	NPD interactome size	p value of overlap	# Genes common to both interactomes	# Novel interactors common to both	# Genes differentially expressed in the disorder	# Genes differentially expressed in the cilia interactome	p value of overlap	# Differentially expressed novel genes
ADHD	406	1.20E−16	88	17	n/a	n/a	n/a	n/a
Alzheimer's disease	417	2.20E−24	104	19	1,103	106	4.70E−05	46
Autism spectrum disorder	53	2.20E−05	15	5	3,692	314	2.40E−04	119
Bipolar disorder	764	3.10E−29	163	31	1,188	101	3.40E−02	38
Major depressive disorder	974	3.70E−23	177	32	187	21	2.50E−02	8
Parkinson's disease	520	1.20E−20	112	18	2,487	258	6.20E−03	104
Schizophrenia	688	1.50E−16	125	26	1,320	118	5.60E−03	40
Intellectual disability	n/a	n/a	n/a	n/a	706	75	6.50E−03	32

Table 3. Overlap of neuropsychiatric disease (NPD) interactomes and genes differentially expressed in NPDs with the cilia interactome. The significance of the overlap along with the number of genes common to the NPD interactome/expression datasets and the cilia interactome are shown.

We then computed the overlap of genes differentially expressed in neuropsychiatric disorders with the genes in the cilia interactome. We analyzed gene expression datasets of MDD (GSE53987)[44], schizophrenia (GSE17612)[45], bipolar disorder (GSE12679)[46], autism spectrum disorder (GSE18123)[47], Alzheimer's disease (GSE29378)[48], Parkinson's disease (GSE28894) and non-syndromic intellectual disability (GSE39326)[50]. The analysis showed the overlap to be statistically significant (Table 3). For example, the cilia interactome has an overlap of 106 genes with genes differentially in the Alzheimer's disease dataset (p value = 4.7E−05) of which 46 are novel interactors of cilia.

Cilia and nervous system drug targets. Given the strong connection between the cilia interactome and neuronal pathways, we tested the possibility of repurposing drugs targeting proteins in the cilia interactome for treating neurological disorders. Identifying new uses for drugs shortens the time of drug discovery and approval[88]. For example, the drug amantadine which is used to treat influenza infection was successfully repurposed to treat dyskinesia and Parkinson's disease[88]. This analysis identified 548 drugs targeting 184 genes in the cilia interactome. These fall into 3 major Anatomic Therapeutic Chemical (ATC) classification system categories, nervous system with 99 drugs, 102 drugs in the respiratory system, and 98 drugs in the cardiovascular system (Fig. 2, Supplementary File 4). This finding points at therapeutics targeting the cilia proteins which may provide a novel strategy for treating neurological disorders.

Overall, 76 nervous system drugs targeted 7 novel interactors: HRH1, SLC6A2, CHRNA9, NQO2, ORM1, CACNA1I and CACNA1G. 57 drugs targeting 22 genes in the interactome are used in the treatment of at least one among the following neurological disorders- Parkinson's disease, Alzheimer's disease, attention deficit hyperactivity disorder (ADHD), major depressive disorder (MDD), autism spectrum disorder, schizophrenia and bipolar disorder- out of which 35 drugs target 6 novel interactors, namely CACNA1G, CACNA1I, CHRNA9, HRH1, SLC6A2 and ORM1. 10 out of these 57 drugs targeted cilia genes as well as known and novel interactors of cilia genes: asenapine, chlorpromazine, clozapine, loxapine and paliperidone are schizophrenia drugs, olanzapine is used in the treatment of Alzheimer's disease and schizophrenia, amphetamine in ADHD, imipramine in ADHD and MDD, mirtazapine in MDD and nortriptyline in schizophrenia, ADHD, MDD and bipolar disorder.

Among other novel interactors targeted by nervous system drugs is SLC6A2 which is involved in neurotransmission and is associated with ADHD[89,90]. SLC6A2 interacts with RPGRIP1L, a ciliary protein known to cause Joubert syndrome, MKS and bipolar disorder[91,92]. The novel interactors CACNA1I and CACNA1G targeted by nervous system drugs are calcium channels that are known to be associated with Alzheimer's disease and schizophrenia, respectively[93,94]. These novel interactors which are drug targets may have significant impact on the nervous system, and the pathogenesis of neurological disorders.

In an independent study, we proposed that the drug acetazolamide which targets the genes CA2 and CA4, having known interactions with the cilia genes, DYNLL1 and CDK3 respectively, may be repurposed for schizophrenia based on negative correlation of drug-induced versus disease-associated gene expression profiles and other biological evidences[95]. Acetazolamide is currently under consideration for funding for clinical trial. Several cancer drugs with reported effects on ciliogenesis target known and novel interactors in the cilia interactome. Vinblastine targeting JUNN, a known interactor of BBS7 and TSG101, and TUBB, a known interactor of NPHP1 and DYNLL1, inhibits cilia regeneration in partially deciliated *Tetrahymena* (a unicellular ciliate)[96]. Valproic acid targeting HDAC9, a known interactor of PKD1, restores ciliogenesis in pancreatic ductal adenocarcinoma cells[97]. Gefitinib targeting EGFR, a known interactor of PDCD6IP, inhibits the smoking-induced loss of ciliated cells in the airway[98]. Gefitinib also increases the percentage of ciliated cells in human pancreatic cancer cell lines[99]. Geldanamycin targeting HSP90AB1, a novel interactor of CETN3, induces lengthening of cilia in 3T3-L1, a fibroblast cell line[100].

Conclusion

We identified novel PPIs of cilia proteins and their associated pathways, their enriched Gene Ontology term associations, and drugs that target the interactors. This cilia interactome analysis reveals a link between cilia function, neuronal function and neurological disorders. We also demonstrated the interconnections of Alzheimer's disease, cilia and aging genes. The predicted interactions will have to be validated at the level of network perturbations in the disease state by comparing neuropsychiatric patients with healthy controls. However, one has to be aware of a few caveats while studying the role of ciliary genes in neuropsychiatric disorders (NPDs). Association of a ciliary gene with a NPD can be unequivocally ascertained only if this association is discovered within the ciliary compartment in the context of the particular NPD, i.e. a mechanistic link between ciliary function and the disorder has to be demonstrated. It may not be a true association if a ciliary gene was shown to be associated with a NPD in a cellular context not connected with cilia; a protein may perform its function at different subcellular locations. Mapping the interactome of cilia genes would be useful in carrying out network-based systems biology studies, which will help elucidate the contribution of these novel PPIs to nervous system disease pathology as well as to develop novel therapeutics for these disorders.

Data availability

We will make the cilia interactome publicly available on our web application Wiki-Pi[101]. Novel PPIs will be highlighted in yellow on the website. The number of novel and known PPIs of the cilia genes are given in Supplementary File 1. Interactome network diagram that is shown in Fig. 1 is also being made available in PDF format and in Cytoscape file format as Supplementary File 6 and Supplementary File 7 respectively. PDF file would be suitable for printing in high resolution and for electronically searching for specific genes, and Cytoscape would allow further processing and data analysis. Wiki-Pi allows users to search for interactions by specifying biomedical associations of one or both proteins involved. Thus, queries can be customized to include/exclude gene symbol, gene name, GO annotations, diseases, drugs, and/or pathways for either gene involved in an interaction. For example, researchers can search for interactions by giving at least one cilia gene and a pathway of interest, say "IFT20 interactions where the interactor is involved in *immunity*"; this query would match 5 PPIs out of a total of 19 PPIs of IFT20. Another example is the search "find interactions where one protein's annotation contains the word *ciliary* and the other protein's annotation contain the word *neuronal*". The search returns 353 PPIs, out of which 13 are novel PPIs.

References

1. Satir, P. & Christensen, S. T. Structure and function of mammalian cilia. *Histochem. Cell Biol.* **129**, 687–693. https://doi.org/10.1007/s00418-008-0416-9 (2008).
2. Gerdes, J. M., Davis, E. E. & Katsanis, N. The vertebrate primary cilium in development, homeostasis, and disease. *Cell* **137**, 32–45. https://doi.org/10.1016/j.cell.2009.03.023 (2009).
3. Enuka, Y., Hanukoglu, I., Edelheit, O., Vaknine, H. & Hanukoglu, A. Epithelial sodium channels (ENaC) are uniformly distributed on motile cilia in the oviduct and the respiratory airways. *Histochem. Cell Biol.* **137**, 339–353. https://doi.org/10.1007/s00418-011-0904-1 (2012).
4. Banizs, B. *et al.* Dysfunctional cilia lead to altered ependyma and choroid plexus function, and result in the formation of hydrocephalus. *Development* **132**, 5329–5339. https://doi.org/10.1242/dev.02153 (2005).
5. Waters, A. M. & Beales, P. L. Ciliopathies: an expanding disease spectrum. *Pediatric Nephrol.* **26**, 1039–1056. https://doi.org/10.1007/s00467-010-1731-7 (2011).
6. Pedersen, L. B. & Rosenbaum, J. L. Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling. *Curr. Top. Dev. Biol.* **85**, 23–61. https://doi.org/10.1016/S0070-2153(08)00802-8 (2008).
7. Jin, H. *et al.* The conserved Bardet–Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. *Cell* **141**, 1208–1219. https://doi.org/10.1016/j.cell.2010.05.015 (2010).
8. Barker, A. R., Renzaglia, K. S., Fry, K. & Dawe, H. R. Bioinformatic analysis of ciliary transition zone proteins reveals insights into the evolution of ciliopathy networks. *BMC Genom.* **15**, 531. https://doi.org/10.1186/1471-2164-15-531 (2014).
9. Lee, J. E. & Gleeson, J. G. Cilia in the nervous system: linking cilia function and neurodevelopmental disorders. *Curr. Opin. Neurol.* **24**, 98–105. https://doi.org/10.1097/WCO.0b013e3283444d05 (2011).
10. Louvi, A. & Grove, E. A. Cilia in the CNS: the quiet organelle claims center stage. *Neuron* **69**, 1046–1060. https://doi.org/10.1016/j.neuron.2011.03.002 (2011).
11. Guemez-Gamboa, A., Coufal, N. G. & Gleeson, J. G. Primary cilia in the developing and mature brain. *Neuron* **82**, 511–521. https://doi.org/10.1016/j.neuron.2014.04.024 (2014).
12. Guo, J. *et al.* Developmental disruptions underlying brain abnormalities in ciliopathies. *Nat. Commun.* **6**, 7857. https://doi.org/10.1038/ncomms8857 (2015).
13. Gazea, M. *et al.* Primary cilia are critical for Sonic hedgehog-mediated dopaminergic neurogenesis in the embryonic midbrain. *Dev. Biol.* **409**, 55–71. https://doi.org/10.1016/j.ydbio.2015.10.033 (2016).
14. Murdoch, J. N. & Copp, A. J. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birth Defects Res. Part A Clin. Mol. Teratol.* **88**, 633–652. https://doi.org/10.1002/bdra.20686 (2010).
15. Marley, A. & von Zastrow, M. A simple cell-based assay reveals that diverse neuropsychiatric risk genes converge on primary cilia. *PLoS ONE* **7**, e46647. https://doi.org/10.1371/journal.pone.0046647 (2012).
16. Alvarez Retuerto, A. I. *et al.* Association of common variants in the Joubert syndrome gene (AHI1) with autism. *Hum. Mol. Genet.* **17**, 3887–3896. https://doi.org/10.1093/hmg/ddn291 (2008).
17. Higginbotham, H. *et al.* Arl13b in primary cilia regulates the migration and placement of interneurons in the developing cerebral cortex. *Dev. Cell* **23**, 925–938. https://doi.org/10.1016/j.devcel.2012.09.019 (2012).
18. Torri, F. *et al.* Fine mapping of AHI1 as a schizophrenia susceptibility gene: from association to evolutionary evidence. *FASEB J.* **24**, 3066–3082. https://doi.org/10.1096/fj.09-152611 (2010).
19. Marley, A. & von Zastrow, M. DISC1 regulates primary cilia that display specific dopamine receptors. *PLoS ONE* **5**, e10902 (2010).

20. Kamiya, A. *et al.* Recruitment of PCM1 to the centrosome by the cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses. *Arch. Gen. Psychiatry* **65**, 996–1006. https://doi.org/10.1001/archpsyc.65.9.996 (2008).
21. Poelmans, G. *et al.* Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **150**, 140–147. https://doi.org/10.1002/ajmg.b.30787 (2009).
22. Massinen, S. *et al.* Increased expression of the dyslexia candidate gene DCDC2 affects length and signaling of primary cilia in neurons. *PLoS ONE* **6**, e20580. https://doi.org/10.1371/journal.pone.0020580 (2011).
23. Keryer, G. *et al.* Ciliogenesis is regulated by a huntingtin-HAP1-PCM1 pathway and is altered in Huntington disease. *J. Clin. Investig.* **121**, 4372–4382. https://doi.org/10.1172/JCI57552 (2011).
24. Dietrich, P., Shanmugasundaram, R., Shuyu, E. & Dragatsis, I. Congenital hydrocephalus associated with abnormal subcommissural organ in mice lacking huntingtin in Wnt1 cell lineages. *Hum. Mol. Genet.* **18**, 142–150. https://doi.org/10.1093/hmg/ddn324 (2009).
25. Mukhopadhyay, S. & Jackson, P. K. Cilia, tubby mice, and obesity. *Cilia* **2**, 1. https://doi.org/10.1186/2046-2530-2-1 (2013).
26. Siljee, J. E. *et al.* Subcellular localization of MC4R with ADCY3 at neuronal primary cilia underlies a common pathway for genetic predisposition to obesity. *Nat. Genet.* https://doi.org/10.1038/s41588-017-0020-9 (2018).
27. Sang, L. *et al.* Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell* **145**, 513–528. https://doi.org/10.1016/j.cell.2011.04.019 (2011).
28. Harrison, P. J. & Weinberger, D. R. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol. Psychiatry* **10**, 40–68. https://doi.org/10.1038/sj.mp.4001558 (2005).
29. Millar, J. K. *et al.* Genomic structure and localisation within a linkage hotspot of Disrupted In Schizophrenia 1, a gene disrupted by a translocation segregating with schizophrenia. *Mol. Psychiatry* **6**, 173–178. https://doi.org/10.1038/sj.mp.4000784 (2001).
30. Camargo, L. M. *et al.* Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol. Psychiatry* **12**, 74–86. https://doi.org/10.1038/sj.mp.4001880 (2007).
31. Wang, Q., Jaaro-Peled, H., Sawa, A. & Brandon, N. J. How has DISC1 enabled drug discovery?. *Mol. Cell. Neurosci.* **37**, 187–195. https://doi.org/10.1016/j.mcn.2007.10.006 (2008).
32. Rachel, R. A., Li, T. & Swaroop, A. Photoreceptor sensory cilia and ciliopathies: focus on CEP290, RPGR and their interacting proteins. *Cilia* **1**, 22. https://doi.org/10.1186/2046-2530-1-22 (2012).
33. Ishikawa, H., Thompson, J., Yates, J. R. & Marshall, W. F. Proteomic analysis of mammalian primary cilia. *Curr. Biol.* **22**, 414–419 (2012).
34. Sang, L. *et al.* Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell* **145**, 513–528 (2011).
35. Gupta, G. D. *et al.* A dynamic protein interaction landscape of the human centrosome-cilium interface. *Cell* **163**, 1484–1499 (2015).
36. Toriyama, M. *et al.* The ciliopathy-associated CPLANE proteins direct basal body recruitment of intraflagellar transport machinery. *Nat. Genet.* **48**, 648 (2016).
37. Boldt, K. *et al.* An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. *Nat. Commun.* **7**, 11491 (2016).
38. Pazour, G. J., Agrin, N., Leszyk, J. & Witman, G. B. Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **170**, 103–113 (2005).
39. Peri, S. *et al.* Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.* **32**, D497-501. https://doi.org/10.1093/nar/gkh070 (2004).
40. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535-539. https://doi.org/10.1093/nar/gkj109 (2006).
41. Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for "omics" research on drugs. *Nucleic Acids Res.* **39**, D1035-1041. https://doi.org/10.1093/nar/gkq1126 (2011).
42. Ganapathiraju, M. K. *et al.* Schizophrenia interactome with 504 novel protein–protein interactions. *NPJ Schizophrenia* **2**, 16012. https://doi.org/10.1038/npjschz.2016.12 (2016).
43. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449. https://doi.org/10.1093/bioinformatics/bti551 (2005).
44. Lanz, T. A. *et al.* STEP levels are unchanged in pre-frontal cortex and associative striatum in post-mortem human brain samples from subjects with schizophrenia, bipolar disorder and major depressive disorder. *PLoS ONE* **10**, e0121744 (2015).
45. Maycox, P. R. *et al.* Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol. Psychiatry* **14**, 1083 (2009).
46. Harris, L. W. *et al.* The cerebral microvasculature in schizophrenia: a laser capture microdissection study. *PLoS ONE* **3**, e3964 (2008).
47. Kong, S. W. *et al.* Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS ONE* **7**, e49475 (2012).
48. Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S. & Geschwind, D. H. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med.* **5**, 48 (2013).
49. Blalock, E. M., Buechel, H. M., Popovic, J., Geddes, J. W. & Landfield, P. W. Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J. Chem. Neuroanat.* **42**, 118–126 (2011).
50. Huang, L. *et al.* A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **91**, 694–702 (2012).
51. Consortium, G. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
52. Sunkin, S. M. *et al.* Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* **41**, D996–D1008 (2012).
53. Piñero, J. et al. *DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes*. Database 2015. https://doi.org/10.1093/database/bav028 (2015).
54. Keshava Prasad, T. *et al.* Human protein reference database—2009 update. *Nucleic Acids Res.* **37**, D767–D772 (2008).
55. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
56. Keshava Prasad, T. S. *et al.* human protein reference database—2009 update. *Nucleic Acids Res.* **37**, D767-772. https://doi.org/10.1093/nar/gkn892 (2009).
57. Luck, K. *et al.* A reference map of the human protein interactome. *bioRxiv*, 605451 (2019).
58. Breslow, D. K. *et al.* A CRISPR-based screen for Hedgehog signaling provides insights into ciliary function and ciliopathies. *Nat. Genet.* **50**, 460 (2018).
59. Geremek, M. *et al.* Ciliary genes are down-regulated in bronchial tissue of primary ciliary dyskinesia patients. *PLoS ONE* **9**, e88216 (2014).
60. Wu, G., Dawson, E., Duong, A., Haw, R. & Stein, L. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis [version 2; peer review: 2 approved]. *F1000Research* **3**. https://doi.org/10.12688/f1000research.4431.2 (2014).
61. Masters, C. L. *et al.* Alzheimer's disease. *Nat. Rev. Dis. Prim.* **1**, 15056. https://doi.org/10.1038/nrdp.2015.56 (2015).

62. Mufson, E. J. *et al.* Hippocampal plasticity during the progression of Alzheimer's disease. *Neuroscience* **309**, 51–67 (2015).

63. Smith, T. D., Adams, M. M., Gallagher, M., Morrison, J. H. & Rapp, P. R. Circuit-specific alterations in hippocampal synaptophysin immunoreactivity predict spatial learning impairment in aged rats. *J. Neurosci.* **20**, 6587–6593 (2000).

64. Jessberger, S. *et al.* Dentate gyrus-specific knockdown of adult neurogenesis impairs spatial and object recognition memory in adult rats. *Learn. Mem.* **16**, 147–154 (2009).

65. Breunig, J. J. *et al.* Primary cilia regulate hippocampal neurogenesis by mediating sonic hedgehog signaling. *Proc. Natl. Acad. Sci.* **105**, 13127–13132 (2008).

66. Whitfield, J. F. & Chakravarthy, B. R. The neuronal primary cilium: driver of neurogenesis and memory formation in the hippocampal dentate gyrus?. *Cell. Signal.* **21**, 1351–1355 (2009).

67. De Magalhães, J. P., Curado, J. & Church, G. M. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* **25**, 875–881 (2009).

68. Davis, R. E. *et al.* A knockin mouse model of the Bardet–Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. *Proc. Natl. Acad. Sci.* **104**, 19422–19427 (2007).

69. Zhang, Q. *et al.* BBS7 is required for BBSome formation and its absence in mice results in Bardet–Biedl syndrome phenotypes and selective abnormalities in membrane protein trafficking. *J. Cell. Sci.* **126**, 2372–2380 (2013).

70. Campos, Y. *et al.* Alix-mediated assembly of the actomyosin–tight junction polarity complex preserves epithelial polarity and epithelial barrier. *Nat. Commun.* **7**, 11876 (2016).

71. Berchtold, N. C. *et al.* Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc. Natl. Acad. Sci.* **105**, 15605–15610 (2008).

72. Hokama, M. *et al.* Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. *Cereb. Cortex* **24**, 2476–2488 (2013).

73. Blalock, E. M. *et al.* Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci.* **101**, 2173–2178 (2004).

74. Berridge, M. J. Calcium signalling and Alzheimer's disease. *Neurochem. Res.* **36**, 1149–1156 (2011).

75. Shaw, D. K. *et al.* Intracellular calcium mobilization is required for sonic hedgehog signaling. *Dev. Cell* **45**, 512–525 (2018).

76. Stefanova, N. A., Maksimova, K. Y., Rudnitskaya, E. A., Muraleva, N. A. & Kolosova, N. G. Association of cerebrovascular dysfunction with the development of Alzheimer's disease-like pathology in OXYS rats. *BMC Genom.* **19**, 75 (2018).

77. Renault, M.-A. *et al.* Sonic hedgehog induces angiogenesis via Rho kinase-dependent signaling in endothelial cells. *J. Mol. Cell. Cardiol.* **49**, 490–498 (2010).

78. Lai, K., Kaspar, B. K., Gage, F. H. & Schaffer, D. V. Sonic hedgehog regulates adult neural progenitor proliferation in vitro and in vivo. *Nat. Neurosci.* **6**, 21 (2003).

79. Schultz, W. Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* **30**, 259–288. https://doi.org/10.1146/annurev.neuro.28.061604.135722 (2007).

80. Karam, C. S. *et al.* Signaling pathways in schizophrenia: emerging targets and therapeutic strategies. *Trends Pharmacol. Sci.* **31**, 381–390. https://doi.org/10.1016/j.tips.2010.05.004 (2010).

81. Panaccione, I. *et al.* Neurodevelopment in schizophrenia: the role of the wnt pathways. *Curr. Neuropharmacol.* **11**, 535–558. https://doi.org/10.2174/1570159X113119990037 (2013).

82. De, A. Wnt/Ca2+ signaling pathway: a brief overview. *Acta Biochim. Biophys. Sin.* **43**, 745–756. https://doi.org/10.1093/abbs/gmr079 (2011).

83. Kaalund, S. S. *et al.* Contrasting changes in DRD1 and DRD2 splice variant expression in schizophrenia and affective disorders, and associations with SNPs in postmortem brain. *Mol. Psychiatry* **19**, 1258–1266. https://doi.org/10.1038/mp.2013.165 (2014).

84. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458. https://doi.org/10.1038/ng.2802 (2013).

85. Chen, D. *et al.* Genetic analysis of the ATG7 gene promoter in sporadic Parkinson's disease. *Neurosci. Lett.* **534**, 193–198 (2013).

86. Sharma, M. *et al.* The sepiapterin reductase gene region reveals association in the PARK3 locus: analysis of familial and sporadic Parkinson's disease in European populations. *J. Med. Genet.* **43**, 557–562 (2006).

87. Won, H. *et al.* GIT1 is associated with ADHD in humans and ADHD-like behaviors in mice. *Nat. Med.* **17**, 566 (2011).

88. Hubsher, G., Haider, M. & Okun, M. S. Amantadine: the journey from fighting flu to treating Parkinson disease. *Neurology* **78**, 1096–1099. https://doi.org/10.1212/WNL.0b013e31824e8f0d (2012).

89. Tellioglu, T. & Robertson, D. Genetic or acquired deficits in the norepinephrine transporter: current understanding of clinical implications. *Expert Rev. Mol. Med.* **2001**, 1–10. https://doi.org/10.1017/S1462399401003878 (2001).

90. Pacholczyk, T., Blakely, R. D. & Amara, S. G. Expression cloning of a cocaine- and antidepressant-sensitive human noradrenaline transporter. *Nature* **350**, 350–354. https://doi.org/10.1038/350350a0 (1991).

91. Jiang, Y. & Zhang, H. Propensity score-based nonparametric test revealing genetic variants underlying bipolar disorder. *Genet. Epidemiol.* **35**, 125–132. https://doi.org/10.1002/gepi.20558 (2011).

92. Delous, M. *et al.* The ciliary gene RPGRIP1L is mutated in cerebello-oculo-renal syndrome (Joubert syndrome type B) and Meckel syndrome. *Nat. Genet.* **39**, 875–881. https://doi.org/10.1038/ng2039 (2007).

93. Sherva, R. *et al.* Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimer's Dement.* **10**, 45–52. https://doi.org/10.1016/j.jalz.2013.01.008 (2014).

94. Irish Schizophrenia Genomics, C. & the Wellcome Trust Case Control, C. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol. Psychiatry* **72**, 620–628. https://doi.org/10.1016/j.biopsych.2012.05.035 (2012).

95. Karunakaran, K. B., Chaparala, S. & Ganapathiraju, M. K. Potentially repurposable drugs for schizophrenia identified from its interactome. *bioRxiv*, 442640 (2018).

96. Rannestad, J. The regeneration of cilia in partially deciliated Tetrahymena. *J. Cell Biol.* **63**, 1009–1017 (1974).

97. Kobayashi, T. *et al.* HDAC2 promotes loss of primary cilia in pancreatic ductal adenocarcinoma. *EMBO Rep.* **18**, 334–343 (2017).

98. Valencia-Gattas, M., Conner, G. E. & Fregien, N. L. Gefitinib, an EGFR tyrosine kinase inhibitor, prevents smoke-mediated ciliated airway epithelial cell loss and promotes their recovery. *PLoS ONE* **11**, e0160216 (2016).

99. Khan, N. A. *et al.* Identification of drugs that restore primary cilium expression in cancer cells. *Oncotarget* **7**, 9975 (2016).

100. Wang, H. *et al.* Hsp90α forms a stable complex at the cilium neck for the interaction of signalling molecules in IGF-1 receptor signalling. *J. Cell. Sci.* **128**, 100–108 (2015).

101. Orii, N. & Ganapathiraju, M. K. Wiki-pi: a web-server of annotated human protein-protein interactions to aid in discovery of protein function. *PLoS ONE* **7**, e49029 (2012).

Acknowledgements

Author contributions

M.K.G.: Conceptualization, design, interactome construction and computational analysis. K.B.K. and S.C.: Study of functional enrichment and pathway associations, and literature review. Also, K.B.K.: Study of association to Alzheimer's Disease and Aging. C.W.L.: Inputs related to cilia biology. Manuscript preparation: K.B.K., S.C., M.K.G. and C.W.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-72024-4.

Correspondence and requests for materials should be addressed to M.K.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Interactome-based framework to translate disease genetic data into biological and clinical insights

3. Novel protein–protein interactions highlighting the crosstalk between hypoplastic left heart syndrome, ciliopathies and neurodevelopmental delays

Summary of this chapter

In this chapter, I demonstrate how the interactomic framework was used to derive biological insights into hypoplastic left heart syndrome (HLHS), a rare and severe subtype of congenital heart disease. Despite the discovery of HLHS-associated genes, an integrative approach to elucidate their functional consequences is still lacking. To address this, I constructed the HLHS interactome containing the experimentally verified and computationally predicted protein-protein interactions (PPIs) of HLHS-associated genes identified from a mouse mutagenesis screen and ascertained its biological validity using transcriptomes of HLHS patients. ~60% of the interactome consisted of housekeeping genes that could harbour large-effect mutations showing limited transmission, possibly explaining the genetic heterogeneity of HLHS. Netbox analysis allowed the identification of topological modules enriched with genes pertinent to HLHS biology. I further showed the network proximity of HLHS-associated genes to genes associated with diseases that occur as comorbidities with HLHS and the tissue-specificity of interactome genes for sites of extracardiac anomalies in HLHS. Finally, I showed the statistically significant intersection of the HLHS interactome with ciliopathy and microcephaly interactomes. Further characterisation of the intersecting genes revealed their biological associations, which supported clinical observations in HLHS patients. Altogether, I demonstrated the utility of the HLHS interactome as a functional landscape to integrate and analyse publicly available HLHS-related multi-omics data and derive novel insights into HLHS biology.

Contribution to this chapter (75%)

- Developed the methodology of the project, which included HLHS interactome construction, validation, functional characterisation, and network proximity and intersection analyses
- Curated all the datasets, performed all the analyses and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

Article

Novel Protein–Protein Interactions Highlighting the Crosstalk between Hypoplastic Left Heart Syndrome, Ciliopathies and Neurodevelopmental Delays

Kalyani B. Karunakaran [1] **, George C. Gabriel** [2] **, Narayanaswamy Balakrishnan** [1] **, Cecilia W. Lo** [2] **and Madhavi K. Ganapathiraju** [3,4,*]

1   Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore 560012, India; kalyanik@iisc.ac.in (K.B.K.); balki@iisc.ac.in (N.B.)
2   Department of Developmental Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15201, USA; gcg9@pitt.edu (G.C.G.); cel36@pitt.edu (C.W.L.)
3   Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15206, USA
4   Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA 15260, USA
*   Correspondence: madhavi@pitt.edu

Abstract: Hypoplastic left heart syndrome (HLHS) is a severe congenital heart disease (CHD) affecting 1 in 5000 newborns. We constructed the interactome of 74 HLHS-associated genes identified from a large-scale mouse mutagenesis screen, augmenting it with 408 novel protein–protein interactions (PPIs) using our High-Precision Protein–Protein Interaction Prediction (HiPPIP) model. The interactome is available on a webserver with advanced search capabilities. A total of 364 genes including 73 novel interactors were differentially regulated in tissue/iPSC-derived cardiomyocytes of HLHS patients. Novel PPIs facilitated the identification of TOR signaling and endoplasmic reticulum stress modules. We found that 60.5% of the interactome consisted of housekeeping genes that may harbor large-effect mutations and drive HLHS etiology but show limited transmission. Network proximity of diabetes, Alzheimer's disease, and liver carcinoma-associated genes to HLHS genes suggested a mechanistic basis for their comorbidity with HLHS. Interactome genes showed tissue-specificity for sites of extracardiac anomalies (placenta, liver and brain). The HLHS interactome shared significant overlaps with the interactomes of ciliopathy- and microcephaly-associated genes, with the shared genes enriched for genes involved in intellectual disability and/or developmental delay, and neuronal death pathways, respectively. This supported the increased burden of ciliopathy variants and prevalence of neurological abnormalities observed among HLHS patients with developmental delay and microcephaly, respectively.

Keywords: protein–protein interactions; interactome; congenital heart disease; developmental disorder; hypoplastic left heart syndrome; web application

1. Introduction

Hypoplastic left heart syndrome (HLHS) is a severe form of congenital heart disease (CHD), which is one of the most common birth defects affecting ~1% of live births and a major driver of infant mortality resulting from congenital defects [1]. CHD constitutes structural abnormalities that can affect any cardiac structure including the atria, ventricles, aorta, and pulmonary artery or the valves connecting these chambers. Examples of CHD include atrial and ventricular septal defects, conotruncal defects affecting the ventricular septum and the outflow tract, complex CHD involving disturbance of left–right patterning (e.g., transposition of the great arteries), and valvular defects including inflow (mitral and tricuspid) and outflow (aortic/pulmonic) valves.

CHD classified as left ventricular outflow tract obstructive (LVOTO) lesions comprise a constellation of structural heart defects involving obstruction of flow from the left ventricle (LV). Clinical studies have provided strong evidence of a shared genetic etiology for LVOTO lesions, such as hypoplastic left heart syndrome (HLHS), bicuspid aortic valve (BAV) and coarctation (CoA) [1]. HLHS is a complex CHD, constituting ~1.4 to 3.8% of the CHD cases and estimated to affect 1 in 5000 newborns [2]. It is characterized by underdevelopment of the structures on the left side of the heart, namely, atresia or critical stenosis of the mitral oraortic valves and hypoplasia of the left ventricle, ascending aorta and aortic arch [2]. Until ~30 years ago, infants born with this condition would have died within the first few weeksof life; 23% of the deaths occurring in the first week of life due to cardiac abnormalities have been attributed to HLHS [2,3]. The incidence of HLHS during the fetal stage could be higher. Currently, surgical palliative techniques and improved post-operative care have significantly improved survival, with ~60–70% HLHS neonates surviving for at least 5 years following repair [4–6]. Mortality, however, is highest in the first year of life, with 30% of the infants dying or requiring heart transplant before turning one year old. Nevertheless, 90% of those surviving to one year will survive long-term up to 18 years old and beyond [7].

Brain comorbidities such as corpus callosum agenesis, holoprosencephaly, microcephaly and white matter injury have been identified in HLHS neonates, and cognitive, motor and behavioral adverse outcomes such as attention-deficit hyperactivity disorder, learning disabilities, and global developmental delay have been noted among HLHS survivors [8,9]. Given the high mortality and comorbidities associated with HLHS, there is a critical need to investigate the molecular mechanism(s) of disease pathogenesis in HLHS,as only then can therapies be developed to improve outcomes.

In humans, a genetic etiology for HLHS is demonstrated by high familial aggregationof HLHS with other LVOTO defects. Thus, using a statistical framework to calculate genetic effect size, >0.9 heritability was observed for HLHS and >0.7% for HLHS associated withother cardiovascular malformations (p-value $< 1 \times 10^{-5}$) [10]. HLHS is also shown to have a complex multigenic etiology, with clinical studies suggesting a digenic etiology being the most likely [11]. Supporting such complex genetics, a large-scale mutagenesis screen in mice for mutations causing CHD recovered eight mutant lines with HLHS. None shared any genes in common, and none showed Mendelian pattern of inheritance. Together, these findings indicated HLHS has a multigenic etiology and is genetically heterogeneous.Interestingly, the recovery of the HLHS causing mutations in one HLHS mouse line, *Ohia*,confirmed a digenic etiology with mutations in two genes, *Sap130* and *Pcdha9*, shown to cause HLHS [12]. Further supporting a multigenic etiology is the finding that five of the eight HLHS mutant mouse lines had two or more genes in 10 of 14 human chromosome linkage intervals associated with HLHS, with significant enrichment observed when two or more of the mouse HLHS-associated genes were interrogated across these linkage intervals (OR 322.5; CI 24.9–4177.2; $p = 5.6 \times 10^{-10}$) [13].

CHD-associated de novo mutations in histone-modifying genes such as *KMT2D*, *CHD7*, *KDM5A*, *KDM5B*, *WDR5*, *RNF20*, *UBE2B* and *USP44* were identified in an exome sequencing study conducted with 60 HLHS cases and 264 controls by the Pediatric Cardiac Genomics Consortium (PCGC) [14]. Genome sequencing studies and genome-wide screening by comparative genomic hybridization have identified HLHS-associated variants in cardiomyopathy-associated genes such as *MYBPC3*, *RYR2* and *MYH6* [15], as well as genes associated with mechanotransduction in cardiomyocytes such as *VASP* and *TLN2* [16]. Other genes implicated in HLHS included *RBFOX2*, which mediates RNA metabolism [17], the cardiac transcription factor *PROX1* [18], the endocytic receptor *LRP2* [19], and the transcriptional regulator *POGZ* found in patients with HLHS and developmental delay [20]. However, despite the recovery of genes associated with HLHS, an integrative approach to elucidate their functional consequences is still lacking.

In the current study, we examined HLHS within the mechanistic framework of the protein–protein interaction (PPI) network or protein 'interactome'. Proteins fuel the cellular machinery, and their interactions reflect the functions that they subserve. This can be infor-

mative of disease mechanisms and may also help uncover higher-order relationships in the genetic architecture of complex disorders [21–23]. However, only ~145,000 PPIs (25%) out of the estimated ~600,000 PPIs estimated to exist are known from public repositories such as HPRD [24] and BioGRID [25]. Detecting these PPIs using experimental techniques such as the yeast two-hybrid system and co-immunoprecipitation is prohibitively time consuming and expensive. Hence, we have developed a machine learning computational method to predict PPIs called HiPPIP (High-Precision Protein–Protein Interaction Prediction). HiPPIP computes features of protein pairs such as cellular localization, molecular function, biological process membership, genomic location of the gene and gene expression in microarray experiments, and classifies the pairwise features as interacting or non-interacting based on a random forest model [21]. This method has been validated as accurate by computational evaluations [21] and experimental validations [21,26,27]. The novel PPIs predicted using HiPPIP have yielded discoveries with translational impact, including identifying the central role of cilia in CHD [12,21,28]. Here we constructed an 'HLHS interactome' with over 400 novel PPIs predicted by HiPPIP and over 1400 known PPIs. We further developed a web resource with the novel PPIs on Wiki-HLHS, an interactive webserver for exploring novel interactions relevant to HLHS proteins or pathways of interest. We demonstrate the utility of the HLHS interactome for discovering higher-order genetic architecture of HLHS based on network analysis, functional enrichment, and transcriptome analyses.

2. Materials and Methods

2.1. Compilation of HLHS-Associated Genes and Prediction of Novel Interactions

A list of 74 HLHS-associated genes was compiled from HLHS mutant mice, specifically, from 8 independent mouse lines recovered from a large-scale mouse mutagenesisscreen [12,28]. This includes all homozygous mutations identified in the 8 HLHS mouse lines and heterozygous mutations also found in the HLHS human linkage intervals. Novel PPIs of the proteins encoded by these genes were predicted using the HiPPIP model thatwe described in our earlier work [21]. Each HLHS protein (say N1) was paired with each of the other human proteins say, (M1, M2, . . . , Mn), and each pair was evaluated with the HiPPIP model. The predicted interactions of each of the HLHS proteins were extracted (namely, the pairs whose score is >0.5, a threshold which through computational evaluations and experimental validations was revealed to indicate interacting partners with high confidence). The interactome figures were created using Cytoscape [29].

2.2. Identification of Network Modules

Network modules among the HLHS proteins and their interactors were identified using Netbox [30]. Netbox reports modularity and a scaled modularity score, as compared with the modularity observed in 1000 random permutations of the subnetwork. Scaled modularity refers to the standard deviation difference between the observed subnetwork and the mean modularity of the random networks [31].

2.3. Functional Enrichment Analysis

Pathway associations of genes in the HLHS interactome were computed using Ingenuity Pathway Analysis (IPA) [32]. Statistical significance of the overlaps between genes in the HLHS interactome and pathways in the Ingenuity Knowledge Base (IKB) was computed with Fisher's exact test based on hypergeometric distribution. Biological process, cellularcomponent and molecular function (Gene Ontology [33]), pathway (Reactome [34]), disease (OMIM [35] and DisGeNET [36]) and transcription factor target (MSigDB [37]) enrichments were computed using WebGestalt [38]. WebGestalt computes the distribution of genes belonging to a particular functional category in the input list and compares it with the background distribution of genes belonging to this functional category among all the genes that belongs to any functional category in the database selected by the user. Statistical significance of functional category enrichment is computed using Fisher's exact test and

corrected using the Benjamini–Hochberg method for multiple test adjustment. Annotations with FDR-corrected p-value < 0.05 were considered significant.

2.4. Gene Expression Enrichment Analysis

The enrichment of the HLHS interactome in genes expressed in specific tissues was computed using RNA-sequencing data from 53 postnatal human tissues extracted from GTEx [39]. Two gene sets were compiled for the analysis. The first set contained genes showing high or medium expression (transcripts per million (TPM) ≥ 9) in 53 tissues, provided that they were not housekeeping genes, i.e., genes detected in all the tissues with transcripts per million 1, as identified in the Human Protein Atlas [40]. The second set contained all the genes that showed high or medium expression in the 53 tissues, irrespective of whether they were housekeeping genes or not. TPM is a metric for quantifying gene expression; it directly measures the relative abundance of transcripts. GMT files servedas inputs for the gene over-representation analysis (GSEA) that was conducted based on hypergeometric distribution. Tissue-specificity of the genes in the HLHS interactome was checked using TissueEnrich [41]. The analysis was based on tissue-specific genes compiled from GTEx [39], Human Protein Atlas [40], and Mouse ENCODE [42]. This included 'tissue- enriched genes' with at least 5-fold higher mRNA levels in a particular tissue compared to all the other tissues, 'group-enriched genes' with at least 5-fold higher mRNA levels in a group of 2–7 tissues, and 'tissue-enhanced genes' with at least 5-fold higher mRNA levels in a particular tissue compared to average levels in all tissues.

2.5. Network Overlap Analysis

Statistical significance of the overlaps between genes in the HLHS interactome and in the SARS-CoV-2-modulated host protein interactome, the ciliary interactome, the ciliopathy interactome and the microcephaly interactome was computed based on hypergeometric test.

3. Results

We compiled a list of 74 genes associated with HLHS that were previously identified from eight independent mouse lines with HLHS [12,28]. The protein interactome of theseHLHS-associated genes (or 'HLHS candidate genes') were assembled by collecting the known protein–protein interactions (PPIs) from the Human Protein Reference Database (HPRD) [24] and the Biological General Repository for Interaction Datasets (BioGRID) [25]. Additionally, we predicted novel PPIs by applying the HiPPIP algorithm, described in our earlier work [21], which is a machine learning algorithm that computes features of protein pairs, such as whether they are co-expressed, or have common biological processor molecular functional annotations, are within the same gene neighborhood, etc., and classifies the feature vector as interacting or non-interacting, using a random forest model. In a recent study, other state-of-the-art algorithms for PPI prediction were extensively evaluated and it was found that none of them reached the superior performance achievedby HiPPIP [43]. Further, seventeen novel PPIs predicted by HiPPIP in relation to other diseases were tested through experimental methods by different research groups, and all those tested were shown to be true PPIs ([21,27] and some unpublished results). Thus, we assembled the HLHS interactome with 1496 previously known PPIs (blue edges in Figure 1) and 408 novel computationally predicted PPIs (red edges in Figure 1), which altogether connected 72 of the 74 HLHS-associated genes with 1248 known interactors and 377 novel interactors (Figure 1 and Supplementary Data File S1). Among the 74 genes, only *WFDC11* and *XKR9* had neither known nor novel PPIs. HiPPIP predicted 644 PPIs of which 236 PPIs were previously known, leaving 408 PPIs to be considered as novel PPIs of the HLHS candidate genes; thus, of the 1496 known PPIs in this interactome, 236 (15.8%) were also predicted by HiPPIP, while 1260 (84.2%) were not (which is as expected, because each method, whether computational or biotechnology based, discovers some of the true PPIs in the interactome, but not all). Four genes identified to be associated

with HLHS in independent studies [15,17,18,20] were retrieved as known interactors of our HLHS candidate genes (shown in bold): *EP300*-*PROX1*, *TSC1*-*POGZ*, *HNRNPAB*-*RBFOX2* and *PSEN1*-*RYR2*.
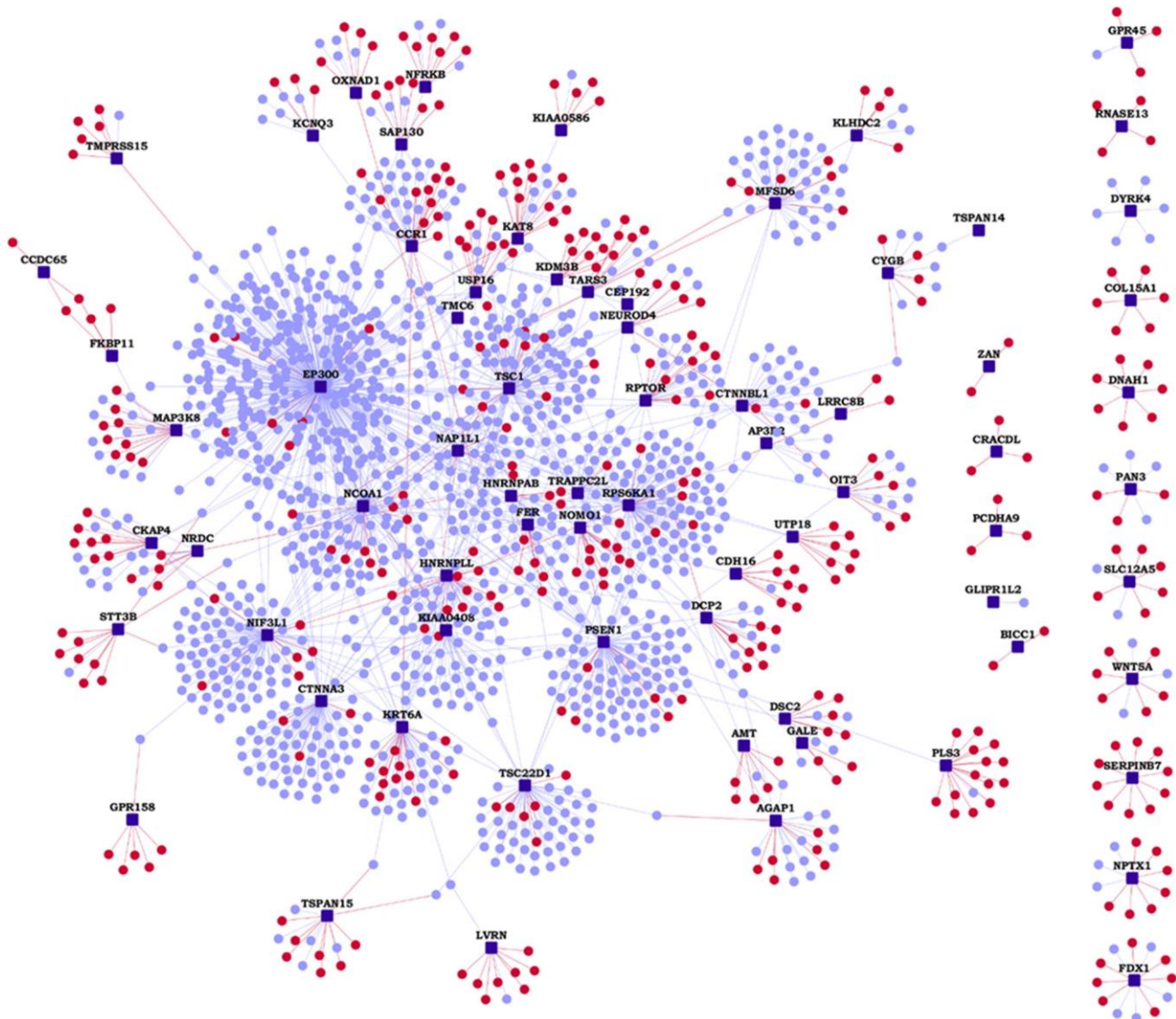
Figure 1. Hypoplastic left heart syndrome (HLHS) protein–protein interactome: Network view of the HLHS interactome is shown as a graph, where genes are shown as nodes and protein–protein interactions (PPIs) as edges connecting the nodes. HLHS-associated genes are shown as dark blue square-shaped nodes, novel interactors and known interactors as red and light blue colored circular nodes, respectively. Red edges are the novel interactions, whereas blue edges are known interactions.

3.1. Wiki-HLHS: A Webserver of HLHS PPIs

To accelerate biomedical discovery, we made the HLHS interactome publicly acces- sible with the construction of a web application called Wiki-HLHS (http://severus.dbmi.pitt.edu/wiki-HLHS, accessed on 7 January 2022). This webserver has advanced search capabilities, and for each pair of PPIs, there are side-by-side comprehensive Gene Ontology (GO) annotations, and annotations related to diseases, drugs and pathways. Here, a user can query for results such as "show me PPIs where one protein is involved in HLHS and the other is involved in microcephaly", and then see the results with the functional details of the two proteins side-by-side. This pairwise retrieval of PPIs and their biomedical associations is a unique feature of this web application not available in any other PPI webdatabase. The PPIs and their annotations are also indexed in major search engines such as

Google and Bing. A user can browse the genes in the HLHS interactome using list view of HLHS-associated genes. Novel PPIs are shown in a different color in search results.

3.2. Identification of Network Modules from the HLHS Interactome

We identified network modules in the HLHS interactome using Netbox [30], starting with the HLHS genes as core genes and adding nodes from the human interactome. The number of edges of node to core genes is statistically significant compared to its degree of interactions in the human interactome. It includes all edges between these nodes and the core genes and identifies highly interconnected modules in this network. Netbox connected 143 proteins (48 HLHS candidate genes and 95 linker proteins) into 19 modules, of which 11 modules had 4 or more nodes each (Supplementary Data File S2). Three modules had statistically significant enrichment of GO biological process terms: *TOR* signaling (p-value = 6.97×10^{-4}, odds ratio = 27.79), response to endoplasmic reticulum (ER) stress (p-value = 2.00×10^{-3}, odds ratio = 55.26) and intracellular receptor signaling pathway (p-value = 9.44×10^{-14}, odds ratio = 20.58) (Figure 2). The novel PPIs facilitated the identification of two functional modules that may be critical to HLHS pathology, namely, TOR signaling and ER stress [44,45].

Figure 2. Modules identified from the hypoplastic left heart syndrome (HLHS) interactome: Three modules that
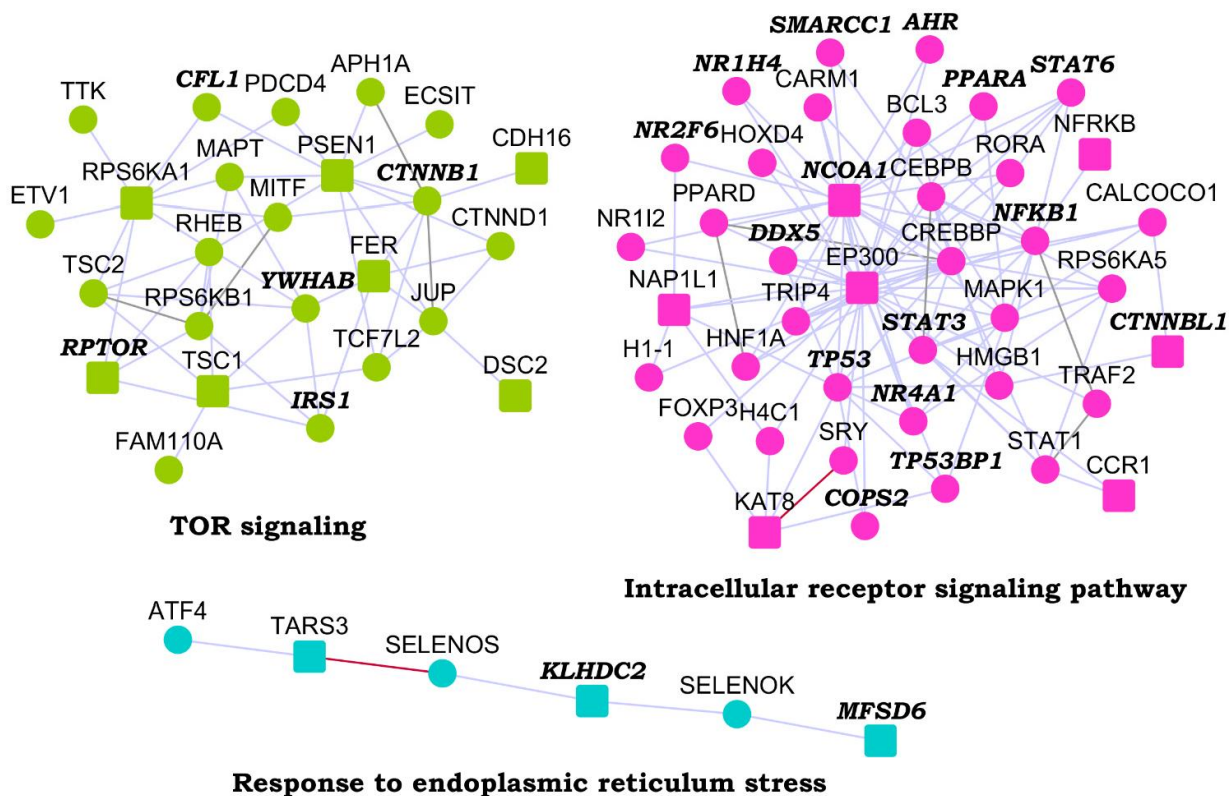
were enriched in specific GO biological processes are shown. Within each module, nodes with bold italicized labels depict genes with at least one transcriptomic evidence relevant to HLHS. HLHS-associated genes are shown as square-shaped nodes and novel interactors and known interactors are shown as circular nodes. Red edges are the novel interactions, whereas blue edges are known interactions.

3.3. Functional Enrichment for Human Diseases in the HLHS Interactome

We compiled the list of pathways for proteins in the HLHS interactome that are associated with the Ingenuity Pathway Analysis suite [32]. Selected pathways that are significantly associated with HLHS are shown in Figure 3 (complete list in Supplementary Data File S3). The Gene Ontology (GO) terms and diseases from OMIM and DisGeNET that are significantly associated with the HLHS interactome at p-value < 0.05 are shown in Supplementary Data Files S4–S8. Examination of OMIM-related genes (Figure 4A; Supplementary Data File S7) showed enrichment

associated with *non-insulin-dependent diabetes mellitus* (p-value = 1.44×10^{-7}, odds ratio = 14.49) and *insulin-dependent diabetes mellitus* (p-value = 0.02, odds ratio = 15.57) in the HLHS interactome, indicating a mechanistic link between HLHS and disease processes related to energy metabolism. Nine diabetes-associated genes that had direct interactions with ten HLHS candidates were responsible for this enrichment, including the novel interaction of the diabetes- associated *IRS1* with the HLHS candidate *NRDC* (Figure 5A; Supplementary Data File S7).Supporting this association, 8.5% of infants born to diabetic mothers have been shown tohave congenital heart defects including HLHS, double-outlet right ventricle, truncus arteriosus, transposition of the great arteries and ventricular septal defects [46]. Interestingly, also significantly enriched are genes associated with Alzheimer's disease (AD) (p-value = 1.84×10^{-5},odds ratio = 21.23) (Figure 4A; Supplementary Data File S7), with five genes associated with AD exhibiting direct interactions with nine HLHS candidate genes (Figure 5B), supporting a recent study showing increased risk of dementia among patients with congenital heart disease [47]. Finally, examination for enrichment in DisGeNET showed marked enrichment for many different types of cancer, with mammary neoplasms, adenocarcinoma and liver carcinoma being the top three diseases recovered from DisGeNET (Figure 4B; Supplementary Data File S8).
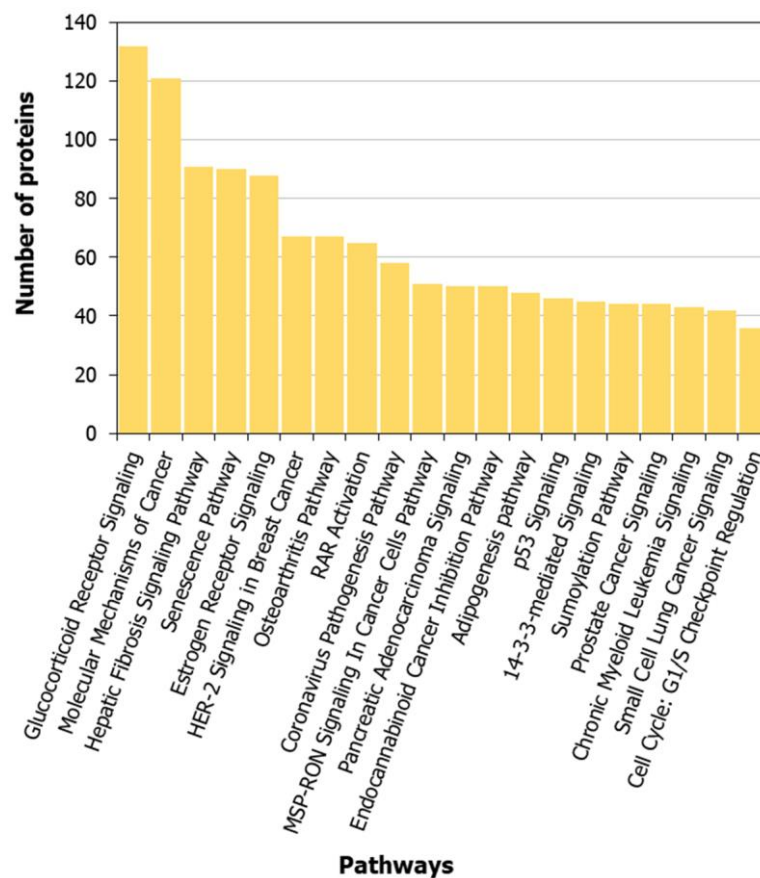
Figure 3. Pathways associated with the hypoplastic left heart syndrome (HLHS) interactome: The number of proteins from the HLHS interactome that are involved in the top 30 pathways most significantly associated with the interactome are shown.

A previous autopsy study showed 43% of HLHS patients have hepatic necrosis [48],and more recent studies have indicated a high incidence of hepatocellular carcinoma among patients having had the Fontan procedure, a third stage surgical palliation that allHLHS patients must undergo [49]. A total of 67 genes associated with liver carcinoma exhibited direct interactions with 28 HLHS candidate genes, including 7 novel interactions (listed in the format HLHS candidate gene-liver carcinoma-associated gene: *NRDC-IRS1*,*RPS6KA1-CXCL12*, *NIF3L1-CXCL12*, *STT3B-GNMT*, *ZAN-TFPI2*, *COL15A1-PDGFB* and *MFSD6-STAT1*). We noted that of the top ten diseases recovered from DisGeNET, the

top nine are cancer related, but the tenth is "insulin resistance", further supporting a link between HLHS and diabetes.
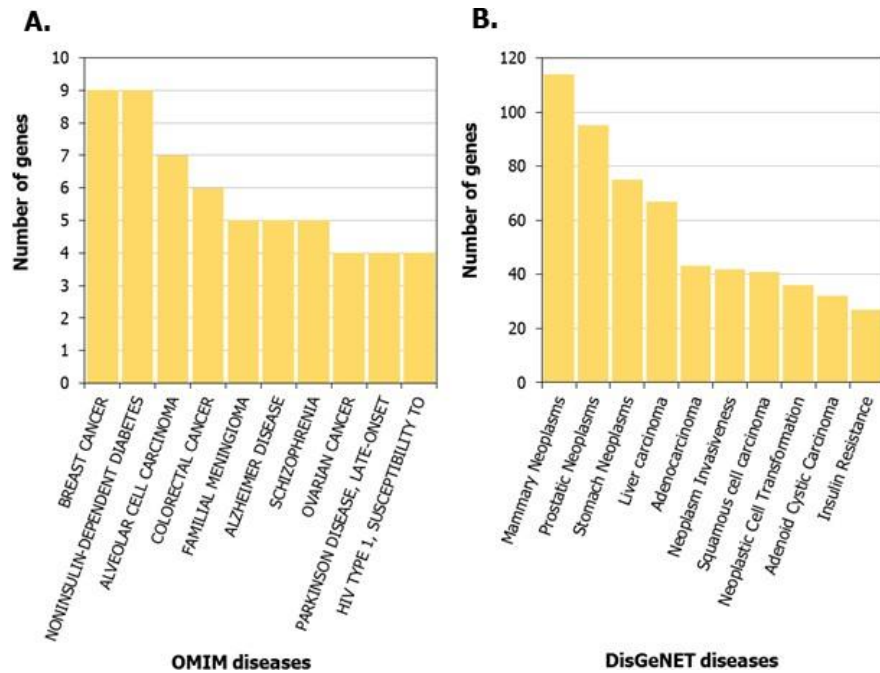
Figure 4. Diseases associated with the hypoplastic left heart syndrome (HLHS) interactome: The number of genes from the HLHS interactome that are involved in the top 10 (**A**) OMIM diseases and (**B**) DisGeNET diseases most significantly associated with the interactome are shown.
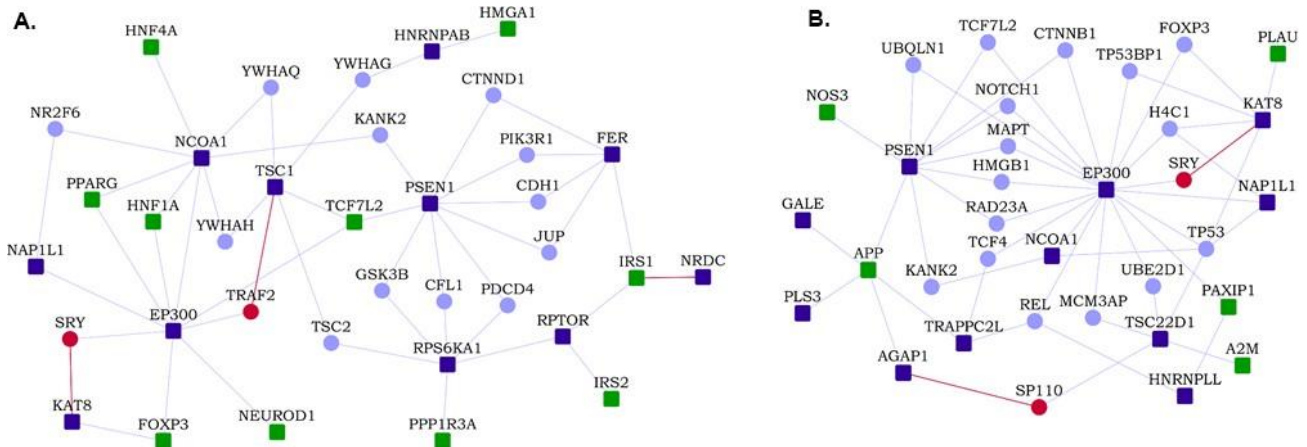
Figure 5. Network proximity of other disease-associated genes to genes associated with hypoplastic left heart syndrome (HLHS): Dark blue square-shaped nodes are HLHS-associated genes and green square-shaped nodes are diabetes-associated genes in (**A**) and Alzheimer's disease-associated genes in (**B**). Light blue nodes are known interactors and red nodes are novel interactors. Red edges are the novel interactions, whereas blue edges are known interactions.

3.4. GO Biological Process Enrichment and Overlap with HLHS Transcriptome Datasets

Among GO biological processes (Supplementary Data File S4), the most significantly enriched in the HLHS interactome was *covalent chromatin modification* (p-value $< 1 \times 10^{-15}$, odds ratio = 2.41). This observation was corroborated by the finding that the most significantly enriched GO subcellular locations included *transcription factor complex* (p-value $< 1 \times 10^{-6}$, odds ratio = 3.14) and *nuclear chromatin* (p-value $< 1 \times 10^{-6}$, odds ratio = 2.82), and among the molecular functions (Supplementary Data File S6), *DNA-binding transcription activator activity, RNA polymerase II-specific* (p-value $< 1 \times 10^{-14}$, odds ratio = 3.33) and *transcription coactivator*

activity (p-value $< 1 \times 10^{-14}$, odds ratio = 3.38). Motivated by the enrichment of transcriptional regulatory processes in our interactome and several previous studies suggesting transcriptomic changes associated with HLHS [19,50,51], we further investigated the overlap of the HLHS interactome with four HLHS-related transcriptomic datasets. We studied whether the genes in the HLHS interactome were differentially expressed or alternatively spliced in four different RNA-seq datasets comprising either tissues or cardiomyocytes derived from induced pluripotent stem cells (iPSCs) (Supplementary Data File S9).

Our analysis identified 73 novel interactors (19%) (Table 1) and overall 364 genes (21%) in the HLHS interactome that had one or more transcriptomic association (Supplementary Data File S9). Each of these four RNA-seq datasets showed consider- able, albeit statistically non-significant, overlap with the interactome. The datasets include differential expression in cardiomyocytes differentiated from iPSCs of five HLHS patients versus two controls (GSE92447 [51]), yielding 131 genes present in the interactome (odds ratio = 1.09), HLHS-right ventricle versus control-left ventricle/control-right ventricle (GSE23959 [50]), yielding seven genes in the interactome (odds ratio = 2.43), iPSC-derived cardiomyocytes at 25 days from one HLHS proband versus parents yielding 131 genes (odds ratio = 1.01) [19], and genes affected by alternative splicing in HLHS-right ventricle versus control-right ventricle/control-left ventricle (GSE23959 [50]), yielding 136 overlapping genes (odds ratio = 1.02). Though these overlaps are not statistically significant at the systems level, the individual genes and their transcriptomic evidence may provide biologically relevant information about the etiology of HLHS. We did not observe statistically significant overlaps between HLHS transcriptomic data and the HLHS interactome, despite examining all the available RNA-seq datasets. This could be attributed to the tendency of iPSC-derived cell lines to exhibit donor-specific gene expression patterns [52] and sample sizes (in these transcriptomic studies) that are not large enough to capture the genetic heterogeneity of HLHS [53]. Additionally, transcriptomic, proteomic and phenotypic equivalences between these external datasets and the murine-derived gene set used for interactome construction should be interpreted cautiously, unless the biological levels are comprehensively characterized and a clear equivalence of factors is demonstrated in both the species [54]. Nevertheless, it has been shown that the HLHS mouse lines (from which the gene set used for interactome construction was identified) had mutations in two or more genes in 10 of 14 human chromosome intervals associated with HLHS and left ventricle outflow obstruction [12]. In addition, essential features that characterize HLHS, such as hypoplasia of the left ventricle, aorta, and mitral valve, were confirmed in the recovered mouse mutant lines [12].

We studied the tissue-specific expression of the HLHS interactome genes using RNA- seq data of 53 postnatal human tissues obtained from GTEx [39], with and without the inclusion of housekeeping genes from the Human Protein Atlas [40]. An expression of more than nine transcripts per million (TPM) is considered high/medium expression. A total of 9634 genes detected in all the tissues with TPM ≥ 1 were considered as housekeeping genes. Statistical significance of the enrichment was computed using Fisher's exact test and corrected using the Benjamini–Hochberg multiple test adjustment. Compared to when housekeeping genes were excluded, as shown in Figure 6A, the HLHS interactome genes were significantly enriched in several tissues (Figure 6B)—including in heart-related tissues such as the atrial appendage, coronary artery, aorta and the left ventricle—when housekeeping genes were included in the analysis. This could indicate that a large number of genes in the HLHS interactome were housekeeping genes. In line with this, we found that 60.5% (1028 genes) of the interactome was comprised of housekeeping genes, a highly statistically significant over-enrichment (p-value $= 2.03 \times 10^{-10}$) of 1.14-fold compared to expectations (906 genes). We also noted that the left ventricle showed a lower statistical significance of enrichment compared with the other three heart-related tissues (atrial appendage, coronary artery and aorta).

Table 1. Novel interactors in the hypoplastic left heart syndrome (HLHS) interactome with biological evidence related to HLHS: The table shows those novel interactors of HLHS-associated genes that have 2 or more HLHS-related biological evidence associated with them. The complete list of biological evidence for all the genes in the interactome can be found in Data File S9.

Gene	Differentially Expressed in Cardiomyocytes from iPSCS of 5 HLHS Patients Versus 2 Controls (GSE92447)	Differentially Expressed in HLHS-Right Ventricle Versus Control-Left Ventricle/Control-Right Ventricle (GSE23959)	Affected by Alternative Splicing in HLHS-Right Ventricle Versus Control-Right Ventricle/Control-Left Ventricle (GSE23959)	Differentially Expressed in 25 Days old iPSC-Derived Cardiomyocytes from1 HLHS Proband Versus Parents	Total Count
DBN1	✓		✓	✓	3
MYL9		✓		✓	2
ASCC3			✓	✓	2
CDH5	✓			✓	2
CKB	✓			✓	2
GART	✓		✓		2
PWP1			✓	✓	2
TFPI2	✓			✓	2
THBS1	✓			✓	2

Figure 6. Tissue enrichment of the genes in the hypoplastic left heart syndrome (HLHS) interactome: The tissue enrichment patterns of the HLHS interactome were identified using the gene expression profiles of 53 postnatal human tissues extracted from GTEx. Enrichment was assessed by considering
(**A**) any gene that showed high/medium expression in the tissues (transcripts per million (TPM) ≥ 9) and (**B**) any gene that showed high/medium expression in the tissues, except for housekeeping genes (detected in all the tissues with TPM ≥ 1). Statistical significance of tissue enrichment was computed using Fisher's exact test and corrected using the Benjamini–Hochberg method for multiple test adjustment. It can be observed that the HLHS interactome genes showed a higher statistical significance of enrichment in several tissues (including the heart-related tissues shown as red data points) when housekeeping genes were considered as shown in (**B**) compared to when housekeeping genes were excluded as shown in (**A**), indicating that housekeeping genes could be over-represented in the interactome.

We further employed the TissueEnrich tool to examine the tissue-specificity of the genes in the HLHS interactome based on expression data from GTEx [39], Human Protein Atlas [40] and Mouse ENCODE [42] (Figure 7A–C). Genes with an expression level greater than 1 TPM (transcripts per million) and relative expression at least 5-fold higher in a particular tissue (tissue-enriched) or a group of two to seven tissues (group-enriched) were considered [55]. As expected from an interactome showing an over-enrichment of housekeeping genes, the HLHS interactome did not show any statistically significant tissue-

specific enrichment. However, it was noteworthy that six tissues—placenta, skin, liver, lung, brain and testis—showed large overlaps with the HLHS interactome according to data from at least two of the databases and appeared among their lists of top ten tissues (in terms of the number of tissue-specific genes found in the interactome, and not the statistical significance of this overlap) (Figure 7A–C). Ten HLHS candidates had novel PPIs with eleven heart- specific proteins across the three databases (HLHS candidates are shown in bold): *NFRKB*- *OPCML*, *NEUROD4*-*IL23A*, *TSC1*-*PAEP*, *OIT3*-*PLA2G12B*, *CDH16*-*CDH5*, *SLC12A5*-*JPH2*, *SLC12A5*-*MYL9*, *DSC2*-*FHOD3*, *GALE*-*PLA2G5*, *PLS3*-*NRSN1* and *TSPAN15*-*ADAMDEC1*.
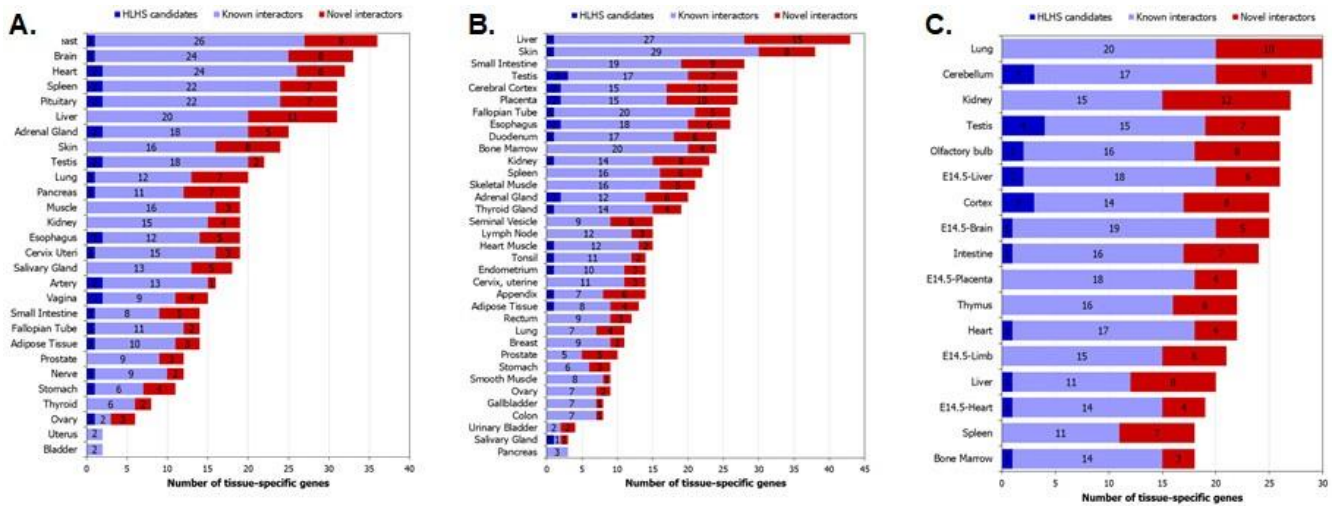
Figure 7. Tissue-specificity of hypoplastic left heart syndrome (HLHS) interactome genes: The graphs show the number of genes from the interactome that exhibit tissue-specificity according to data from (**A**) GTEx, (**B**) Human Protein Atlas and (**C**) mouse ENCODE. The genes show at least 5-fold higher expression in a tissue ('tissue-enriched') or a group of 2–7 tissues compared to all the other tissues('group-enriched').

3.5. HLHS and Developmental Delay

Cilia are dynamic projections on cellular surfaces, which detect a wide variety of cues from the environment and transduce signals into the cell to regulate physiological and developmental processes. A genetic screen for recessive CHD-associated mutations had highlighted the role of cilia-transduced cell signaling in CHD [28]. Thirty-four of the sixty-one CHD-associated genes recovered in this screen were cilia-related. To examine for possible ciliary connection to HLHS, we computed the overlap of the HLHS interactome with the interactome of ciliary proteins containing a total of 1665 proteins and 1776 PPIs [56].The interactomes shared a highly statistically significant overlap (p-value = 3.97×10^{-25}) of 284 genes, and 30% of the overlapping genes were differentially expressed or were affected by alternative splicing in at least one of the four HLHS RNA-seq datasets described in the previous section [19,50,51]. The Reactome pathways *gene expression*, *SUMOylation* and *cell cycle* were enriched among the shared genes by 2.4-fold, 6.4-fold and 3.4-fold, respectively.Next, we collected a list of 187 genes that have been implicated in 35 ciliopathies from Reiter et al. [57], and assembled its interactome containing 2486 proteins and 3022 interactions. We found that 28% of the HLHS interactome overlapped with 19% of the ciliopathy interactome (473 genes), a highly statistically significant overlap (p-value = 1.18×10^{-58})with an enrichment ratio of 2-fold compared to expectations (234 genes) (Figure 8 and Supplementary Data File S10). We also found that 67 HLHS-associated genes, 157 ciliopathy- associated genes and 3 genes associated with both (*CCDC65*, *KIAA0586* and *DNAH1*) were connected via 841 intermediate interactors. Eight direct known interactions were found between HLHS candidates and ciliopathy-associated genes (HLHS candidates are shown in bold): *TSC22D1*-*UNC119*, *RPTOR*-*CILK1*, *MFSD6*-*TMEM237*, *EP300*-*CRX*, *CTNNA3*-*CRX*, *CTNNA3*-*FAM161A*, *NIF3L1*-*NME7* and *TSC1*-*GLIS2*, and one direct novel interaction, *RP-*

TOR-*CCDC40*. We identified the top 30 GO biological processes that were significantly associated with the HLHS and the ciliopathy interactomes and computed the number of genes that were exclusively found in the HLHS/ciliopathy interactomes or shared between the two interactomes in each of these processes (Supplementary Figures S1 and S2). *Regulation of DNA-binding transcription factor activity* was enriched 4-fold among the shared genes between the two interactomes (*p*-value = 3.96×10^{-12}). Speculating that transcription fac- tor (TF) activity could be a major factor in the crosstalk between HLHS and ciliopathy, wesought to identify the TFs whose target genes were significantly enriched among the genes shared by the HLHS and ciliopathy interactomes. The enrichment analysis was performed using WebGestalt [38] and based on curated TF-target gene sets in MSigDB [37]. The targets of *CREBP1* and *ALX4* showed significant over-enrichment (*p*-values of 2.57×10^{-4} and 1.18×10^{-3}) of 8.68-fold and 38.85-fold (five genes and two genes) compared to expectations. The targets of *CREBP1* found from among the genes shared between the HLHS and ciliopathy interactomes were *EP300*, *PRNP*, *SMAD3*, *SUMO1* and *TBX6*, while the targets of *ALX4* were *JUN* and *TCF7L2*.
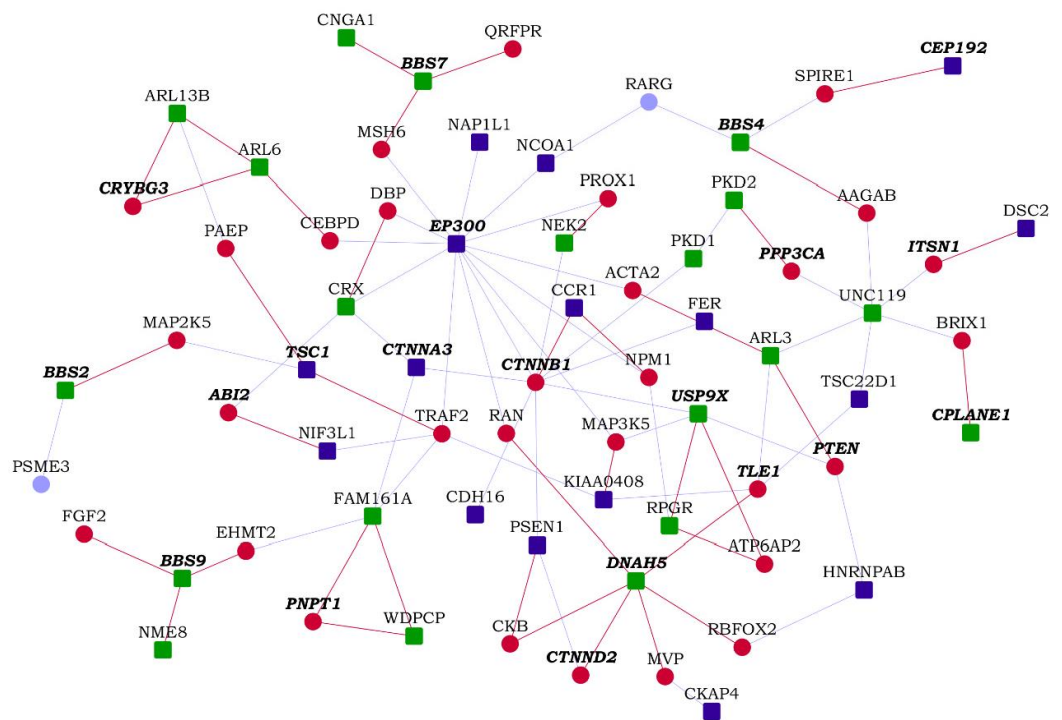
Figure 8. A partial network view of novel protein–protein interactions (PPIs) interconnecting hypoplastic left heart syndrome (HLHS) genes with ciliopathy-associated genes: Genes are shown as nodes and PPIs as edges. As the integrated HLHS and ciliopathy interactome is very large, only a partial view incorporating genes that are associated with intellectual disability and/or developmental delay (ID/DD) and the novel interactors of HLHS-associated genes/ciliopathy-associated genes are shown. Legend— square-shaped dark blue nodes: HLHS-associated genes; square-shaped green nodes: ciliopathy-associated genes; nodes with bold and italicized labels: ID/DD-associated genes; red nodes/edges: novel interactors/interactions; light blue nodes/edges: known interactors/interactions.

HLHS patients having developmental delay as a comorbidity have been shown to have a higher burden of ciliopathy variants compared to HLHS patients without developmental delay, with a summative C-score of 4.05 versus 2.02 (*p*-value of the observed difference < 0.01). Summative C-score is a standardized value used to assess the level of gene disruption in a condition; in this specific study, the C-score of 4.07 was identified as a threshold at which 50% of pathogenic variants and 3% of benign variants were retained [58].This prompted us to compare the enrichment of genes implicated in developmental delay

in the HLHS and ciliopathy interactomes, and specifically among the genes uniquely found in the HLHS/ciliopathy interactomes and those shared between the HLHS and ciliopathy interactomes. A total of 703 genes harboring loss-of-function and missense variants linked to intellectual disability and/or developmental delay (ID/DD) were collected from the Developmental Brain Disorder Gene Database [59]. These ID/DD genes showed an overlap of higher significance with the HLHS interactome (p-value = 3.77×10^{-8}, odds ratio = 1.67) compared with the ciliopathy interactome (p-value = 7.61×10^{-3}, odds ratio = 1.23). Additionally, significant enrichment for ID/DD was shown by genes uniquely found in the HLHS interactome (p-value = 7.16×10^{-6}, odds ratio = 1.65) and genes shared by the HLHS and ciliopathy interactomes (p-value = 1.93×10^{-3}, odds ratio = 1.73) (Figure 8), but not by the genes uniquely found in the ciliopathy interactome. The HLHS candidate *KIAA0586* was associated with ciliopathy as well as ID/DD. Eight other HLHS candidates were linked to ID/DD (i.e., they also harbored ID/DD-associated variants), namely, *TSC1*, *KDM3B*, *CEP192*, *TRAPPC2L*, *EP300*, *CTNNA3*, *KCNQ3* and *AP3B2*. We identified 19 novel PPIs of HLHS candidates with ID/DD genes (HLHS candidates are shown in bold): **SERPINB7**-*OGDH*, **SERPINB7**-*PIGN*, **NOMO1**-*NDE1*, **NOMO1**-*KCNA1*, **NCOA1**- *PPP1CB*, **FER**-*HSD17B4*, **CKAP4**-*POLR3B*, **AMT**-*IMPDH2*, **RPS6KA1**-*ASCC3*, **STT3B**- *SMARCC1*, **DSC2**-*ITSN1*, **AGAP1**-*SLC19A3*, **OXNAD1**-*SLC6A1*, **MFSD6**-*HIBCH*, **KAT8**- *PRRT2*, **HNRNPAB**-*SYNCRIP*, **NIF3L1**-*ABI2*, **PLS3**-*CUL4B* and **CCR1**-*CTNNB1*.

These results implicate the cilium as a potential focal point for examining HLHS etiology and its comorbid relationships with ciliopathy, intellectual disability and developmental delay.

3.6. HLHS and Microcephaly

Severe neurological outcomes such as seizure activity, ischemia, and hemorrhage in HLHS patients are more prevalent with neonatal microcephaly than without (43% versus 4%, p-value = 0.02); the prevalence is 33% in HLHS patients with fetal microcephaly (p-value = 0.06) [60]. This prompted us to examine their interconnections. A total of 84 genes associated with microcephaly were collected from the MONARCH database [61] and the microcephaly interactome containing 1867 proteins and 2081 interactions was assembled. Sixty-two HLHS candidates were connected to 77 microcephaly genes via 652 intermediate interactors. Five direct known interactions were found between HLHS candidates and microcephaly-associated genes (HLHS candidates are shown in bold): **TSC1**-*POGZ*, **TSC1**-*CDK6*, **CTNNBL1**-*STAMBP*, **TRAPPC2L**-*TRAPPC6B* and **TSC22D1**-*QARS1*. We found that 24% of the HLHS interactome overlapped with 22% of the microcephaly interactome (405 genes), a highly statistically significant overlap (p-value = 2.39×10^{-65}) with an enrichment ratio of 2.31-fold compared to expectations (176 genes) (Figure 9 and Supplementary Data File S11). We identified the top 30 GO biological processes that were significantly associated with the HLHS and the microcephaly interactomes and computed the number of genes that were exclusively found in the HLHS/microcephaly interactomes or shared between the two interactomes in each of these processes (Supplementary Figures S3 and S4).

Neuron death was enriched 4-fold (32 genes) among the genes common to the HLHS and microcephaly interactomes (p-value = 5.13×10^{-10}). It was also interesting to note that neuron death, neurodegeneration and other related processes and diseases appeared as enriched terms in the HLHS interactome across several functional categories. We extracted a total of 95 genes from the HLHS interactome belonging to these categories, specifically, 5 genes linked to Alzheimer's disease (OMIM ID: 104300), 4 genes in late-onset Parkinson's disease (OMIM ID: 168600), 5 genes linked to neurofibrillary degeneration (DisGeNET ID: C0085400), 17 genes in neurodegenerative disorders (DisGeNET ID: C0524851) and 79 genes linked to neuron death (GO ID: 0070997). Two pathways showed high statistical significance and enrichment ratio among these 95 genes, namely, *constitutive signaling by AKT1 E17K in cancer* (p-value = 5.94×10^{-7}, odds ratio = 32.83) and *intrinsic pathway for apoptosis* (p-value = 6.94×10^{-7}, odds ratio = 21.32 folds). A total of 10 novel interactors of HLHS candidates were found among these 95 genes (*ITSN1*, *RHOA*, *NQO1*, *CTNNB1*, *TRAF2*,

MAP3K5, *PRPH*, *UBQLN2*, *UNC5B* and *FUS*). Five of these novel interactors (*ITSN1*, *RHOA*, *TRAF2*, *CTNNB1* and *UNC5B*) seemed to be involved in *death receptor signaling/apoptosis*, *axon guidance/EPH-Ephrin signaling* and/or *developmental biology*, and their novel PPIs with HLHS candidates were as follows (HLHS candidates are shown in bold): **DSC2**-*ITSN1*, **AMT**-*RHOA*, **CCR1**-*CTNNB1*, **TSC1**-*TRAF2* and **OIT3**-*UNC5B*.
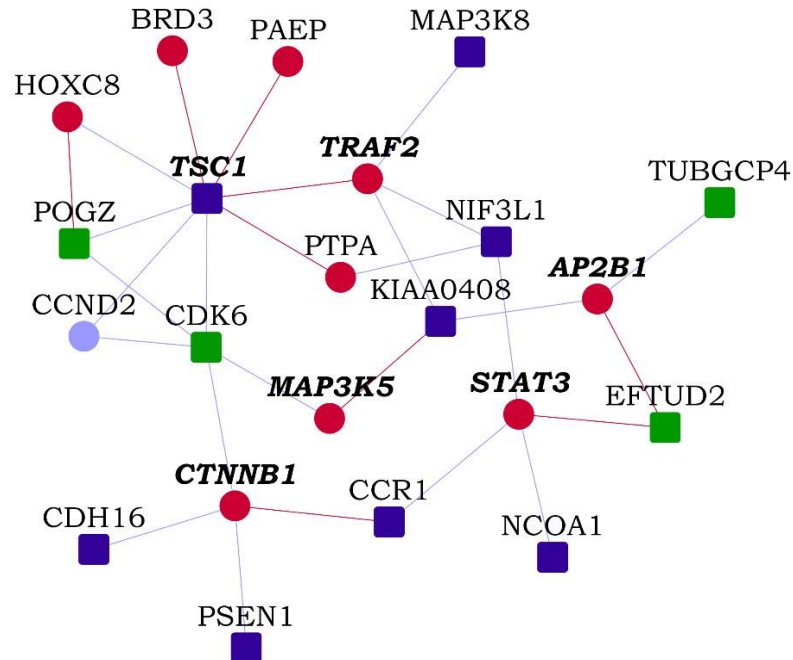
Figure 9. A partial network view of novel protein–protein interactions (PPIs) interconnecting hypoplastic left heart syndrome (HLHS) genes with microcephaly-associated genes: Genes are shown as nodes and PPIs as edges. As the integrated HLHS and microcephaly interactome is very large, only apartial view incorporating genes that are involved in neuronal death processes and the novel interactors of HLHS-associated genes/microcephaly-associated genes are shown. Legend—square-shaped dark blue nodes: HLHS-associated genes; square-shaped green nodes: microcephaly-associated genes; nodes with bold and italicized labels: genes involved in neuron death; red nodes/edges: novel interactors/interactions; light blue nodes/edges: known interactors/interactions.

Genes in the HLHS interactome linked to neuronal death processes may serve as potential candidates for examining the genetic basis of microcephaly in HLHS patients and the increased prevalence of poor neurological outcomes in these patients. The intriguing link between neurodegenerative processes and HLHS is another result that warrants closer inspection especially in light of the recent finding that adults with congenital heart disease show an increased risk of dementia and early onset dementia, particularly amongst patients with complex lesions [47].

4. Discussion

In this study, we adopted a protein interactome analysis approach to study HLHS- associated genes. The interactome analysis framework postulates that diseases develop when PPIs are perturbed by genetic mutations or aberrant expression of genes/proteins,ultimately leading to disrupted cellular functions [62]. Extensive interconnectivity and intraconnectivity of the network components in the PPI network suggest that the effects of such perturbations may spread to other proteins, encoded by genes that do not harbor any disease-associated alterations, through the network of their interactions, posing deeper im- plications for disease development [62]. In this study, we assembled the HLHS interactome by supplementing previously known protein PPIs with computationally predicted PPIs, which are deemed accurate, and provided valuable insights into etiology through network and enrichment analysis.

We made the PPIs, including the novel PPIs, available on a searchable webserver to enable biologists to study the PPI of their interest (http://severus.dbmi.pitt.edu/wiki- HLHS, accessed on 7 January 2022). Our website provides advanced search capabilities, which allows a user to ask questions that will help generate testable hypotheses around individual PPIs. The full text of the PPIs and their annotations will also be indexed in internet search engines, so that biologists searching in Google, Bing, etc., will find this content. System-level analysis of the interactome with transcriptomic or proteomic data may help to identify its functional landscape. Investigation of individual PPIs will accelerate the understanding of disease biology by several years.

More than 60% of the HLHS interactome, including 51% (38) of the HLHS-associated genes (or 'core genes' used for interactome construction), was composed of genes that are constitutively expressed in all the tissues (i.e., housekeeping genes). It has been reported that none of the HLHS-associated mutations harbored by the core genes were shared among the HLHS mutant mouse lines [12]. The preponderance of housekeeping genes among the core genes as well as the HLHS interactome as a whole could explain this genetic heterogeneity. The transmission of mutations in housekeeping genes may be stymied due to their roles in sustaining essential cellular functions, whose perturbation may result in lethality or reduction in reproductive fitness [1]. Although mutations in housekeeping genes are expected to give rise to phenotypes that affect multiple tissues, it is possible that they give rise to cardiac-restricted phenotypes due to complex regulatory influences stemming from epigenetic, epistatic and protein interactions. Further interactome-based investigations driven by this observation (i.e., the enrichment of housekeeping genes in the HLHS interactome) such as those examining vulnerability to network perturbations, and compensatory mechanisms counteracting them, may provide interesting insights.

The HLHS interactome did not show statistically significant enrichment for specific ex- pression in any tissues, as can be expected from the over-enrichment of housekeeping genes in the interactome. However, three out of the six tissues that contained the greatest number of tissue-specific interactome genes have also been documented as sites of extracardiac anomalies in HLHS, namely, the placenta, liver and brain. Specifically, placental abnormalities have been noted in pregnancies that involved fetal HLHS [63]. Increased occurrence of hepatic necrosis has been noted among patients with infantile coarctation of the aorta and HLHS compared with patients having other cardiac defects (38% vs. 6%) [64]. The prevalence of brain abnormalities among HLHS neonates and survivors is well-documented [8,9]. The total number of tissue-elevated genes was highly variable among the six tissues that showed the largest overlaps with the HLHS interactome. For example, 2709, 1987, 981, 593, 288 and 197 tissue-elevated genes were found in these tissues, i.e., brain, testis, liver, skin, placenta and lung, respectively, according to the Human Protein Atlas data [40]. Tissues showing a relatively lower number of tissue-elevated genes such as the placenta and lung, as well as those with a higher number of tissue-elevated genes such as the brain, testis, liver and skin, showed overlaps with HLHS interactome. Hence, the overlaps may not have been skewed in relation to the total number of tissue-elevated genes in the tissues. Nevertheless, these statistically non-significant results, which are derived from transcriptomic analysis of the human orthologs of mouse genes, must be interpreted with caution (and after further analysis) in the context of HLHS.

We showed that the interactome of ciliopathy-associated genes shared a significant overlap with the HLHS interactome and that transcription regulation may be over-enriched among these common genes. The targets of transcription factors *CREBP1* and *ALX4* were identified to be significantly enriched among the shared genes. *CREBP1* (also known as *ATF2*) has been shown to regulate the expression of five genes, namely, *EP300* (an HLHS-associated gene sharing a direct interaction with the ciliopathy gene *CRX*), *SUMO1* (a shared interactor having a novel PPI with the ciliopathy-associated gene *MAK* and a known PPI with the HLHS-associated gene *NCOA1*) and three known shared interactors of HLHS and ciliopathy genes (*PRNP*, *SMAD3* and *TBX6*). *ATF2* is involved in cardiomyocyte

differentiation [65]. Future studies could concentrate on the role played by *ATF2* and its targets in the shared etiology of ciliopathies and HLHS.

We also showed the preferential enrichment of genes involved in intellectual disability and/or developmental delay (ID/DD) among genes unique to the HLHS interactome and genes shared between the HLHS and ciliopathy interactomes (in comparison with genes unique to the ciliopathy interactome). This finding was in line with the observation of increased ciliopathy variant burden among HLHS patients with developmental delay [58]. Additionally, we provided a list of 19 direct novel PPIs between HLHS-associated and ID/DD genes that may be biochemically validated. For example, *OGDH* is an ID/DD gene that is critical to the tricarboxylic acid cycle and found in the mitochondrial matrix. Loss of *OGDH* has been shown to lead to neurodegeneration [66]. This gene shows high expression in the left ventricle and in brain regions such as the olfactory bulb, hippocampus, cerebellum and pons [67]. This evidence supports *OGDH* as a potential candidate for future studies on the comorbidity of HLHS and ID/DD.

We predicted five direct novel PPIs between HLHS- and microcephaly-associated genes. In addition, genes associated with HLHS and microcephaly share several common interactors that are significantly enriched for neuronal death pathways. This suggests a mechanistic basis for their comorbidity and the increased prevalence of neurological abnormalities among HLHS patients with microcephaly [60]. The over-representation of neurodegenerative disease-associated genes and processes in the HLHS interactome should be investigated, with a focus on the potentially pleiotropic roles of the AKT1-mediated pathways and the intrinsic apoptotic pathway in HLHS and neurodegeneration. The ten direct PPIs between HLHS- and diabetes-associated genes can be used to examine their joint genetic basis and the increased risk of developing HLHS seen among infants born to diabetic mothers [46].

We studied the associations of ciliopathy and microcephaly to the HLHS interactome (on a case-by-case basis) because of their specific relevance to HLHS, namely, (a) increased burden of ciliopathy variants among HLHS patients with developmental delay [58] and
(b) increased prevalence of neurological abnormalities among HLHS patients with microcephaly [60]. Further studies may be required to systematically compare the associations of all the phenotypes relevant to HLHS. Nevertheless, we constructed the interactomes of two other disorders that are comorbid with HLHS, namely, chronic kidney disease [68] and cardiovascular disease [69], and compared their overlaps with that exhibited by ciliopathy and microcephaly interactomes. The interactomes of 12 expert-curated chronic kidney disease (CKD)-associated genes and 43 cardiovascular disease (CVD)-associated genes compiled from DisGeNET (with a gene–disease association score > 0.2) showed statistically significant overlaps with the HLHS interactomes (*p*-values of 6.97×10^{-6} and 5.47×10^{-30}). However, fewer genes were shared by the HLHS interactome with CKD (33 genes) and CVD (179 genes) interactomes in comparison with the ciliopathy (473 genes) and microcephaly (405 genes) interactomes. In summary, our study provides evidence for the utility of the HLHS interactome in investigating various HLHS comorbidities and the functional consequences of the genes harboring HLHS-associated mutations. These results will directly inform and catalyze future investigations on the molecular basis of HLHS and biomedical studies seeking to improve clinical interventions in HLHS.

5. Conclusions

Knowledge on the exact mechanistic basis of HLHS is limited despite a steady increase in the generation of CHD- and HLHS-related data. In this scenario, the HLHS interactome will serve as a functional landscape to integrate and analyze publicly available HLHS-related multi-omics data and generate new hypotheses that will allow biologists to prioritize pathways and drugs for experimental testing and the developmental of new avenues for therapeutic interventions. To facilitate analysis by both computational and biomedical scientists, the HLHS interactome is being released via an interactive webserver called Wiki-HLHS.

References

1. Zaidi, S.; Brueckner, M. Genetics and genomics of congenital heart disease. *Circ. Res.* **2017**, *120*, 923–940. [CrossRef] [PubMed]
2. Gobergs, R.; Salputra, E.; Lubaua, I. Hypoplastic left heart syndrome: A review. *Acta Med. Litu.* **2016**, *23*, 86–98. [CrossRef] [PubMed]
3. Šamánek, M.; Slavík, Z.; Zbořilová, B.; Hroboňová, V.; Voříšková, M.; Skovranek, J. Prevalence, treatment, and outcome of heart disease in live-born children: A prospective analysis of 91,823 live-born children. *Pediatr. Cardiol.* **1989**, *10*, 205–211. [CrossRef] [PubMed]
4. Hamzah, M.; Othman, H.F.; Baloglu, O.; Aly, H. Outcomes of hypoplastic left heart syndrome: Analysis of National Inpatient Sample Database 1998–2004 versus 2005–2014. *Eur. J. Pediatr.* **2019**, *179*, 309–316. [CrossRef] [PubMed]

5. D'Udekem, Y.; Iyengar, A.J.; Galati, J.C.; Forsdick, V.; Weintraub, R.G.; Wheaton, G.R.; Bullock, A.; Justo, R.N.; Grigg, L.E.; Sholler, G.F. Redefining expectations of long-term survival after the Fontan procedure: Twenty-five years of follow-up from the entire population of Australia and New Zealand. *Circulation* **2014**, *130*, S32–S38. [CrossRef] [PubMed]

6. Alsoufi, B.; Mori, M.; Gillespie, S.; Schlosser, B.; Slesnick, T.; Kogon, B.; Kim, D.; Sachdeva, R.; Kanter, K. Impact of patient characteristics and anatomy on results of norwood operation for hypoplastic left heart syndrome. *Ann. Thorac. Surg.* **2015**, *100*, 591–598. [CrossRef]

7. Siffel, C.; Riehle-Colarusso, T.; Oster, M.E.; Correa, A. Survival of children with hypoplastic left heart syndrome. *Pediatrics* **2015**, *136*, e864–e870. [CrossRef] [PubMed]

8. Marino, B.S.; Lipkin, P.H.; Newburger, J.W.; Peacock, G.; Gerdes, M.; Gaynor, J.W.; Mussatto, K.A.; Uzark, K.; Goldberg, C.S.; Johnson, W.H., Jr. Neurodevelopmental outcomes in children with congenital heart disease: Evaluation and management: A scientific statement from the American Heart Association. *Circulation* **2012**, *126*, 1143–1172. [CrossRef] [PubMed]

9. Hinton, R.B.; Andelfinger, G.; Sekar, P.; Hinton, A.C.; Gendron, R.L.; Michelfelder, E.C.; Robitaille, Y.; Benson, D.W. Prenatal head growth and white matter injury in hypoplastic left heart syndrome. *Pediatr. Res.* **2008**, *64*, 364–369. [CrossRef]

10. Hinton, R.B.; Martin, L.J.; Tabangin, M.E.; Mazwi, M.L.; Cripe, L.H.; Benson, D.W. Hypoplastic left heart syndrome is heritable. *J. Am. Coll. Cardiol.* **2007**, *50*, 1590–1595. [CrossRef] [PubMed]

11. McBride, K.; Pignatelli, R.; Lewin, M.; Ho, T.; Fernbach, S.; Menesses, A.; Lam, W.; Leal, S.M.; Kaplan, N.; Schliekelman, P.; et al. Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability. *Am. J. Med. Genet. Part A* **2005**, *134A*, 180–186. [CrossRef] [PubMed]

12. Liu, X.; Yagi, H.; Saeed, S.; Bais, A.S.; Gabriel, G.C.; Chen, Z.; Peterson, K.A.; Li, Y.; Schwartz, M.C.; Reynolds, W.T.; et al. The complex genetics of hypoplastic left heart syndrome. *Nat. Genet.* **2017**, *49*, 1152–1159. [CrossRef] [PubMed]

13. McBride, K.L.; Zender, G.A.; Fitzgerald-Butt, S.M.; Koehler, D.; Menesses-Diaz, A.; Fernbach, S.; Lee, K.; Towbin, J.A.; Leal, S.; Belmont, J. Linkage analysis of left ventricular outflow tract malformations (aortic valve stenosis, coarctation of the aorta, and hypoplastic left heart syndrome). *Eur. J. Hum. Genet.* **2009**, *17*, 811–819. [CrossRef] [PubMed]

14. Zaidi, S.; Choi, M.; Wakimoto, H.; Ma, L.; Jiang, J.; Overton, J.D.; Romano-Adesman, A.; Bjornson, R.D.; Breitbart, R.E.; Brown, K.K.; et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **2013**, *498*, 220–223. [CrossRef]

15. Theis, J.L.; Hu, J.J.; Sundsbak, R.S.; Evans, J.M.; Bamlet, W.R.; Qureshi, M.Y.; O'Leary, P.W.; Olson, T.M. Genetic Association Between Hypoplastic Left Heart Syndrome and Cardiomyopathies. *Circ. Genom. Precis. Med.* **2021**, *14*, e003126. [CrossRef]

16. Reuter, M.S.; Chaturvedi, R.R.; Liston, E.; Manshaei, R.; Aul, R.B.; Bowdin, S.; Cohn, I.; Curtis, M.; Dhir, P.; Hayeems, R.Z.; et al. The Cardiac Genome Clinic: Implementing genome sequencing in pediatric heart disease. *Genet. Med.* **2020**, *22*, 1015–1024. [CrossRef]

17. Verma, S.K.; Deshmukh, V.; Nutter, C.A.; Jaworski, E.; Jin, W.; Wadhwa, L.; Abata, J.; Ricci, M.; Lincoln, J.; Martin, J.F.; et al. Rbfox2 function in RNA metabolism is impaired in hypoplastic left heart syndrome patient hearts. *Sci. Rep.* **2016**, *6*, 30896. [CrossRef]

18. Gill, H.K.; Parsons, S.R.; Spalluto, C.; Davies, A.F.; Knorz, V.J.; Burlinson, C.E.; Ng, B.L.; Carter, N.P.; Ogilvie, C.M.; Wilson, D.I.; et al. Separation of the PROX1 gene from upstream conserved elements in a complex inversion/translocation patient with hypoplastic left heart. *Eur. J. Hum. Genet.* **2009**, *17*, 1423–1431. [CrossRef] [PubMed]

19. Theis, J.L.; Vogler, G.; Missinato, M.A.; Li, X.; Nielsen, T.; Zeng, X.-X.I.; Martinez-Fernandez, A.; Walls, S.M.; Kervadec, A.; Kezos, J.N.; et al. Patient-specific genomics and cross-species functional analysis implicate LRP2 in hypoplastic left heart syndrome. *eLife* **2020**, *9*, e59554. [CrossRef] [PubMed]

20. Homsy, J.; Zaidi, S.; Shen, Y.; Ware, J.S.; Samocha, K.E.; Karczewski, K.J.; DePalma, S.R.; McKean, D.; Wakimoto, H.; Gorham, J.; et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **2015**, *350*, 1262–1266. [CrossRef]

21. Ganapathiraju, M.K.; Thahir, M.; Handen, A.; Sarkar, S.N.; Sweet, R.A.; Nimgaonkar, V.L.; Loscher, C.E.; Bauer, E.M.; Chaparala, S. Schizophrenia interactome with 504 novel protein–protein interactions. *NPJ Schizophr.* **2016**, *2*, 16012. [CrossRef] [PubMed]

22. Lim, J.; Hao, T.; Shaw, C.; Patel, A.J.; Szabó, G.; Rual, J.-F.; Fisk, C.J.; Li, N.; Smolyar, A.; Hill, D.E.; et al. A Protein–protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration. *Cell* **2006**, *125*, 801–814. [CrossRef]

23. Sakai, Y.; Shaw, C.A.; Dawson, B.C.; Dugas, D.V.; Al-Mohtaseb, Z.; Hill, D.E.; Zoghbi, H.Y. Protein interactome reveals converging molecular pathways among autism disorders. *Sci. Transl. Med.* **2011**, *3*, 86ra49. [CrossRef] [PubMed]

24. Prasad, T.S.K.; Goel, R.; Kandasamy, K.; Keerthikumar, S.; Kumar, S.; Mathivanan, S.; Telikicherla, D.; Raju, R.; Shafreen, B.; Venugopal, A.; et al. Human protein reference database-2009 update. *Nucleic Acids Res.* **2008**, *37*, D767–D772. [CrossRef] [PubMed]

25. Stark, C.; Breitkreutz, B.-J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34*, D535–D539. [CrossRef] [PubMed]

26. Zhu, J.; Zhang, Y.; Ghosh, A.; Cuevas, R.A.; Forero, A.; Dhar, J.; Ibsen, M.S.; Schmid-Burgk, J.L.; Schmidt, T.; Ganapathiraju, M.; et al. Antiviral activity of human oasl protein is mediated by enhancing signaling of the RIG-I RNA Sensor. *Immunity* **2014**, *40*, 936–948. [CrossRef] [PubMed]

27. Karunakaran, K.B.; Yanamala, N.; Boyce, G.; Becich, M.J.; Ganapathiraju, M.K. Malignant pleural mesothelioma interactome with 364 novel protein-protein interactions. *Cancers* **2021**, *13*, 1660. [CrossRef] [PubMed]

28. Li, Y.; Klena, N.T.; Gabriel, G.C.; Liu, X.; Kim, A.J.; Lemke, K.; Chen, Y.; Chatterjee, B.; Devine, W.; Damerla, R.R.; et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature* **2015**, *521*, 520–524. [CrossRef] [PubMed]

29. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]

30. Cerami, E.; Demir, E.; Schultz, N.; Taylor, B.S.; Sander, C. Automated network analysis identifies core pathways in glioblastoma. *PLoS ONE* **2010**, *5*, e8918. [CrossRef] [PubMed]

31. Wang, Z.; Zhang, J. In search of the biological significance of modular structures in protein networks. *PLoS Comput. Biol.* **2007**, *3*, e107. [CrossRef] [PubMed]

32. Krämer, A.; Green, J.; Pollard, J., Jr.; Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014**, *30*, 523–530. [CrossRef] [PubMed]

33. Consortium, G.O. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **2004**, *32*, D258–D261. [CrossRef] [PubMed]

34. Croft, D.; Mundo, A.F.; Haw, R.; Orlic-Milacic, M.; Weiser, J.; Wu, G.; Caudy, M.; Garapati, P.V.; Gillespie, M.; Kamdar, M.R.; et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **2013**, *42*, D472–D477. [CrossRef] [PubMed]

35. Hamosh, A.; Scott, A.F.; Amberger, J.S.; Bocchini, C.A.; Valle, D.; McKusick, V.A. Online Mendelian Inheritance in Man (OMIM), aknowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **2002**, *30*, 52–55. [CrossRef]

36. Piñero, J.; Bravo, À.; Queralt-Rosinach, N.; Gutiérrez-Sacristán, A.; Deu-Pons, J.; Centeno, E.; García-García, J.; Sanz, F.; Furlong, L.I. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **2016**, *45*, gkw943. [CrossRef] [PubMed]

37. Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdóttir, H.; Tamayo, P.; Mesirov, J.P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **2011**, *27*, 1739–1740. [CrossRef] [PubMed]

38. Liao, Y.; Wang, J.; Jaehnig, E.J.; Shi, Z.; Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **2019**, *47*, W199–W205. [CrossRef]

39. Consortium, G. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **2015**, *348*, 648–660. [CrossRef]

40. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef] [PubMed]

41. Jain, A.; Tuteja, G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **2019**, *35*, 1966–1967. [CrossRef] [PubMed]

42. Stamatoyannopoulos, J.A.; Snyder, M.; Hardison, R.; Ren, B.; Gingeras, T.; Gilbert, D.M.; Groudine, M.; Bender, M.; Kaul, R.; Canfield, T. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **2012**, *13*, 1–5. [CrossRef]

43. Dunham, B.; Ganapathiraju, M.K. Benchmark Evaluation of Protein–Protein Interaction Prediction Algorithms. *Molecules* **2020**, *27*, 41. [CrossRef]

44. Gaber, N.; Gagliardi, M.; Patel, P.; Kinnear, C.; Zhang, C.; Chitayat, D.; Shannon, P.; Jaeggi, E.; Tabori, U.; Keller, G.; et al. Fetal Reprogramming and Senescence in Hypoplastic Left Heart Syndrome and in Human Pluripotent Stem Cells during Cardiac Differentiation. *Am. J. Pathol.* **2013**, *183*, 720–734. [CrossRef]

45. Xu, X.; Jin, K.; Bais, A.S.; Zhu, W.; Yagi, H.; Feinstein, T.N.; Nguyen, P.; Criscione, J.; Liu, X.; Beutner, G.; et al. Uncompensated mitochondrial mediated oxidative stress underlies heart failure in an iPSC-derived model of congenital heart disease. *Cell Stem Cell*, 2022; *in press*. [CrossRef]

46. Becerra, J.E.; Khoury, M.J.; Cordero, J.F.; Erickson, J.D. Diabetes mellitus during pregnancy and the risks for specific birth defects: A population-based case-control study. *Pediatrics* **1990**, *85*, 1–9. [CrossRef]

47. Bagge, C.N.; Henderson, V.W.; Laursen, H.B.; Adelborg, K.; Olsen, M.; Madsen, N.L. Risk of dementia in adults with congenital heart disease: Population-based cohort study. *Circulation* **2018**, *137*, 1912–1920. [CrossRef]

48. Komatsu, H.; Inui, A.; Kishiki, K.; Kawai, H.; Yoshio, S.; Osawa, Y.; Kanto, T.; Fujisawa, T. Liver disease secondary to congenital heart disease in children. *Expert Rev. Gastroenterol. Hepatol.* **2019**, *13*, 651–666. [CrossRef] [PubMed]

49. Kogiso, T.; Tokushige, K. Fontan-associated liver disease and hepatocellular carcinoma in adults. *Sci. Rep.* **2020**, *10*, 21742. [CrossRef]

50. Ricci, M.; Xu, Y.; Hammond, H.L.; Willoughby, D.A.; Nathanson, L.; Rodríguez, M.M.; Vatta, M.; Lipshultz, S.E.; Lincoln, J. Myocardial alternative RNA splicing and gene expression profiling in early stage hypoplastic left heart syndrome. *PLoS ONE* **2012**, *7*, e29784. [CrossRef]

51. Yang, C.; Xu, Y.; Yu, M.; Lee, D.; Alharti, S.; Hellen, N.; Shaik, N.A.; Banaganapalli, B.; Mohamoud, H.S.A.; Elango, R.; et al. Induced pluripotent stem cell modelling of HLHS underlines the contribution of dysfunctional NOTCH signalling to impaired cardiogenesis. *Hum. Mol. Genet.* **2017**, *26*, 3031–3045. [CrossRef] [PubMed]

52. Carcamo-Orive, I.; Hoffman, G.E.; Cundiff, P.; Beckmann, N.D.; D'Souza, S.L.; Knowles, J.W.; Patel, A.; Papatsenko, D.; Abbasi, F.; Reaven, G.M.; et al. Analysis of transcriptional variability in a large human ipsc library reveals genetic and non-genetic determinants of heterogeneity. *Cell Stem Cell* **2016**, *20*, 518–532.e519. [CrossRef] [PubMed]

53. Yagi, H.; Liu, X.; Gabriel, G.C.; Wu, Y.; Peterson, K.; Murray, S.A.; Aronow, B.J.; Martin, L.J.; Benson, D.W.; Lo, C.W. The Genetic Landscape of Hypoplastic Left Heart Syndrome. *Pediatr. Cardiol.* **2018**, *39*, 1069–1081. [CrossRef] [PubMed]

54. Breschi, A.; Gingeras, T.R.; Guigó, A.B.R. Comparative transcriptomics in human and mouse. *Nat. Rev. Genet.* **2017**, *18*, 425–440. [CrossRef] [PubMed]

55. Fagerberg, L.; Hallström, B.M.; Oksvold, P.; Kampf, C.; Djureinovic, D.; Odeberg, J.; Habuka, M.; Tahmasebpoor, S.; Danielsson, A.; Edlund, K. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteom.* **2014**, *13*, 397–406. [CrossRef]

56. Karunakaran, K.B.; Chaparala, S.; Lo, C.W.; Ganapathiraju, M.K. Cilia interactome with predicted protein–protein interactions reveals connections to Alzheimer's disease, aging and other neuropsychiatric processes. *Sci. Rep.* **2020**, *10*, 15629.

57. Reiter, J.F.; Leroux, M.R. Genes and molecular pathways underpinning ciliopathies. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 533–547. [CrossRef]

58. Geddes, G.C.; Stamm, K.; Mitchell, M.; Mussatto, K.A.; Tomita-Mitchell, A. Ciliopathy variant burden and developmental delay in children with hypoplastic left heart syndrome. *Genet. Med.* **2017**, *19*, 711–714. [CrossRef]

59. Gonzalez-Mantilla, A.J.; Moreno-De-Luca, A.; Ledbetter, D.H.; Martin, C.L. A cross-disorder method to identify novel candidate genes for developmental brain disorders. *JAMA Psychiatry* **2016**, *73*, 275–283. [CrossRef]

60. Hangge, P.T.; Cnota, J.F.; Woo, J.G.; Hinton, A.C.; Divanovic, A.A.; Manning, P.B.; Ittenbach, R.F.; Hinton, R.B. Microcephaly is associated with early adverse neurologic outcomes in hypoplastic left heart syndrome. *Pediatr. Res.* **2013**, *74*, 61–67. [CrossRef]

61. Cacheiro, P.; Haendel, M.A.; Smedley, D.; Consortium, I.M.P.; Initiative, M. New models for human disease from the International Mouse Phenotyping Consortium. *Mamm. Genome* **2019**, *30*, 143–150. [CrossRef] [PubMed]

62. Barabási, A.-L.; Gulbahce, N.; Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **2010**, *12*, 56–68. [CrossRef] [PubMed]

63. Jones, H.N.; Olbrych, S.K.; Smith, K.L.; Cnota, J.F.; Habli, M.; Ramos-Gonzales, O.; Owens, K.J.; Hinton, A.C.; Polzin, W.J.; Muglia, L.J.; et al. Hypoplastic left heart syndrome is associated with structural and vascular placental abnormalities and leptin dysregulation. *Placenta* **2015**, *36*, 1078–1086. [CrossRef] [PubMed]

64. Weinberg, A.G.; Bolande, R.P. The liver in congenital heart disease. Effects of infantile coarctation of the aorta and the hypoplastic left heart syndrome in infancy. *Am. J. Dis. Child.* **1970**, *119*, 390–394. [CrossRef]

65. Monzen, K.; Hiroi, Y.; Kudoh, S.; Akazawa, H.; Oka, T.; Takimoto, E.; Hayashi, D.; Hosoda, T.; Kawabata, M.; Miyazono, K.; et al. Smads, Tak1, and Their Common Target Atf-2 Play a Critical Role in Cardiomyocyte Differentiation. *J. Cell Biol.* **2001**, *153*, 687–698. [CrossRef]

66. Yoon, W.H.; Sandoval, H.; Nagarkar-Jaiswal, S.; Jaiswal, M.; Yamamoto, S.; Haelterman, N.A.; Putluri, N.; Putluri, V.; Sreekumar, A.; Tos, T. Loss of nardilysin, a mitochondrial co-chaperone for α-ketoglutarate dehydrogenase, promotes mTORC1 activation and neurodegeneration. *Neuron* **2017**, *93*, 115–131. [CrossRef]

67. Sadakata, T.; Furuichi, T. Identification and mRNA expression of Ogdh, QP-C, and two predicted genes in the postnatal mouse brain. *Neurosci. Lett.* **2006**, *405*, 217–222. [CrossRef] [PubMed]

68. Morgan, C.; Al-Aklabi, M.; Guerra, G.G. Chronic kidney disease in congenital heart disease patients: A narrative review of evidence. *Can. J. Kidney Health Dis.* **2015**, *2*, 27. [CrossRef]

69. Wang, T.; Chen, L.; Yang, T.; Huang, P.; Wang, L.; Zhao, L.; Zhang, S.; Ye, Z.; Chen, L.; Zheng, Z.; et al. Congenital heart disease and risk of cardiovascular disease: A meta-analysis of cohort studies. *J. Am. Heart Assoc.* **2019**, *8*, e012030. [CrossRef]

Interactome-based framework to translate disease genetic data into biological and clinical insights

4. Novel malignant pleural mesothelioma interactome with 364 novel protein-protein interactions

The experimental chapter is based on the following peer-reviewed publication:

Karunakaran, Kalyani B., Naveena Yanamala, Gregory Boyce, Michael J. Becich, and Madhavi K. Ganapathiraju. Malignant pleural mesothelioma interactome with 364 novel protein-protein interactions. Cancers 13, no. 7 (2021): 1660.

Summary of this chapter

In this chapter, I demonstrate how the interactomic framework was used to gain biological and clinically translatable insights into malignant pleural mesothelioma (MPM), an aggressive cancer affecting the pleural lining of the lungs. MPM has a median survival of ~1 year after diagnosis in the invasive phase, warranting the expeditious discovery of the underlying molecular mechanisms and therapeutics. I constructed the MPM interactome using genes that affect MPM through gene expression changes or genetic variants, or by being targeted by drugs clinically active against MPM, as the starting points, and assembling their experimentally determined and computationally predicted protein-protein interactions (PPIs), including five which were experimentally validated in this study. I demonstrated the biological validity of the MPM interactome through comparison with ten MPM-related multi-omics datasets. This effectively showed how the interactome pieces together an integrated view of the functional links among MPM-associated genes from various studies. The interactome showed enrichment for cancer-related pathways. Further, I performed a comparative analysis of the differential expression profiles of lung cancer patients and the profiles induced by drugs targeting proteins in the MPM interactome and further downstream analyses to identify five repurposable drugs for MPM. Altogether, this study provided an integrative and mechanistic framework for the functional translation of MPM-related multi-omics data.

Contribution to this chapter (75%)

- Developed the methodology of the project, which included MPM interactome construction, validation, functional characterisation and drug repurposing analysis
- Curated all the datasets, performed all the analyses (except for experimental validation of novel PPIs) and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures (except for Fig. 3), tables, supplementary files and appendix

Article

Malignant Pleural Mesothelioma Interactome with 364 Novel Protein-Protein Interactions

Kalyani B. Karunakaran [1], Naveena Yanamala [2], Gregory Boyce [2], Michael J. Becich [3] and Madhavi K. Ganapathiraju [3,4,*] (ID)

[1] Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore 560012, India; kalyanik@iisc.ac.in
[2] Exposure Assessment Branch, National Institute of Occupational Safety and Health, Center for Disease Control, Morgantown, WV 26506, USA; yanamala.naveena@gmail.com (N.Y.); omu0@cdc.gov (G.B.)
[3] Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15206, USA; becich@pitt.edu
[4] Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA 15213, USA
* Correspondence: madhavi@pitt.edu

Simple Summary: Internal organs like the heart and lungs, and body cavities like the thoracic and abdominal cavities, are covered by a thin, slippery layer called the mesothelium. Malignant pleural mesothelioma (MPM) is an aggressive cancer of the lining of the lung, where genetics and asbestos exposure play a role. It is not diagnosable until it becomes invasive, offering only a short survival time to the patient. To help understand the role of the genes that relate to this disease most of which are poorly understood, we constructed the 'MPM interactome', including in it the protein-protein interactions that we predicted computationally and those that are previously known in the literature. Five novel protein-protein interactions (PPIs) were tested and validated experimentally. 85.65% of the interactome is supported by genetic variant, transcriptomic, and proteomic evidence. Comparative transcriptome analysis revealed 5 repurposable drugs targeting the interactome proteins. We make the interactome available on a freely accessible web application, Wiki-MPM.

Abstract: Malignant pleural mesothelioma (MPM) is an aggressive cancer affecting the outer lining of the lung, with a median survival of less than one year. We constructed an 'MPM interactome' with over 300 computationally predicted protein-protein interactions (PPIs) and over 2400 known PPIs of 62 literature-curated genes whose activity affects MPM. Known PPIs of the 62 MPM associated genes were derived from Biological General Repository for Interaction Datasets (BioGRID) and Human Protein Reference Database (HPRD). Novel PPIs were predicted by applying the HiPPIP algorithm, which computes features of protein pairs such as cellular localization, molecular function, biological process membership, genomic location of the gene, and gene expression in microarray experiments, and classifies the pairwise features as interacting or non-interacting based on a random forest model. We validated five novel predicted PPIs experimentally. The interactome is significantly enriched with genes differentially ex-pressed in MPM tumors compared with normal pleura and with other thoracic tumors, genes whose high expression has been correlated with unfavorable prognosis in lung cancer, genes differentially expressed on crocidolite exposure, and exosome-derived proteins identified from malignant mesothelioma cell lines. 28 of the interactors of MPM proteins are targets of 147 U.S. Food and Drug Administration (FDA)-approved drugs. By comparing disease-associated versus drug-induced differential expression profiles, we identified five potentially repurposable drugs, namely cabazitaxel, primaquine, pyrimethamine, trimethoprim and gliclazide. Preclinical studies may be con-ducted in vitro to validate these computational results. Interactome analysis of disease-associated genes is a powerful approach with high translational impact. It shows how MPM- associated genes identified by various high throughput studies are functionally linked, leading to clinically translatable results such as repurposed drugs. The PPIs are made available on a webserver with interactive user interface, visualization and advanced search capabilities.

---

1. Introduction

Internal organs such as heart and lung, and body cavities such as thoracic and abdominal cavities, are covered by a thin slippery layer of cells called the "mesothelium". This protective layer prevents organ adhesion and plays a number of important roles in inflammation and tissue repair [1]. The mesothelia that line the heart, lung and abdominal cavity are called pericardium, pleura and peritoneum, respectively. Mesothelioma is the cancer that originates from this lining (described in detail in a recent review article [2]). Most types of mesothelioma metastasize to different locations in the body [3]. Pleural mesotheliomas account for ~90% of malignant mesotheliomas and have a short median survival, of less than 1 year [4].

Malignant pleural mesothelioma (MPM) is associated with exposure to asbestos; it has a long latency period after exposure and is conclusively diagnosable only after reaching the invasive phase [3]. It tends to cluster in families and occurs only in a small fraction of the population exposed to asbestos, suggesting the involvement of a genetic component [5]. These factors necessitate expeditious discovery of genetic predispositions, molecular mechanisms and therapeutics for the disease.

The molecular mechanisms of disease are often revealed by the protein-protein interactions (PPIs) of disease-associated genes. For example, the involvement of transcriptional deregulation in MPM pathogenesis was identified through mutations detected in *BAP1* and its interactions with proteins such as *HCF1*, *ASXL1*, *ASXL2*, *ANKRD1*, *FOXK1* and *FOXK2* [6]. PPI of *BAP1* with *BRCA1* was central to understanding the role of *BAP1* in growth-control pathways and cancer; *BAP1* was suggested to play a role in *BRCA1* stabilization [7,8]. Studies on *BAP1* and *BRCA1* later led to clinical trials of the drug vinorelbine as a second line therapy for MPM patients, and the drug was shown to have rare or moderate effects in MPM patients [9,10]. *BAP1* expression was shown to be necessary for vinorelbine activity; 40% of MPM patients in a study showed low *BRCA1* expression and vinorelbine resistance [11–13]. Further, 60% of the disease-associated missense mutations perturb PPIs in human genetic disorders [14].

Despite their importance, only about 10–15% of expected PPIs in the human protein interactome are currently known; for nearly half of the human proteins, not even a single PPI is currently known [15]. Due to the sheer number of PPIs remaining to be discovered in the human interactome, it becomes imperative that biological discovery be accelerated by computational and high-throughput biotechnological methods. We developed a computational model, called HiPPIP (high-precision protein-protein interaction prediction) that is deemed accurate by computational evaluations and experimental validations of 18 predicted PPIs, where all the tested pairs were shown to be true PPIs ([16,17] and current work, and other unpublished works). HiPPIP computes features of protein pairs such as cellular localization, molecular function, biological process membership, genomic location of the gene, and gene expression in microarray experiments, and classifies the pairwise features as interacting or non-interacting based on a random forest model [16]. Though each of the features by itself is not an indicator of an interaction, a machine learning model was able to use the combined features to make predictions with high precision. The threshold of HiPPIP to classify a protein-pair as "a PPI" was set high in such a way that it yields very high-precision predictions, even if low recall. Novel PPIs predicted using this model are making translational impact. For example, they highlighted the role of cilia and mitochondria in congenital heart disease [18,19], that oligoadenylate synthetase-like protein (*OASL*) activates host response during viral infections through RIG-I signaling via its PPI with retinoic acid-inducible gene I (*RIG-I*) [17], and led to the identification of drugs

potentially repurposable for schizophrenia [20], one of which is currently under clinical trials.

In this work, we studied MPM-associated genes and their PPIs assembled with HiPPIP and analyzed the MPM interactome to draw translatable results. We demonstrate the various ways in which systems-level analysis of this interactome could lead to biologically insightful and clinically translatable results. We made the interactome available to the cancer biology research community on a webserver with comprehensive annotations, so as to accelerate biomedical research on MPM.

2. Results

We collected 62 MPM-associated genes from the Ingenuity Pathway Analysis (IPA) suite, which will be referred to as 'MPM genes' here; these genes have been reported to affect MPM through gene expression changes or genetic variants, or by being targeted bydrugs clinically active against MPM (see details in Data File S1) [21]. Previously known PPIs of the 62 MPM genes were collected from Human Protein Reference Database (HPRD), version 9 [22] and Biological General Repository for Interaction Datasets (BioGRID) version 4.3.194 [23]. Novel (hitherto unknown) PPIs were predicted with HiPPIP, a computational model. We discovered 364 novel PPIs of MPM genes (Table 1), which are deemed highly accurate according to prior evaluation of the HiPPIP model including experimental validations [16]. The MPM interactome thus assembled has 2459 known PPIs and 364 novel PPIs among the 62 MPM-associated genes and 1911 interactors (Figure 1 and Data File S2). Nearly half of the MPM genes had 10 or less known PPIs each, and about 130 novel PPIs have been predicted for these (Figure 2). HiPPIP predicted 920 PPIs of which 556 PPIs were previously known, leaving 364 PPIs to be considered as novel PPIs of the MPM genes. There were an additional 1903 PPIs that are known and not predicted by HiPPIP. This is as expected because the HiPPIP prediction threshold has been fixed to achieve *high precision* by compromising *recall*, which is required for adoption into biology; in other words, it is set to predict only a few PPIs out of the hundreds of thousands of unknown PPIs, but those that are predicted will be highly accurate. It has to be noted that neither PPI prediction nor high throughput PPI screening can be performed with high-precision *and* high-recall. Co-immunoprecipitation (Co-IP) based methods show high-precision and extremely-low recall (detecting only one PPI at a time), whereas multi-screen high-quality yeast 2-hybrid methods show high-precision with low recall (detecting a few tens of thousands of PPIs). Thus, HiPPIP is on par with other methods in terms of precision and the number of new PPIs detected. 18 novel PPIs predicted by HiPPIP were validated to be true (validations have been reported in [16,17], the current work and other unpublished works); the experiments were carried out by diverse research labs.

Table 1. Novel Interactors of each of the malignant pleural mesothelioma (MPM) Genes: Number of known (K) and computationally predicted novel (N) protein-protein interactions (PPIs) and liststhe novel interactors. Bold genes in the 4th column are Novel Interactors that were experimentallyvalidated in the current study.

Gene	K	N	Novel Interactors
ATP1B1	21	7	HCRTR1, SERPINC1, TM4SF1, PRRX1, CD84, CREG1, THOC1
ATIC	5	5	MAP3K7, CPS1, KIAA1524, VWC2L, DES
ATXN1	287	5	CNOT6L, XPO7, C7, PITX3, RPL19
BAP1	27	2	PLN, **PARP3**
CDKN2A	168	5	NFX1, DNAI1, GLIPR2, SIT1, CA9
CTLA4	17	10	PLCL1, DCTD, SKP1, GLP1R, AOX1, CD28, ATP5G3, CLK1, BCS1L, CDC26

Table 1. *Cont.*

Gene	K	N	Novel Interactors
DHFR	10	7	*RHOQ, SCZD1, TOMM7, EXOC4, DTYMK, COPS8, CRHBP*
FGFR1	67	7	*ZFYVE1, NRG1, TPMT, OR51B4, SHB, PPP2CB, EIF4EBP1*
FGFR2	46	8	*PTPRE, OAT, PLXNA1, SEC23IP, MDM2, MGMT, PLSCR1, ELK4*
FGFR3	43	6	*GRK4, GMPS, STK32B, IDUA, IRF2BPL, ADD1*
FLT1	25	8	*MIPEP, RASSF9, HMGB1, FLT3, LATS2, ALOX5AP, ARL2BP, CDK8*
FLT3	17	8	*FMO1, SNRPA1, PNPLA3, NFIB, GPR12, SHC1, FLT1, CDK8*
FLT4	16	4	*NKX2-5, HNRNPH1, GRIA1, PNPLA8*
FOXO3	27	4	*GPR6, HDAC2, PRDM13, SIM1*
GART	4	5	*TIAM1, NMI, TMPRSS15, JUN, IFNAR1*
GIPR	2	0	*None*
HLA-DQA1	9	6	*HLA-DQA2, KLHDC3, TAL2, NXF1, BRD2, HLA-DPB1*
HSP90AA1	158	6	*IGHA2, MED28, PHLDA2, TCIRG1, IGHD, USP13*
HSP90AB1	59	10	*SLC25A27, PENK, ZFP36L2, MTX2, TPSAB1, PROS1, GPRC5B, CCR7, GNPDA1, CETN3*
HSP90B1	36	2	*MMP17, EPB41L4B*
IL4R	23	5	*RBBP6, NPIPB5, SLC20A1, ERN2, HDGFRP3*
KAZN	12	6	*KIF1B, NPPA, CELA2A, CELA2B, CTRC, FBLIM1*
KDR	60	8	*UTP3, SRP72, SHOX2, KIT, **ALB**, CACNA1S, CHIC2, GSTA2*
KRT5	25	10	*SORD, KRT6A, NADSYN1, SAP18, KRT7, TARBP2, KRT6B, KRT4, DCTN1, GPD1*
KRT72	19	8	*SP7, KRT78, KRT80, LARP4, MYL6B, KRT74, BCDIN3D, GRASP*
LCK	143	5	*NCDN, ZSCAN20, YBX1, CITED4, CAMK1D*
LY6E	6	8	*PIP, GLI4, HSF1, AKR1B1, EIF3H, JRK, GML, GPAA1*
LYN	125	12	*NEK7, SGK3, PDCD4, TRPA1, TERF1, PNMA2, IL7, CLCF1, AGXT, ARFGEF1, CRH, KLHL41*
NTRK2	34	3	*NXNL2, KCNS1, CDK20*
PDCD1	2	3	*COPS8, MCL1, OR6B3*
PDGFRA	64	4	*SPOCK1, RAPGEF1, **ALB**, CD244*
PDGFRB	76	8	*PLAUR, TUFM, CDX1, CHRM3, FAXDC2, ITK, CDK14, MITF*
PDPN	2	5	*PRDM2, PRMT1, ZBTB48, CELA2B, LHX1*
POLE	12	7	*SCARB1, RAN, VSIG4, ULK1, EIF2B1, MMP17, NOS1*
POLE2	19	6	*SAV1, PYGL, NID2, PARK7, DRD3, ATOH1*
POLE3	7	7	*TNC, TRIM32, EIF4G2, ASTN2, GSN, CST3, ALAD*
POLE4	7	4	*REG3G, SGOL1, EVA1A, B4GALT4*
PRR5	5	3	*WNT7B, TTC38, SCUBE1*

Table 1. *Cont.*

Gene	K	N	Novel Interactors
RRM1	10	12	*SLC22A18AS, SIRPA, SLC22A18, STIM1, SPINK1, ZFPM2, SH2D3A, PSMD13, RNH1, NUP98, CUZD1, RGS4*
RRM2	9	10	*TAF1B, ST3GAL3, NPBWR2, LPIN1, GCG, MGAT4A, BARX1, ASAP2, ITSN2, LAPTM4A*
SP1	146	5	*HNRNPA1, REG1A, RAPGEF3, GRIN1, ENDOU*
SRC	300	9	*ZNF687, ENPP7, FMR1, PI3, PTPRT, CUL4B, DPYD, BARD1, PLTP*
TARP	1	4	*TBX20, GGCT, IL6, CPVL*
TBCE	2	3	*SERTAD3, EIF2B2, PRDM2*
TTF1	6	3	*AMPH, DFNB31, QRFP*
TUBA1A	63	3	*TUBA1C, AMHR2, ACVR1B*
TUBA1C	63	8	*PRKAG1, SHMT2, AMHR2, SCAF11, ACVR1B, AQP5, KMT2D, TUBA1A*
TUBA3C	12	3	*XPO4, EIF3FP2, PARP4*
TUBA3D	1	6	*TUBA3E, WTH3DI, CCDC74B, FAM168B, LOC151121, IMP4*
TUBA4A	51	14	*WNT6, ETV6, ATP5G3, CAPN2, CXCR1, SLC11A1, CDK5R2, ALPP, IL1RL1, NUPR1, HPCA, SKP1, DPYSL2, STK16*
TUBA8	7	2	*POTEH, CCT8L2*
TUBB1	1	2	*C20orf85, SLMO2*
TUBB2A	27	0	*None*
TUBB3	34	6	*PRDM7, SLC7A5, PIEZO1, MVD, TRAPPC2L, TCF25*
TUBB4A	10	7	*UQCR11, APC2, ABCA7, PLIN3, KDM4B, SBNO2, HMG20B*
TUBB4B	19	4	*TSC1, NELFB, C9orf9, PTPRE*
TUBD1	1	6	*TMED1, PTRH2, TRPV1, GJB3, EPX, RFX5*
TUBE1	0	6	*DPAGT1, NUDC, RPS20, CDC40, GOPC, C6orf203*
TUBG1	28	6	*WNT3, PHB, RND2, CTRL, SGCA, RARA*
TUBG2	3	3	*NBR2, IKZF3, CLMP*
TYMS	3	9	*YES1, TAF3, ITGAM, NDUFV2, EPB41L3, SMCHD1, OCRL, THOC1, NAPG*
WT1	64	8	*FJX1, PEX3, CAPRIN1, PAX6, BST2, B3GNT3, CALML5, HIPK3*

Figure 1. Malignant pleural mesothelioma (MPM) Protein-Protein Interactome: Network view of the MPM interactome is shown as a graph, where genes are shown as nodes and protein-protein interactions (PPIs) as edges connecting the nodes. MPM-associated genes are shown as dark blue square-shaped nodes, novel interactors and known interactors as red and light blue colored circular nodes respectively. Red edges are the novel interactions, whereas blue edges are known interactions.

Figure 2. Number of protein-protein interactions (PPIs) in the malignant pleural mesothelioma (MPM) Interactome: The 62 MPM genes are shown along the X-axis, arranged in ascending order of their number of known PPIs. Blue line, right-side axis: Number of known PPIs is shown. Red bars, left-side axis: Number of novel PPIs.

2.1. Experimental Validation of Selected Protein-Protein Interactions (PPIs)

We carried out experimental validations of five predicted PPIs chosen for their biological relevance and proximity to MPM genes, namely, *BAP1-PARP3*, *KDR-ALB*, *PDGFRA-ALB*,*CUTA-HMGB1* and *CUTA-CLPS*. They were validated using protein pull-down followed by protein identification using mass spectrometry (Table S1) or size-based protein detectionassay (Figure 3). Each bait protein was also paired with a random prey protein serving as control (specifically, *BAP1*-phospholambin, *ALB-FGFR2* and *CUTA-FGFR2*). All predicted PPIs were validated to be true, while control pairs tested negative. In addition to these five, another PPI from the MPM interactome, namely *HMGB1-FLT1* was validated in our prior work through co-immunoprecipitation [16]. Three novel PPIs, namely *HLA-DQA1— HLA-DQB1*, *FGFR2—FGF2* and *CDKN2A—CDKN2B*, that we reported in the preprint of this work [24], have since been reported as known PPIs in a recent version of BioGRID (downloaded February 2021); these three are treated as known PPIs in the remaining description.
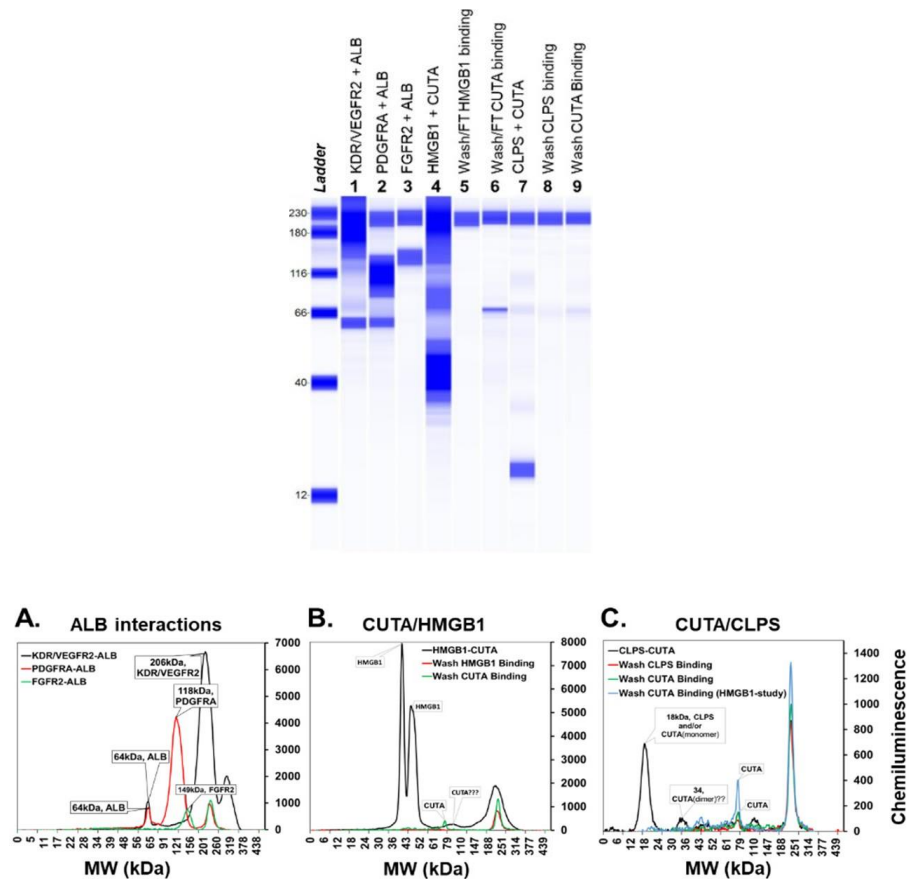
Figure 3. Validation of predicted *ALB* interactions and *CUTA* interactions using Wes™ Simple West- ern total protein detection assay: Pseudo-gel or virtual-blot like image of the validated interactions of *ALB* (lanes 1–2) and *CUTA* (lanes 4, 7) along with negative control (lane 3). In addition to the final pull-down samples, wash and/or flow through after binding 'bait' and 'prey' proteins for the *CUTA*interactions are also shown (lanes 5, 6, 8 and 9). The electropherogram image of Simple Westernresults using Total protein size-based assay. (**A**) *ALB* interactions with true positives *KDR/VEGFR2*, *PDGFRA* and false positive *FGFR2*. (**B**) *CUTA* interactions with *HMGB1*. (**C**) *CUTA* interactions with*CLPS*. An overlay of the electropherogram of the wash from *HMGB1* after *CUTA* binding is also shown in (**C**) for comparison.

2.2. Functional Interactions of Malignant Pleural Mesothelioma (MPM) Genes with Predicted Novel Interactors

We used ReactomeFIViz [25], a Cytoscape plugin, to extract known functional interactions between MPM-associated genes and their novel interactors. Seven novel PPIs hadsuch functional interactions, namely (MPM genes are shown in bold), **PDGFRB**-*RAPGEF1* ('*part of the same complex*', '*bound by the same set of ligands*'), **SP1**→*HNRNPA1* ('*expression regulation*'), **HLA-DQA1**→*HLA-DPB1*, *HLA-DQA2*→**HLA-DQA1** ('*part of the same complex*', '*catalysis*'), **CTLA4**-*CD28*, **PDGFRB**-*PLAUR* ('*bound by the same set of ligands*') and **FGFR2**-*MDM2* ('*ubiquitination*').

2.3. Web Server

We made the MPM interactome available on a webserver called *Wiki-MPM* (http://severus.dbmi.pitt.edu/wiki-MPM). It has advanced-search capabilities, and presents comprehensive annotations, namely Gene Ontology, diseases, drugs and pathways, of thetwo proteins of each PPI side-by-side. Here, a user can query for results such as "PPIs where one protein is involved in mesothelioma and the other is involved in immunity", and then see the results with the functional details of the two proteins side-by-side. The PPIs and their annotations also get indexed in major search engines like Google and Bing; thus a user searching for '*KDR and response to starvation*' would find the PPIs *KDR-CAV1* and *KDR-ALB*, where the interactors are each involved in 'response to starvation'. Querying by biomedical associations is a unique feature which we developed in Wiki-Pi that presents known interactions of all human proteins [26]. Wiki-MPM is a specialized version for disseminating the MPM interactome with its novel PPIs, visualizations and browse features. The novel PPIs have a potential to accelerate biomedical discovery in mesothelioma and making them available on this web server brings them to the biologists inan easily-discoverable and usable manner. Wiki-MPM will be integrated into the National Mesothelioma Virtual Bank [27,28], and will be available to the mesothelioma research community as part of our translational support of cancer research.

2.4. Pathway Analysis

We compiled the list of pathways that any of the proteins of MPM interactome are associated with, using Ingenuity Pathway Analysis suite [29]. Top 30 pathways by statistical significance of association are shown in Figure 4A. A number of pathways such as *NF-κB signaling*, *PI3/AKT signaling*, *VEGF signaling* and *natural killer cell signaling,* are highly relevant to mesothelioma etiology. They are found to be connected to MPM genes through novel PPIs that were previously unknown. For example, the PI3K/AKT signaling pathway regulating the cell cycle is aberrantly active in MPM, and the mesothelioma gene *FGFR1* is connected to this pathway via its novel predicted PPIs with *EIF4EBP1* and *PRP2CB* (Figure 4B) [30]. Statistical significance of association to the interactome, and various MPM genes and novel interactors belonging to these pathways are shown in Table 2 and Data File S3. A cancer biologist may utilize the Supplementary Data (Data Files S2 and S3) to study novel PPIs that connect MPM genes to a pathway that they are interested in studying.

Table 2. Pathways that are relevant to the pathophysiology and genetics of malignant pleural mesothelioma: Pathway analysis revealed that molecular mechanisms underlying various types of cancers, axonal guidance signaling, PI3/AKT signaling, VEGF signaling, natural killer cell signaling and inflammation signaling pathways may be pertinent to the development of MPM. A list of all the associated pathways is shown in Data File S3.

Pathway	p-Value	MPM Genes	Novel Interactors
Glucocorticoid ReceptorSignaling	6.13×10^{-56}	KRT72, HSP90B1, FGFR3, HSP90AB1, FGFR1, KRT5, FOXO3,FGFR2, HSP90AA1	KRT74, HMGB1, PRKAG1, IL6, KRT6B, KRT78, KRT80, KRT7, KRT4,TAF3, NPPA, MAP3K7, KRT6A
Molecular Mechanisms of Cancer	5.01×10^{-53}	CDKN2A, SRC, FGFR3, FGFR1,FGFR2	CDK14, CDK20, CDKN2B, PRKAG1,WNT7B, RND2, WNT6, CDK8, RHOQ, RAPGEF3, MAP3K7, WNT3
NF-κB Signaling	1.26×10^{-39}	FGFR1, LCK, FLT1, KDR, PDGFRA, FGFR2, NTRK2, FGFR3, PDGFRB, FLT4	MAP3K7
Small Cell Lung Cancer Signaling	2.00×10^{-37}	FGFR1, FGFR2, FGFR3	CDKN2B
Axonal Guidance Signaling	2.51×10^{-37}	TUBB1, TUBA1A, TUBA4A, TUBA8,TUBB2A, NTRK2, FGFR3, FGFR1, TUBB3, TUBG1, TUBA1C, TUBB4B,FGFR2, TUBB4A	MYL6B, DPYSL2, PRKAG1, PLCL1, WNT7B, WNT6, PLXNA1, TUBA3E, WNT3
PI3K/AKT Signaling	1.58×10^{36}	HSP90B1, FOXO3, HSP90AA1,HSP90AB1	OCRL, PPP2CB, MCL1, EIF4EBP1
VEGF Signaling	3.98×10^{-36}	FGFR1, FLT1, SRC, KDR,FOXO3, FGFR2, FGFR3,	EIF2B1, EIF2B2
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	6.31×10^{-36}	SRC, FGFR3, FGFR1, FGFR2	IL1RL1, IL6, PLCL1, WNT7B, IL7, WNT6, CALML5, MAP3K7, WNT3, APC2
Natural Killer Cell Signaling	6.31×10^{-32}	FGFR1, LCK, FGFR2, FGFR3	OCRL, CD244
Actin Cytoskeleton Signaling	1.58×10^{-30}	FGFR1, FGFR2, FGFR3	MYL6B, GSN, APC2
eNOS Signaling	3.16×10^{-30}	FGFR1, FLT1, KDR, HSP90B1, FGFR2, HSP90AA1, FGFR3, FLT4, HSP90AB1	PRKAG1, CALML5, AQP5, CHRM3
Neuroinflammation Signaling Pathway	3.98×10^{-30}	FGFR1, HLA-DQA1, FGFR2, FGFR3	HMGB1, HLA-DQB1, ACVR1B, GRIN1, GRIA1
Gap Junction Signaling	1.00×10^{-29}	IL6,FGFR1, TUBB3, TUBG1, TUBB1, TUBA1C, TUBA1A, SRC, TUBB4B, TUBA4A, FGFR2, TUBA8, TUBB2A,FGFR3, SP1, TUBB4A	GJB3, PRKAG1, TUBA3E, PLCL1, GRIA1
Integrin Signaling	1.58×10^{-28}	FGFR1, SRC, FGFR2, FGFR3	GSN, ITGAM, RHOQ,CAPN2, RND2
IL-6 Signaling	1.58×10^{-28}	FGFR1, FGFR2, FGFR3	IL1RL1, MCL1, IL6, MAP3K7
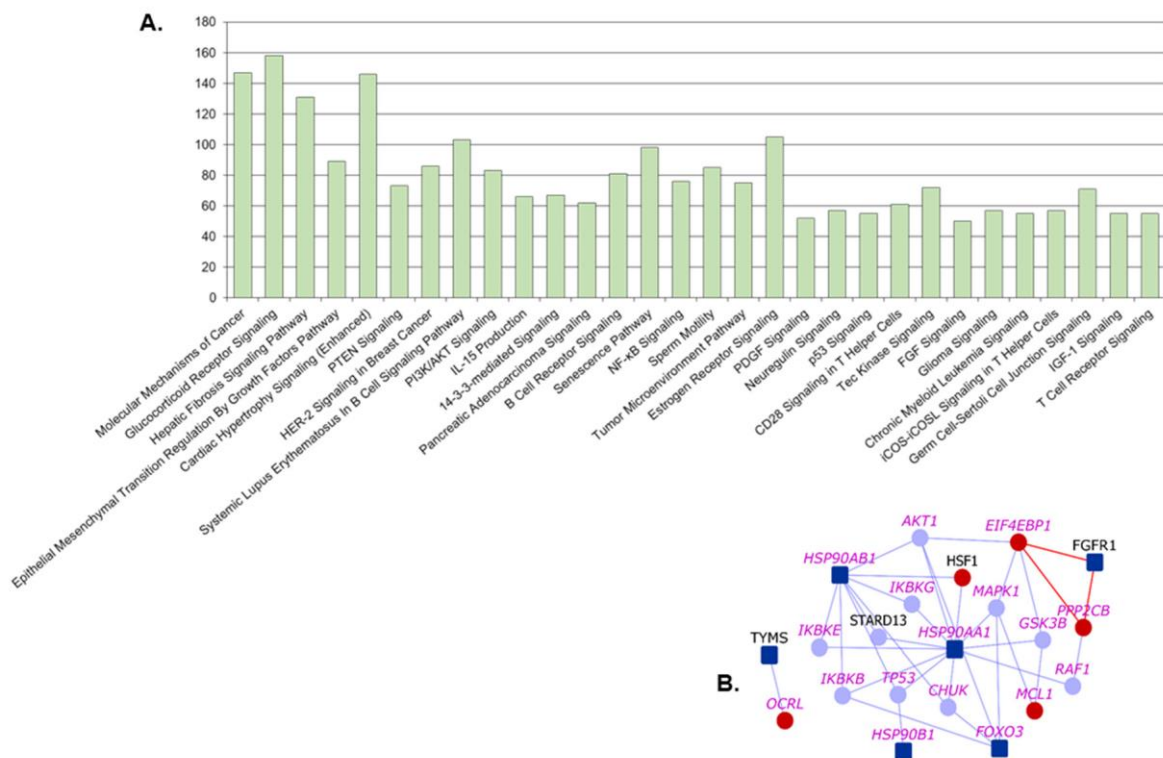
Figure 4. Pathways associated with malignant pleural mesothelioma (MPM) interactome: (**A**) Number of genes from MPM interactome associated with various pathways are shown. Top 30 pathways based on significance of association with the interactome are shown. (**B**) PI3K/AKT Signaling Pathway: Dark blue nodes are MPM genes, light blue nodes are known interactors and red nodes are novel interactors. Nodes with purple labels are genes involved in the PI3K/AKT signaling pathway.

2.5. Potentially Repurposable Drugs

We previously identified drugs potentially repurposable for schizophrenia through interactome analysis, and one of them is currently in clinical trials (ClinicalTrials.gov Identifier: NCT03794076) and another clinical trial has been funded and is yet to start [20]. Following this methodology, we constructed the MPM drug-protein interactome that shows the drugs that target any protein in the MPM interactome. This analysis has been carried out on an earlier version of BioGRID (3.4.159), which had fewer known PPIs, as reported in the preprint version of the paper [24], and has not been recomputed with the latest version of BioGRID unlike the other analyses presented here. There are 513 unique drugs that target 206 of these proteins (of which 28 are novel interactors that are targeted by 147 drugs)(Figure 5 and Data File S4). We adopted the established approach of comparing drug-induced versus disease-associated differential expression using the BaseSpace correlation software (previously called NextBio) [31,32], to identify five drugs that could be potentially repurposable for MPM (Table 3; the table also shows corresponding information for two known MPM drugs). These are: *cabazitaxel*, used in the treatment of refractory prostate cancer; *primaquine* and *pyrimethamine*, two anti-parasitic drugs; *trimethoprim*, an antibiotic; and *gliclazide*, an anti-diabetic drug (See Appendix A, titled 'Repurposable Drugs for Treatment of Malignant Pleural Mesothelioma'). The drugs were selected based on whether they induced a differential expression (DE) in genes that showed a negative correlation with lung cancer associated DE, and affected genes of high DE in MPM tumors/cell lines (GSE51024 [33] and GSE2549 [34]), or underwent prior clinical testing in lung cancer.Lung cancers share common pathways with mesothelioma initiated on asbestos exposure. Therefore, drugs targeting lung cancers can potentially be used in MPM [35]. Table 3 shows pharmacokinetic details of the drugs as reported in Drug Bank [36].
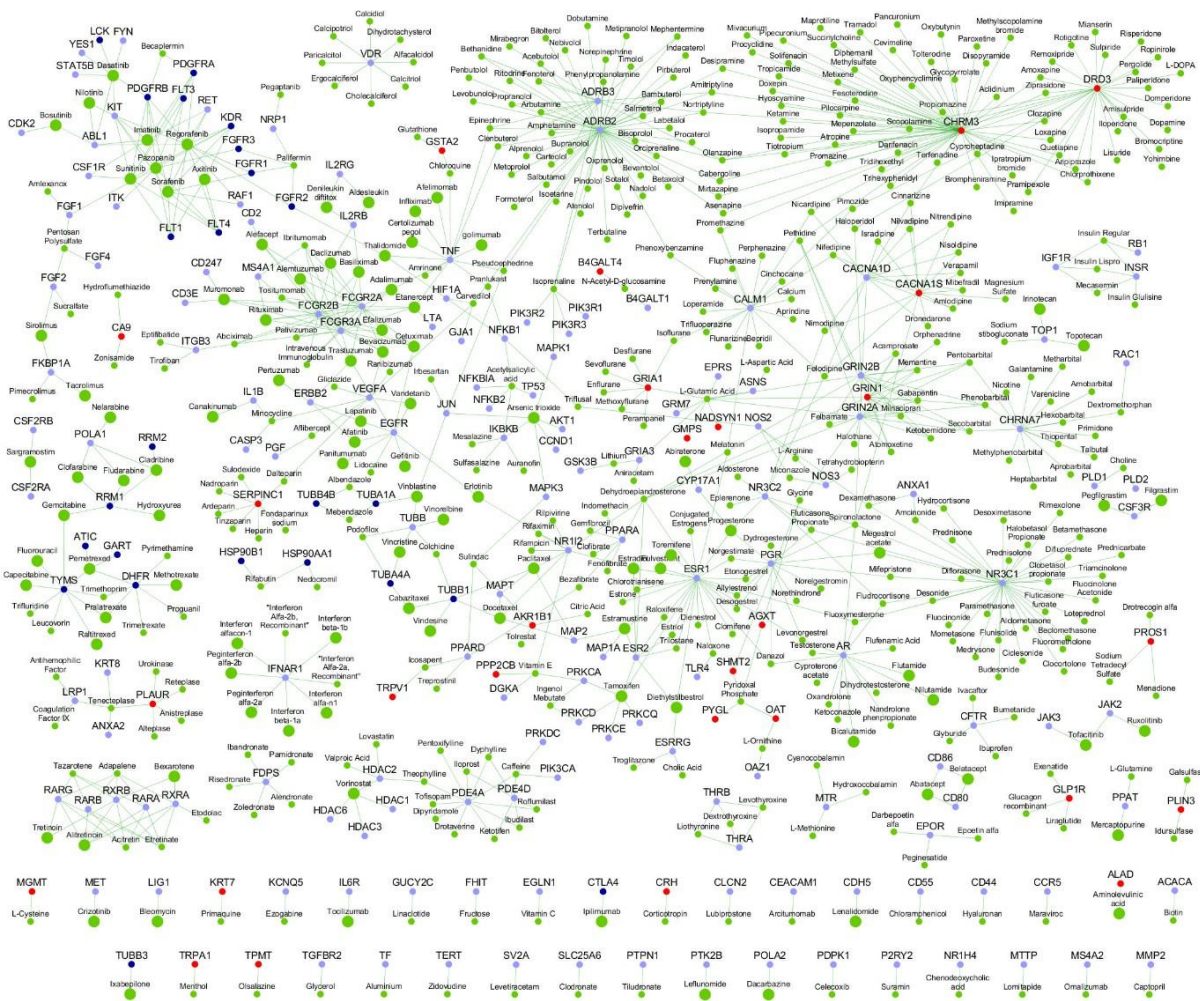
Figure 5. Malignant pleural mesothelioma (MPM) Drug-Protein Interactome: The network shows the drugs (green color nodes) that target the proteins in the MPM interactome. Larger green nodes correspond to drugs that target the anatomic category 'antineoplastic and immunomodulating agents'. The color legend for genes (proteins) is as shown in Figure 1, with MPM genes in dark blue, their known interactors in light blue and novel interactors in red.

Table 3. Pharmacokinetic details of known mesothelioma drugs and the drugs that are presented as candidates for repurposing. Known mesothelioma drugs are shown in bold italics. Score corresponds to scaled correlation score with lung cancer expression studies from BaseSpace (NextBio) analysis.

Drug Name & Score	Original Therapeutic Purpose(s)	Delivery	Half-Life	Toxicity	Targets
Pemetrexed negative 79	Chemotherapeutic drug for pleural mesothelioma and non-small cell lung cancer	Powder for solution; Intravenous	3.5 h	Data not available	*ATIC, DHFR, GART, TYMS*
Mitomycin negative 64	Chemotherapeutic drug for breast, bladder, esophageal, stomach, pancreas, mesothelioma, lung and liver cancers	Injection, powder or lyophilized for solution; Intravenous	8–48 min	Nausea and vomiting	-
Cabazitaxel negative 79	Anti-neoplastic agent in hormone-refractory metastatic prostate cancer	Solution; Intravenous	Rapid initial-phase of 4 min, intermediate-phase of 2 h and prolonged terminal-phase of 95 h	Neutropenia, hypersensitivity reactions, gastrointestinal symptoms, renal failure	*TUBB1, TUBA4A*
Pyrimethamine negative 83	Anti-parasitic agent in toxoplasmosis and acute malaria	Tablet; Oral	4 days	Data not available	*DHFR*

Table 3. *Cont.*

Drug Name & Score	Original Therapeutic Purpose(s)	Delivery	Half-Life	Toxicity	Targets
Trimethoprim negative 63	Anti-bacterial agent/antibiotic in urinary tract, respiratory tract and middle-ear infections and traveler's diarrhea	Tablet/solution; Oral	8 to 11 h	Oral toxicity in mice at LD50 = 4850 mg/kg	*DHFR, TYMS*
Primaquine negative 71	Anti-malarial agent	Tablet; Oral	3.7 to 7.4 h	Data not available	*KRT7*
Gliclazide negative 56	Anti-diabetic/hypoglycemic medication in type 2 diabetes mellitus	Tablet; Oral	10.4 h	Oral toxicity in mice at LD50 = 3000 mg/kg, accumulation in people with severe hepatic and/or renal dysfunction, side-effects of hypoglycemia including dizziness, lack of energy, drowsiness, headache and sweating	*VEGFA*

Although in each case, there would be some genes that are differentially expressed in the same direction for both the drug and the disorder (for e.g., both the drug and the disease cause some genes to overexpress), the overall effect on the entire transcriptome has an anti-correlation. A correlation score is generated based on the strength of the overlap between the drug and the disease datasets. Statistical criteria such as correction for multiple hypothesis testing are applied and the correlated datasets are then ranked by statistical significance. A numerical score of 100 is assigned to the most significant result, and the scores of the other results are normalized with respect to this top-ranked result. We excluded drugs with unacceptable toxicity (e.g., minocycline) or unsuitable pharmacokinetics. The final list comprised 15 drugs, out of which 10 have already been tested against mesothelioma in clinical trials/animal models, and several of them were found to display clinical activity [37–53] (Table S2). Gemcitabine and pemetrexed are being used as first-line therapy for mesothelioma, in combination with cisplatin [45,53]. Ipilimumab has been identified to be a potential second-line or third-line therapy in combination with nivolumab [47]. Ixabepilone stabilizes cancer progression for up to 28 months [49]. Zoledronate, which showed modest activity in MPM, induced apoptosis and S-phase arrest in human mesothelioma cells and inhibited tumor growth in an orthotopic animal model [54,55]. Sirolimus/cisplatin increased cell death and decreased cell proliferation in MPM cell lines [56]. α-Tocopheryl succinate increased the survival of orthotopic animal models of malignant peritoneal mesothelioma [57]. Pre-clinical testing of vitamin E and its analogs are in progress [58,59].

Primaquine targets *KRT7*, a novel interactor of *KRT5*, whose high expression has been correlated with tumour aggressiveness and drug resistance in malignant mesothelioma [60–62]. Primaquine may be re-purposed for MPM treatment at least as an adjunctive drug with pemetrexed, the drug currently used for first-line therapy. Primaquine enhanced the sensitivity of the multi-drug resistant cell line KBV20C to cancer drugs [63]. Gliclazide is an anti-diabetic drug inhibiting *VEGFA* [64], a known interactor of *KDR*, and is significantly upregulated in MPM tumour ($Log_2FC = 1.83$, *p*-value = 0.0018). Gliclazide inhibits VEGF-mediated neovascularization [64]. High levels of VEGF have been correlated with both asbestos exposure in MPM and advanced cancer [65,66]. Glibenclamide, a drug with a similar mechanism of action as that of gliclazide, increases caspase activity in MPM cell lines and primary cultures, leading to apoptosis mediated by *TRAIL* (TNF-related apoptosis inducing ligand) [67].

Eliminating those drugs which are being/have already been tested in mesothelioma with varying results, we arrived at a list of five potentially repurposable drugs in the descending order of negative correlation scores: pyrimethamine, cabazitaxel, primaquine, trimethoprim and gliclazide (Table 3). Cabazitaxel targets the MPM genes, *TUBB1* and

TUBA4A, and was effective in treating non-small cell lung cancer (NSCLC) that was resistant to docetaxel, a drug that targets *TUBB1* along with other known interactors of MPM genes [37]. Pyrimethamine and trimethoprim target the MPM gene *TYMS* involvedin folate metabolism, which was found to be differentially expressed in MPM tumors (GSE51024 [33]) ($\log_2$FC = 1.82, *p*-value = 4.10 10^{-17}). MPM tumors have been shown to be responsive to anti-folates [68].

2.6. Analysis with Other High-Throughput Data

This section describes the overlap of the MPM interactome with various types of MPM- related biological evidence. 1690 (85.65%) proteins in the interactome were supported by genetic variant, transcriptomic, and proteomic evidence, and are listed in Data File S5. Table 4 shows 48 novel interactors that had three or more pieces of biological evidence.

Table 4. Novel interactors in the malignant pleural mesothelioma (MPM) interactome with biological evidences related to MPM. The table shows the following data in columns labeled A to F. (A) 48 novel interactors of MPM associated genes that have been linked to four or more biological evidences related to MPM, namely, **B1**: high or medium gene expression in lungs, **B2**: differential gene expression in MPM tumor versus other thoracic tumors, **B3**: differential gene expression in MPM tumor versus normal adjacent pleural tissue, **B4**: differential gene expression in MPM tumors of epithelioid, biphasic and sarcomatoid types, **B5**: differential gene methylation in MPM, **B6:** gene expression correlated with unfavorable lung cancer prognosis, **B7**: differential gene expression on exposure to asbestos or asbestos-like particles, **C**: isolation asexosome-derived proteins from malignant mesothelioma cell lines, **D**: differential protein abundance levels in epithelioidand sarcomatoid types of malignant mesothelioma, and **E**: genetic variants in MPM. Last column, **F**, gives the total numberof sources of evidences for each gene. The complete list of biological evidence for all the genes in the interactome can befound in Data File S5.

A	B							C	D	E	F
Novel Interactor	Differential Gene Expression							Exosome-Derived Proteins	Differential Protein Levels	Genetic Variants	Total
	B1	B2	B3	B4	B5	B6	B7				
CAPRIN1	yes	yes	yes	yes				yes		yes	6
RAN	yes	yes	yes	yes				yes	yes		6
TNC	yes	yes			yes		yes	yes		yes	6
CUL4B	yes	yes	yes	yes				yes			5
GMPS	yes	yes	yes	yes						yes	5
IL6	yes	yes		yes	yes			yes			5
MGMT	yes	yes	yes		yes					yes	5
NFIB	yes		yes	yes			yes			yes	5
NUDC	yes	yes		yes				yes		yes	5
PLAUR	yes	yes		yes	yes	yes					5
PLIN3	yes	yes		yes		yes		yes			5
PLXNA1	yes	yes	yes	yes				yes			5
PRMT1	yes		yes					yes	yes	yes	5
RNH1	yes	yes		yes				yes	yes		5
SCARB1	yes	yes		yes		yes		yes			5
SLC7A5	yes		yes	yes		yes		yes			5
SMCHD1	yes		yes	yes				yes		yes	5
ASAP2	yes	yes		yes				yes			4
B4GALT4	yes	yes		yes			yes				4
CAPN2	yes	yes		yes			yes				4
CDC40	yes		yes	yes			yes				4
DTYMK	yes	yes	yes	yes							4
EIF3H	yes			yes				yes		yes	4
EPB41L3	yes	yes		yes						yes	4

Table 4. *Cont.*

A	B							C	D	E	F
Novel Interactor	Differential Gene Expression							Exosome-Derived Proteins	Differential Protein Levels	Genetic Variants	Total
	B1	B2	B3	B4	B5	B6	B7				
EXOC4	yes	yes		yes				yes			4
GNPDA1	yes	yes		yes				yes			4
HNRNPA1	yes			yes			yes	yes			4
HNRNPH1	yes		yes	yes				yes			4
LARP4	yes		yes	yes			yes				4
MGAT4A	yes	yes				yes		yes		yes	4
MITF	yes	yes	yes	yes							4
NDUFV2	yes	yes							yes	yes	4
OAT	yes	yes			yes			yes			4
PHB	yes	yes	yes					yes			4
PHLDA2	yes	yes			yes			yes			4
PLCL1		yes	yes	yes						yes	4
PRKAG1	yes	yes						yes		yes	4
PROS1	yes	yes		yes				yes			4
PTRH2		yes		yes		yes				yes	4
PYGL	yes	yes		yes				yes			4
RBBP6	yes		yes	yes						yes	4
SEC23IP	yes	yes	yes	yes							4
SGK3		yes	yes	yes						yes	4
SHMT2	yes	yes	yes	yes							4
SLC20A1	yes	yes		yes						yes	4
TCIRG1	yes			yes				yes			4
XPO4				yes		yes		yes		yes	4
YBX1	yes	yes	yes					yes			4

We compiled the list of genes harboring MPM-associated genetic variants from Bueno et al. [5],and compared this list with all the genes in the MPM interactome (i.e., MPM-associated genes, their known and novel interactors) to identify overlaps. 275 genes in the MPM interactome harbored either germline mutations, or somatic single nucleotide variants (SNVs) or indels (insertions or deletions) (Figure 6, Table 4 and Data File S5) associated with MPM tumors. Of these 275 genes, 37 were novel interactors of MPM genes. *MGMT* carried germline mutations while the following carried somatic mutations: *ASTN2, BARX1, BRD2, CALML5, CAPRIN1, CLK1, CPS1, DPYD, EIF3H, EPB41L3, GMPS, GPR12, ITGAM, KIAA1524, KMT2D, KRT4, MGAT4A, NBR2, NDUFV2, NFIB, NFX1, NUDC, PLCL1, PRDM2,PRKAG1, PRMT1, PTPRT, PTRH2, RBBP6, SGK3, SLC20A1, SMCHD1, SPOCK1, TMPRSS15, TNC* and *XPO4*. Fourteen of these interact with MPM genes that also harbored a genetic variant (MPM genes are shown in bold): ***CDKN2A**-NFX1,* ***FLT1**-LATS2,* ***TUBA3C**-XPO4,* ***PDGFRA**-SPOCK1,* ***TYMS**-SMCHD1,* ***TYMS**-EPB41L3,* ***GART**-TMPRSS15,* ***TYMS**-NDUFV2,* ***TYMS**-ITGAM,* ***RRM2**-BARX1,* ***RRM2**-MGAT4A* and ***ATIC**-CPS1,* ***ATIC**-KIAA1524* and ***POLE**-NOS1.*
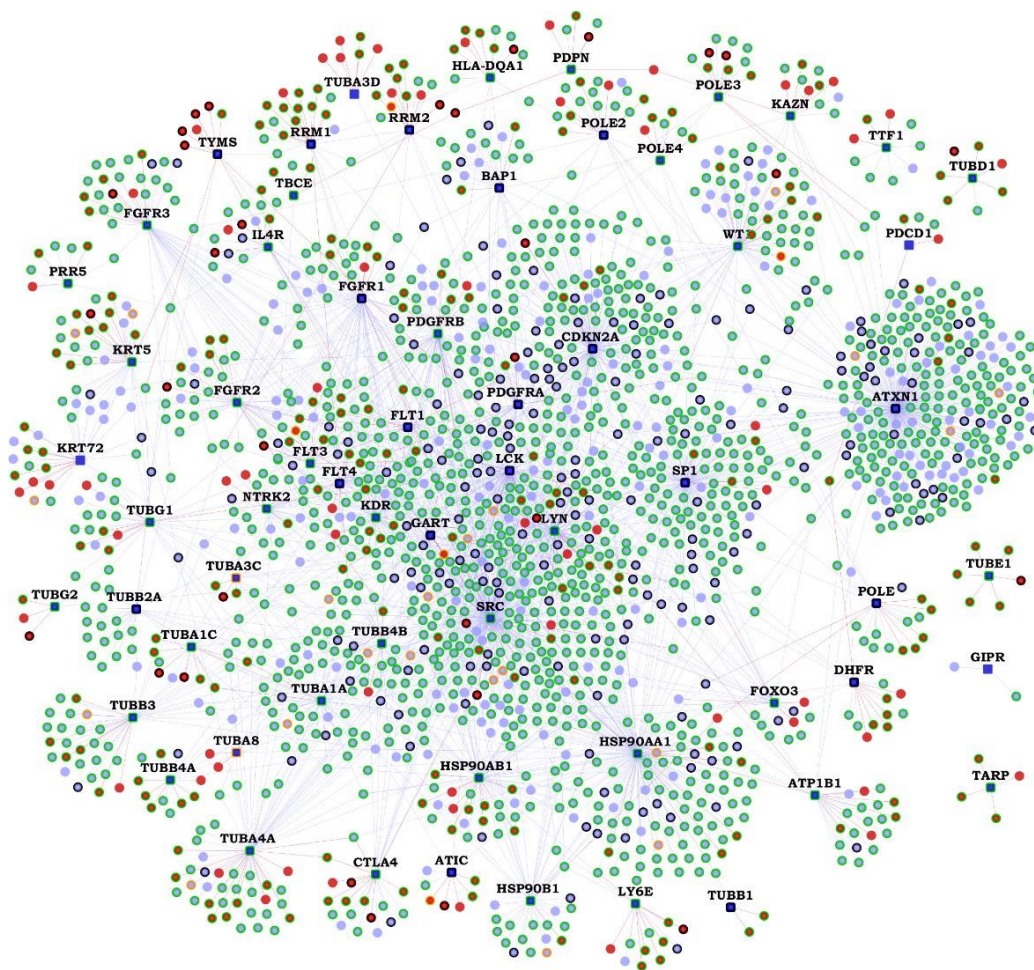
Figure 6. Genes with biological evidences in the malignant pleural mesothelioma (MPM) Protein-Protein Interactome: On the interactome network shown in Figure 1, various biological evidences of relation to malignant pleural mesothelioma (MPM) are shown as node border colors. Genes with variants associated with MPM have orange borders, genes with MPM/lung cancer/asbestos exposure-associated gene/protein expression changes have light green-colored borders and genes with black border have both genetic variants and gene/protein expression changes associated with them. The gene expression-associated features include differential expression in MPM tumors versus normal adjacent pleura, MPM tumors versus other thoracic tumors, differential gene methylation (affecting gene expression) in MPM tumors, gene expression correlated with unfavorable lung cancer prognosis, differential gene expression on exposure to asbestos or asbestos-like particles and high/medium expression in normal lungs. The protein expression-associated features include isolation as exosome-derived proteins from malignant mesothelioma cell lines and differential protein abundance levels in epithelioid and sarcomatoid types of malignant mesothelioma. The complete list of genes in the interactome and their corresponding evidence can be found in Data File S5.

We collected the methylation profile of pleural mesothelioma [69], and found 8 novel interactors to be hypomethylated in pleural mesothelioma versus non-tumor pleural tissue, namely, *ACVR1B*, *IL6*, *MGMT*, *NRG1*, *OAT*, *PHLDA2*, *PLAUR* and *TNC* (Table S3). Some of them have little or no expression in lung tissue but are overexpressed in MPM. *PLAUR* is a prognostic biomarker of MPM [70]. Similarly, *FGFR1* and its novel interactor *NRG1* had elevated mRNA expression in H2722 mesothelioma cell lines and in MPM tissue, both contributing to increased cell growth under tumorigenic conditions [71,72]. *TNC*, involved in invasive growth, is a prognostic biomarker overexpressed in MPM, having low expression in normal lung tissues [73,74]. Thus, these novel interactors, which are not normally expressed in lung tissue, may be hypomethylated in MPM leading to their overexpression, contributing to MPM etiology.

Three hundred and ninety three (393) genes in the MPM interactome were also differentially expressed in mesothelioma tumors versus normal pleural tissue adjacent to tumor (GSE12345 [75]) (p-value of overlap = 9.525×10^{-19}, odds ratio = 1.51). 52 out of the 314 novel interactors in the interactome were differentially expressed in this dataset(p-value = 0.046, odds ratio = 1.26). 938 genes, including 132 novel interactors, in the inter- actome were found to be differentially expressed in MPM tumors of epithelioid, biphasic and sarcomatoid types versus paired normal tissues (GSE51024 [33]) (p-value of overlap = 1.415×10^{-18}, odds ratio = 1.24). Genes with fold-change >2 or $< 1/2$ were considered as overexpressed and underexpressed, respectively, at a p-value < 0.05. Similarly, 744 genes in the MPM interactome were differentially expressed in MPM tumors versus other thoracic cancers such as thymoma and thyroid cancer (GSE42977 [76]) (p-value = 3.04×10^{-41}, odds ratio = 1.53). 112 out of the 314 novel interactors in the interactome were differentiallyexpressed in this dataset (p-value = 7.77×10^{-6}, odds ratio = 1.45). This shows that the MPM interactome is enriched with genes whose expression helps in distinguishing MPM from other thoracic tumors and also with genes differentially expressed in mesothelioma tumors versus normal pleural tissue (Figure 6 and Data File S5). From RNA-seq data in GTEx, we found that 1311 genes, including 189 novel interactors, in the interactome have high/medium expression in normal lung tissue (median transcripts-per-million (TPM) > 9) (Figure 6 and Data File S5) [77].

A recent study had examined the gene expression profiles from the lungs of mice exposed to asbestos fibers (crocidolite and tremolite), an asbestiform fiber (erionite) and a mineral fiber (wollastonite) [78]. Crocidolite, tremolite and erionite are capable of inducing lungcancer and mesothelioma in humans and animal models [78]. On the other hand, wollastonite is a low pathogenicity fiber that shows no association with the incidence of lung cancer and mesothelioma in humans, or carcinogenesis in animal models [79]. The MPM interactome showed significant enrichment with all the 4 fibers (Figure 6 and Data File S5). The highest statistical significance was shown for the human orthologs of the mouse genes that were differentially expressed upon crocidolite exposure (199 genes, p-value = 1.16×10^{-18}, odds ratio = 1.88). This was followed by tremolite (47 genes, p-value = 2.445×10^{-5}, odds ratio = 1.87), wollastonite (16 genes, p-value = 0.0037, odds ratio = 2.09) and erionite (10 genes, p-value = 0.025, odds ratio = 2.01). Altogether, 245 genes in the interactome, including 29 novel interactors, have transcriptomic evidence with respect to exposure to asbestos or asbestos-like fibers. These novel interactors are: *ALB*, *B4GALT4*, *CAPN2*, *CDC40*, *DES*, *FMO1*, *FMR1*, *GML*, *GRIA1*, *HMG20B*, *HNRNPA1*, *ITSN2*, *LARP4*, *LPIN1*, *MGAT4A*, *NEK7*, *NFIB*, *NRG1*, *OCRL*, *PAX6*, *PDCD4*, *PITX3*, *PTRH2*, *REG3G*, *TAF1B*, *THOC1*, *TMED1*,*TNC* and *XPO4*.

From data in Pathology Atlas, we found that high expression of 73 genes, including that of 10 novel interactors, in the interactome has been positively correlated with unfavor- able prognosis for lung cancer (p-value = 1.72×10^{-9}, odds ratio = 2.05) [80]. These novel interactors are: *SPOCK1*, *SLC7A5*, *SCARB1*, *PLIN3*, *PLAUR*, *PIEZO1*, *KRT6A*, *GJB3*, *B3GNT3* and *ARL2BP*. We predicted *ARL2BP* to interact with *FLT1*, a VEGF receptor expressed in MPM cells. VEGF level in MPM patients is a biomarker for unfavorable prognosis, and lung cancer tumors expressing *FLT1* have been associated with poor prognosis [81,82].

Exosomes are extracellular vesicles secreted into the tumor microenvironment. They facilitate immunoregulation and metastasis by shuttling cellular cargo and directing inter- cellular communication. In a proteomic profiling study, 2176 proteins were identified inexosomes of at least one of the four human malignant mesothelioma cell lines (JO38, JU77, OLD1612 and LO68) [83]. 324 proteins in the MPM interactome appeared among these exosome-derived proteins (p-value = 8.86×10^{-10}, odds ratio = 1.36), out of which 47 were novel interactors. Six hundred and thirty one (631) exosome-derived proteins were identi- fied in all four malignant mesothelioma cell lines. Out of these, 127 occurred in the MPM interactome (p-value = $4.54\ 10^{-12}$, odds ratio = 1.84), out of which 15 were novel interac- tors (*PRKAG1*, *HNRNPA1*, *HNRNPH1*, *SORD*, *RNH1*, *RAN*, *PYGL*, *SLC7A5*, *RPS20*, *PARP4*, *YBX1*, *DCTN1*, *TUFM*, *EXOC4* and *GNPDA1*). In the following novel PPIs, both proteins

involved in the interaction appeared among exosome-derived proteins (MPM gene in the interaction is shown in bold): ***TUBB3***-SLC7A5, ***HSP90AB1***-PROS1, ***HSP90AB1***-GNPDA1, ***TUBB4A***-PLIN3, ***LYN***-ARFGEF1, ***HSP90AA1***-PHLDA2, ***HSP90AA1***-TCIRG1, ***TUBG1***-PHB, *GART-NMI*, ***SRC***-CUL4B and ***ATIC***-CPS1.

We computed the overlap of the interactome with 142 proteins that showed significant differences in abundance levels between epithelioid and sarcomatoid types of diffuse malignant mesothelioma [84]. In that study, a Fourier transform infrared (FTIR) imaging approach was employed to identify pathologic regions from diffuse malignant mesothelioma tissue samples [84]. These pathologic regions were then harvested using laser capture microdissection for proteomic analysis. 32 proteins in the interactome were more abundant in either epithelioid or sarcomatoid subtypes (p-value = 5.16×10^{-5}, odds ratio = 2.06), including six novel interactors (p-value = 0.038, odds ratio = 2.43). The novel interactors *KRT78*, *NDUFV2*, *PRMT1*, *RAN* and *RNH1*—predicted to interact with the MPM genes *KRT72*, *TYMS*, *PDPN*, *POLE* and *RRM1*, respectively—had higher abundance in epithelioid samples, whereas *IGHA2*—predicted to interact with *HSP90AA1*—had higher abundance in sarcomatoid samples. The predicted interactions of these protein biomarkers with MPM- associated genes provide a mechanistic basis for experimental dissection of their ability to act as factors differentiating epithelioid tumors from sarcomatoid tumors (and vice versa).

3. Discussion

Currently, mesothelioma biologists only study a handful of genes, such as *BAP1*, *CDKN2A* and *NF2*. To shed light onto the other MPM-associated genes, whose functions remain poorly characterized, we assembled the 'MPM interactome' with ~2400 previously known PPIs and 364 computationally predicted PPIs (five of which have been validated in this work), which along with their biological annotations are being made available to researchers. We demonstrate the power of interactome-scale analyses to generate biologically insightful and clinically translatable results. The interactome has highly significant overlaps with MPM-associated genetic variants, genes differentially expressed or methylated in MPM or upon asbestos exposure, genes whose expression has been correlated with lung cancer prognosis, and with exosome-derived proteins in malignant mesothelioma cell lines. The interactome was enriched in cancer-related pathways. We extended the MPM interactome to include the drugs that target any of its proteins and analyzed it to identify a shortlist of 5 drugs that can potentially be repurposed for MPM—an example of a clinically translatable result.

We validated in vitro five novel PPIs in the interactome, namely, *BAP1-PARP3*, *ALB- KDR*, *ALB-PDGFRA*, *CUTA-HMGB1* and *CUTA-CLPS*. Literature evidence shows that these PPIs may be viable candidates for further experimentation in MPM cell lines or animal models. We hypothesize that the *BAP1-PARP3* interaction may enhance cancer growth in MPM. *BAP1* is a tumor suppressor protein playing a role in cell cycle progression, repair of DNA breaks, chromatin remodeling, and gene expression regulation; variants in *BAP1* have been implicated in hereditary and sporadic mesothelioma [85]. *PARP3* is involved in DNA repair, regulation of apoptosis, and maintenance of genomic stability and telomere integrity [86]. Interaction of *BAP1* with *BRCA1* has been shown to inhibit breast cancer growth [7]. In the absence of *BRCA1* activity or with a perturbation in its interaction with *BAP1*, cancerous growth is enhanced [87]. Loss of *BRCA1* protein expression has been noted in MPM [12]. In this scenario, it is possible that the novel interaction of *BAP1* with *PARP3* in cancerous cells may be promoting cancerous growth, possibly through regulation of DNA repair and apoptosis. *BAP1* and *PARP3* were found to be moderately overexpressed in sarcomatoid MPM tumors compared with normal pleural tissue (log$_2$FC = 0.575, p-value = 0.028, and log$_2$FC = 0.695, p-value = 0.0212, respectively) (GSE42977 [76]). Perturbation of the interaction of *BAP1* with *PARP3*, using *PARP3* inhibitors, may then suppress cancerous growth, at least in sarcomatoid MPM. Several studies and clinical trials [87], have shown that PARP inhibitors influence cancers in which mutations in *BRCA1* or *BRCA2* are observed, which led us to assume that the

cancerous growth-inhibiting interaction of *BAP1* with *BRCA1* may already be perturbed in this case, and that PARP inhibitors may actually be blocking the novel interaction of *BAP1* with *PARP3* which enhances cancer growth. It has been pointed out that upon inhibiting PARP activity, cancerous cells that lack *BRCA1* or *BRCA2* activity may undergo cell cycle arrest and apoptosis, possibly due to an accumulation of chromatid aberrations and an inability to perform DNA repair in the absence of BRCA [7,87]. Thus, we suspect that the novel interaction of *BAP1* and *PARP3* may also be perturbed by PARP inhibitors, leadingto inhibition of cancer growth.

Low levels of *ALB* have been correlated with poor prognosis in MPM patients [88]. Thetwo MPM genes, *KDR* and *PDGFRA*, that *ALB* is predicted to interact with, are members of the PI3K/AKT pathway which has been shown to be aberrantly active in mesothelioma [89].High expression of *CUTA* has been correlated with favorable prognosis in lung cancer (Pathology Atlas). It was found to be overexpressed in MPM tumors versus normal pleura ($\log_2$FC = 0.871, *p*-value = 0.0039) (GSE2549 [34]). *CLPS* inhibits metastasis of the melanoma cell line, B16F10, to lungs by blocking the signaling pathway involving β1 integrin, *FAK* and paxillin [90]. *CLPS* has a novel interaction with *NEDD9*, which has been shown to mediate β1 integrin signaling and promote metastasis of non-small lung cancer cells [91].*CD26*, a cancer stem cell marker of malignant mesothelioma, has been shown to associate with the integrin α5β1 (or *ITGA5*, a novel interactor of the MPM gene, *FGFR2*) and promotecell migration and invasion in mesothelioma cells [91]. Another cancer stem cell marker of malignant mesothelioma, *CD9*, inhibits this metastatic effect mediated by *CD26*. Depletion of *CD26* and *CD9* was shown to respectively lead to decreased and increased expression of *NEDD9* and *FAK* in mesothelioma cells lines, hinting at the involvement of *NEDD9* in mesothelioma tumor invasiveness [91]. *NEDD9* has a known interaction with *LYN*,an MPM gene, shown to play a negative role in the regulation of integrin signaling in neutrophils [92]. *CUTA* has a novel interaction with *HMGB1*, which has been shown toactivate the integrin αMβ2 (or *ITGAM*, a novel interactor of the MPM gene, *TYMS*) and the cell adhesion and migratory function of neutrophils mediated by αMβ2 [93]. *HMGB1* also has a novel interaction with the MPM gene, *FLT1*, shown to be involved in the migration of multiple myeloma cells by associating with β1 integrin, and mediating PKC activation [94]. A recent bioinformatics study identified the genes differentially expressed in epithelioid MPM tissues versus normal pleural tissues (GSE42977 [76]), and extracted the known PPIs interconnecting these genes from the STRING database [95]. They identified 10 hub genes from this network and shortlisted 31 drugs targeting the proteins in the networkbased on scores from the Drug-Gene Interaction Database (DGIdb). The DGIdb score takes into account the literature evidence for a particular drug-protein interaction, the numberof proteins in the network that interact with the given drug, and the ratio of the average number of known protein interactors for all drugs compared to the number of known protein interactors for the given drug. *CDK1*, which is one of the hub genes identified in their study, is a known interactor of three MPM-associated genes, namely, *LYN*, *SP1* and *RRM2*, and we showed that it has association to MPM in three omics datasets: high expression correlated with unfavorable lung cancer prognosis, differential expression inMPM tumors versus adjacent pleural tissue, and isolation as an exosome-derived protein in malignant mesothelioma cell lines. Our work overall presents a more comprehensive study in terms of a larger number of MPM genes analyzed, which were compiled from multiple sources by IPA, and analysis of a larger number of MPM associated omics data sets, and presents transcriptomic-driven shortlisting of repurposable drugs for which additional evidence is presented from clinical trial data, literature, and differential expression of target genes in MPM datasets.

Our study provides an integrative and mechanistic framework for functional translation of mesothelioma-related multi-omics data. The novelty of our work stems from two key factors: (a) we present computationally predicted PPIs of high precision, which link MPM-related genes from disparate genetic-variant / transcriptomic/proteomic studies in hitherto unknown ways within the functional landscape of the interactome, and (b) the

richly annotated MPM interactome is made available on a webserver to facilitate analysisby biologists and computational systems biologists. Our approach has some limitations. The drug-associated expression profiles analyzed in this study were induced in a diverseset of cell lines rather than in mesothelioma cell lines. The effect of the proposed drugs should be examined in MPM cell lines or animal models. We reported the overlap of mouse genes differentially expressed upon asbestos exposure [78] with corresponding human orthologs in the interactome. Mouse models have been routinely used to study pathologic changes associated with asbestos exposure, including gene expression, and these findings have been extrapolated to human diseases such as mesothelioma [96–99].Nevertheless, our results should be interpreted with caution. It is not possible to draw direct transcriptomic/proteomic/phenotypic equivalences between mice and humans, unless these levels are comprehensively characterized in both the species, and a clear equivalence of factors defining a condition such as asbestos exposure is demonstrated inboth the species [100]. Next, it is beyond the scope of our expertise to validate the large number of computationally predicted PPIs in a tissue or cell line of interest. However, we demonstrated the validity of computational predictions on a small number of PPIson purified proteins with appropriate controls. The computational model has also been validated through additional experiments previously; some of the novel PPIs predicted previously by our method have translated into results of biomedical significance [17–19].

4. Methods

4.1. Data Collection

A search using the keyword "malignant pleural mesothelioma" on IPA (Ingenuity Pathway Analysis) retrieved genes causally related to the disease. IPA retrieves genes from the Ingenuity Knowledge Base which has ~5 million experimental findings expert-curated from biomedical literature or incorporated from other databases [29].

4.2. High-Precision Protein-Protein Interaction Prediction (HiPPIP) Model

PPIs were predicted by computing features of protein pairs, namely, cellular localiza- tion, molecular function and biological process membership, genomic location of the gene, gene expression from microarray experiments, protein domains and tissue membershipof proteins, as described in Thahir et al. [101], and developing a random forest model to classify the pairwise features as interacting or non-interacting. A random forest with 30 trees was trained using the feature offering maximum information gain out of four random features to split each node; minimum number of samples in each leaf node was set to be 10. The random forest outputs a continuous valued score in the range of [0,1]. The threshold to assign a final label was varied over the range of the score for positive class (i.e., 0 to 1) to find the precision and recall combinations that are observed.

4.3. Evaluation of PPI Prediction Model

Evaluations on a held-out test data showed a precision of 97.5% and a recall of 5% ata threshold of 0.75 on the output score. Next, we created ranked lists for each of the hubgenes (i.e., genes that had >50 known PPIs), where we considered all pairs that receiveda score >0.5 to be novel interactions. The predicted interactions of each of the hub genesare arranged in descending order of the prediction score, and precision versus recall is computed by varying the threshold of predicted score from 1 to 0. Next, by scanning these ranked lists from top to bottom, the number of true positives versus false positives was computed.

4.4. Novel PPIs in the MPM Interactome

Each MPM gene, say Z, is paired with each of the other human genes (G_1, G_2 . . . G_N), and each pair is evaluated with the HiPPIP model. The predicted interactions of each of the MPM genes (namely, the pairs whose score is >0.5) were extracted. These PPIs, combined

with the previously known PPIs of MPM genes collectively form the 'MPM interactome'. Interactome figures were created using Cytoscape [102].

Note that 0.5 is the threshold chosen not because it is the midpoint between the two classes, but because the evaluations with hub proteins showed that the pairs that received a score >0.5 are highly confident to be interacting pairs. This was further validated through experiments for a few novel PPIs above this score.

4.5. Previously Known PPIs in the MPM Interactome

Previously known PPIs of the 62 MPM genes were collected from Human Protein Reference Database (HPRD) version 9 [22] and Biological General Repository for Interaction Datasets (BioGRID) version 4.3.194 [23]. The data behind our web-server will be updated once in a year with recent versions of BioGRID, and if novel PPIs are shown validated by such updates to known PPIs, the information will be posted on the web-server.

4.6. In Vitro Pull-Down Assays

An initial screening to find physical interactions was performed using an in vitro pull- down assay for some of the predicted novel PPIs. This technique utilizes a His/biotin tag-fused protein immobilized on an affinity column as the bait protein and a passing-through solution containing the 'prey' protein that binds to the 'bait' protein. The subsequent elution will pull down both the target (prey) and tagged-protein (bait) for further analysis by immunoblotting to confirm the predicted interactions. The pull-down assays were conducted using the Pull-Down PolyHis Protein:Protein Interaction Kit (Pierce™, Rockford, IL, USA) according to the manufacturer's instructions.

4.7. Protein Identification Methods

Peptide sequencing experiments were performed using an EASY-nLC 1000 coupled to a Q Exactive Orbitrap Mass Spectrometer (Thermo Scientific, San Jose, CA, USA) operating in positive ion mode. An EasySpray C18 column (2 μm particle size, 75 μm diameter by 15 cm length) was loaded with 500 ng of protein digest in 22 μL of solvent A (water, 0.1% formic acid) at a pressure of 800 bar. Separations were performed using a linear gradient ramping from 5% solvent B (75% acetonitrile, 25% water, 0.1% formic acid) to 30% solvent B over 120 min, flowing at 300 nL/min.

The mass spectrometer was operated in data-dependent acquisition mode. Precursor scans were acquired at 70,000 resolution over 300–1750 m/z mass range (3e6 AGC target, 20 ms maximum injection time). Tandem MS spectra were acquired using HCD of the top 10 most abundant precursor ions at 17,500 resolution (NCE 28, 1e5 AGC target, 60 ms maximum injection time, 2.0 m/z isolation window). Charge states 1, 6–8 and higher were excluded for fragmentation and dynamic exclusion was set to 20.0 s.

Mass spectra were searched for peptide identifications using Proteome Discoverer 2.1 (Thermo Scientific, Waltham, MA, USA) using the Sequest HT and MSAmanda algorithms, peptide spectral matches were validated using Percolator (target FDR 1%). Initial searches were performed against the complete UniProt database (downloaded 19 March 2018). Pep- tide matches were restricted to 10 ppm MS1 tolerance, 20 mmu MS2 tolerance, and 2 missed tryptic cleavages. Fixed modifications were limited to cysteine carbamidomethylation, and dynamic modifications were methionine oxidation and protein N-terminal acetylation. Peptide and protein grouping and results validation was performed using Scaffold 4.8.4 (Proteome Software, Portland, OR, USA) along with the X! Tandem algorithm against the previously described database. Proteins were filtered using a 99% FDR threshold.

4.8. Ingenuity Pathway Analysis

Pathway associations of genes in the MPM interactome were computed using Ingenuity Pathway Analysis (IPA). Statistical significance of the overlaps between genes in the MPM interactome and pathways in the Ingenuity Knowledge Base (IKB) was computed with Fisher's exact test based on hypergeometric distribution. In this method, p-value is

computed from the probability of k successes in n draws (without replacement) from a finite population of size N containing exactly k objects with an interesting feature, where N = total number of genes associated with pathways in IKB, K = number of genes associated with a particular pathway in IKB, n = number of genes in the MPM interactome and k = K $\cap$ n. This value was further adjusted for multiple hypothesis correction using the Benjamini-Hochberg procedure.

4.9. Analysis of Differential Gene Expression in Pleural Mesothelioma Tumors and Lungs of Asbestos-Exposed Mice Versus Normal Tissue in Lungs

The overlap of the MPM interactome with genes differentially expressed in pleural mesothelioma tumors compared with normal pleural tissue adjacent to mesothelioma was computed using the dataset GSE12345 [75]. Genes differentially expressed in the lungs of mice exposed to crocidolite and erionite fibers were obtained from the dataset GSE100900 [78]. Genes with fold change >2 or $^1/_2$ were considered as significantly overexpressed and underexpressed respectively at p-value < 0.05.

4.10. Analysis of DNA Methylation in MPM Tumors

The dataset GSE16559 [69] was used to analyze the methylation profile of pleural mesotheliomas. In this study, genes found to be differentially methylated in mesothelioma were identified from a set of 773 cancer-related genes associated with 1413 autosomal CpG loci. Methylation values (M-values) were computed as M = log2 (β (1 β)) for both control (non-tumor pleural tissue) and test (pleural mesothelioma) cases, where β is the ratio of methylated probe intensity and overall intensity. Difference between M-values of test and control cases was then computed, and genes with M-value > 1 and M-value < 1 were considered to be hypermethylated and hypomethylated respectively at p-value < 0.05.

4.11. Correlating Expression of MPM Genes with Lung Cancer Prognosis

Data for correlation of gene expression and fraction of patient population surviving after treatment for lung cancer was taken from the Pathology Atlas [80]. Genes with log-rank p-value < 0.001 were considered to be prognostic. Unfavorable prognosis indicates positive correlation of high gene expression with reduced patient survival.

4.12. Identification of Repurposable Drugs in the MPM Drug-Protein Interactome

Negative correlation between lung cancer and drugs were studied using the BaseSpace correlation software, which uses a non-parametric rank-based approach to compute the extent of enrichment of a particular set of genes (or 'bioset') in another set of genes [31]. Readers may refer to Appendix A, titled 'Repurposable Drugs for Treatment of Malignant Pleural Mesothelioma (MPM)' for more details on the methodology used to identify repurposable drugs.

5. Conclusions

Biomedical discovery in the field of MPM research has to be accelerated to fuel clinically translatable results due to an urgent need to diagnose MPM preemptively, prevent its post-treatment recurrence, and curb its predicted increase in incidence in western and economically emerging nations [103]. In this study, we presented the MPM interactome as a valuable resource for mesothelioma biologists. We demonstrated its biological validity through comparison with MPM-related multi-omics data, which served to contextualize the novel PPIs within the mesothelioma landscape. Making novel MPM PPIs available freely on a webserver will catalyze investigations into these by cancer biologists and may lead to biologically or clinically translatable results. The MPM interactome with disease-associated proteins and their interacting partners will help biologists, bioinformaticians and clinicians to piece together an integrated view on how MPM-associated genes from various studies are functionally linked. Biological insights from this 'systems-level' view will help generate testable hypotheses and clinically translatable results. Future work

will focus on expanding this interactome by including interactions from additional PPI repositories, other mesothelioma types and mesothelioma datasets.

Appendix A Repurposable Drugs for Treatment of Malignant Pleural Mesothelioma (MPM)

We present here five drugs (*cabazitaxel*, *pyrimethamine*, *trimethoprim*, *primaquine* and *glicazide*) that could potentially be repurposed for the treatment of malignant pleural mesothelioma (MPM). These drugs were shortlisted through three types of analysis: (A) considering those that were already tested in non-small cell lung cancer (NSCLC), (B) gene expression analysis of drugs that target MPM genes or novel interactors in the MPM interactome, or (C) gene expression analysis of drugs that target known interactors in the malignant pleural mesothelioma (MPM) interactome. Drugs were selected based on whether they were already tested against lung cancer in clinical trials and/or showed overall negative correlation with lung cancer expression studies, because both mesothelioma and lung cancers have been shown to share common pathways that are initiated on exposure to asbestos fibres in mesothelial cells and lung epithelial cells respectively [35].

Another criterion used was whether the genes targeted by the drugs showed high differential expression in MPM tumours/cell lines. The details of these methods and observations are presented below.

Appendix A.1 Repurposable Drugs Already Tested in Non-Small Cell Lung Cancer

Nine overlapping drugs were found between drugs tested in NSCLC and drugs occurring in the MPM drug-protein interactome, that were negatively correlated with lung cancer expression studies, namely, cabazitaxel, dasatinib, docetaxel, gemcitabine, ipilimumab, ixabepilone, minocycline, pazopanib and pemetrexed. Minocycline was eliminated due to its toxicity. All of the remaining eight drugs were found to be effective in treatment of NSCLC (Table S2). Out of these eight drugs, cabazitaxel was the only drug that was not tested for treatment of mesothelioma. The fact that the other seven drugs were already tested against mesothelioma in clinical trials demonstrates the validity of our approach. It was interesting to note that drugs that targeted known interactors in addition to some MPM genes were found to have either no effect or limited clinical activity in mesothelioma, for e.g., dasatinib, docetaxel and pazopanib. On the other hand, drugs that targeted only MPM genes were found to be effective in treatment of mesothelioma or were capable of preventing disease progression, for e.g., gemcitabine, ipilimumab, ixabepilone and pemetrexed. This raises the suspicion that drugs that do not act on "off-target" genes (known interactors, in this case) may be more effective. In this respect, cabazitaxel, which targets the MPM genes *TUBB1* and *TUBA4A*, may be a suitable candidate for mesothelioma. Cabazitaxel was found to be effective in treatment of NSCLC resistant to docetaxel, a drug that targets *TUBB1* and other known interactors [37].

Appendix A.2 Repurposable Drugs Targeting MPM Genes and Novel Interactors

The MPM genes that were most differentially expressed with high significance in MPM tumors (GSE51024 [33]) were *TYMS* ($\log_2$FC = 1.82, *p*-value = 4.10×10^{-17}) and *DHFR* ($\log_2$FC = 0.89, *p*-value = 1.20×10^{-14}), and the drugs that target these genes (also having negative correlation with lung cancer expression) were pyrimethamine and trimethoprim. The first line drug currently used to treat mesothelioma is premetrexed, which targets proteins in the folate metabolic pathway, namely, *DHFR*, *TYMS* and *GART* [104]. Since MPM tumors have been shown to be responsive to anti-folates [68], both pyrimethamine (which targets only *DHFR*) and trimethoprim (which targets both *DHFR* and *TYMS*), seem to be interesting candidates. Pyrimethamine, an anti-parasitic drug commonly used to treat toxoplasmosis and cystoisosporiasis, has shown anti-tumor activity in metastatic melanoma cells and in murine models of breast cancer [105,106]. Trimethoprim, an anti-bacterial drug commonly used in the treatment of urinary bladder and respiratory tract infections, is also used to treat bacterial infections in cancer patients [107,108].

Keratin proteins form important components of the cell cytoskeleton, called intermediate filaments, in epithelial cells, and are commonly used as diagnostic markers in cancer [60].In addition to their role as cancer markers, their involvement in cellular functions such as cell motility, proliferation, cell polarity, protein synthesis, membrane trafficking and most importantly, tumour invasion and metastasis make them attractive as candidates for drug development [60]. *KRT7* is a keratin primarily expressed in mesothelial cells, apart from cells lining ducts and the intestine [60]. In a patient with malignant mesothelioma of the epithelioid type (which spreads to mediastinum and breast), *KRT7* was found to be significantly overexpressed when she developed resistance to pemetrexed/carboplatin, provided as a second line therapy [61]. The cancer cells showed a drastic increase in theirimmunoreactivity to CK7, the protein encoded by *KRT7* [61]. At the last stage of cancer progression (which was followed by her death), the patient showed dyspnoea (difficultyin breathing), increased tumour volume and pleural fluid [61]. In another case, *KRT7* was found be significantly overexpressed in an aggressive state of MPM, prior to treatment [61]. Two-thirds of malignant mesothelioma cases have been reported to be K7$^+$/K20$^-$ (pos- itive for expression of *KRT7* and negative for expression of *KRT20*) [60]. Expression of

KRT7 in three histological types of mesothelioma, namely, epithelioid, sarcomatid and biphasic, has been used to distinguish them from synovial sarcoma that metastasizes to the lungs and pleura [62]. *KRT7* has been identified as marker of circulating tumour cells in lung cancer [109]. *KRT7* was also found to be significantly upregulated in MPM tumours ($\log_2$FC = 3.80, *p*-value = 0.0002), and in cell line models of MPM ($\log_2$FC = 2.266, *p*-value = 0.029) (GSE2549 [34]). Positive expression of *KRT7* was noted in various types of non-small cell lung cancers, including large cell neuroendocrine carcinoma and lung adenocarcinoma [110,111]. In the MPM interactome, *KRT7* was predicted to interact with *KRT5*, an MPM gene that serves as a marker for malignant mesothelioma, along with vimentin, and is specifically used to distinguish pleural mesothelioma of the epithelioid type from pulmonary adenocarcinoma and non-pulmonary adenocarcinoma metastasizing to pleura [60,112]. *KRT7* is a target of primaquine, an-antimalarial agent known to destroy the malarial parasites, *Plasmodium vivax* and *Plasmodium ovale*, inside the liver [113,114]. The exact mechanism of action has not been elucidated for this drug. However, in independent studies, primaquine has been shown to bind to keratin in a concentration-dependent manner, and also mediate strong membrane perturbations in cell membrane models [113,115]. Since high expression of *KRT7* has been shown to be related to tumour aggressiveness and drug resistance in malignant mesothelioma, and its high expression was also noted in MPM tumours and cell lines, primaquine may be re-purposed for treatment of MPM atleast as an adjunctive drug with pemetrexed, the drug currently used for first line therapy.It is interesting to note that primaquine enhanced the sensitivity of KBV20C cells to cancer drugs, namely, vinblastine, vinorelbine, paclitaxel, docetaxel, vincristine and halaven [63]. KVB20C is a multi-drug resistant cell line derived from oral squamous carcinoma. Primaquine compounds (substituted quinolines) have also been shown to exert anti-tumor activity in breast cancer cells [116].

Appendix A.3 Repurposable Drugs Targeting Known Interactors

Out of the four drugs targeting known interactors in the MPM interactome and showing negative correlation with lung cancer associated gene expression, three drugs were al- ready known to exhibit anti-tumour activity in pre-clinical models of mesothelioma, namely,zoledronate, sirolimus and the vitamin E analog, alpha-tocopheryl succinate, which shows the validity of our approach. Zoledronate, which showed modest activity in MPM, induced apoptosis and S-phase arrest in human mesothelioma cells and inhibited tumor growth in the pleural cavity of an orthotopic animal model [54,55]. Sirolimus/cisplatin increased cell death and decreased cell proliferation in cell lines of MPM [56]. Alpha-tocopheryl succinate increased survival of orthotopic animal models of malignant peritoneal mesothelioma [57]. Zoledronate has demonstrated modest clinical activity in patients with advanced MPM [54]. Sirolimus has not been tested against MPM in clinical trials, but everolimus, a drug derived from sirolimus sharing similar properties with it, has shown only limited clinical activityin MPM, and further testing as a single-agent was not advised based on the results from this study [117]. Both vitamin E and its analog, alpha-tocopheryl succinate have not beentested against MPM in clinical trials. However, testing of vitamin E and its analogs are being carried out in various pre-clinical settings [58,59]. Hence, it was the drug gliclazide that emerged as a potentially repurposable drug, untested against MPM.

Gliclazide, an anti-diabetic drug, inhibits *VEGFA*, which has been shown to be significantly upregulated ($\log_2$FC = 1.83, *p*-value = 0.0018) in MPM tumour (GSE2549 [34]). This drug inhibits VEGF expression induced by advanced glycation end products in bovine reticular endothelial cells, and VEGF expression induced by ischemia in retinal tissueof mice [64,118]. In the latter case, glicazide also inhibits neovascularization, a process known to be mediated by VEGF. VEGF has been identified as a prognostic marker for MPM. High levels of VEGF have been correlated with both asbestos exposure in MPM, and an advanced stage of the disease [65,66]. It is interesting to note that glibenclamide, adrug whose mechanism of action is similar to that of glicazide, has been shown to increase caspase activity in MPM cell lines and primary cultures, leading to apoptosis mediated by

TNF-related apoptosis inducing ligand (*TRAIL*) [67]. Hence, glicazide may be repurposedto inhibit neovascularization and perhaps enhance apoptosis in MPM.

References

1. Mutsaers, S.E. The mesothelial cell. *Int. J. Biochem. Cell Biol.* **2004**, *36*, 9–16. [CrossRef]
2. Carbone, M.; Adusumilli, P.S.; Alexander, H.R., Jr.; Baas, P.; Bardelli, F.; Bononi, A.; Bueno, R.; Felley-Bosco, E.; Galateau-Salle, F.;Jablons, D.; et al. Mesothelioma: Scientific clues for prevention, diagnosis, and therapy. *CA Cancer J. Clin.* **2019**, *69*, 402–429. [CrossRef] [PubMed]
3. Wang, Z.J.; Reddy, G.P.; Gotway, M.B.; Higgins, C.B.; Jablons, D.M.; Ramaswamy, M.; Hawkins, R.A.; Webb, W.R. Malignant Pleural Mesothelioma: Evaluation with CT, MR Imaging, and PET. *Radiographics* **2004**, *24*, 105–119. [CrossRef] [PubMed]
4. Lang-Lazdunski, L. Malignant pleural mesothelioma: Some progress, but still a long way from cure. *J. Thorac. Dis.* **2018**, *10*, 1172–1177. [CrossRef]
5. Bueno, R.; Stawiski, E.W.; Goldstein, L.D.; Durinck, S.; De Rienzo, A.; Modrusan, Z.; Gnad, F.; Nguyen, T.T.; Jaiswal, B.S.; Chirieac, L.R.; et al. Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. *Nat. Genet.* **2016**, *48*, 407–416. [CrossRef]
6. Bott, M.; Brevet, M.; Taylor, B.S.; Shimizu, S.; Ito, T.; Wang, L.; Creaney, J.; Lake, R.A.; Zakowski, M.F.; Reva, B.; et al. The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. *Nat. Genet.* **2011**, *43*, 668–672. [CrossRef]
7. Jensen, D.E.; Proctor, M.; Marquis, S.T.; Gardner, H.P.; Ha, S.I.; Chodosh, L.A.; Ishov, A.M.; Tommerup, N.; Vissing, H.; Sekido, Y.; et al. BAP1: A novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **1998**, *16*, 1097–1112. [CrossRef] [PubMed]
8. Hakiri, S.; Osada, H.; Ishiguro, F.; Murakami, H.; Murakami-Tonami, Y.; Yokoi, K.; Sekido, Y. Functional differences between wild-type and mutant-type BRCA1-associated protein 1 tumor suppressor against malignant mesothelioma cells. *Cancer Sci.* **2015**, *106*, 990–999. [CrossRef]
9. Zauderer, M.G.; Kass, S.L.; Woo, K.; Sima, C.S.; Ginsberg, M.S.; Krug, L.M. Vinorelbine and gemcitabine as second- or third-line therapy for malignant pleural mesothelioma. *Lung Cancer* **2014**, *84*, 271–274. [CrossRef]
10. Zucali, P.; Perrino, M.; Lorenzi, E.; Ceresoli, G.; De Vincenzo, F.; Simonelli, M.; Gianoncelli, L.; De Sanctis, R.; Giordano, L.; Santoro, A. Vinorelbine in pemetrexed-pretreated patients with malignant pleural mesothelioma. *Lung Cancer* **2014**, *84*, 265–270. [CrossRef]
11. Zauderer, M.G.; Bott, M.; McMillan, R.; Sima, C.S.; Rusch, V.; Krug, L.M.; Ladanyi, M. Clinical characteristics of patients with malignant pleural mesothelioma harboring somatic BAP1 mutations. *J. Thorac. Oncol.* **2013**, *8*, 1430–1433. [CrossRef]
12. Busacca, S.; Sheaff, M.; Arthur, K.; Gray, S.G.; O'Byrne, K.J.; Richard, D.J.; Soltermann, A.; Opitz, I.; Pass, H.; Harkin, D.P.; et al. BRCA1 is an essential mediator of vinorelbine-induced apoptosis in mesothelioma. *J. Pathol.* **2012**, *227*, 200–208. [CrossRef]
13. Toyokawa, G.; Takenoyama, M.; Hirai, F.; Toyozawa, R.; Inamasu, E.; Kojo, M.; Morodomi, Y.; Shiraishi, Y.; Takenaka, T.; Yamaguchi, M.; et al. Gemcitabine and vinorelbine as second-line or beyond treatment in patients with malignant pleural mesothelioma pretreated with platinum plus pemetrexed chemotherapy. *Int. J. Clin. Oncol.* **2013**, *19*, 601–606. [CrossRef]
14. Sahni, N.; Yi, S.; Taipale, M.; Bass, J.I.F.; Coulombe-Huntington, J.; Yang, F.; Peng, J.; Weile, J.; Karras, G.I.; Wang, Y.; et al. Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell* **2015**, *161*, 647–660. [CrossRef]
15. Rolland, T.; Taşan, M.; Charloteaux, B.; Pevzner, S.J.; Zhong, Q.; Sahni, N.; Yi, S.; Lemmens, I.; Fontanillo, C.; Mosca, R.; et al. A Proteome-Scale Map of the Human Interactome Network. *Cell* **2014**, *159*, 1212–1226. [CrossRef]
16. Ganapathiraju, M.K.; Thahir, M.; Handen, A.; Sarkar, S.N.; Sweet, R.A.; Nimgaonkar, V.L.; Loscher, C.E.; Bauer, E.M.; Chaparala, S. Schizophrenia interactome with 504 novel protein–protein interactions. *Npj Schizophr.* **2016**, *2*, 16012. [CrossRef]
17. Zhu, J.; Zhang, Y.; Ghosh, A.; Cuevas, R.A.; Forero, A.; Dhar, J.; Ibsen, M.S.; Schmid-Burgk, J.L.; Schmidt, T.; Ganapathiraju, M.K.; et al. Antiviral Activity of Human OASL Protein Is Mediated by Enhancing Signaling of the RIG-I RNA Sensor. *Immunity* **2014**, *40*, 936–948. [CrossRef] [PubMed]
18. Li, Y.; Klena, N.T.; Gabriel, G.C.; Liu, X.; Kim, A.J.; Lemke, K.; Chen, Y.; Chatterjee, B.; Devine, W.; Damerla, R.R.; et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nat. Cell Biol.* **2015**, *521*, 520–524. [CrossRef] [PubMed]
19. Liu, X.; Yagi, H.; Saeed, S.; Bais, A.S.; Gabriel, G.C.; Chen, Z.; Peterson, K.; Li, Y.; Schwartz, M.C.; Reynolds, W.T.; et al. The complex genetics of hypoplastic left heart syndrome. *Nat. Genet.* **2017**, *49*, 1152–1159. [CrossRef] [PubMed]
20. Karunakaran, K.B.; Chaparala, S.; Ganapathiraju, M.K. Potentially repurposable drugs for schizophrenia identified from its interactome. *Sci. Rep.* **2019**, *9*, 1–14. [CrossRef] [PubMed]
21. Cedrés, S.; Montero, M.; Martinez, P.; Rodríguez-Freixinós, V.; Torrejon, D.; Gabaldon, A.; Salcedo, M.; Cajal, S.R.Y.; Felip, E. Exploratory analysis of activation of PTEN–PI3K pathway and downstream proteins in malignant pleural mesothelioma (MPM). *Lung Cancer* **2012**, *77*, 192–198. [CrossRef] [PubMed]
22. Prasad, T.S.K.; Goel, R.; Kandasamy, K.; Keerthikumar, S.; Kumar, S.; Mathivanan, S.; Telikicherla, D.; Raju, R.; Shafreen, B.; Venugopal, A.; et al. Human Protein Reference Database—2009 update. *Nucleic Acids Res.* **2008**, *37*, D767–D772. [CrossRef] [PubMed]

23. Stark, C.; Breitkreutz, B.-J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34*, D535–D539. [CrossRef] [PubMed]

24. Karunakaran, K.B.; Yanamala, N.; Boyce, G.; Ganapathiraju, M.K. Mesothelioma Interactome with 367 Novel Protein-Protein Interactions. *bioRxiv* **2018**, 459065. [CrossRef]

25. Wu, G.; Dawson, E.; Duong, A.; Haw, R.; Stein, L. ReactomeFIViz: A Cytoscape app for pathway and network-based data analysis. *F1000Research* **2014**, *3*, 146. [CrossRef]

26. Orii, N.; Ganapathiraju, M.K. Wiki-Pi: A Web-Server of Annotated Human Protein-Protein Interactions to Aid in Discovery of Protein Function. *PLoS ONE* **2012**, *7*, e49029. [CrossRef]

27. Amin, W.; Parwani, A.V.; Melamed, J.; Flores, R.M.; Pennathur, A.; Valdivieso, F.A.; Whelan, N.B.; Landreneau, R.; Luketich, J.D.; Feldman, M.; et al. National Mesothelioma Virtual Bank: A Platform for Collaborative Research and Mesothelioma Biobanking Resource to Support Translational Research. *Lung Cancer Int.* **2013**, *2013*, 1–9. [CrossRef]

28. Amin, W.; Singh, H.; Pople, A.K.; Winters, S.; Dhir, R.; Parwani, A.V.; Becich, M.J. A decade of experience in the development and implementation of tissue banking informatics tools for intra and inter-institutional translational research. *J. Pathol. Inform.* **2010**, *1*, 12. [CrossRef]

29. Krämer, A.; Green, J.; Pollard, J.; Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014**, *30*, 523–530. [CrossRef]

30. LoPiccolo, J.; Granville, C.A.; Gills, J.J.; Dennis, P.A. Targeting Akt in cancer therapy. *Anti-Cancer Drugs* **2007**, *18*, 861–874. [CrossRef]

31. Kupershmidt, I.; Su, Q.J.; Grewal, A.; Sundaresh, S.; Halperin, I.; Flynn, J.; Shekar, M.; Wang, H.; Park, J.; Cui, W.; et al. Ontology-Based Meta-Analysis of Global Collections of High-Throughput Public Data. *PLoS ONE* **2010**, *5*, e13066. [CrossRef]

32. Chattopadhyay, A.; Ganapathiraju, M.K. Demonstration Study: A Protocol to Combine Online Tools and Databases for Identifying Potentially Repurposable Drugs. *Data* **2017**, *2*, 15. [CrossRef]

33. Suraokar, M.B.; Nunez, M.I.; Diao, L.; Chow, C.W.; Kim, D.; Behrens, C.; Lin, H.; Lee, S.; Raso, G.; Moran, C.; et al. Expression profiling stratifies mesothelioma tumors and signifies deregulation of spindle checkpoint pathway and microtubule network with therapeutic implications. *Ann. Oncol.* **2014**, *25*, 1184–1192. [CrossRef] [PubMed]

34. Gordon, G.J.; Rockwell, G.N.; Jensen, R.V.; Rheinwald, J.G.; Glickman, J.N.; Aronson, J.P.; Pottorf, B.J.; Nitz, M.D.; Richards, W.G.; Sugarbaker, D.J.; et al. Identification of Novel Candidate Oncogenes and Tumor Suppressors in Malignant Pleural Mesothelioma Using Large-Scale Transcriptional Profiling. *Am. J. Pathol.* **2005**, *166*, 1827–1840. [CrossRef]

35. Heintz, N.H.; Janssen-Heininger, Y.M.; Mossman, B.T. Asbestos, lung cancers, and mesotheliomas: From molecular approaches to targeting tumor survival pathways. *Am. J. Respir. Cell Mol. Biol.* **2010**, *42*, 133–139. [CrossRef] [PubMed]

36. Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nuc. Acids Res.* **2008**, *36*, D901–D906. [CrossRef] [PubMed]

37. Kotsakis, A.; Matikas, A.; Koinis, F.; Kentepozidis, N.; Varthalitis, I.I.; Karavassilis, V.; Samantas, E.; Katsaounis, P.; Dermitzaki, E.K.; Hatzidaki, D.; et al. A multicentre phase II trial of cabazitaxel in patients with advanced non-small-cell lung cancer progressing after docetaxel-based chemotherapy. *Br. J. Cancer* **2016**, *115*, 784–788. [CrossRef] [PubMed]

38. Johnson, F.M.; Bekele, B.N.; Feng, L.; Wistuba, I.; Tang, X.M.; Tran, H.T.; Erasmus, J.J.; Hwang, L.-L.; Takebe, N.; Blumenschein, G.R.; et al. Phase II Study of Dasatinib in Patients with Advanced Non–Small-Cell Lung Cancer. *J. Clin. Oncol.* **2010**, *28*, 4609–4615. [CrossRef]

39. Tsao, A.S.; Lin, H.; Carter, B.W.; Lee, J.J.; Rice, D.; Vaporcyan, A.; Swisher, S.; Mehran, R.; Heymach, J.; Nilsson, M.; et al. Biomarker-Integrated Neoadjuvant Dasatinib Trial in Resectable Malignant Pleural Mesothelioma. *J. Thorac. Oncol.* **2018**, *13*, 246–257. [CrossRef]

40. Comer, A.M.; Goa, K.L. Docetaxel. A review of its use in non-small cell lung cancer. *Drugs Aging* **2000**, *17*, 53–80. [CrossRef]

41. Belani, C.P.; Adak, S.; Aisner, S.; Stella, P.J.; Levitan, N.; Johnson, D.H. Docetaxel for malignant mesothelioma: Phase II study of the Eastern Cooperative Oncology Group. *Clin. Lung Cancer* **2004**, *6*, 43–47. [CrossRef]

42. Ralli, M.; Tourkantonis, I.; Makrilia, N.; Gkini, E.; Kotteas, E.; Gkiozos, I.; Katirtzoglou, N.; Syrigos, K. Docetaxel plus gemcitabine as first-line treatment in malignant pleural mesothelioma: A single institution phase II study. *Anticancer Res.* **2009**, *29*, 3441–3444.

43. Tourkantonis, I.; Makrilia, N.; Ralli, M.; Alamara, C.; Nikolaidis, I.; Tsimpoukis, S.; Charpidou, A.; Kotanidou, A.; Syrigos, K. Phase II study of gemcitabine plus docetaxel as second-line treatment in malignant pleural mesothelioma: A single institution study. *Am. J. Clin. Oncol.* **2011**, *34*, 38–42. [CrossRef]

44. Manegold, C. Gemcitabine (Gemzar®) in non-small cell lung cancer. *Expert Rev. Anticancer Ther.* **2004**, *4*, 345–360. [CrossRef]

45. Kindler, H.L.; van Meerbeeck, J.P. The role of gemcitabine in the treatment of malignant mesothelioma. *Semin. Oncol.* **2002**, *29*, 70–76. [CrossRef]

46. Malhotra, J.; Jabbour, S.K.; Aisner, J. Current state of immunotherapy for non-small cell lung cancer. *Transl. Lung Cancer Res.* **2007**, *6*, 196–211. [CrossRef]

47. Scherpereel, A.; Mazieres, J.; Greillier, L.; Dô, P.; Bylicki, O.; Monnet, I.; Corre, R.; Audigier-Valette, C.; Locatelli-Sanchez, M.; Molinier, O. Second-or third-line nivolumab (Nivo) versus nivo plus ipilimumab (Ipi) in malignant pleural mesothelioma (MPM) patients: Results of the IFCT-1501 MAPS2 randomized phase II trial. *Am. Soc. Clin. Oncol.* **2017**, *35*, LBA8507. [CrossRef]

48. Spigel, D.R.; Greco, F.A.; Waterhouse, D.M.; Shipley, D.L.; Zubkus, J.D.; Bury, M.J.; Webb, C.D.; Hart, L.L.; Gian, V.G.; Infante, J.R.; et al. Phase II trial of ixabepilone and carboplatin with or without bevacizumab in patients with previously untreated advanced non-small-cell lung cancer. *Lung Cancer* **2012**, *78*, 70–75. [CrossRef] [PubMed]

49. Puhalla, S.; Brufsky, A. Ixabepilone: A new chemotherapeutic option for refractory metastatic breast cancer. *Biol. Targets Ther.* **2008**, *2*, 505.

50. Altorki, N.; Lane, M.E.; Bauer, T.; Lee, P.C.; Guarino, M.J.; Pass, H.; Felip, E.; Peylan-Ramu, N.; Gurpide, A.; Grannis, F.W.; et al. Phase II Proof-of-Concept Study of Pazopanib Monotherapy in Treatment-Naive Patients with Stage I/II Resectable Non–Small-Cell Lung Cancer. *J. Clin. Oncol.* **2010**, *28*, 3131–3137. [CrossRef] [PubMed]

51. Hiddinga, B.I.; Rolfo, C.; Van Meerbeeck, J.P. Mesothelioma treatment: Are we on target? A review. *J. Adv. Res.* **2015**, *6*, 319–330. [CrossRef]

52. Scagliotti, G.; Parikh, P.; Von Pawel, J.; Biesma, B.; Vansteenkiste, J.; Manegold, C.; Serwatowski, P.; Gatzemeier, U.; Digumarti, R.; Zukin, M.; et al. Phase III Study Comparing Cisplatin Plus Gemcitabine with Cisplatin Plus Pemetrexed in Chemotherapy-Naive Patients with Advanced-Stage Non–Small-Cell Lung Cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin.* **2008**, *26*, 3543–3551. [CrossRef]

53. Vogelzang, N.J.; Rusthoven, J.J.; Symanowski, J.; Denham, C.; Kaukel, E.; Ruffie, P.; Gatzemeier, U.; Boyer, M.; Emri, S.; Manegold, C.; et al. Phase III Study of Pemetrexed in Combination with Cisplatin Versus Cisplatin Alone in Patients with Malignant Pleural Mesothelioma. *J. Clin. Oncol.* **2003**, *21*, 2636–2644. [CrossRef]

54. Jamil, M.O.; Jerome, M.S.; Miley, D.; Selander, K.S.; Robert, F. A pilot study of zoledronic acid in the treatment of patients with advanced malignant pleural mesothelioma. *Lung Cancer Targets Ther.* **2017**, *8*, 39–44. [CrossRef]

55. Okamoto, S.; Kawamura, K.; Li, Q.; Yamanaka, M.; Yang, S.; Fukamachi, T.; Tada, Y.; Tatsumi, K.; Shimada, H.; Hiroshima, K.; et al. Zoledronic Acid Produces Antitumor Effects on Mesothelioma Through Apoptosis and S-Phase Arrest in p53-Independent and Ras prenylation-Independent Manners. *J. Thorac. Oncol.* **2012**, *7*, 873–882. [CrossRef]

56. Hartman, M.-L.; Esposito, J.M.; Yeap, B.Y.; Sugarbaker, D.J. Combined treatment with cisplatin and sirolimus to enhance cell death in human mesothelioma. *J. Thorac. Cardiovasc. Surg.* **2010**, *139*, 1233–1240. [CrossRef]

57. Tomasetti, M.; Gellert, N.; Procopio, A.; Neuzil, J. A vitamin E analogue suppresses malignant mesothelioma in a preclinical model: A future drug against a fatal neoplastic disease? *Int. J. Cancer* **2004**, *109*, 641–642. [CrossRef] [PubMed]

58. Kovarova, J.; Bajzikova, M.; Vondrusová, M.; Stursa, J.; Goodwin, J.; Nguyen, M.; Zobalova, R.; Pesdar, E.A.; Truksa, J.; Tomasetti, M.; et al. Mitochondrial targeting of α-tocopheryl succinate enhances its anti-mesothelioma efficacy. *Redox Rep.* **2014**, *19*, 16–25. [CrossRef] [PubMed]

59. Sato, A.; Virgona, N.; Sekine, Y.; Yano, T. The evidence to date: A redox-inactive analogue of tocotrienol as a new anti-mesothelioma agent. *J. Rare Dis. Res. Treat.* **2016**, *2*, 38–42.

60. Karantza, V. Keratins in health and cancer: More than mere epithelial cell markers. *Oncogene* **2010**, *30*, 127–138. [CrossRef]

61. Røe, O.D.; Szulkin, A.; Anderssen, E.; Flatberg, A.; Sandeck, H.; Amundsen, T.; Erlandsen, S.E.; Dobra, K.; Sundstrøm, S.H. Molecular resistance fingerprint of pemetrexed and platinum in a long-term survivor of mesothelioma. *PLoS ONE* **2012**, *7*, e40521. [CrossRef] [PubMed]

62. Miettinen, M.; Limon, J.; Niezabitowski, A.; Lasota, J. Calretinin and other mesothelioma markers in synovial sarcoma: Analysis of antigenic similarities and differences with malignant mesothelioma. *Am. J. Surg. Pathol.* **2001**, *25*, 610–617. [CrossRef] [PubMed]

63. Choi, A.-R.; Kim, J.-H.; Woo, Y.H.; Kim, H.S.; Yoon, S. Anti-malarial Drugs Primaquine and Chloroquine Have Different Sensitization Effects with Anti-mitotic Drugs in Resistant Cancer Cells. *Anticancer Res.* **2016**, *36*, 1641–1648. [CrossRef]

64. Kimura, T.; Takagi, H.; Suzuma, K.; Kita, M.; Watanabe, D.; Yoshimura, N. Comparisons between the beneficial effects of different sulphonylurea treatments on ischemia-induced retinal neovascularization. *Free Radic. Biol. Med.* **2007**, *43*, 454–461. [CrossRef]

65. Yasumitsu, A.; Tabata, C.; Tabata, R.; Hirayama, N.; Murakami, A.; Yamada, S.; Terada, T.; Iida, S.; Tamura, K.; Fukuoka, K.; et al. Clinical Significance of Serum Vascular Endothelial Growth Factor in Malignant Pleural Mesothelioma. *J. Thorac. Oncol.* **2010**, *5*, 479–483. [CrossRef]

66. Hirayama, N.; Tabata, C.; Tabata, R.; Maeda, R.; Yasumitsu, A.; Yamada, S.; Kuribayashi, K.; Fukuoka, K.; Nakano, T. Pleural effusion VEGF levels as a prognostic factor of malignant pleural mesothelioma. *Respir. Med.* **2011**, *105*, 137–142. [CrossRef]

67. Pasello, G.; Urso, L.; Conte, P.; Favaretto, A. Effects of Sulfonylureas on Tumor Growth: A Review of the Literature. *Oncologist* **2013**, *18*, 1118–1125. [CrossRef]

68. Krug, L.M.; Heelan, R.T.; Kris, M.G.; Venkatraman, E.; Sirotnak, F.M. Phase II Trial of Pralatrexate (10-Propargyl-10-deazaaminopterin, PDX) in Patients with Unresectable Malignant Pleural Mesothelioma. *J. Thorac. Oncol.* **2007**, *2*, 317–320. [CrossRef] [PubMed]

69. Christensen, B.C.; Marsit, C.J.; Houseman, E.A.; Godleski, J.J.; Longacker, J.L.; Zheng, S.; Yeh, R.-F.; Wrensch, M.R.; Wiemels, J.L.; Karagas, M.R.; et al. Differentiation of Lung Adenocarcinoma, Pleural Mesothelioma, and Nonmalignant Pulmonary Tissues Using DNA Methylation Profiles. *Cancer Res.* **2009**, *69*, 6315–6321. [CrossRef]

70. Wang, S.; Jiang, L.; Han, Y.; Chew, S.H.; Ohara, Y.; Akatsuka, S.; Weng, L.; Kawaguchi, K.; Fukui, T.; Sekido, Y.; et al. Urokinase-type plasminogen activator receptor promotes proliferation and invasion with reduced cisplatin sensitivity in malignant mesothelioma. *Oncotarget* **2016**, *7*, 69565–69578. [CrossRef] [PubMed]

71. Marek, L.A.; Hinz, T.K.; Von Mässenhausen, A.; Olszewski, K.A.; Kleczko, E.K.; Boehm, D.; Weiser-Evans, M.C.; Nemenoff, R.A.; Hoffmann, H.; Warth, A.; et al. Nonamplified FGFR1 Is a Growth Driver in Malignant Pleural Mesothelioma. *Mol. Cancer Res.* **2014**, *12*, 1460–1469. [CrossRef]

72. Wilson, T.R.; Lee, D.Y.; Berry, L.; Shames, D.S.; Settleman, J. Neuregulin-1-Mediated Autocrine Signaling Underlies Sensitivity to HER2 Kinase Inhibitors in a Subset of Human Cancers. *Cancer Cell* **2011**, *20*, 158–172. [CrossRef] [PubMed]

73. Kaarteenaho-Wiik, R.; Soini, Y.; Pöllänen, R.; Pääkkö, P.; Kinnula, V. Over-expression of tenascin-C in malignant pleural mesothelioma. *Histopathology* **2003**, *42*, 280–291. [CrossRef] [PubMed]

74. Lin, C.-C.; Chen, L.-C.; Tseng, V.S.; Yan, J.-J.; Lai, W.-W.; Su, W.-P.; Huang, C.-Y.F. Malignant pleural effusion cells show aberrant glucose metabolism gene expression. *Eur. Respir. J.* **2010**, *37*, 1453–1465. [CrossRef] [PubMed]

75. Crispi, S.; Calogero, R.A.; Santini, M.; Mellone, P.; Vincenzi, B.; Citro, G.; Vicidomini, G.; Fasano, S.; Meccariello, R.; Cobellis, G.; et al. Global Gene Expression Profiling of Human Pleural Mesotheliomas: Identification of Matrix Metalloproteinase 14 (MMP-14) as Potential Tumour Target. *PLoS ONE* **2009**, *4*, e7016. [CrossRef]

76. De Rienzo, A.; Richards, W.G.; Yeap, B.Y.; Coleman, M.H.; Sugarbaker, P.E.; Chirieac, L.R.; Wang, Y.E.; Quackenbush, J.; Jensen, R.V.; Bueno, R. Sequential Binary Gene Ratio Tests Define a Novel Molecular Diagnostic Strategy for Malignant Pleural Mesothelioma. *Clin. Cancer Res.* **2013**, *19*, 2493–2502. [CrossRef]

77. Consortium, G. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **2015**, *348*, 648–660. [CrossRef] [PubMed]

78. Yanamala, N.; Kisin, E.R.; Gutkin, D.W.; Shurin, M.R.; Harper, M.; Shvedova, A.A. Characterization of pulmonary responses in mice to asbestos/asbestiform fibers using gene expression profiles. *J. Toxicol. Environ. Health Part A* **2017**, *81*, 60–79. [CrossRef] [PubMed]

79. Maxim, L.D.; McConnell, E.E. A Review of the Toxicology and Epidemiology of Wollastonite. *Inhal. Toxicol.* **2005**, *17*, 451–466. [CrossRef]

80. Uhlén, M.; Zhang, C.; Lee, S.; Sjöstedt, E.; Fagerberg, L.; Bidkhori, G.; Benfeitas, R.; Arif, M.; Liu, Z.; Edfors, F.; et al. A pathology atlas of the human cancer transcriptome. *Science* **2017**, *357*, eaan2507. [CrossRef]

81. Strizzi, L.; Catalano, A.; Vianale, G.; Orecchia, S.; Casalini, A.; Tassi, G.; Puntoni, R.; Mutti, L.; Procopio, A. Vascular endothelial growth factor is an autocrine growth factor in human malignant mesothelioma. *J. Pathol.* **2001**, *193*, 468–475. [CrossRef]

82. Seto, T.; Higashiyama, M.; Funai, H.; Imamura, F.; Uematsu, K.; Seki, N.; Eguchi, K.; Yamanaka, T.; Ichinose, Y. Prognostic value of expression of vascular endothelial growth factor and its flt-1 and KDR receptors in stage I non-small-cell lung cancer. *Lung Cancer* **2006**, *53*, 91–96. [CrossRef] [PubMed]

83. Greening, D.W.; Ji, H.; Chen, M.; Robinson, B.W.S.; Dick, I.M.; Creaney, J.; Simpson, R.J. Secreted primary human malignant mesothelioma exosome signature reflects oncogenic cargo. *Sci. Rep.* **2016**, *6*, 32643. [CrossRef]

84. Großerueschkamp, F.; Bracht, T.; Diehl, H.C.; Kuepper, C.; Ahrens, M.; Kallenbach-Thieltges, A.; Mosig, A.; Eisenacher, M.; Marcus, K.; Behrens, T.; et al. Spatial and molecular resolution of diffuse malignant mesothelioma heterogeneity by integrating label-free FTIR imaging, laser capture microdissection and proteomics. *Sci. Rep.* **2017**, *7*, srep44829. [CrossRef]

85. Cigognetti, M.; Lonardi, S.; Fisogni, S.; Balzarini, P.; Pellegrini, V.; Tironi, A.; Bercich, L.; Bugatti, M.; De Rossi, G.; Murer, B.; et al. BAP1 (BRCA1-associated protein 1) is a highly specific marker for differentiating mesothelioma from reactive mesothelial proliferations. *Mod. Pathol.* **2015**, *28*, 1043–1057. [CrossRef]

86. Lupo, B.; Trusolino, L. Inhibition of poly(ADP-ribosyl)ation in cancer: Old and new paradigms revisited. *Biochim. Biophys. Acta (BBA) Bioenerg.* **2014**, *1846*, 201–215. [CrossRef]

87. Nasu, M. *Identification of bap1 as a Predisposing Gene for Malignant Mesothelioma*; University of Hawaii: Manoa, HI, USA, 2012.

88. Yao, Z.-H.; Tian, G.-Y.; Yang, S.-X.; Wan, Y.-Y.; Kang, Y.-M.; Liu, Q.-H.; Yao, F.; Lin, D.-J. Serum albumin as a significant prognostic factor in patients with malignant pleural mesothelioma. *Tumor Biol.* **2014**, *35*, 6839–6845. [CrossRef] [PubMed]

89. Iacono, M.L.; Monica, V.; Righi, L.; Grosso, F.; Libener, R.; Vatrano, S.; Bironzo, P.; Novello, S.; Musmeci, L.; Volante, M.; et al. Targeted Next-Generation Sequencing of Cancer Genes in Advanced Stage Malignant Pleural Mesothelioma: A Retrospective Study. *J. Thorac. Oncol.* **2015**, *10*, 492–499. [CrossRef]

90. Liu, Y.; Zou, X.; Sun, G.; Bao, Y. Codonopsis lanceolata polysaccharide CLPS inhibits melanoma metastasis via regulating integrin signaling. *Int. J. Biol. Macromol.* **2017**, *103*, 435–440. [CrossRef] [PubMed]

91. Okamoto, T.; Iwata, S.; Yamazaki, H.; Hatano, R.; Komiya, E.; Dang, N.H.; Ohnuma, K.; Morimoto, C. CD9 Negatively Regulates CD26 Expression and Inhibits CD26-Mediated Enhancement of Invasive Potential of Malignant Mesothelioma Cells. *PLoS ONE* **2014**, *9*, e86671. [CrossRef] [PubMed]

92. Pereira, S.; Lowell, C. The Lyn Tyrosine Kinase Negatively Regulates Neutrophil Integrin Signaling. *J. Immunol.* **2003**, *171*, 1319–1327. [CrossRef] [PubMed]

93. Orlova, V.V.; Choi, E.Y.; Xie, C.; Chavakis, E.; Bierhaus, A.; Ihanus, E.; Ballantyne, C.M.; Gahmberg, C.G.; Bianchi, M.E.; Nawroth, P.P.; et al. A novel pathway of HMGB1-mediated inflammatory cell recruitment that requires Mac-1-integrin. *EMBO J.* **2007**, *26*, 1129–1139. [CrossRef] [PubMed]

94. Podar, K.; Tai, Y.-T.; Lin, B.K.; Narsimhan, R.P.; Sattler, M.; Kijima, T.; Salgia, R.; Gupta, D.; Chauhan, D.; Anderson, K.C. Vascular Endothelial Growth Factor-induced Migration of Multiple Myeloma Cells Is Associated with β1 Integrin and Phosphatidylinositol 3-Kinase-dependent PKCα Activation. *J. Biol. Chem.* **2002**, *277*, 7875–7881. [CrossRef] [PubMed]

95. Zhang, X.; Yang, L.; Chen, W.; Kong, M. Identification of Potential Hub Genes and Therapeutic Drugs in Malignant Pleural Mesothelioma by Integrated Bioinformatics Analysis. *Oncol. Res. Treat.* **2020**, *43*, 656–671. [CrossRef]

96. Cheresh, P.; Morales-Nebreda, L.; Kim, S.-J.; Yeldandi, A.; Williams, D.B.; Cheng, Y.; Mutlu, G.M.; Budinger, G.R.S.; Ridge, K.; Schumacker, P.T.; et al. Asbestos-Induced Pulmonary Fibrosis Is Augmented in 8-Oxoguanine DNA Glycosylase Knockout Mice. *Am. J. Respir. Cell Mol. Biol.* **2015**, *52*, 25–36. [CrossRef] [PubMed]

97. Bozelka, B.; Sestini, P.; Gaumer, H.; Hammad, Y.; Heather, C.; Salvaggio, J. A murine model of asbestosis. *Am. J. Pathol.* **1983**, *112*, 326.

98. Rehrauer, H.; Wu, L.; Blum, W.; Pecze, L.; Henzi, T.; Serre-Beinier, V.; Aquino, C.; Vrugt, B.; De Perrot, M.; Schwaller, B.; et al. How asbestos drives the tissue towards tumors: YAP activation, macrophage and mesothelial precursor recruitment, RNA editing, and somatic mutations. *Oncogene* **2018**, *37*, 2645–2659. [CrossRef] [PubMed]

99. Altomare, D.A.; Vaslet, C.A.; Skele, K.L.; De Rienzo, A.; Devarajan, K.; Jhanwar, S.C.; McClatchey, A.I.; Kane, A.B.; Testa, J.R. A Mouse Model Recapitulating Molecular Features of Human Mesothelioma. *Cancer Res.* **2005**, *65*, 8090–8095. [CrossRef] [PubMed]

100. Breschi, A.; Gingeras, T.R.; Guigó, A.B.R. Comparative transcriptomics in human and mouse. *Nat. Rev. Genet.* **2017**, *18*, 425–440. [CrossRef]

101. Thahir, M.; Sharma, T.; Ganapathiraju, M.K. An efficient heuristic method for active feature acquisition and its application to protein-protein interaction prediction. *BMC Proc.* **2012**, *6*, S2. [CrossRef]

102. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2013**, *13*, 2498–2504. [CrossRef] [PubMed]

103. Tagawa, M.; Tada, Y.; Shimada, H.; Hiroshima, K. Gene therapy for malignant mesothelioma: Current prospects and challenges. *Cancer Gene Ther.* **2013**, *20*, 150–156. [CrossRef] [PubMed]

104. Shimizu, T.; Nakanishi, Y.; Nakagawa, Y.; Tsujino, I.; Takahashi, N.; Nemoto, N.; Hashimoto, S. Association between expression of thymidylate synthase, dihydrofolate reductase, and glycinamide ribonucleotide formyltransferase and efficacy of pemetrexed in advanced non-small cell lung cancer. *Anticancer Res.* **2012**, *32*, 4589–4596. [PubMed]

105. Giammarioli, A.M.; Maselli, A.; Casagrande, A.; Gambardella, L.; Gallina, A.; Spada, M.; Giovannetti, A.; Proietti, E.; Malorni, W.; Pierdominici, M. Pyrimethamine Induces Apoptosis of Melanoma Cells via a Caspase and Cathepsin Double-Edged Mechanism. *Cancer Res.* **2008**, *68*, 5291–5300. [CrossRef] [PubMed]

106. Khan, M.W.; Saadalla, A.; Ewida, A.H.; Al-Katranji, K.; Al-Saoudi, G.; Giaccone, Z.T.; Gounari, F.; Zhang, M.; Frank, D.A.; Khazaie, K. The STAT3 inhibitor pyrimethamine displays anti-cancer and immune stimulatory effects in murine models of breast cancer. *Cancer Immunol. Immunother.* **2017**, *67*, 13–23. [CrossRef] [PubMed]

107. Grose, W.E.; Bodey, G.P.; Rodriguez, V. Sulfamethoxazole-Trimethoprim for Infections in Cancer Patients. *JAMA* **1977**, *237*, 352–354. [CrossRef]

108. Bodey, G.P.; Grose, W.E.; Keating, M.J. Use of Trimethoprim-Sulfamethoxazole for Treatment of Infections in Patients with Cancer. *Clin. Infect. Dis.* **1982**, *4*, 579–585. [CrossRef]

109. Lu, S.-H.; Tsai, W.-S.; Chang, Y.-C.; Chou, T.-Y.; Pang, S.-T.; Lin, P.-H.; Tsai, C.-M. Identifying cancer origin using circulating tumor cells. *Cancer Biol. Ther.* **2016**, *17*, 430–438. [CrossRef]

110. Nitadori, J.-i.; Ishii, G.; Tsuta, K.; Yokose, T.; Murata, Y.; Kodama, T.; Nagai, K.; Kato, H.; Ochiai, A. Immunohistochemical differential diagnosis between large cell neuroendocrine carcinoma and small cell carcinoma by tissue microarray analysis with a large antibody panel. *Am. J. Clin. Pathol.* **2006**, *125*, 682–692. [CrossRef]

111. Camilo, R.; Capelozzi, V.L.; Siqueira, S.A.C.; Bernardi, F.D.C. Expression of p63, keratin 5/6, keratin 7, and surfactant-A in non–small cell lung carcinomas. *Hum. Pathol.* **2006**, *37*, 542–546. [CrossRef]

112. Ordóñez, N.G. Value of cytokeratin 5/6 immunostaining in distinguishing epithelial mesothelioma of the pleura from lung adenocarcinoma. *Am. J. Surg. Pathol.* **1998**, *22*, 1215–1221. [CrossRef] [PubMed]

113. Heard, C.; Monk, B.; Modley, A. Binding of primaquine to epidermal membranes and keratin. *Int. J. Pharm.* **2003**, *257*, 237–244. [CrossRef]

114. Kimura, T.; Shirakawa, R.; Yaoita, N.; Hayashi, T.; Nagano, K.; Horiuchi, H. The antimalarial drugs chloroquine and primaquine inhibit pyridoxal kinase, an essential enzyme for vitamin B6 production. *FEBS Lett.* **2014**, *588*, 3673–3676. [CrossRef] [PubMed]

115. Basso, L.G.; Rodrigues, R.Z.; Naal, R.M.; Costa-Filho, A.J. Effects of the antimalarial drug primaquine on the dynamic structure of lipid model membranes. *Biochim. Biophys. Acta (BBA) Biomembr.* **2011**, *1808*, 55–64. [CrossRef] [PubMed]

116. Gakhar, G.; Ohira, T.; Battina, S.; Hua, D.H.; Nguyen, T.A. Anti-Tumor Effect of Primaquine Compounds in Human Breast Cancer Cells. In Proceedings of the AACR Annual Meeting, Las Angeles, CA, USA, 14–18 April 2007; American Association for Cancer Research: Philadelphia, PA, USA, 2007.

117. Ou, S.-H.I.; Moon, J.; Garland, L.L.; Mack, P.C.; Testa, J.R.; Tsao, A.S.; Wozniak, A.J.; Gandara, D.R. SWOG S0722: Phase II Study of mTOR Inhibitor Everolimus (RAD001) in Advanced Malignant Pleural Mesothelioma (MPM). *J. Thorac. Oncol.* **2015**, *10*, 387–391. [CrossRef] [PubMed]

118. Mamputu, J.-C.; Renier, G. Advanced glycation end products increase, through a protein kinase C-dependent pathway, vascular endothelial growth factor expression in retinal endothelial cells: Inhibitory effect of gliclazide. *J. Diabetes Complicat.* **2002**, *16*, 284–293. [CrossRef]

Interactome-based framework to translate disease genetic data into biological and clinical insights

5. Malignant peritoneal mesothelioma interactome with 417 novel protein-protein interactions

The experimental chapter is based on the following pre-print publication:
Karunakaran, Kalyani B., and Madhavi K. Ganapathiraju. Malignant peritoneal mesothelioma interactome with 417 novel protein-protein interactions. Research Square (2021).

Summary of this chapter

In this chapter, I demonstrate how the interactomic framework was used to gain biological and clinically translatable insights into malignant peritoneal mesothelioma (MPeM). MPeM is a mesothelioma subtype affecting the peritoneal lining of the abdominal cavity and intra-abdominal organs and having a median survival of ~2.5 years. To understand its underlying molecular mechanisms, I constructed the MPeM interactome using genes that showed genetic aberrations in MPeM or expression changes correlated with MPeM prognosis or drug response. I assembled their experimentally determined and computationally predicted protein-protein interactions (PPIs). I demonstrated the biological validity of the interactome by showing that MPeM-related transcriptomic evidence supports ~76% of the constituent genes. I identified functional gene modules and interactome genes highly expressed in extramedullary hematopoiesis sites and genes correlated with unfavourable prognosis in various cancers, all of which are predicted to have functional consequences for MPeM. I further showed the extensive overlap shared by the MPeM interactome with the malignant pleural mesothelioma (MPM) interactome (described in chapter 4) and MPM cell line expression profiles. Lastly, I performed a comparative analysis of peritoneal mesothelioma-associated and drug-induced gene expression profiles and identified 29 repurposable drugs for MPeM. In summary, this study provided valuable insights into MPeM biology and helped identify potentially repurposable drugs, such as irinotecan, paclitaxel and sirolimus.

Contribution to this chapter (90%)

- Developed the methodology of the project, which included MPeM interactome construction, validation, functional characterisation intersection analysis with MPM interactome and drug repurposing analysis
- Curated all the datasets, performed all the analyses and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

Malignant Peritoneal Mesothelioma Interactome with 417 Novel Protein-Protein Interactions

Kalyani B. Karunakaran

Indian Institute of Science   https://orcid.org/0000-0003-1279-7750

Madhavi K. Ganapathiraju (☐ madhavi@pitt.edu)

University of Pittsburgh   https://orcid.org/0000-0002-3825-0924

Malignant Peritoneal Mesothelioma Interactome
with 417 Novel Protein-Protein Interactions

Kalyani B. Karunakaran[1] and Madhavi K. Ganapathiraju[2]*

[1]Supercomputer Education and Research Centre, Indian Institute of Science, Bengaluru, 560012, India

[2]Department of Biomedical Informatics, School of Medicine,

Intelligent Systems Program, School of Computing and Information,

University of Pittsburgh, 5607 Baum Blvd, 5th Floor, Pittsburgh, PA, 15206, USA

Corresponding author: madhavi@pitt.edu

Abstract

Malignant peritoneal mesothelioma (MPeM) is an aggressive cancer affecting the abdominal peritoneal lining and intra-abdominal organs, with a median survival of ~2.5 years. We constructed an 'MPeM interactome' with over 400 computationally predicted protein-protein interactions (PPIs) and over 4,700 known PPIs of 59 literature-curated genes whose activity affects MPeM. The interactome integrated known PPIs of these genes from BioGRID and HPRD databases. Novel PPIs were predicted using the HiPPIP algorithm, which computes features of protein pairs such as cellular localization, molecular function, biological process, genomic location, and gene expression, and classifies the pairwise features as interacting or non-interacting based on a random forest model. Transcriptomic evidence from rodent and human peritoneal mesothelioma samples validated 75.6% of the interactome and 65% of the novel interactors. 152 drugs targeting 427 proteins within the MPeM interactome were identified. Comparative transcriptome analysis of peritoneal mesothelioma-associated vs. drug-induced gene expression profiles, revealed 39 repurposable drugs, with 29 effective against peritoneal and pleural mesothelioma and primary peritoneal cancer in clinical trials, animal models, or cell lines. Functional modules related to chromosomal segregation, transcriptional deregulation, IL-6 production and hematopoiesis were identified from the interactome. Furthermore, genes with tissue-specific expression in extramedullary hematopoietic sites and those correlated with unfavourable prognoses in various cancers were detected. Lastly, the MPeM interactome significantly overlapped with the malignant pleural mesothelioma (MPM) interactome, revealing shared molecular pathways. Our findings demonstrate the utility of the MPeM interactome in uncovering functional links among MPeM genes and generating clinically translatable results such as repurposed drugs.

Keywords: malignant mesothelioma, protein-protein interactions, systems biology, network analysis, drug repurposing

Introduction

Internal organs, such as the heart and lungs, and body cavities, such as the thoracic and abdominal cavities, are covered by a thin, slippery layer of cells called the "mesothelium". Mesothelioma is a rare but highly aggressive cancer that originates from this lining, constituting the subtypes pericardial (heart), pleural (lung) and peritoneal (abdomen) mesothelioma; it is usually found in advanced stages and has a median survival of one year.[1] Mesothelioma is intricately linked with exposure to asbestos but with a long latency period of a few decades between exposure and the occurrence of the disease and does not have a non-invasive pre-malignant phase, unlike other cancers. The focus of this work is on the genetics and biological mechanisms of malignant peritoneal mesothelioma (MPeM).

The peritoneum is a serosal membrane made up of two layers of mesothelial cells lining the abdominal cavity and intra-abdominal organs. MPeM affects this peritoneal lining and grows within the peritoneal space.[1] Patients may exhibit symptoms such as weight loss, shortness of breath, chest and abdominal pain, increased abdominal girth and peritoneal effusion between the ages of 40-65 years.[1] MPeM constitutes a substantial percentage (15%-20%) of all mesothelioma diagnoses and is distinct from malignant pleural mesothelioma (MPM) due to its limited association with asbestos exposure (8% compared to 80% for MPM).[2] MPeM was more apparent among patients with a history of abdominal surgeries rather than asbestos exposure.[1,2] Peritoneal cases are also becoming increasingly prevalent among mesothelioma patients without occupational exposure, given the current scenario in which the population of asbestos-exposed individuals is diminishing.[1] MPeM exhibits a highly variable pattern of disease progression and patients often develop risk of postoperative morbidity and mortality.[1] MPeM has a higher median survival rate than pleural mesothelioma (31 months versus 14 months),[3] and is currently treated with a combination of pemetrexed and cisplatin.[1] Given the unique features of MPeM and its fatal nature, it is imperative that the molecular mechanisms underlying this disease are expeditiously discovered.

Factors predisposing patients to MPeM remain unclear.[2] However, MPeM is known to be proportionally more prevalent than MPM among patients with germline mutations and without a history of asbestos exposure (25% versus 7%[4]).[1] Multiple studies have examined the genetic underpinnings of MPeM (see **Table 1**). Altogether, these studies reported 59 MPeM-associated genes that harboured mutations, copy number aberrations, and rearrangements, or whose expression correlated with poor prognosis in MPeM patients, and reduced cell survival or unfavourable drug responses in MPeM surgical specimens.[5-13] The Cancer Genome Atlas (TCGA) also provided comprehensive genomic datasets of mesothelioma, describing mutations in BAP1, CDKN2A, LATS1,

LATS2, MSH2, NF2, PBRM1, PTCH1, RBFOX1, SETD2, SETDB1 and TP53 (included in **Table 1**).[14] The next step to discovering biological mechanisms is to understand how these genes play a role in the disease. To address this, we constructed the protein-protein interaction network (or the 'interactome') of these MPeM-associated genes. Using various bioinformatics methods, we gained insights into the biological processes underlying MPeM and identified repurposable drugs.

Table 1: Publications examining the genetic basis of MPeM, related details, and the gene lists from each used in our study

Study	Study methods	Patient samples	MPeM-associated genes
Hung et al.[13]	Immunohistochemistry, fluorescence in situ hybridization (FISH), targeted next-generation sequencing of tumour DNA and RNA	88 consecutive patients with peritoneal mesotheliomas diagnosed at a single institution between 2005 and 2015	ALK
Joseph et al.[5]	Next-generation sequencing 510 cancer-related genes, immunohistochemistry	13 patients with malignant mesothelioma arising in the peritoneal cavity	ARID1A, BAP1, DDX3X, NF2, SETD2, TERT, WT1
Ugurluer et al.[6]	Next-generation sequencing testing, descriptive and Kaplan-Meier statistics	11 patients with somatic cancer-related mutations	AR, ASXL1, BAP1, BRIP1, CDK12, DAXX, EPHB1, ESR1, FGF6, IRS2, JAK1, KDM6A, KDR, KEAP1, KMT2A, MET, MRE11, MTOR, NSD1, SETD2, TSC1
Chirac et al.[7]	Comparative genomic hybridization using the Agilent Human Genome CGH 180 K array	MPeM samples from 33 patients	ADAM3A, ARHGAP22, BAP1, CDH5, CDKN2A, CHEK2, CTNNB1, DPYD, EGFR, HRAS, IGKC, JUN, MAPK8, NF2, NR2F2, PTEN, RASSF1, RB1, RHEB, RICTOR, SDHB, SMARCB1, STK11, TRIO, VEGFB
Foster et al.[8]	Evaluation of patient tumours for mutations in the catalytic TK-domain, treatment of patients with cytoreductive surgery, COS-7 cell expression model to determine mutation activating profiles and response to erlotinib	MPeM tumours from 29 patients, 25 of whom were treated with cytoreductive surgery	EGFR

Hung et al.[9]	Targeted next-generation sequencing, immunohistochemistry	Diffuse peritoneal mesotheliomas from 26 patients	ARID1B, BAP1, CDKN2A, CHEK2, NF2, PBRM1, PRDM1, SETD2, SUZ12, TP53, TRAF7
Pillai et al.[10]	Immunohistochemistry, prognostic significance using the Kaplan-Meier method	MPeM samples from 42 patients	MUC1
Varghese et al.[11]	Gene expression analyses, pathway-specific inhibition	Fresh pre-treatment MPeM tumour samples collected from 41 patients who underwent surgical cytoreduction and received regional intraoperative chemotherapy perfusion	PIK3CA, RICTOR
Zaffaroni et al.[12]	Immunohistochemistry	32 MPeM surgical specimens	BIRC5

Protein-protein interactions (PPIs) drive the biological processes in cells including signal transduction, formation of cellular structures and enzymatic complexes. The molecular mechanisms of disease are often revealed by the PPIs of disease-associated genes. For example, the involvement of transcriptional deregulation in pleural mesothelioma pathogenesis was identified through mutations detected in BAP1 and its interactions with proteins such as HCF1, ASXL1, ASXL2, ANKRD1, FOXK1 and FOXK2.[15] PPI of BAP1 with BRCA1 was central to understanding the role of BAP1 in growth-control pathways and cancer; BAP1 was suggested to play a role in BRCA1 stabilization.[16,17] Studies on BAP1 and BRCA1 later led to clinical trials of the drug vinorelbine as a second line therapy for MPM patients, and the drug was shown to have rare or moderate effects in MPM patients.[18,19]

Despite their crucial role in understanding disease mechanisms and discovering drugs, ~75% of estimated PPIs are unknown, and several disease-associated genes lack known PPIs. The human interactome may contain more than 600,000 PPIs,[20] but only ~150,000 PPIs are known from PPI repositories such as HPRD[21] and BioGRID.[22] Experimental detection of PPIs using techniques such as co-immunoprecipitation (Co-IP)[23,24] is time-consuming at large scale. Although systematic high throughput studies with yeast two-hybrid (Y2H) system[25] and affinity purification–mass spectrometry (AP–MS)[26] have helped discover tens of thousands of PPIs, a large part of the interactome remains unknown. We developed **HiPPIP** (high-precision protein-protein interaction prediction), a computational model deemed highly accurate by computational evaluations, and experimental validations of 18 predicted PPIs, where all the tested pairs were shown to be true PPIs.[27,28]

We derived valuable insights from the analysis of disease-specific protein interactomes that included PPIs predicted by HiPPIP. Notably, we identified 2,156 novel PPIs for diseases such as MPM,[29] schizophrenia,[27] rheumatoid arthritis,[30] and congenital heart disease.[31,32] Our previous study that demonstrated the functional links of MPM-associated genes from various high throughput investigations within the MPM interactome, underscored the importance of interactome analysis in understanding the molecular basis of mesothelioma.[29] More than 85% of the genes in the interactome were supported by MPM-related genetic variant, transcriptomic and proteomic evidence. Furthermore, we experimentally validated 5 novel PPIs of MPM-associated genes and identified 5 repurposable drugs targeting the interactome proteins. This collective evidence motivated us to extend our interactome-based approach to the exploration of the genetic basis of MPeM.

In this work, we constructed the 'MPeM interactome' by assembling the known and computationally predicted PPIs of the genes associated with MPeM. Analysing this interactome within the context of peritoneal mesothelioma transcriptomic data, gene tissue specificity, prognostic relevance of genes in other cancers, and interconnections to MPM, we expanded our understanding of MPeM. We then investigated the pathways and functional modules associated with the interactome. Finally, we integrated drugs sourced from the Drug Bank repository[33] targeting at least one of the interactome proteins, and performed comparative transcriptome analysis of drug-induced and MPeM-associated profiles to identify 29 repurposable drugs for MPeM.

Results

PPIs of the MPeM-associated genes (or 'core' genes) shown in **Table 1** were collected from HPRD[21] (Human Protein Reference Database) and BioGRID[22] (Biological General Repository for Interaction Datasets); see **Supplementary Data File 1** for the reported gene alterations. The HiPPIP algorithm described in our earlier work was applied to MPeM genes to discover hitherto unkown PPIs.[34] HiPPIP computes features of protein pairs such as cellular localization, molecular function, biological process membership, genomic location of the gene, and gene expression in microarray experiments, and classifies the pairwise features as *interacting* or *non-interacting* based on a random forest model.[27] The 'MPeM interactome' assembled in this manner contained 4,747 known PPIs and 417 novel PPIs connecting 58 MPeM-associated genes to 2,747 known interactors and 306 novel interactors (**Fig. 1** and **Supplementary Data File 2**). The 59th MPeM-associated gene ADAM3A had neither known nor novel PPIs.

The number of known and computationally predicted novel PPIs for each of the MPeM genes are shown in **Fig. 2** and **Supplementary Data File 3**; the novel interactors are listed in **Supplementary Data File 3**. Thirteen genes had 10 or less interactions each and 73 novel PPIs were predicted for all of the genes combined. There are 21 hub-genes that had more than 75 known PPIs each and 160 novel PPIs were predicted for all of the genes combined.
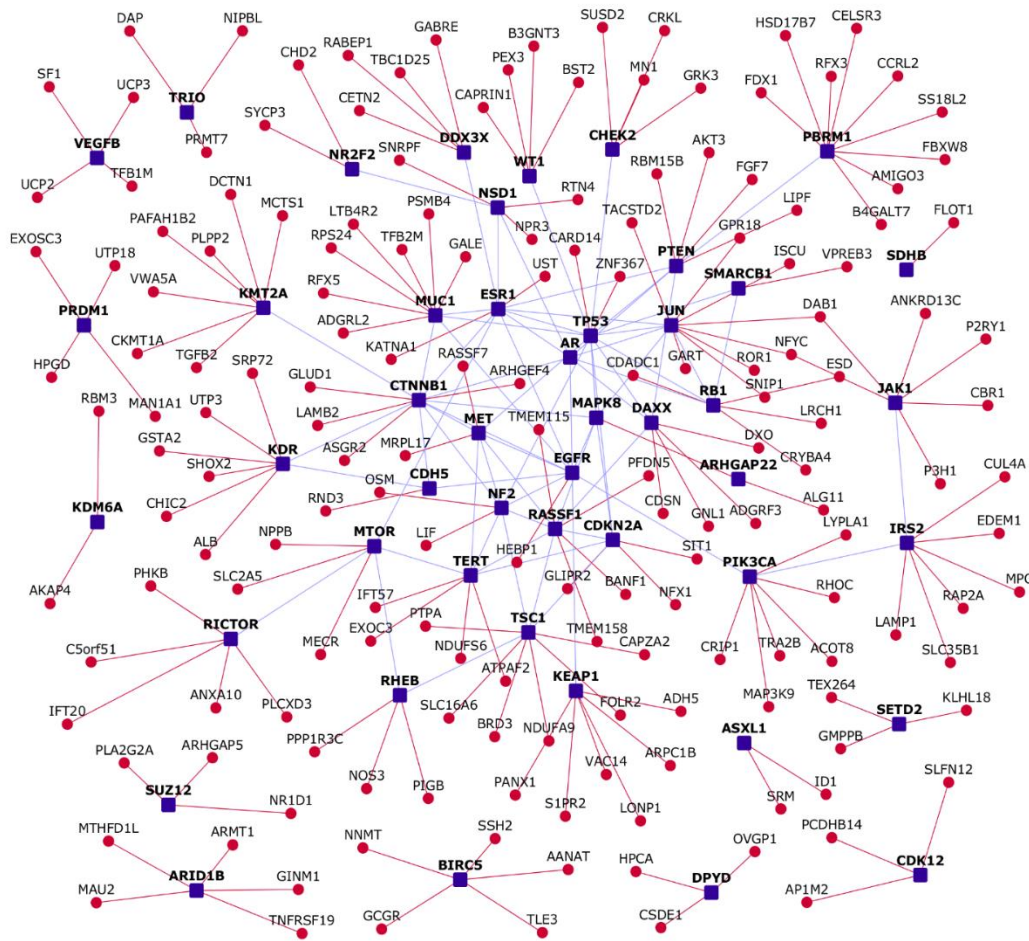
Figure 1. A partial network view of protein-protein interactions (PPIs) in the malignant peritoneal mesothelioma (MPeM) interactome: Genes are shown as nodes and PPIs as edges. As the complete MPeM network is very large, only a partial view showing a large connected component of MPeM candidate genes and their novel interactors, all of which have MPeM-related transcriptomic evidence (Supplementary Data File 4), is shown. Legend: dark blue square-shaped nodes: MPeM candidate genes; red nodes/edges: novel interactors/interactions; light blue nodes and
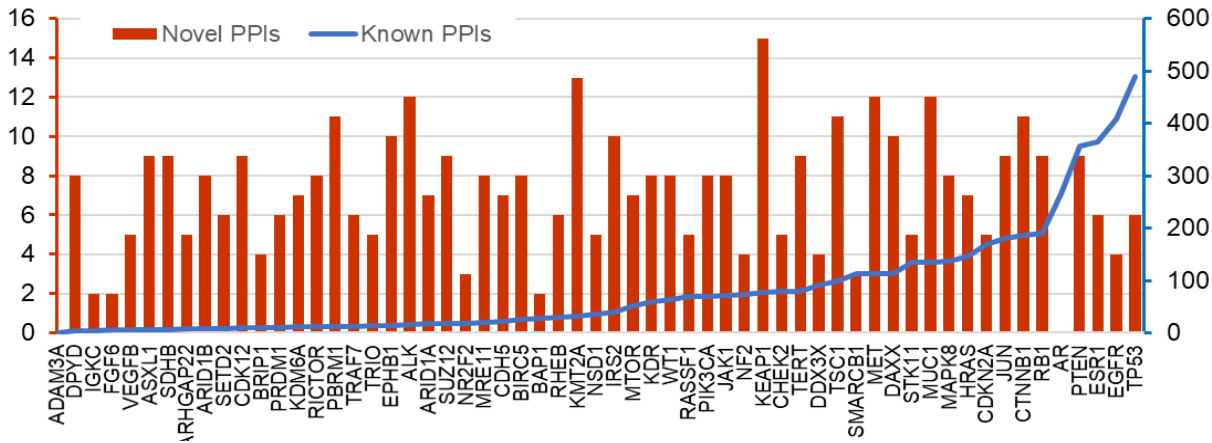
Figure 2. Number of protein-protein interactions: The MPeM associated genes are listed along the x-axis, arranged in the ascending order of their number of known protein-protien interactions. The number of novel predicted PPIs and previously known PPIs are shown as red bars on primary axis (left) and blue line on secondary axis (right). For example, DPYD has three known PPIs and 8 novel PPIs, and AR has 265 known and zero novel PPIs.

Overlap of the MPeM interactome with transcriptomic data

198 out of 306 (65%) novel interactors, and 2,353 (75.6%) proteins overall of the MPeM interactome, showed differential gene expression in pre-clinical models and human tumour specimens of peritoneal mesothelioma (see **Table 2** and **Supplementary Data File 4**). These included human orthologues of genes differentially expressed in MPeM tumour specimens from patients, peritoneal mesotheliomas of rats, either spontaneously or chemically induced, mouse peritoneal mesothelioma cell lines resulting from crocidolite asbestos, and human peritoneal mesothelial lines exposed to crocidolite. These overlap studies confirmed the biological validity of the MPeM interactome by demonstrating its relevance in the context of rodent xenograft/cell line mesothelioma models and human mesothelial cell lines.

Table 2: Transcriptomic datasets enriched in the MPeM interactome, with the number of differentially expressed genes (DEGs) and *p*-value and odds ratio of enrichment.

Transcriptomic dataset (with reference to source publication)	Number of DEGs in the interactome	P-value of enrichment	Odds ratio of enrichment
Granulocytic myeloid-derived suppressor cells (G-MDSCs) from spleens of mice bearing AB12 mesothelioma grafts versus naive neutrophils (GSE43254[35])	975	2.02E-14	1.21
Neutrophils infiltrating AB12 mesothelioma tumour grafts versus naive bone marrow derived neutrophils (GSE43254[35])	1006	1.97E-17	1.24
BCA induced peritoneal mesothelioma versus non-transformed mesothelial cell line	533	1.08E-04	1.15

O-Nitrotoluene (O-NT) induced peritoneal mesothelioma versus non-transformed mesothelial cell line (GSE4682[36])*	332	-	-
Spontaneous malignant mesotheliomas from 2-year-old rats versus normal mesothelial Fred-PE cells (GSE47581[37])*	794	-	-
LP9 mesothelial cells exposed for 8 hours to *5ug/cm2 crocidolite asbestos* versus untreated mesothelial cells (GSE14034[38])	303	6.39E-08	1.32
LP9 mesothelial cells exposed for 8 hours to *5ug/cm2 crocidolite asbestos* versus untreated mesothelial cells (GSE63966[39])	560	1.24E-05	1.16
LP9 mesothelial cells exposed for 8 hours to *1ug/cm2 crocidolite asbestos* (GSE14034[38])	85	8.76E-04	1.38
Primary peritoneal mesothelial HM3 cells exposed to *5ug/cm2 crocidolite asbestos* for 8 hours (GSE63966[39])	797	3.99E-12	1.22
Biphasic versus epithelial peritoneal mesothelioma tumour specimens[40]	118	2.17E-19	2.25
Lungs of mice exposed to crocidolite fibers[41]	322	3.5E-13	1.44
Lungs of mice exposed to wollastonite fibers[41]	23	0.044	1.43

*A considerable number of genes in the interactome were differentially expressed in rat models of spontaneous and induced peritoneal mesothelioma, but their overlaps were not statistically significant.

In order to examine whether the interactome showed preferential enrichment for any specific subtype of peritoneal mesothelioma, we computed its overlap with genes found to be differentially expressed in biphasic versus epithelial peritoneal mesothelioma tumour specimens and vice versa.[40] Significant enrichment was found with biphasic mesothelioma, but not with epithelial mesothelioma (**Table 2** and **Supplementary Data File 4**). This overlap included 4 genes predicted as novel interactors of 4 MPeM-associated genes (MPeM genes are shown in bold): **ARID1A**-TAF12, **PIK3CA**-LYPLA1, **EPHB1**-MRPL3 and **KEAP1**-LONP1. Hence, the interactome with over 100 genes specific to the biphasic subtype of MPeM will prove valuable for investigating this relatively rarer subtype compared to the epithelioid and sarcomatoid subtypes.[42]

Diffuse MPeM is known to share similar clinical presentation, morphology and immunostaining profiles with ovarian/primary peritoneal serous carcinoma (OC/PPC), and may hence be indistinguishable from the latter.[43] Gene expression signatures characterizing these two tumours have been identified in an attempt to elucidate the molecular differences distinguishing them from one another.[43] We computed the overlap of the MPeM interactome with these expression profiles (see **Supplementary Data File 4**). Out of the 12 genes in the interactome found to be differentially expressed in OC/PPC versus diffuse MPeM (including the MPeM-associated gene ESR1), 3 were predicted as novel interactors of MPeM-associated genes: **HRAS**-IGF2, **JUN**-TACSTD2 and **CHEK2**-SUSD2. Eight genes including the MPeM-associated gene KDR were found to be

differentially expressed in diffuse MPeM versus OC/PPC. This analysis helped pinpoint the genes that distinguish MPeM from other morphologically and histogenetically similar tumours.

In summary, these overlap studies validated the relevance of MPeM interactome to MPeM tumours in rodent models and human patients, identified genes specific to MPeM subtypes and those aiding in differential diagnosis from other cancers. The interactome can be used as a mechanistic framework for investigating MPeM-related genes.

Tissue-specificity of the genes in the MPeM interactome

We studied tissue specific expression of the interactome genes using mouse ENCODE and GTEx data.[44,45] Genes with an expression > 1 TPM (transcripts per million) and 5-folds higher in a single tissue (tissue-enriched) or 2-7 tissues (group-enriched) were included.[46] Unexpectedly, the top enriched organs were spleen and small intestine, and not abdominal organs lined by the peritoneum (**Fig. 3A**). Instead, the human organs that shared many genes with the interactome were brain, testis, skin, lung, heart, oesophagus, artery and muscle (**Fig. 3A**). Similar trends were observed with mouse expression data, with the intestine, cortex, cerebellum, olfactory bulb, testis and bone marrow, and embryonic tissues such as E14.5 brain, E14.5 placenta and E14.5 heart, showing enrichment in the interactome (**Fig. 3B**). The interactome exhibited notable enrichment in human orthologues of mouse genes specific to spleen (81 genes, *P*-value = 0.019, odds ratio = 1.39) and thymus (57 genes, *P*-value = 0.028, odds ratio = 1.42) (**Fig. 3B**). Ten MPeM-associated genes had novel PPIs with the orthologues of 10 spleen-specific mouse genes, namely, **SMARCB1**-VPREB3, **JAK1**-VNN2, **RHEB**-NOS3, **ALK**-NLRC4, **IRS2**-MPO, **TSC1**-FCN1, **RICTOR**-CTSW, **HRAS**-CCL4 and **BIRC5**-AANAT (i.e. 10 novel interactors had spleen-specificity; MPeM genes are shown in bold).

We used BaseSpace Correlation Engine[47,48] to identify human peritoneum-specific genes from the interactome. A gene was deemed specific to peritoneum if its expression decrease in other tissues compared to the tissue of interest was > 0.8 (i.e. specificity index > 0.8). Based on this, only 6 genes showed peritoneum-specific expression. OVGP1, a predicted interactor of the MPeM-associated gene DPYD, had moderate peritoneal specificity (specificity index = 0.57).

Altogether, the analysis of tissue-specific expression in the interactome revealed scarce peritoneum-specific expression and unexpected enrichment in lymphatic organs such as spleen and thymus.

Functional modules and pathways enriched in the MPeM interactome
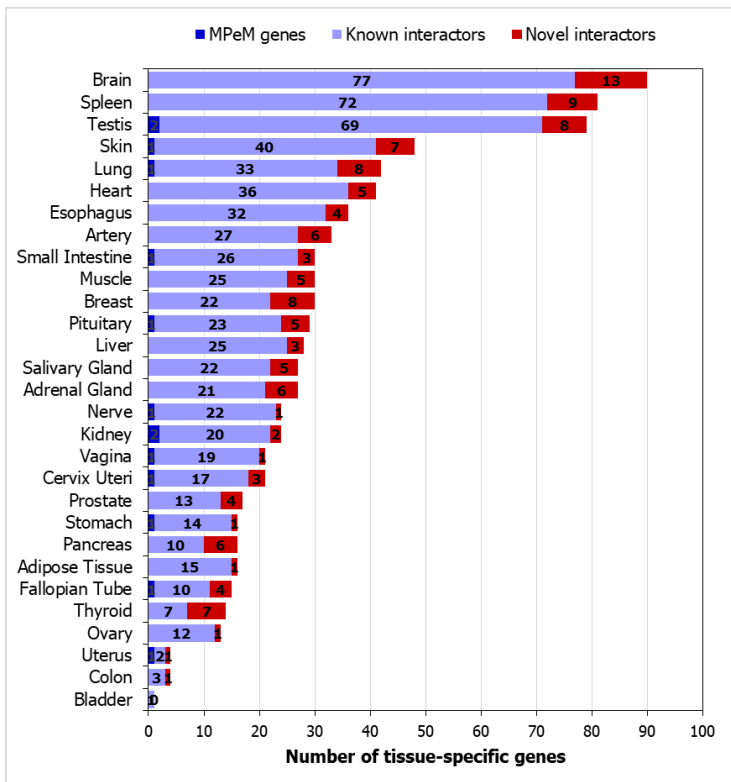
Figure 3. (A) Tissue-specificity of malignant peritoneal mesothelioma (MPeM) interactome genes in human organs: Tissue-specific expression of the genes in the interactome was examined using GTEx data. The graph shows the number of genes from the interactome that exhibit tissue specificity. The genes show at least 5-fold higher expression in a tissue ('tissue-enriched') or a group of 2-7 tissues compared to all the other tissues ('group-enriched').
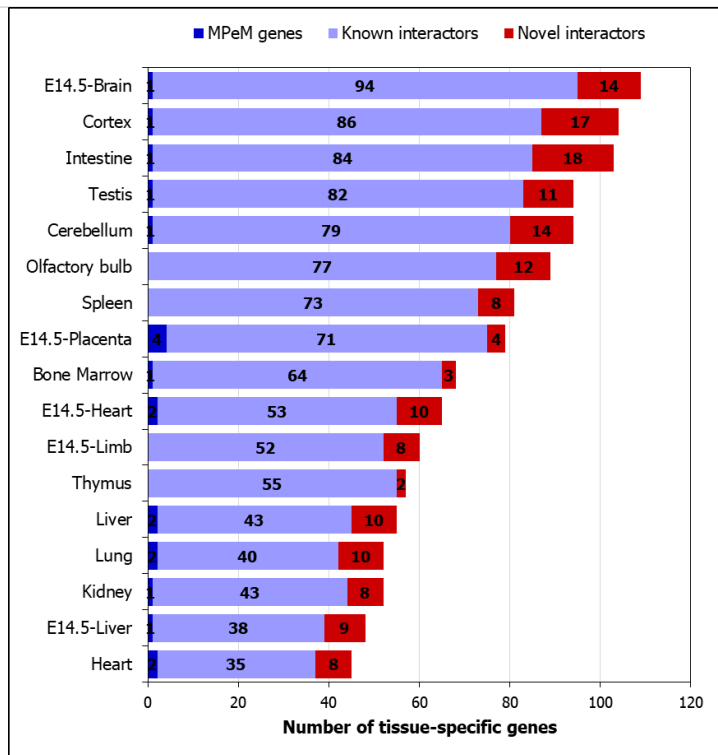
Figure 3. (B) Tissue-specificity of malignant peritoneal mesothelioma (MPeM) interactome genes in mouse organs: Tissue-specific expression of the genes in the interactome was examined using mouse ENCODE data. The graph shows the number of genes from the interactome that exhibit tissue specificity. The genes show at least 5-fold higher expression in a tissue ('tissue-enriched') or a group of 2-7 tissues compared to all the other tissues ('group-enriched').

We used the HumanBase toolkit[49] (https://hb.flatironinstitute.org/) to identify functional modules in the MPeM interactome. HumanBase employs shared k-nearest-neighbours and the Louvain community-finding algorithm to cluster the genes sharing the same network neighbourhoods and similar Gene Ontology (GO) biological processes, into functional modules. Fourteen modules were detected of which 11 had more than 4 proteins each (**Table 3**).

133

Table 3: Functional modules in the MPeM Interactome (with FDR-corrected *p*-value)

Module	Enriched GO Biological Process	FDR-corrected *p*-value
M1	Chromosome Segregation	<1E-08
M2	Translation	<1E-08
M3	Hematopoiesis	<1E-08
M4	Covalent Chromatin Modification	<1E-08
M5	Transmembrane Receptor Protein Tyrosine Kinase Pathway	<1E-08
M6	Histone Modification	<1E-08
M7	mRNA Metabolic Process	<1E-08
M8	Cell-Cell Adhesion	2.86E-05
M9	Transmembrane Receptor Protein Tyrosine Kinase Pathway	2.82E-04
M10	Transmembrane Receptor Protein Tyrosine Kinase Pathway	2.64E-03
M11	Negative Regulation of Intracellular Signal Transduction	3.98E-03
M12	Negative Regulation of Hydrolase Activity	5.84E-03
M13	Cell-Cell Junction Assembly	5.97E-03
M14	Positive Regulation of Interleukin-6 Production	0.023

Next, we identified the Reactome pathways enriched in the MPeM interactome using the gene set analysis toolkit called WebGestalt (**Supplementary Data File 5**).[50] WebGestalt computes the statistical significance of the association of the genes with a specific functional group (e.g. a Reactome Pathway) using Fischer's exact test and Benjamini-Hochberg method for multiple test adjustment. The top-30 pathways associated with the MPeM are shown in **Table 4**.[51]

The identified modules and pathways could contribute to peritoneal mesothelioma development and progression (see **Discussion**), including dysregulated chromosome segregation, covalent chromatin modification, altered mRNA metabolic processes, disrupted translation, post-translational events, activation of transmembrane receptor protein tyrosine kinase pathways, disrupted cell-cell junction assembly, and cytokine signalling, particularly interleukin-6 production. The identification of hematopoiesis as an enriched module aligns with the enrichment of genes specific to the extramedullary hematopoietic sites, spleen and thymus, in the interactome (**Fig. 3**).

Table 4: Selected pathways associated with MPeM interactome (FDR-corrected *p*-value < 1E-15)

Pathway	Number of genes	MPeM genes	Novel interactors

Immune System	578	BIRC5, CTNNB1, DDX3X, HRAS, IRS2, JAK1, JUN, KEAP1, MAPK8, MRE11, MTOR, MUC1, NF2, PIK3CA, PTEN, RICTOR, TP53, TRAF7	AKT3, AP1M2, ARPC1B, BST2, BTBD1, CALML5, CAPZA2, CCL4, CENPE, CFHR3, CPB2, CRKL, DCTN1, DEFB115, DEFB116, FBXW8, FCN1, GM2A, GSTA2, KLRC1, KLRC2, LAMP1, LIF, MADCAM1, MPO, NCSTN, NLRC4, NOS3, OSM, PAFAH1B2, PANX1, PLA2G2A, PSMB4, PTPRN2, PYGB, RAP1GAP, REG3A, SIAH1, SLC2A5
Gene expression (Transcription)	517	AR, ARID1A, ARID1B, BIRC5, BRIP1, CDK12, CDKN2A, CHEK2, CTNNB1, DAXX, EGFR, ESR1, JUN, KMT2A, MET, MRE11, MTOR, PBRM1, PRDM1, PTEN, RB1, RHEB, RICTOR, SMARCB1, STK11, SUZ12, TP53, TSC1	AKT3, DNMT3A, GTF2E1, ICE1, KIT, NLRC4, NR1D1, PCBP4, PSMB4, SNRPF, TAF12, TFB2M, ZNF157, ZNF195, ZNF266, ZNF79
Developmental Biology	361	CTNNB1, EGFR, EPHB1, HRAS, IRS2, JUN, KDM6A, MAPK8, MET, NR2F2, PIK3CA, SUZ12, TRIO	AKT3, ARPC1B, CACNA1S, CDSN, DAB1, FOXA3, NCSTN, PKLR, PSMB4, RAP1GAP, RHOC, RPS24, SIAH1, SLC2A2
Cytokine Signaling in Immune system	269	BIRC5, IRS2, JAK1, JUN, MAPK8, MUC1, PIK3CA, TP53	BST2, CCL4, CRKL, GSTA2, LIF, OSM, PSMB4
Cell Cycle	247	BIRC5, BRIP1, CDKN2A, CHEK2, MRE11, RB1, TERT, TP53	AKT3, BANF1, CENPE, CETN2, DCTN1, MAU2, NIPBL, PCBP4, PCNT, POLD1, PSMB4, SYCP3
Cellular responses to stress	170	AR, CDKN2A, JUN, MAPK8, MRE11, MTOR, RB1, SUZ12, TP53	CAPZA2, DCTN1, ID1, PSMB4
DNA Repair	133	BAP1, BRIP1, CHEK2, MAPK8, MRE11, TP53	CETN2, CUL4A, MBD4, POLD1
Deubiquitination	125	AR, ASXL1, BAP1, ESR1, KEAP1, PTEN, TP53	PSMB4
MAPK family signaling cascades	153	EGFR, FGF6, HRAS, IRS2, JAK1, JUN, MET	DLG2, FGF7, KIT, NRG2, PSMB4
SUMOylation	104	AR, BIRC5, CDKN2A, DAXX, ESR1, SUZ12, TP53	CETN2, DNMT3A, SENP1

Association with other cancers

The prolonged survival of carriers of MPeM-associated mutations (e.g. in BAP1 and TP53) has been linked to the occurrence of other cancers.[1] This connection between cancer prognosis and comorbidities in MPeM patients prompted us to explore the presence of prognostic genes from various cancers within the MPeM interactome. We systematically examined the overlap between the MPeM interactome and prognostic genes from 20 cancer types, using data from Pathology Atlas for gene expression and patient survival correlation.[52] Genes with log-rank P-value < 0.001 were deemed prognostic, where high expression correlated with low patient survival was unfavourable prognosis, and increased survival was favourable prognosis. In the MPeM interactome, we identified a significant enrichment of genes that exhibited elevated expression, positively correlating with (i) unfavourable prognosis in liver, renal, pancreatic and lung cancers and (ii) favourable prognosis in testis, breast, thyroid and skin cancers (**Supplementary Data File 6**).

Next, we explored the relationship between interactome genes and other diseases using the DisGeNET database.[53] The top-5 diseases associated with MPeM were prostatic, mammary, stomach, liver and lung neoplasms, all at *P*-value < 1E-15 (**Supplementary Data File 7**). Notably, numerous novel interactors were linked to these diseases. For example, 13 novel interactors of MPeM-associated genes were associated with prostatic neoplasms (MPeM

genes are shown in bold): **MET**-SLC26A4, **DPYD**-SULT2A1, **CTNNB1**-LAMB2, **IRS2**-MPO, **HRAS**-ZFP36L2, **VEGFB**-UCP3, **PRDM1**-HPGD, **NSD1**-NPR3, **KEAP1**-SLC5A5, **MET**-FOXA3, **RHEB**-NOS3, **HRAS**-HBG1 and **JAK1**-CBR1.

We then utilized Phenogrid from the MONARCH toolkit[54] to identify diseases most phenotypically akin to MPeM. Phenogrid, an algorithm in the toolkit, determines shared phenotypes between two diseases. It gauges the information content of each phenotype (gene and disease associations) to quantify the observed similarity observed between the diseases. Ovarian fibroma (OF), desmoplastic small round cell tumour (DSRCT), Budd-Chiari syndrome (BCS) and primary peritoneal carcinoma (PPC) exhibited high phenotypic similarity to MPeM (similarity score > 80). We compiled 6, 43, 24 and 49 genes associated with OF, DSRCT, BCS and PPC, respectively, and examined their enrichment in the MPeM interactome. Notably, significant enrichment was found for genes associated with DSRCT (P-value = 4.16E-04, odds ratio = 2.31) and PPC (P-value = 5.72E-08, odds ratio = 2.98).

Altogether, the significant gene enrichment patterns for diverse cancers uncovered from the interactome offers the potential to improve prognosis predictions and customize treatment strategies. Moreover, our identification of diseases closely resembling MPeM can be leveraged to advance the diagnosis of MPeM.

Interconnections to pleural mesothelioma interactome

We sought to uncover the shared biological aspects between MPeM and malignant pleural mesothelioma (MPM). We compared the overlap of MPeM interactome with the MPM interactome,[29] revealing 989 shared genes, a highly significant overlap (P-value = 3.18E-289, odds ratio = 2.92). This overlap included 4 core genes linked to both MPM and MPeM (BAP1, CDKN2A, KDR and WT1), 29 PPIs between MPM and MPeM core genes (one novel), and 21 novel interactors of MPM and MPeM core genes, alongside known interactors. Thirty-eight MPM-associated genes, 41 MPeM genes and the 4 genes common between them formed an intricately interconnected network of PPIs (**Fig. 4**). Six of these were novel PPIs (FLT1-FLT3, TUBA1A-TUBA1C, RHGAP22-MAPK8, DPYD-SRC, JUN-GART and TSC1-TUBB4B).

Of the genes shared between the MPM and MPeM interactomes, 62.5% displayed differential expression in both MPeM and MPM differential gene expression datasets, i.e. in at least one of seven MPeM (**Supplementary Data File 4**) and one of six MPM datasets.[29] Notably, 33% of these differentially expressed genes were involved in the immune system (P-value < 1E-16). We identified a compact network (**Supplementary Fig. 1**) interconnecting 5 MPeM-associated and 5 MPM-associated genes via 19 known PPIs and 5 novel PPIs. This network highlighted the potential shared immune pathways that could contribute to tumour invasion and metastasis in both subtypes,[55] notably IL-17 signalling via its modulator IKBKE[56] and T_H17 cell differentiation via four genes, namely, MPeM-
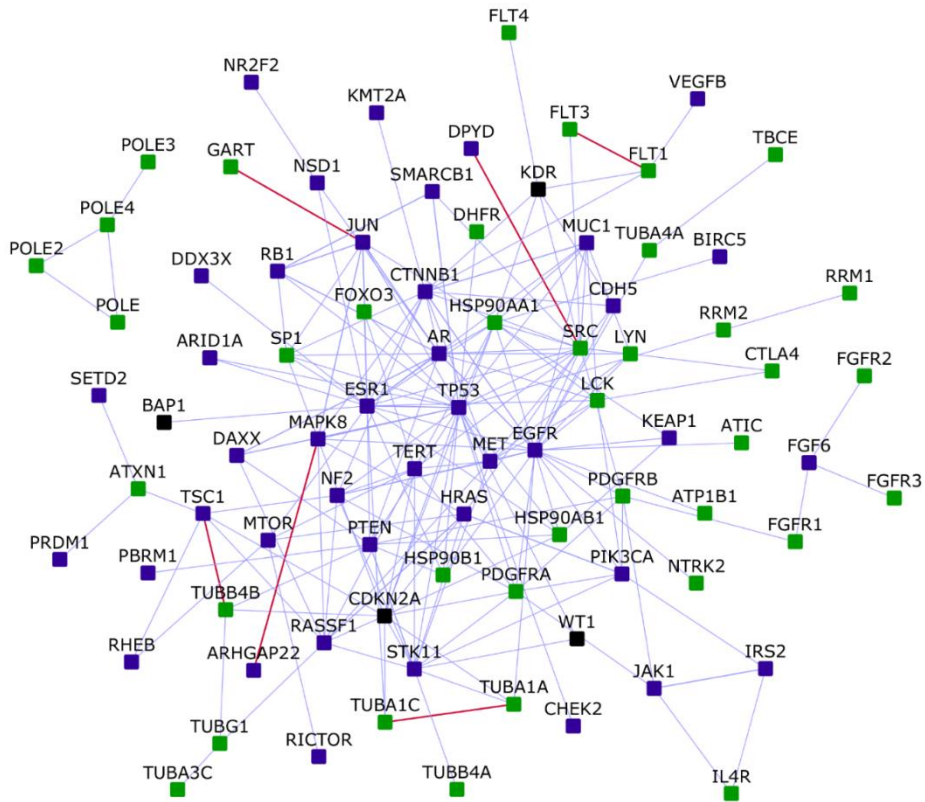
Figure 4. Interconnections of malignant peritoneal and pleural mesothelioma candidate genes: Square-shaped blue nodes: malignant peritoneal mesothelioma (MPeM) candidates, square-shaped green nodes: malignant pleural mesothelioma (MPM) candidates, square-shaped black nodes: genes that are MPeM as well as MPM candidates. Light blue and red colored edges indicate known and novel interactions respectively.

associated HSP90AA1 and HSP90AB1, MPM-associated JUN and MAPK8, and MPM-associated membrane protein MUC1 widely implicated in mesothelioma malignancy.[57]

In summary, we uncovered a substantial shared gene pool between MPeM and MPM upon exploring their interactome overlaps. The majority of these genes showed altered expression in both MPeM and MPM transcriptomic datasets, predominantly in immune-related pathways. This underscores the significant role played by immune pathways in the progression of both mesothelioma subtypes, holding crucial implications for future research and therapeutic approaches.

Potentially repurposable drugs for MPeM

We followed the established approach of comparing drug-induced versus disease-associated differential expression[58] to identify potential drugs for MPeM treatment. Utilizing the BaseSpace Correlation software suite (https://www.nextbio.com),[47,48] which previously helped identify repurposable drug candidates for schizophrenia[59] (currently undergoing clinical trials[60,61]) and mesothelioma,[29] we analysed pre-processed gene expression datasets. We constructed the MPeM drug-protein interactome that shows the drugs that target any protein in the MPeM interactome. In total, 152 drugs (collected from Drug Bank[33]) were found to target 427

proteins, encompassing 16 MPeM-associated genes, 361 known interactors and 50 novel interactors. Our focus then turned to selecting 5 gene expression datasets pertinent to peritoneal mesothelioma. These included granulocytic myeloid-derived suppressor cells (G-MDSCs) sourced from spleens of mice with AB12 mesothelioma grafts versus naive neutrophils, as well as neutrophils infiltrating AB12 mesothelioma tumour grafts versus naive bone marrow derived neutrophils (GSE43254[35]). Additionally, datasets covering BCA induced peritoneal mesothelioma versus non-transformed mesothelial cell line, O-NT induced peritoneal mesothelioma versus non-transformed mesothelial cell line (GSE4682[36]), and spontaneous malignant mesotheliomas from 2-year-old rats versus normal mesothelial Fred-PE cells (GSE47581[37]) were included.

Then, we curated a list of chemical compounds with differential gene expression profiles (drug vs. no drug) that exhibited negative correlations with at least one of the 5 peritoneal mesothelioma differential gene expression datasets (disease vs. control). The rationale for choosing drugs that show a negative correlation with at least one of the 5 expression datasets is rooted in the complexity of the MPeM genetic landscape. This approach recognizes the heterogeneous nature of MPeM and the inherent variability across its associated expression datasets. At the
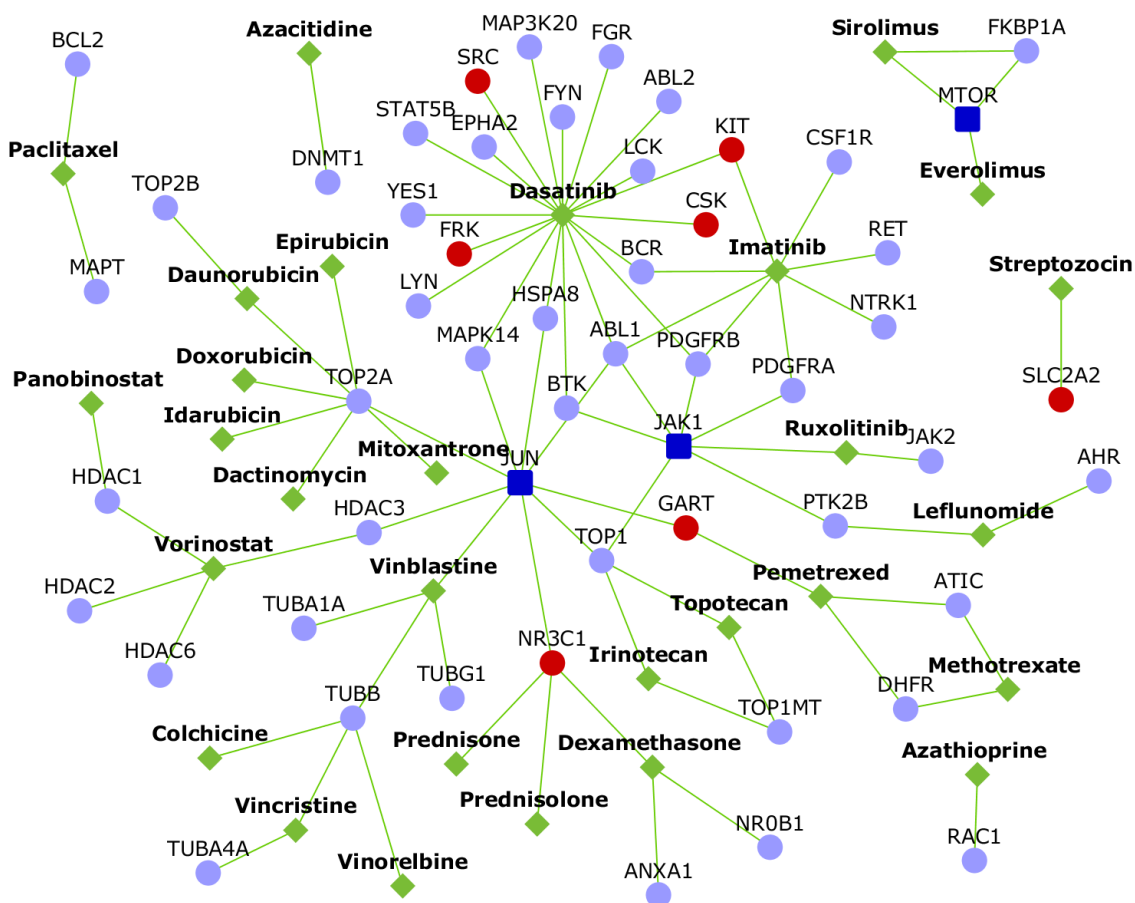
Figure 5. Repurposable drugs for malignant peritoneal mesothelioma (MPeM): The network shows 29 repurposable drugs (diamond-shaped green colored nodes) that target the proteins in the MPeM interactome. MPeM candidates are shown as dark blue nodes, their known interactors are light blue and novel interactors are red.

same time, it acknowledges the potential of drugs – even those that display correlation with only a single MPeM expression profile – to effectively target specific genes that might not exhibit uniform dysregulation across datasets.

Overall, we identified 39 drugs as potentially repurposable candidates for MPeM, including 23 that showed negative correlation with two or more gene expression datasets and 16 negatively correlated with a single dataset (**Supplementary Data Files 8-12**).

Literature review supported the biological validity of 29 (74%) out of these 39 drugs. These 29 drugs are shown in **Fig. 5**. Notably, 2 of these drugs (paclitaxel: NCT04000906 and imatinib: NCT00402766) are already in clinical trials for MPeM, and 2 others (pemetrexed and vinorelbine) are part of the standard therapy for mesothelioma.[62] In addition to this, the other shortlisted drugs exhibited activity relevant to MPeM (see **Supplementary Note 1** for details). In short, irinotecan has exhibited effectiveness against peritoneal mesothelioma, pleural mesothelioma, and peritoneal metastasis. Clinical trials and tests in cell lines have demonstrated the efficacy of paclitaxel and sirolimus against peritoneal mesothelioma and peritoneal metastasis. Clinical trials, animal models, and cell lines have validated the efficacy of twelve drugs against malignant pleural mesothelioma, namely, epirubicin, panobinostat, doxorubicin, imatinib, vinblastine, idarubicin, azacitidine, vorinostat, dactinomycin, acetylcysteine, staurosporine, and quercetin. Six drugs have shown effectiveness against primary peritoneal cancer and peritoneal metastasis in other cancers, namely, ruxolitinib, daunorubicin, dasatinib, topotecan, dexamethasone, and nintedanib. Methotrexate, resveratrol, everolimus, and genistein have demonstrated efficacy against both malignant pleural mesothelioma and peritoneal metastasis or sclerosis. Mitoxantrone and vincristine have been proven effective in managing pleural/peritoneal effusions.

Discussion

While multiple studies have examined MPeM genetics,[5-13] this study employs the protein interactome to uncover the biological themes underlying the MPeM-associated genes. The MPeM interactome, constructed from over 4,700 known and over 400 novel interactions of MPeM-associated genes from 9 studies, is examined using functional enrichment and transcriptome-based analyses to confirm its biological significance and to gain valuable insights into MPeM etiology, as well as to identify potentially repurposable drugs. Although a study centered on a single biological hypothesis would have been advantageous, the absence of mechanistic research on MPeM compelled us to conduct an exploratory analysis, resulting in a comprehensive understanding of its functional landscape. The hypotheses generated herein can be explored further through in vitro and in vivo studies.

Given the limited biological information available for MPeM, integrating the MPeM interactome with transcriptomic evidence becomes crucial to distinguish true disease-associated genes from those unrelated to the

disease, going beyond the core MPeM genes. We found that, within the interactome, over 75% — including more than 60% of novel interactors predicted for MPeM-associated genes — exhibited MPeM-related transcriptomic changes in humans and rodent models. Notably, 70% of these genes (1654 in total) had two or more pieces of supporting evidence. This integration of transcriptomic proof and the MPeM interactome effectively helped discern disease-associated genes from others. By overlaying disease-specific transcriptomic and genomic data onto the interactome, we could uncover an active sub-network of MPeM-associated genes (see **Supplementary Data File 4**) that possibly drives disease phenotypes.[63] The validity of our interactome-based approach is ascertained further by two factors: first, the unbiased identification of additional genes from the MPeM interactome, previously appearing in MPeM-related transcriptomic datasets, and second, their close functional proximity and interconnectedness with curated core genes harbouring MPeM-associated variants.

Biphasic MPeM, a rare subtype combining the common yet milder epithelioid type with the rarer, more severe sarcomatoid type, remains challenging to diagnose and manage due to limited reporting and occurrence.[64,65] The enrichment of the interactome with genes unique to biphasic MPeM implies distinctive molecular mechanisms underlying this subtype, operating at the network level. The identification of new interactors specific to this subtype suggests unexplored pathways and potential treatment targets. The interactome – encompassing over 100 biphasic subtype genes – can be used as a resource for biomarker discovery and tailored therapies. This underscores the broader potential of interactome-based methods for uncovering complexities in rare cancer subtypes.

Differential diagnosis of MPeM is challenging due to its non-specific clinical symptoms and histological patterns, often leading to misdiagnosis of other cancers.[66] Three novel interactors of MPeM-associated genes – TACSTD2, IGF2 and SUSD2 – can help differentiate OC/PPC from MPeM. Given that MPeM diagnosis currently relies on pathological evaluations,[66] resources such as our interactome can be leveraged to develop immunohistochemical diagnostic panels, thereby improving diagnosis and treatment outcomes.

The interactome showed enrichment for spleen and thymus-specific genes. This corresponded to the identification of a hematopoiesis module in the MPeM interactome. Both spleen and thymus regulate extramedullary hematopoiesis, i.e. the production of blood cells outside the bone marrow, a phenomenon crucial for cancer progression, albeit less reported in solid tumours compared to myeloproliferative neoplasms.[67] When reported, the phenomenon often manifests as organomegaly (enlarged organ).[67] Expansion of myeloid cells in the spleen through the process of extramedullary hematopoiesis resulting in splenomegaly has been observed in BAP1 (a core mesothelioma gene) knockout mice.[68] Although further investigations may be necessary to understand the functional implications, our finding suggests a potential link between extramedullary hematopoiesis and MPeM development.

The lack of enrichment of peritoneum-specific genes or genes specific to abdominal organs in the interactome is consistent with the absence of a distinct primary site for MPeM.[69] Pathological assessments commonly depict MPeM as diffusely spread throughout the abdominal cavity. However, additional research is necessary to determine if this lack of a primary site arises from the heterogeneity of MPeM-associated genes.

The modules and pathways identified from the interactome provide insights into processes spanning multiple biological levels that could contribute to the development and progression of peritoneal mesothelioma. Note that the majority of supporting evidence stems from pleural mesothelioma studies. Dysregulated covalent chromatin modification, including histone modifications and SUMOylation, can lead to genetic instability and epigenetic changes driving malignant transformation.[70] Altered mRNA metabolic and transcriptional processes might impact gene expression profiles,[71,72] while disrupted translation and post-translational events like deubiquitination could influence cellular proteomes.[40,73] Furthermore, the activation of transmembrane receptor protein tyrosine kinase pathways, coupled with downstream MAPK cascades,[74] and disrupted cell-cell junction assembly can enhance tumour cell survival and invasiveness,[75] thereby promoting cancer progression.

The enrichment of cytokine signalling underscores the potential impact of inflammation, particularly the positive regulation of interleukin-6 (IL-6) production, on the progression of peritoneal mesothelioma. Indeed, elevated expression of an anti-apoptotic factor called survivin (BIRC5) induced by the cytokine IL-6 has been reported in MPeM patients; knockdown of this gene led to increased (spontaneous and drug-induced) apoptosis.[12] The IL-6 production module contained 13 novel interactors of MPeM core genes: **NR2F2**-SYCP3, **ESR1**-DDX43, **RB1**-LRCH1, **RB1**-PCDHB5, **MRE11**-GPR83, **PBRM1**-FBXW8, **RB1**-CDADC1, **MET**-FOXA3, **RB1**-CNTN3, **SMARCB1**-MYO18B, **TRIO**-DNAH5, **ARHGAP22**-ZNF488 and **SDHB** (and **MTOR**)-SLC45A1. Future studies could concentrate on examining these novel PPIs. This is particularly important because both chronic inflammation induced by abdominal surgeries and persistent peritoneal inflammation (i.e. chronic peritonitis) confer a risk of developing MPeM.[1]

Lastly, among the MPeM core genes used for interactome construction, 56% (33 in total) were linked to chromosomal events such as copy number gain/loss, gene loss, deletion and gene rearrangement. Correspondingly, the interactome revealed enrichment for chromosome segregation as a functional module. Notably, we identified 7 novel PPIs that can be examined in experimental studies, with both the MPeM core genes and their novel interactors involved in chromosomal events (MPeM genes are shown in bold): **RASSF1**-LARS2, **ARID1B**-MTHFD1L, **RHEB**-CENPE, **VEGFB**-TFB1M, **JUN**-GART, **PTEN**-KIF20B and **KEAP1**-SENP1.

The pleural and peritoneal mesothelioma subtypes differ in their association with germline mutations, history of asbestos exposure, and post-operative complications and have different median survival rates.[2,3] Although several

studies have examined the genomic features distinguishing them,[76,77] none have identified their underlying biological themes. We showed that more than 950 genes co-occurred commonly in MPM and MPeM interactomes, which is an approximately 3-fold higher enrichment of high statistical significance than expected. Moreover, this shared interactomic subspace that underlies the two distinct mesothelioma subtypes is likely driven by immune pathways. This observation is particularly relevant given the emerging potential of gene signatures from the mesothelioma tumour immune microenvironment to predict therapy responses.[78]

Currently, the first-line chemotherapy regimen for MPeM involves pemetrexed/cisplatin, resulting in complete or partial responses in merely 26% of patients and disease stabilization in only 45% of patients.[1] We identified several repurposable drugs for MPeM treatment, with over 70% showing effectiveness against peritoneal mesothelioma, pleural mesothelioma, peritoneal metastasis and/or primary peritoneal cancer in clinical trials, animal models or cell lines, confirming the credibility of the approach. However, the drug-associated expression profiles analyzed in our study were induced in a wide variety cell lines. Therefore, to advance clinical translation in MPeM, the effect of the proposed drugs should be examined in human peritoneal mesothelioma cell lines or animal models.

Overall, the study allows us to conceptualise MPeM as originating from disrupted interactions within the MPeM interactome due to genetic mutations or aberrant expression of MPeM-associated genes, yielding broader implications for comorbid conditions and drug responses.[79] The disturbances capable of influencing this interactome can manifest across multiple levels. The genetic underpinnings of MPeM manifest across several organs outside of the peritoneum and abdominal organs and are linked to processes operating at the genomic, transcriptomic, and proteomic levels. Perturbations within the immunological system also contribute to MPeM development, with immune-mediated pathways playing a critical role in the shared origins of pleural and peritoneal subtypes of mesothelioma. Additionally, MPeM shares genetic attributes with other malignancies, including (but not limited to) genes predictive of patient prognosis. It could be difficult to differentially diagnose some of these malignancies from MPeM upon phenotypic assessment. Overall, MPeM is a complex disorder warranting investigations from various perspectives.

Our study has a few limitations. For several analyses, we have used genetic data from animal models due to the absence of human patient data. Results from these should be interpreted with caution. Direct correlations of genes/proteins/phenotypes between animal models and humans require thorough characterization in both species.[80] Also, our bioinformatics-based conclusions should be confirmed through experimental validation in pertinent tissues or cell lines.

In summary, our study provides a network-level view of MPeM-associated genes and their functional consequences. The MPeM interactome can serve as a functional landscape to integrate multi-omics data, informing genetic studies and biomedical studies seeking to improve clinical interventions in MPeM.

Methods

Compilation of MPeM-associated genes and prediction of novel interactions

A list of 59 MPeM-associated genes that harboured mutations, copy number aberrations, rearrangements or showed expression correlated with poor prognosis in MPeM patients or reduced cell survival or less favourable response to drugs in MPeM surgical specimens was compiled from eight studies.[5-13] Novel PPIs of the proteins encoded by these genes were predicted using the HiPPIP model that we developed.[34] Each MPeM protein (say N1) was paired with each of the other human proteins say, (M1, M2,…Mn), and each pair was evaluated with the HiPPIP model.[34] The predicted interactions of each of the MPeM proteins were extracted (namely, the pairs whose score is >0.5, a threshold which through computational evaluations and experimental validations was revealed to indicate interacting partners with high confidence). The interactome figures were created using Cytoscape.[81]

Identification of functional modules

Functional gene modules were extracted using the HumanBase toolkit[49] (https://hb.flatironinstitute.org/). HumanBase uses shared k-nearest-neighbors and the Louvain community-finding algorithm to cluster the genes sharing the same network neighborhoods and similar GO biological processes into functional modules. The p-values of the terms enriched in the modules are calculated using Fisher's exact test and Benjamini–Hochberg method.

Functional enrichment analysis

Biological process (Gene Ontology[82]), pathway (Reactome[83]) and disease (DisGeNET[53]) enrichments were computed using WebGestalt.[50] WebGestalt computes the distribution of genes belonging to a particular functional category in the input list and compares it with the background distribution of genes belonging to this functional category among all the genes that belongs to any functional category in the database selected by the user. Statistical significance of functional category enrichment is computed using Fisher's exact test, and corrected using the Benjamini-Hochberg method for multiple test adjustment. Annotations with FDR-corrected p-value < 0.05 were considered significant.

Tissue-specific expression analysis

Tissue-specificity of the genes in the MPeM interactome were checked using TissueEnrich.[84] The analysis was based on tissue-specific genes compiled from GTEx and Mouse ENCODE.[44,45] This included 'tissue-enriched genes' with at least 5-folds higher mRNA levels in a particular tissue compared to all the other tissues, 'group-enriched genes' with at least 5-folds higher mRNA levels in a group of 2-7 tissues and 'tissue-enhanced genes' with at least 5-folds higher mRNA levels in a particular tissue compared to average levels in all tissues.

Network overlap analysis

Statistical significance of the overlaps between genes in the MPeM and MPeM interactomes was computed based on hypergeometric distribution.

Identification of prognostic cancer genes

Data for correlation of gene expression and fraction of patient population surviving after treatment of 20 cancer types was taken from Pathology Atlas.[52] Genes with log-rank P-value < 0.001 were considered to be prognostic. Unfavourable prognosis indicates positive correlation of high gene expression with reduced patient survival.

Identification of repurposable drugs

The list of chemical compounds whose gene expression profiles correlated negatively with 5 gene expression datasets associated with peritoneal mesothelioma were compiled using the BaseSpace correlation software (https://www.nextbio.com) (List 1). The datasets considered were granulocytic myeloid-derived suppressor cells (G-MDSCs) from spleens of mice bearing AB12 mesothelioma grafts versus naive neutrophils, neutrophils infiltrating AB12 mesothelioma tumour grafts versus naive bone marrow derived neutrophils (GSE43254[35]), BCA induced peritoneal mesothelioma versus non-transformed mesothelial cell line, O-NT induced peritoneal mesothelioma versus non-transformed mesothelial cell line (GSE4682[36]) and spontaneous malignant mesotheliomas from 2-year-old rats versus normal mesothelial Fred-PE cells (GSE47581[37]). Next, we identified drugs that targeted at least one gene in in the MPeM interactome using Drug Bank.[33] We then compared list 1 and list 2 to identify the drugs that not only target proteins in the interactome but are also negatively correlated with MPeM-associated gene expression profiles.

Acknowledgements

for joint effort in securing funding that supported this work. Authors thank Prof. N. Balakrishnan of Indian Institute of Science for supporting KBK.

Author contributions

MKG conceptualized and supervised the study and carried out interactome prediction and analysis. KBK carried out studies of overlap of the interactome with high-throughput data, literature-based evidence gathering and identification of repurposable drugs. Manuscript has been written by KBK and edited by MKG. Manuscript has been read and approved by all authors.

Competing interests

None.

Data and materials availability

The MPeM core genes used for interactome construction, the complete list of PPIs in the MPeM interactome and the list of novel interactors in the interactome have been made available as **Supplementary Data File 1**, **Supplementary Data File 2** and **Supplementary Data File 3**, respectively.

References

1   Carbone, M. *et al.* Mesothelioma: scientific clues for prevention, diagnosis, and therapy. *CA: a cancer journal for clinicians* **69**, 402-429 (2019).

2   Robinson, B. W. & Lake, R. A. Advances in malignant mesothelioma. *New England Journal of Medicine* **353**, 1591-1603 (2005).

3   Amin, W. *et al.* Factors influencing malignant mesothelioma survival: a retrospective review of the National Mesothelioma Virtual Bank cohort. *F1000Research* **7** (2018).

4   Panou, V. *et al.* Frequency of germline mutations in cancer susceptibility genes in malignant mesothelioma. *Journal of Clinical Oncology* **36**, 2863 (2018).

5   Joseph, N. M. *et al.* Genomic profiling of malignant peritoneal mesothelioma reveals recurrent alterations in epigenetic regulatory genes BAP1, SETD2, and DDX3X. *Modern Pathology* **30**, 246-254 (2017).

6   Ugurluer, G. *et al.* Genome-based mutational analysis by next generation sequencing in patients with malignant pleural and peritoneal mesothelioma. *Anticancer research* **36**, 2331-2338 (2016).

7   Chirac, P. *et al.* Genomic copy number alterations in 33 malignant peritoneal mesothelioma analyzed by comparative genomic hybridization array. *Human pathology* **55**, 72-82 (2016).

8   Foster, J. M. *et al.* Clinical implications of novel activating EGFR mutations in malignant peritoneal mesothelioma. *World journal of surgical oncology* **8**, 88 (2010).

9   Hung, Y. P. *et al.* Molecular characterization of diffuse malignant peritoneal mesothelioma. *Modern Pathology*, 1-11 (2020).

10   Pillai, K., Pourgholami, M. H., Chua, T. C. & Morris, D. L. MUC1 has prognostic significance in malignant peritoneal mesothelioma. *The International journal of biological markers* **28**, 303-312 (2013).

11   Varghese, S. *et al.* Activation of the phosphoinositide-3-kinase and mammalian target of rapamycin signaling pathways are associated with shortened survival in patients with malignant peritoneal mesothelioma. *Cancer* **117**, 361-371 (2011).

12   Zaffaroni, N. *et al.* Survivin is highly expressed and promotes cell survival in malignant peritoneal mesothelioma. *Analytical Cellular Pathology* **29**, 453-466 (2007).

13   Hung, Y. P. *et al.* Identification of ALK rearrangements in malignant peritoneal mesothelioma. *JAMA oncology* **4**, 235-238 (2018).

14   Hmeljak, J. *et al.* Integrative molecular characterization of malignant pleural mesothelioma. *Cancer discovery* **8**, 1548-1565 (2018).

15   Bott, M. *et al.* The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21. 1 losses in malignant pleural mesothelioma. *Nature genetics* **43**, 668 (2011).

16   Jensen, D. E. *et al.* BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16**, 1097 (1998).

17   Hakiri, S. *et al.* Functional differences between wild-type and mutant-type BRCA1-associated protein 1 tumour suppressor against malignant mesothelioma cells. *Cancer science* **106**, 990-999 (2015).

18   Zauderer, M. G. *et al.* Vinorelbine and gemcitabine as second-or third-line therapy for malignant pleural mesothelioma. *Lung Cancer* **84**, 271-274 (2014).

19   Zucali, P. *et al.* Vinorelbine in pemetrexed-pretreated patients with malignant pleural mesothelioma. *Lung Cancer* **84**, 265-270 (2014).

20   Keskin, O., Tuncbag, N. & Gursoy, A. Predicting protein–protein interactions from the molecular to the proteome level. *Chemical reviews* **116**, 4884-4909 (2016).

21   Keshava Prasad, T. *et al.* Human protein reference database—2009 update. *Nucleic acids research* **37**, D767-D772 (2008).

22   Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic acids research* **34**, D535-D539 (2006).

23   Blasche, S. & Koegl, M. Analysis of Protein–Protein Interactions Using LUMIER Assays. *Virus-Host Interactions: Methods and Protocols*, 17-27 (2013).

24   Trepte, P. *et al.* DULIP: a dual luminescence-based co-immunoprecipitation assay for interactome mapping in mammalian cells. *Journal of molecular biology* **427**, 3375-3388 (2015).

25   Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature*, 1-7 (2020).

26   Huttlin, E. L. *et al.* Dual Proteome-scale Networks Reveal Cell-specific Remodeling of the Human Interactome. *bioRxiv* (2020).

27   Ganapathiraju, M. K. *et al.* Schizophrenia interactome with 504 novel protein–protein interactions. *npj Schizophrenia* **2**, 16012 (2016).

28   Zhu, J. *et al.* Antiviral activity of human OASL protein is mediated by enhancing signaling of the RIG-I RNA sensor. *Immunity* **40**, 936-948 (2014).

29   Karunakaran, K. B., Yanamala, N., Boyce, G. & Ganapathiraju, M. K. Mesothelioma Interactome with 367 Novel Protein-Protein Interactions. *bioRxiv*, 459065 (2018).

30   Malavia, T. *et al.* Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epidemiological, genomic and protein interaction data *npj Schizophrenia* **in Press** (2017).

31   Li, Y. *et al.* Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature* **521**, 520-524, doi:10.1038/nature14269 (2015).

32   Liu, X. *et al.* The complex genetics of hypoplastic left heart syndrome. *Nat Genet* **49**, 1152-1159, doi:10.1038/ng.3870 (2017).

33   Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* **34**, D668-D672 (2006).

34   Ganapathiraju, M. K., Karunakaran, K. B. & Correa-Menendez, J. Predicted protein interactions of IFITMs may shed light on mechanisms of Zika virus-induced microcephaly and host invasion. *F1000Res* **5**, 1919, doi:10.12688/f1000research.9364.2 (2016).

35   Fridlender, Z. G. *et al.* Transcriptomic analysis comparing tumour-associated neutrophils with granulocytic myeloid-derived suppressor cells and normal neutrophils. *PloS one* **7**, e31524 (2012).

36   Kim, Y. *et al.* Major carcinogenic pathways identified by gene expression analysis of peritoneal mesotheliomas following chemical treatment in F344 rats. *Toxicology and applied pharmacology* **214**, 144-151 (2006).

37   Blackshear, P. E. *et al.* Spontaneous mesotheliomas in F344/N rats are characterized by dysregulation of cellular growth and immune function pathways. *Toxicologic pathology* **42**, 863-876 (2014).

38   Shukla, A. *et al.* Alterations in gene expression in human mesothelial cells correlate with mineral pathogenicity. *American journal of respiratory cell and molecular biology* **41**, 114-123 (2009).

39   Dragon, J., Thompson, J., MacPherson, M. & Shukla, A. Differential susceptibility of human pleural and peritoneal mesothelial cells to asbestos exposure. *Journal of cellular biochemistry* **116**, 1540-1552 (2015).

40   Borczuk, A. *et al.* Molecular profiling of malignant peritoneal mesothelioma identifies the ubiquitin–proteasome pathway as a therapeutic target in poor prognosis tumours. *Oncogene* **26**, 610-617 (2007).

41   Yanamala, N. *et al.* Characterization of pulmonary responses in mice to asbestos/asbestiform fibers using gene expression profiles. *Journal of Toxicology and Environmental Health, Part A* **81**, 60-79 (2018).

42   Cunha, P. *et al.* Malignant peritoneal mesothelioma--diagnostic and therapeutic difficulties. *Acta medica portuguesa* **15**, 383-386 (2002).

43   Davidson, B. *et al.* Gene expression signatures differentiate ovarian/peritoneal serous carcinoma from diffuse malignant peritoneal mesothelioma. *Clinical cancer research* **12**, 5944-5950 (2006).

44   Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* **45**, 580-585, doi:10.1038/ng.2653 (2013).

45   Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic acids research* **46**, D794-D801 (2018).

46   Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics* **13**, 397-406 (2014).

47   Kupershmidt, I. *et al.* Ontology-based meta-analysis of global collections of high-throughput public data. *PloS one* **5**, doi:10.1371/journal.pone.0013066 (2010).

48   Chattopadhyay, A. & Ganapathiraju, M. K. Demonstration Study: A Protocol to Combine Online Tools and Databases for Identifying Potentially Repurposable Drugs. *Data* **2**, 15 (2017).

49   Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nature neuroscience* **19**, 1454-1462 (2016).

50   Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic acids research* (2019).

51   Liu, Y. *et al.* Global prevalence of congenital heart disease in school-age children: a meta-analysis and systematic review. *BMC cardiovascular disorders* **20**, 1-10 (2020).

52   Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).

53   Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*, gkw943 (2016).

54   Cacheiro, P., Haendel, M. A., Smedley, D., Consortium, I. M. P. & Initiative, M. New models for human disease from the International Mouse Phenotyping Consortium. *Mammalian Genome* **30**, 143-150 (2019).

55   Whiteside, T. L. in *Seminars in cancer biology.* 3-15 (Elsevier).

56   Bulek, K. *et al.* The inducible kinase IKKi is required for IL-17-dependent signaling associated with neutrophilia and pulmonary inflammation. *Nature immunology* **12**, 844-852 (2011).

57   Creaney, J. *et al.* Overexpression and altered glycosylation of MUC1 in malignant mesothelioma. *British journal of cancer* **98**, 1562-1569 (2008).

58   Sirota, M. *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science translational medicine* **3**, 96ra77-96ra77 (2011).

59   Karunakaran, K. B., Chaparala, S. & Ganapathiraju, M. K. Potentially repurposable drugs for schizophrenia identified from its interactome. *Scientific Reports* **9**, 12682, doi:10.1038/s41598-019-48307-w (2019).

60   Nimgaonkar, V. Cromoglicate Adjunctive Therapy for Outpatients With Schizophrenia (CATOS). *ClinicalTrials.gov* (2019).

61   Nimgaonkar, V. Acetazolamide for Treatment Resistant Schizophrenia (APTS). *ClinicalTrials.gov* (2022).

62   Stahel, R., Weder, W., Lievens, Y. & Felip, E. Malignant pleural mesothelioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology* **21**, v126-v128 (2010).

63   Mitra, K., Carvunis, A.-R., Ramesh, S. K. & Ideker, T. Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics* **14**, 719-732 (2013).

64   Fujishima, F. *et al.* Histological and immunohistochemical characteristics and p16 status studied by FISH in six incidentally detected cases of well-differentiated papillary mesothelioma of the peritoneum. *Indian Journal of Pathology and Microbiology* **64**, 277 (2021).

65   Sato, T. *et al.* Three newly established immortalized mesothelial cell lines exhibit morphological phenotypes corresponding to malignant mesothelioma epithelioid, intermediate, and sarcomatoid types, respectively. *Cancer Cell International* **21**, 1-11 (2021).

66   Kim, J., Bhagwandin, S. & Labow, D. M. Malignant peritoneal mesothelioma: a review. *Annals of translational medicine* **5** (2017).

67   Bao, Y. *et al.* Extramedullary hematopoiesis secondary to malignant solid tumours: a case report and literature review. *Cancer Management and Research*, 1461-1470 (2018).

68   Dey, A. *et al.* Loss of the tumour suppressor BAP1 causes myeloid transformation. *Science* **337**, 1541-1546 (2012).

69   Deraco, M., Bartlett, D., Kusamura, S. & Baratti, D. Consensus statement on peritoneal mesothelioma. *Journal of surgical oncology* **98**, 268-272 (2008).

70   Cakiroglu, E. & Senturk, S. Genomics and functional genomics of malignant pleural mesothelioma. *International Journal of Molecular Sciences* **21**, 6342 (2020).

71   Kettunen, E. *et al.* Gene expression profiling of malignant mesothelioma cell lines: cDNA array study. *International journal of cancer* **91**, 492-496 (2001).

72   Rihn, B. *et al.* Differential gene expression in mesothelioma. *FEBS letters* **480**, 95-100 (2000).

73   Grosso, S. *et al.* The pathogenesis of mesothelioma is driven by a dysregulated translatome. *Nature Communications* **12**, 4920 (2021).

74   Yang, H. *et al.* MEK1 drives oncogenic signaling and interacts with PARP1 for genomic and metabolic homeostasis in malignant pleural mesothelioma. *Cell death discovery* **9**, 55 (2023).

75   Keller, M., Reis, K., Hjerpe, A., Dobra, K. & Aspenström, P. Cytoskeletal organization correlates to motility and invasiveness of malignant mesothelioma cells. *Cancers* **13**, 685 (2021).

76   Takeda, M. *et al.* Comparison of genomic abnormality in malignant mesothelioma by the site of origin. *Journal of clinical pathology* **67**, 1038-1043 (2014).

77   Borczuk, A. C. *et al.* Genome-wide analysis of abdominal and pleural malignant mesothelioma with DNA arrays reveals both common and distinct regions of copy number alteration. *Cancer biology & therapy* **17**, 328-335 (2016).

78   Nair, N. U. *et al.* Genomic and transcriptomic analyses identify a prognostic gene signature and predict response to therapy in pleural and peritoneal mesothelioma. *Cell Reports Medicine* **4** (2023).

79   Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature reviews genetics* **12**, 56-68 (2011).

80   Breschi, A., Gingeras, T. R. & Guigó, R. Comparative transcriptomics in human and mouse. *Nature Reviews Genetics* **18**, 425-440 (2017).

81   Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).

82   Consortium, G. O. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research* **32**, D258-D261 (2004).

83   Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic acids research* **42**, D472-D477 (2014).

84   Jain, A. & Tuteja, G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**, 1966-1967, doi:10.1093/bioinformatics/bty890 (2019).

Interactome-based framework to translate disease genetic data into biological and clinical insights

6. Interactome of SARS-CoV-2 modulated host proteins with computationally predicted PPIs: Insights from illustrative translational systems biology studies

The experimental chapter is based on the following peer-reviewed publication:

Karunakaran, Kalyani B., N. Balakrishnan, and Madhavi K. Ganapathiraju. Interactome of SARS-CoV-2 modulated host proteins with computationally predicted PPIs: Insights from illustrative translational systems biology studies. Frontiers in Systems Biology (2022): 2.

Summary of this chapter

In this chapter, I demonstrate how the neighbourhood network of the human proteins modulated by SARS-CoV-2 – containing both experimentally determined and computationally predicted protein-protein interactions – was constructed to derive biological insights into host invasion and response mechanisms and identify repurposable drugs for COVID-19. The host protein interactome was validated using multiple SARS-CoV/SARS-CoV-2 associated transcriptomic and proteomic datasets. Topological module, tissue-specificity, and functional enrichment analyses helped characterise the interactome extensively. I also showed the network proximity of host proteins to genes associated with common co-morbidities found among critically ill COVID patients and non-survivors. Finally, I identified 24 repurposable drugs for COVID-19 using comparative transcriptome analysis, which included those undergoing COVID-19 clinical trials, showing broad-spectrum antiviral properties or proven activity against SARS-CoV-2 or SARS-CoV/MERS-CoV in cell-based assays. In summary, the interactomic framework helped to integrate rapidly emerging SARS-CoV-2 data and generate biologically insightful and biomedically actionable results.

Contribution to this chapter (85%)

- Designed the study and developed the methodology of the project, which included interactome construction, validation, functional characterisation, network proximity and drug repurposing analysis
- Curated all the datasets, performed all the analyses and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

Interactome of SARS-CoV-2 Modulated Host Proteins With Computationally Predicted PPIs: Insights From Translational Systems Biology Studies

Kalyani B. Karunakaran[1], N. Balakrishnan[1] and Madhavi K. Ganapathiraju[2,3]*

[1]Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, India, [2]Department of Biomedical Informatics, School of Medicine, Pittsburgh, PA, United States, [3]Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, United States

Accelerated efforts to identify intervention strategies for the COVID-19 pandemic caused by SARS-CoV-2 need to be supported by deeper investigations into host invasion and response mechanisms. We constructed the neighborhood interactome network of the 332 human proteins targeted by SARS-CoV-2 proteins, augmenting it with 1,941 novel human protein-protein interactions predicted using our High-precision Protein-Protein Interaction Prediction (HiPPIP) model. Novel interactors, and the interactome as a whole, showed significant enrichment for genes differentially expressed in SARS-CoV-2-infected A549 and Calu-3 cells, postmortem lung samples of COVID-19 patients and blood samples of COVID-19 patients with severe clinical outcomes. The PPIs connected host proteins to COVID-19 blood biomarkers, ACE2 (SARS-CoV-2 entry receptor), genes differentiating SARS-CoV-2 infection from other respiratory virus infections, and SARS-CoV-targeted host proteins. Novel PPIs facilitated identification of the *cilium organization* functional module; we deduced the potential antiviral role of an interaction between the virus-targeted NUP98 and the cilia-associated CHMP5. Functional enrichment analyses revealed promyelocytic leukaemia bodies, midbody, cell cycle checkpoints and tristetraprolin pathway as potential viral targets. Network proximity of diabetes and hypertension associated genes to host proteins indicated a mechanistic basis for these co-morbidities in critically ill/non-surviving patients. Twenty-four drugs were identified using comparative transcriptome analysis, which include those undergoing COVID-19 clinical trials, showing broad-spectrum antiviral properties or proven activity against SARS-CoV-2 or SARS-CoV/MERS-CoV in cell-based assays. The interactome is available on a webserver at http://severus.dbmi.pitt.edu/corona/.

Keywords: interactome analysis, protein-protein interactions, computational prediction, COVID-19, SARS-CoV-2, drugs, drug repurposing

1 INTRODUCTION

COVID-19 (Coronavirus Disease 2019) is an infectious virus outbreak which emerged as an epidemic in one city in December 2019, and within 3 months swept across 220 countries and territories, developing into a pandemic global health crisis with more than 180 million confirmed infected cases and more than 4 million deaths as of 1 July 2021 (WHO). The novel coronavirus (SARS-CoV-2/nCoV-19) has been identified as the causative agent of this disease (Rothan and Byrareddy, 2020). Coronaviruses are a large family of viruses that are pathogenic in mammals and birds. In humans, they cause respiratory infections ranging from the common cold to possibly fatal acute respiratory distress syndrome (ARDS) and acute lung injury (ALI), which are noted in COVID-19 as well as in its predecessors, namely, SARS (severe acute respiratory syndrome, 2002–2003) and MERS (middle east respiratory syndrome, 2012) (Gralinski and Baric, 2015). SARS-CoV-2 is airborne, and causes no symptoms in several infected people who may become silent carriers of the disease to the more vulnerable population. COVID-19 is spreading at an exponential rate globally, prompting scientists across the globe to investigate the mechanisms of its host invasion and host response to viral infection, in hopes of discovering treatment strategies to combat the outbreak.

The viral infection sets off a cascade of interactions among multiple genes and proteins in the host cell. This complex network has the potential to restrict viral replication in host cells, or conversely, to be taken over by the virus for its perpetuation. Several research groups have studied the effects of SARS-CoV-2 on the host from a systems-level perspective (Blanco-Melo et al., 2020a; Gordon et al., 2020a; Zhou et al., 2020a). 332 human proteins that bind to SARS-CoV-2 proteins were identified through affinity purification—mass spectrometry (AP-MS) by Gordon et al. (Gordon et al., 2020b). Melo et al. identified more than 6,000 genes differentially expressed in A549, Calu-3 and NHBE cell lines upon SARS-CoV-2 infection, and in COVID-19 patients (Blanco-Melo et al., 2020b). Bojkova et al. monitored SARS-CoV-2 infection in Caco2 cell line and generated temporal infection profiles of 2,687 genes in the host translatome and 6,258 proteins in the proteome (Denisa Bojkova and Koch, 2020). Data generated by these studies can be employed to conduct systematic, unbiased and data-driven investigations into COVD-19 from the perspective of the host, by constructing the relevant protein interactome (i.e., protein- protein interaction network).

Protein-protein interactions (PPIs) drive the cellular machinery and facilitate biological processes including signal transduction, formation of cellular structures and enzymatic complexes. When viral proteins bind to some proteins in the host cell, this effect may spread along the interactome through regulatory and biophysical interactions, affecting other proteins in the PPI network, posing deeper implications for viral infection, host immunity, and the effect of therapeutics (Barabási et al., 2011). Despite being critical to unravelling novel disease mechanisms and drugs, ~75% of estimated PPIs are currently unknown and several disease-associated genes have no known

PPIs. More than ~600,000 PPIs are said to exist in the human interactome (Keskin et al., 2016) and only ~150,000 PPIs are known from PPI repositories such as HPRD (Keshava Prasadet al., 2008) and BioGRID (Stark et al., 2006). Detecting the PPIs using experimental techniques such as co-immunoprecipitation (Co-IP) (Blasche and Koegl, 2013; Trepte et al., 2015) is prohibitively laborious and time-consuming at large scale. Tens of thousands of PPIs are being added into the interactome through systematic high throughput studies with yeast two hybrid (Y2H) system (Luck et al., 2020) and AP–MS (Huttlin et al., 2020). Despite this, a large part of the interactome remains unknown. Hence, computational algorithms have been developed to predict PPIs in human as well as model organisms (Deng et al., 2003; Raja et al., 2013; You et al., 2013; Emamjomehet al., 2014; Hopf et al., 2014; Jia et al., 2015; Kotlyar et al., 2015; Garzón et al., 2016; Malavia et al., 2017a). We have previously developed a computational model called HiPPIP (High-Precision Protein-Protein Interaction Prediction) that was deemed highly accurate by computational evaluations and experimental validations (Zhu et al., 2014; Ganapathiraju et al., 2016a; Dunham and Ganapathiraju, 2022). HiPPIP computes features of protein pairs such as cellular localization, molecular function, biological process membership, genomic location of the gene, and gene expression in microarray experiments, and classifies the pairwise features as *interacting* or *non-interacting* based on a random forest model (Ganapathiraju et al., 2016a). Though each of the features by itself is not an indicator of an interaction, a machine learning model was able to use the combined features to make predictions with high precision. The threshold of HiPPIP to classify a protein-pair as "a PPI" was set high in such a way that it yields very high-precision predictions even if low recall. Seventeen of the predicted PPIs were tested experimentally and were shown to be true PPIs, namely, 8 PPIs validated by co-immunoprecipitation: DDX58-OASL (Zhu et al., 2014), HMGB1-FLT1 (Ganapathiraju et al., 2016a), HMGB1-KL (Ganapathiraju et al., 2016a), STT3A-RPS25 (Ganapathirajuet al., 2016a), STT3A-SYCP3 (Ganapathiraju et al., 2016a), STT3A-MCAM (Ganapathiraju et al., 2016a), PDCD1-<hidden> (unpublished validation), YWHAE1-<hidden> (unpublished validation), five PPIs validated by *in vitro* pull-down and mass spectrometry: ALB-KDR (Karunakaran et al., 2021), ALB-PDGFRA (Karunakaran et al., 2021), BAP1-PARP3 (Karunakaran et al., 2021), CLPS-CUTA (Karunakaran et al., 2021), HMGB1-CUTA (Karunakaran et al., 2021) and 4 PPIs validated by co-localization: STX3-LPXN (Ganapathiraju et al., 2016a), STX4-MAPK3 (Ganapathiraju et al., 2016a), IFT88-KL (unpublished validation) and WDR5-IGFBP3 (unpublished validation). Some of the predicted PPIs proved to have high translational impact. For example, we predicted that the human OASL protein (*IFN-inducible oligoadenylate synthetases-like*) interacts with RIG-I (*retinoic acid-inducible gene I*); it was validated to be a true PPI. Further investigations conclusively showed that this interaction is responsible for activating cellular innate immunity to virus infections: OASL enhances antiviral signalling mediated by the viral RNA sensor RIG-I by binding through its C-terminal ubiquitin-like domain (Zhu et al., 2014). Other high-impact results from interactome analysis include

FIGURE 1 | Concept diagram of the analysis presented in the paper: Clockwise from top-left: (A). SARS-CoV-2 enters the host cell with the help of an interaction between its surface-anchored "spike protein" and the ACE2 receptor on the host cell. Once the virus gains entry into the host cell, it hijacks the host cellular machinery to promote viral genome replication, and viral mRNA and protein synthesis (B). SARS-CoV-2 viral proteins (purple) produced in this manner interact with a specific set of host cell proteins (dark blue; identified by Gordon et al. (2020a). In this study, (C). we assembled the known (light blue) and computationally predicted/novel (red) interactors of the host proteins, and (D). systematically studied this expanded neighbourhood network of virus-targeted host proteins (E). Various types of analysis carried out in this work. Several more systems studies are possible with the interactome both because of new types of analysis or analysis with new data sources. Cilium, PML bodies and midbody figures are from (Morgan, 2007; Wilson, 2015; Hoischen et al., 2018) (with creative commons license on WikiMedia). Novel PPIs of the SARS-CoV-2 targeted host proteins predicted by HiPPIP can be found at http://severus.dbmi.pitt.edu/corona/.

shared PPIs explaining inverse epidemiological relationship between schizophrenia and rheumatoid arthritis (Malavia et al., 2017b) and cilia-transduced cell signaling in congenital heart disease (Li et al., 2015; Liu et al., 2017), and more (Karunakaran et al., 2019a).

In this work, we present the human protein-protein interactome of the proteins targeted by SARS-CoV-2 (Gordon et al. (2020a)). A concept diagram of the analysis carried out here is shown in Figure 1. Key contributions of this work are about 2,000 previously unknown human PPIs that are computationally predicted with high-precision, and the results of analyzing the network of known and predicted interactions with functional annotations and with SARS-CoV-2-relevant transcriptomic and proteomic data. Importantly, we are making this interactome,

with rich annotations, available on a webserver and in graph formats downloadable for further computational analyses.

2 RESULTS

We collected 332 host proteins that were identified to interact with 27 SARS-CoV-2 viral proteins identified from the 2019-nCoV/USA-WA1/2020 strain by Gordon et al. (2020a). To assemble the interactome of these host proteins, we compiled known PPIs from HPRD (Keshava Prasad et al., 2008) (Human Protein Reference Database) and BioGRID (Stark et al., 2006) (Biological General Repository for Interaction Datasets), and predicted novel PPIs by applying the HiPPIP algorithm

FIGURE 2 | Neighborhood networks of SARS-CoV-2-targeted host proteins expanded by novel PPIs: Genes are shown as nodes and PPIs as edges. The network diagrams illustrate augmentation of the host protein interactome with computationally predicted novel interactors (shown as round red colored nodes). The data pertaining to the interaction (golden yellow edges) of SARS-CoV-2 viral proteins (shown as diamond shaped dark green colored nodes) with host proteins (square shaped dark blue colored nodes) was collected from Gordon et al. (Gordon et al., 2020a). The data pertaining to known PPIs (blue edges) of the host proteins with known interactors (round light blue colored nodes) was collected from BioGRID and HPRD. To present a more complete version of the interactome, we computationally predicted novel PPIs (red edges) of the host proteins with novel interactors (round red colored nodes) using the HiPPIP algorithm. Each of these novel PPIs not only serves as valuable candidates for further experimentation, but also facilitates systems-level analysis of the host protein interactome and generation of new testable hypotheses. Note that the figure only shows the neighbourhood networks of a few selected viral proteins. The complete list of interactions between viral proteins and host proteins, and between host proteins and its interactors can be found in http://severus.dbmi.pitt.edu/corona/.

described in our earlier work (Ganapathiraju et al., 2016b) (Supplementary Table S1). Note that the interactome presented here is human protein interactome, and not a virus- host interactome; the relevance to COVID-19 is that the core proteins for which the interactome is assembled are those that the viral proteins bind to. Specifically, as shown in Figure 2, we assembled the known and novel interactors (round light blue and red colored nodes, respectively) of the host proteins (square- shaped dark blue colored nodes) targeted by SARS-CoV-2 viral proteins (diamond-shaped green colored nodes). HiPPIP predicted ~2,600 PPIs of which ~600 PPIs were previously known, leaving ~2,000 PPIs to be considered as novel PPIs of the host proteins. There were an additional 3,500 PPIs that were known and not predicted by HiPPIP. This is as expected as the HiPPIP prediction threshold has been fixed to achieve *high precision* by compromising *recall*, which is required for adoption into biology; in other words, it is set to predict only a few PPIs out of the hundreds of thousands of unknown PPIs, but those will be highly accurate. It has to be noted that neither PPI prediction nor high throughput PPI screening can be performed with high-precision *and* high-recall. Co-IP based methods show high-precision and extremely-low recall

(detecting only one PPI at a time), whereas multi-screen high-quality yeast 2-hybrid methods show high-precision with low recall (detecting a few tens of thousands of PPIs). Thus, HiPPIP is on par with other methods in terms of precision and the number of new PPIs detected. Recently, state-of-the-art algorithms that were developed after HiPPIP have been extensively evaluated, but none of them reached the superior performance achieved by HiPPIP (Dunham and Ganapathiraju, 2022). Seventeen novel PPIs predicted by HiPPIP in our other studies were tested, and all validated to be true; the experiments were carried out by diverse research labs. Overall, the host protein (HoP) interactome consisted of 4,408 proteins and 6,076 interactions. A partial network of host proteins and their novel interactors is shown in Figure 3A. Several COVID-centric network biology studies (Zhou et al., 2020b; Gysi et al., 2020; Kumar et al., 2020) presented analysis of the "known PPI neighborhood" of the host proteins targeted by SARS-CoV-2. Contrary to this, in our study, we augment this neighborhood with 1,941 computationally predicted PPIs of high precision (Figure 2), so as to 1) present a more complete version of the host protein interactome, 2) facilitate discovery of previously unknown disease mechanisms, and 3) allow characterization of under-
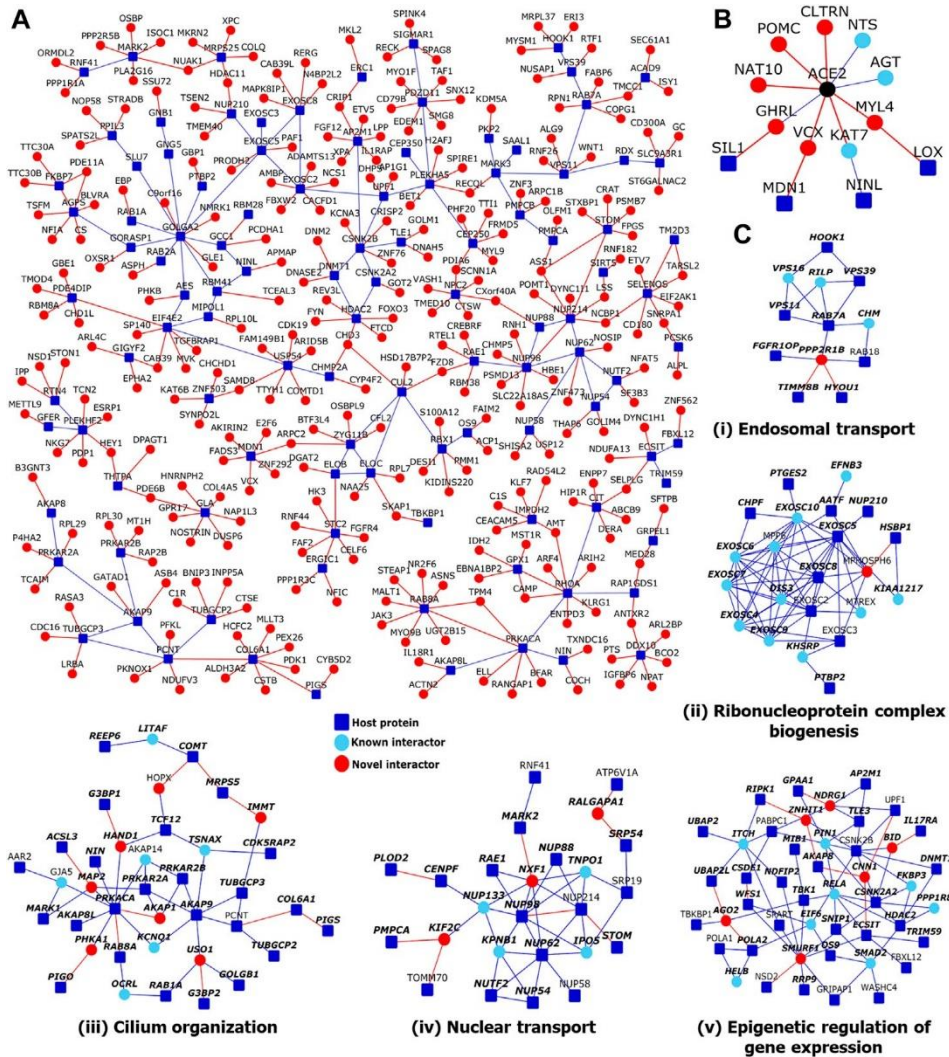
FIGURE 3 | Network views of protein-protein interactions in the host protein interactome: (A) Partial view of the HoP interactome: Genes are shown as nodes and PPIs as edges. As the full network is very large, only a partial view showing a large connected component of novel interactors and their neighbors, all of which have transcriptomic/proteomic evidence related to SARS-CoV-2 (Supplementary Table S2), is shown. Legend: Dark blue square-shaped nodes: host proteins targeted by SARS-CoV-2; red nodes/edges: novel interactors/interactions; light blue nodes and blue edges: known interactors and interactions. (B) ACE2 interactome: PPIs of ACE2 protein, extended to show four host proteins that are two-edges away from it. Color legend is as in (A-C) Modules identified from network topology: Five out of seventeen total modules each with 3 or more nodes are shown. Each module was enriched in a specific GO biological process: (i) Endosomal transport, (ii) Ribonucleoprotein complex biogenesis, (iii) Cilium organization, (iv) Nuclear transport and (v) Epigenetic regulation of gene expression. Within each module shown in (i-v), nodes with bold italicized labels depict genes with at least one transcriptomic/proteomic evidence relevant to SARS-CoV-2.

studied host proteins through functional associations of their predicted interactors. Moreover, the network is made available on an interactive webserver to enable biologists to examine the novel interactions relevant to their specific protein or pathway of interest, and as downloadable files in various formats to facilitate its investigation in conjunction with transcriptomic/ proteomic data by computational systems biologists.

We verified whether any of the 2,000 novel PPIs came up in recently released interactome maps such as HuRI (HI-Union) (Luck et al., 2020) and BioPlex (Huttlin et al., 2020). While there was no overlap with the HI-union dataset, there were 8 PPIs in the BioPlex map (ADAM9-ADAM32, P3H3-OS9, PVR-NECTIN2,

SRRM2-SNIP1, PABPC4-LUC7L2, PRKACA-AKAP1, NDUFA13-ECSIT, and NPTX1-NPTX2). The small overlap is not surprising because even high-throughput biotechnological methods discover different parts of the interactome with only small overlaps with each other, thus demonstrating complementary strengths (Luck et al., 2020).

2.1 Wiki-CORONA: A Web Server of Novel Host PPIs

The HoP interactome is available on a website called Wiki-CORONA (http://severus.dbmi.pitt.edu/corona/). It has advanced

155

search capabilities, and presents comprehensive annotations, namely Gene Ontology, diseases, drugs and pathways, of the two proteins of each PPI side-by-side. Here, a user can query for results such as "PPIs where one protein is anti-viral and the other is involved in immunity," and then see the results with the functional details of the two proteins side-by-side. The PPIs and their annotations also get indexed in major search engines like Google and Bing. Querying by biomedical associations is a unique feature which we developed in Wiki-Pi that presented known interactions of human proteins (Orii and Ganapathiraju, 2012).

2.2 Interconnections of ACE2 With Host Proteins Targeted by SARS-CoV-2

SARS-CoV-2 engages the host receptor ACE2 (*angiotensin-converting enzyme 2*) for cell entry (Hoffmann et al., 2020). The interactions of SARS-CoV-2 viral proteins with host proteins were studied by Gordon et al. in human embryonic kidney cells (HEK-293T/17) (Gordon et al., 2020a), which show very low endogenous expression of ACE2 (Warner et al., 2005); even if HEK-293 cells were transfected with ACE2 to allow heterologous ACE2 expression, its protein product may undergo proteolytic cleavage mediated by ADAM17 (Lambert et al., 2005). Possibly due to this reason, ACE2 was not identified as a host protein in that study (Gordon et al., 2020b). Therefore, we assembled the known and novel PPIs of ACE2 separately, owing to its crucial role in SARS-CoV-2 infection. Then, we extracted the shortest paths in the interactome connecting ACE2 to any of the 332 host proteins using methods described in our prior work, LENS (Lens for Enrichment and Network Studies of human proteins) (Handen and Ganapathiraju, 2015). We found that ACE2 was connected to 4 host proteins targeted by SARS- CoV-2 (SIL1, LOX, MDN1 and NINL) through an intermediate interactor, i.e. separated by two edges, where one or both intermediary PPIs were novel predicted ones (see red edges in Figure 3B). Thus, we showed that novel PPIs connect ACE2 to multiple host proteins through intermediary proteins.

These connections revealed interesting insights: ACE2 is a key player of the renin-angiotensin hormone system that regulates blood pressure and electrolyte balance (Burrell et al., 2004). In line with this, we found that its interactors, AGT (*angiotensin*), GHRL, CLTRN and POMC, were associated with the Reactome Pathway *peptide hormone metabolism* (p-value = 2.9×10^{-5}). ACE2 and its interactors were also enriched in the Gene Ontology (GO) Biological Process *circulatory system process* (ACE2, AGT, NTS, POMC, GHRL and the host protein MYL4; p-value = 1×10^{-3}). Three host proteins were associated with numerous vascular and cardiac phenotypes: LOX with abnormality of blood volume homeostasis, aortic root aneurysm, ascending aortic dissection, carotid artery dilatation, coronary artery atherosclerosis, cystic medial necrosis of the aorta, descending thoracic aorta aneurysm, dilatation of the cerebral artery, left ventricular failure, peripheral arterial stenosis, MYL4 with paroxysmal atrial fibrillation and bradycardia, and SIL1 with abnormal aldolase level. The co-morbidity of hypertension, diabetes and cardiovascular diseases among the group of COVID-19 patients with high fatality rate (Fang et al., 2020a)

warrants a closer look at ACE2 and the host proteins linked to cardiac and vascular phenotypes. We also examined the interconnections of the host proteins with other proteins that facilitate SARS-CoV-2 entry into host cells, namely, TMPRSS2, CTSB, CTSL, NRP1, AGTR2 and OR51E2 (Cantuti-Castelvetriet al., 2020; Cui et al., 2020; Hoffmann et al., 2020; Kerslake et al., 2020). Five out of these 6 proteins—TMPRSS2, CTSB, CTSL, NRP1, and AGTR2—were found to be connected to 33 host proteins *via* 52 intermediate interactors including 12 novel interactors (Supplementary Figure S1). Detailed investigations may be necessary to understand the relationships of these host cellular entry proteins to other host factors targeted by SARS-CoV-2.

2.3 Identification of Network Modules From the Host Protein Interactome

Viruses have been shown to target network modules of host proteins (Jäger et al., 2012; Hafirassou et al., 2017; Yang et al., 2019). These modules could either correspond to 1) protein complexes in which proteins interact within a specific location/time/condition to perform a function in a coordinated manner (e.g., RNA splicing machinery and transcription machinery), or 2) to form dynamic, yet functionally coherent units, in which the proteins interact with one another at different times/conditions to carry out a biological process (e.g., signaling pathways and cell cycle regulation) (Spirin and Mirny, 2003). We employed "Netbox" software implementation with data consisting of SARS-CoV-2 target proteins (core proteins) and all human PPIs (Cerami et al., 2010) to identify network modules. It expands the core proteins by adding nodes from the interactome whose number of links to core proteins are statistically significant compared to its degree in the human interactome. From this network, it identifies highly interconnected modules. It was able to connect 323 proteins (220 host proteins targeted by SARS-CoV-2 and 103 linker proteins) into 21 modules, of which 14 modules had 4 or more nodes each. For comparison, when novel PPIs were not included, it connected 199 proteins (138 host proteins and 61 linker proteins) into 18 modules of which 10 had 4 or more proteins each. Scaled modularity score (*Z*-score compared to random networks) was 17.0 with novel PPIs, and it was 14.5 without novel PPIs (*Z*-score compared to corresponding random networks). Five modules formed with novel interactors had statistically significant enrichment of GO biological process terms: *epigenetic regulation of gene expression* (p-value = 3.3×10^{-4}, odds ratio = 10.4), *nuclear transport* (p-value = 2.4×10^{-12}, odds ratio = 21.6), *cilium organization* (p-value = 1.28×10^{-3}, odds ratio = 7.8), *ribonucleoprotein complex biogenesis* (p-value = 0, odds ratio = 22.4), and *vesicle-mediated transport between endosomal compartments* (p-value = 9.4×10^{-6}, odds ratio = 123.4) (Figures 3Ci–v). When novel PPIs were excluded, some of these associations were missed (*epigenetic regulation of gene expression* and *cilium organization*) and the modules were smaller, but 3 additional functional modules were found: *cell cycle G2/M phase transition* (p-value = 1.9×10^{-3}, odds ratio = 21.7), *DNA replication* (p-value = 4.9×10^{-3}, odds ratio = 55.25)

and *cell-cell signaling by Wnt* (p-value = 4.9×10^{-3}, odds ratio = 9.3). Hence, although several biological processes detected by including the novel PPIs could also be detected using only the known PPIs, functional modules such as *cilium organization* were only uncovered on inclusion of the novel PPIs that we predicted for the host proteins (namely, COMT-HOPX, MRPS5-IMMT, G3BP1-HAND1, ACSL3-MAP2, PRKACA-AKAP1, PIGO-PHKA1 and G3BP2-USO1). In summary, the novel PPIs improved existing COVID-related knowledge by facilitating the identification of functional modules, which would have remained hidden if one had only used known PPIs for module identification.

2.4 Overlap of the Host Protein Interactome With Transcriptome and Proteome Data

We systematically analyzed the overlap of the HoP interactome with gene expression profiles induced by SARS-CoV and SARS-CoV-2. Statistically significant overlaps were found with the genes differentially expressed in A549 (human lung alveolar carcinoma) cell lines transfected with ACE2 and infected with a high load of SARS-CoV-2 (multiplicity of infection/MOI = 2.0) (p-value = 3.67×10^{-17}, odds ratio = 1.26), Calu-3 (human lung epithelial carcinoma) cell line infected with a high SARS-CoV-2 load (MOI = 2.0) (p-value = 1.98×10^{-3}, odds ratio = 1.12) and postmortem lung samples of COVID-19 positive patients (p-value = 8.3×10^{-17}, odds ratio = 1.37) [GSE147507 (Blanco-Melo et al., 2020b)]. Significant enrichment of the novel interactors that were predicted to interact with the host proteins targeted by SARS-CoV-2 was noted in the A549 cell line (p-value = 1.6×10^{-3}, odds ratio = 1.17) and COVID-19 patient (p-value = 1.18×10^{-2}, odds ratio = 1.19) datasets. Many proteins in the interactome, including novel interactors, were differentially expressed in epithelial cells infected with SARS-CoV (GSE17400, Calu-3 cell, 48 h post-infection; p-value = 4.76×10^{-12}). Several proteins also showed differential expression after infection by Urbani strain of SARS-CoV (GSE37827, Calu-3 cells, 72 h post-infection), in peripheral blood mononuclear cells of SARS patients (GSE1739 (Reghunathan et al., 2005)), in A549 cell line infected with a low SARS-CoV-2 load (MOI = 0.2) and in NHBE (normal human bronchial epithelial) cell line infected with a high SARS-CoV-2 load (MOI = 2.0), but their overlaps were not statistically significant. Most importantly, the interactome demonstrated statistically significant overlaps with the genes differentially expressed in the leukocytes of COVID-19 patients with ARDS admitted to the intensive care unit (ICU) versus those receiving non-intensive care (p-value = 4.63E-10, odds ratio = 1.13) (GSE157103 (Overmyer et al., 2021)) and whole blood of COVID-19 patients critical in ICU with ARDS versus non-critical patients on oxygen [p-value = 0.035, odds ratio = 1.04) (GSE172114 (Carapito et al., 2021)). This suggested that the HoP interactome can be used as a framework to contextualize the gene expression signatures differentiating the various clinical outcomes of COVID-19. Additionally, we showed the overlap of the interactome with genes differentially expressed in blood samples of COVID-19 patients admitted to the ICU with ARDS compared with non-critical patients on oxygen (p-value

= 6.03×10^{-7}, odds ratio = 1.11) (GSE172114 (Carapito et al., 2021)) and in peripheral blood mononuclear cells (PBMCs) of COVID-19 patients in ICU versus healthy subjects (p-value = 1.55×10^{-27}, odds ratio = 1.25) and COVID-19 patients with moderate symptoms versus healthy subjects (p-value = 5.08×10^{-30}, odds ratio = 1.35) (GSE152418 (Arunachalam et al., 2020)). Statistically significant enrichments for novel interactors were also found with PBMCs of ICU-admitted COVID-19 patients (p-value = 1.69×10^{-5}, odds ratio = 1.18) and those with moderate symptoms (p-value = 8.01×10^{-4}, odds ratio = 1.18). In summary, the overlaps of the transcriptional profiles induced by SARS-CoV/SARS-CoV-2 with the HoP interactome 1) ascertained the biological validity of the HoP interactome, 2) contextualized the differentially expressed genes within the mechanistic framework of the protein interactome and 3) highlighted novel interactors of the host proteins targeted by SARS-CoV-2 that may be prioritized for further study.

Melo et al. had identified 120 genes differentially expressed upon infection by SARS-CoV-2 in the A549 cell line compared with infection by respiratory syncytial virus and/or influenza A virus (GSE147507 (Blanco-Melo et al., 2020a)). Of these, only 2 differentially expressed genes (DEGs) were among the 332 host proteins targeted by SARS-CoV-2. Our study revealed several interesting links between host proteins and these DEGs (Figure 4A): 1) although only 2 DEGs were found among the host proteins themselves, 31 DEGs were *direct interactors* of 38 host proteins, with some DEGs interacting with multiple host proteins; 2) 13 novel PPIs existed between the two sets: AAR2-SAMHD1, TUBGCP2-C1R, IMPDH2-C1S, GOLGA7-TCIM, RAB8A-STEAP1, GDF15-EHF, REEP5-PDK4 FAM162A-PARP14, STOML2-CDH1, FGA-RAB14, FBXL12-C19orf66, ECSIT-C19orf66 and EIF4H-PTPN12; 3) 108 DEGs and 285 host proteins connected to each other *via* a common interactor (there were 808 such shared interactors between DEGs and host proteins; statistically significant overlap with odds ratio = 1.5, p-value = 7.12×10^{-54}); 4) Pathway enrichment analysis of overlapping interactome (consisting of 808 shared interactors, and the DEGs and the host proteins that they interact with) revealed enrichment of several immune-related pathways (each with FDR-corrected p-value < 0.05).

Messner et al. had identified 27 protein biomarkers whose expression varied according to the WHO severity grades for COVID-19 infection (i.e. no oxygen support, with oxygen support and critical) (Messner et al., 2020). Out of these, 11 biomarkers were identified in our study as interactors of the host proteins targeted by SARS-CoV-2. This included 8 proteins, ACTB, C1R, C1S, CD14, FGA, GSN, ITIH3 and SAA1, which were predicted as novel interactors of the host proteins, TCF12, RALA, TUBGCP2, IMPDH2, REEP5, RAB14, RHOA and GNG5.

Next, we considered the overlap between 65 host proteins that were identified to interact with SARS-CoV proteins by Pfeffer et al. (2011) and the host proteins targeted by SARS-CoV-2. Only 4 proteins were common to both (BZW2, MARK2, MARK3 and SMOC1) (Figure 4B). However, the interactome revealed that 50 host proteins targeted by SARS-CoV-2 had direct interactions with 32 host proteins targeted by SARS-CoV, and that 8 of these
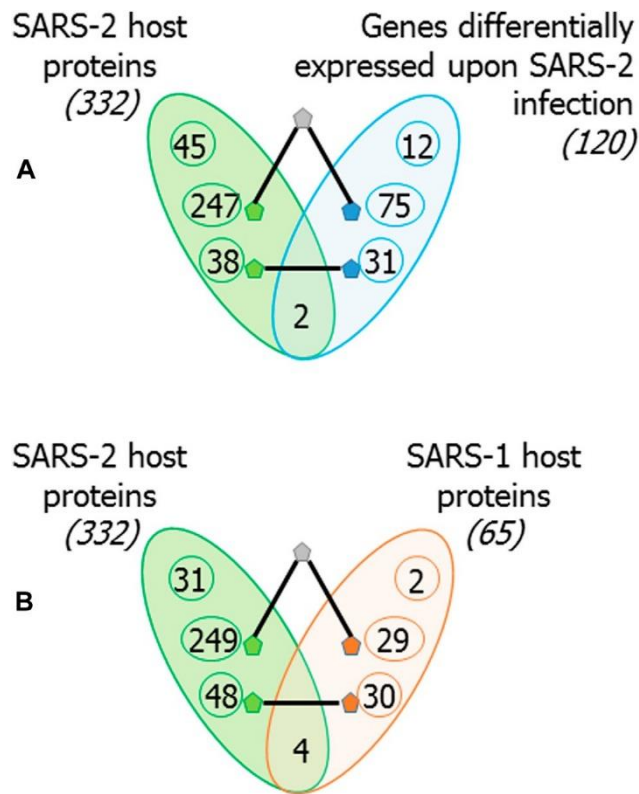
FIGURE 4 | Overlap of the host proteins targeted by SARS-CoV-2 with genes differentially expressed upon SARS-CoV-2 infection and with host proteins targeted by SARS-CoV: (A) 120 genes were identified to be differentially expressed in A549 cell line upon SARS-CoV-2 infection compared with infection by respiratory syncytialvirus and/or influenza virus. Of these, 2 are also host proteins targeted by SARS-CoV-2, 31 have direct PPIs with SARS-CoV-2 host proteins and 75 link to SARS-CoV-2host proteins via an intermediate interactor. Thus, 108 out of 120 genes differentially expressed upon SARS-CoV-2 infections are closely connected to SARS-2 host proteins. Note that only closest connections are shown; i.e., a directly interacting DEG may also be connected via an intermediate interactor to host proteins, but it is counted among former. (B) Similar counts are shown for 65 host proteins targeted by SARS-CoV: 4 are common, 30 have direct PPIs, and 29 link to SARS-CoV-2 host proteins via an intermediate interactor. Thus, 63 out of 65 host proteins targeted by SARS-CoV are closely connected to host proteins targeted by SARS-CoV-2.

were novel PPIs (N4BP2L2-EXOSC8, NMB-MRPS5, MKRN2-MRPS25, HOXC6-BRD2, XPA-AP2M1, VKORC1- DCTPP1, RSRP1-CEP350 and TPSAB1-ADAMTS1). 29 host proteins targeted by SARS-CoV were connected to 249 host proteins targeted by SARS-CoV-2 *via* a common interactor.

GO biological process terms such as *autophagic mechanism* (odds ratio = 4.5, *p*-value = 2.2×10^{-5}) *regulation of mitochondrion organization* (odds ratio = 7.5, *p*-value = 5.5×10^{-5}) and *protein localization to mitochondrion* (odds ratio = 7.8,*p*-value = 3.74×10^{-4}) were enriched in the overlapping interactome, suggesting that these processes may be commonly targeted by both these viruses. Mitochondria may be directly targeted by viral proteins, and may be affected by the cellular changes arising from viral infection. They may also play a crucial role in viral pathogenesis due to their function as immune signalling hubs (Khan et al., 2015). These organelles are constantly eliminated and recycled through a process called mitophagy. Viruses can modulate mitochondrial function and mitophagy to exacerbate infection (Khan et al., 2015).

In summary, 63 out of the 65 host proteins targeted by SARS-CoV, and 108 out of the 120 genes differentially expressed upon

SARS-CoV-2 infection interacted directly or through an intermediate interactor with the host proteins targeted by SARS-CoV-2 (Figure 4).

3,787 (86%) proteins in the interactome are supported by theabove mentioned transcriptomic and proteomic evidence, and arelisted in Supplementary Table S2. In fact, the selected novel interactors shown in Figure 2A all have transcriptomic/proteomic evidence.

We studied tissue-specific expression of the proteins in the interactome using GTEx data (Lonsdale et al., 2013). Genes with an expression level greater than 1 TPM (transcripts per million) and relative expression at least 5-fold higher in a particular tissue (tissue-enriched) or a group of 2-7 tissues (group-enriched) were considered (Fagerberg et al., 2014). As expected, many genes showed specific expression in lung, which is the target tissue of the virus, and in spleen, which regulates the immune response of the host (Figure 5). Host proteins targeted by SARS-CoV-2 had novel PPIs with 37 lung-specific proteins and 49 spleen-specific proteins. Apart from these expected tissue associations, we noted that the host proteins also had novel PPIs with 61 brain-specific and 28 heart-specific proteins, which

FIGURE 5 | Tissue-specific genes in the host protein interactome: Number of genes from the interactome which show tissue specificity are shown. The genes show at least 5-fold higher expression in a tissue or a group of 2-7 tissues compared to all the other tissues.
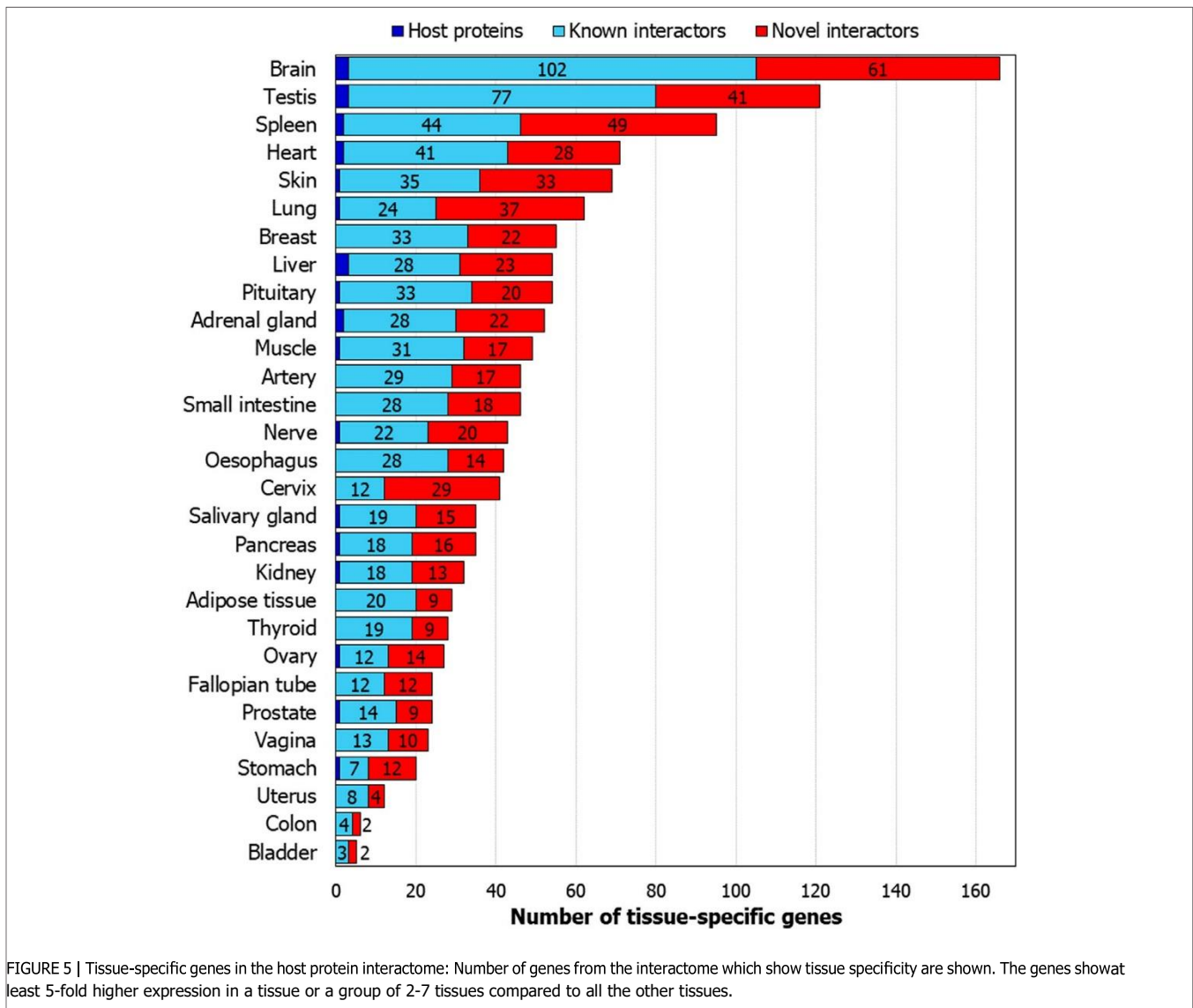
is of importance as cerebrovascular diseases and coronary heart diseases are co-morbidities among COVID-19 non-survivors (Figure 5) (Fang et al., 2020b).

2.5 Functional Enrichment Analysis of the Host Protein Interactome

We identified functional associations of the HoP interactome using the gene set analysis toolkit called WebGestalt (Liao et al., 2019). WebGestalt computes enrichment of specific functional groups (e.g., a Reactome Pathway) in an input list (e.g., genes in the HoP interactome). Statistical significance is computed using Fisher's exact test, and corrected using the Benjamini–Hochberg method for multiple test adjustment. WebGestalt was chosen for its user-friendly interface, intuitive plots, large collection of functional categories from different types of functional databases and multiple enrichment analysis methods. This

analysis yielded information from different biological levels that may potentially be influenced by SARS-CoV-2 infection:

(1) *Co-morbidity relationships:* proteins encoded by the genes associated with diabetes and hypertension showed network proximity to the host proteins targeted by SARS-CoV-2

(2) *Subcellular locations:* PML (promyelocytic leukaemia) bodies and the midbody may function as subcellular targets of SARS-CoV-2, since proteins localizing to these structures were found to be significantly enriched in the HoP interactome

(3) *Cellular processes:* enrichment of proteins involved in cell cycle phase transitions may allude to SARS-CoV-2 modulating critical junctures in the host cell cycle to facilitate viral infection

(4) *Cellular pathways:* the post-transcriptional tristetraproline-mediated regulatory pathway is significantly associated with the interactome and may be targeted by SARS-CoV-2 proteins to weaken host immune response

FIGURE 6 | Network proximity of diabetes and hypertension to host proteins targeted by SARS-CoV-2: Dark blue nodes are host proteins targeted by SARS-CoV- 2, light blue nodes are known interactors and red nodes are novel interactors. Nodes with bold black italicized labels are diabetes-associated genes, whereas nodes withbold orange colored italicized labels are hypertension-associated genes.

2.6 Co-Morbidity Relationships

We studied the association of interactome genes with any genetic disorders/traits in the OMIM database. 155 genes in the interactome, including 9 host protein-encoding genes, and 121 known interactors and 25 novel interactors of host proteins, were found to be associated with 35 disorders (overlap of each disease had p-value < 0.05). This included 13 types of cancers, 7 metabolic disorders, 4 neurological disorders, 3 developmental disorders, 2 eye-related disorders, 2 vascular diseases, 1 infectious disease, 1 inflammatory disorder, 1 respiratory disorder and 1 skin disease (Figure 6; Table 1). Some of these diseases enriched in the interactome are co-morbidities among non-survivors and critically ill COVID patients (e.g., diabetes, hypertension, cerebrovascular events and cancer) (Fang et al., 2020b; Sidaway, 2020). 13 genes in the interactome were associated with *non-insulin dependent diabetes mellitus* (odds ratio = 10.8, p-value = 4.38×10^{-10}), 6 genes with *essential hypertension* (odds ratio = 12, p-value = 2.34×10^{-5}), 3 genes with *ischemic stroke* (odds ratio = 14.4, p-value = 1.7×10^{-3}) and 10 genes with *lung cancer* (odds ratio = 14.1, p-value = 2.36×10^{-9}).

Network proximity of the proteins associated with these co-morbid conditions to the SARS-CoV-2 host proteins may explain why patients with these conditions are increasingly affected by the viral infection. Further investigations are necessary to dissect these co-morbidities. Treatment strategies that prevent the deterioration of the underlying genetic conditions must be devised to combat COVID-19 in susceptible individuals. Additionally, neurological disorders such as *Alzheimer's disease* (odds ratio = 15.3, p-value = 5.13×10^{-7}) and *schizophrenia* (odds ratio = 12, p-value = 4.19×10^{-6}) were also found to be enriched in the interactome, warranting further investigations intothese potential co-morbidities.

2.7 Subcellular locations

Gene Ontology enrichment analysis of the interactome identified several subcellular locations that may be targeted by SARS-CoV-2. Cellular locations included points of virus entry such as the *cell-substrate junction*, *nuclear periphery* and specific sites fromwhere viral proteins may potentiate viral replication, gene

TABLE 1 | List of OMIM diseases enriched in the interactome: Details of enrichment of the disease genes in the interactome including the number of disease-associated genes, odds ratio and statistical significance (p-value) of enrichment are shown.

OMIM disease	Number of disease genes in the interactome	Odds ratio of enrichment	p-value of overlap	Genes
BREAST CANCER	15	15.033683	5.86E-14	PPM1D, NQO2, RB1CC1, AKT1, ATM, BRCA1, CASP8, CDH1, ESR1, RAD51, CHEK2, BARD1, KRAS, TSG101, TP53
LEUKEMIA, ACUTE MYELOID	12	13.120305	2.26E-10	LPP, NSD1, RUNX1, CBFB, DNMT3A, CHIC2, GATA2, KRAS, TERT, NPM1, NSD3, NUP214
COLORECTAL CANCER	10	17.181352	2.26E-10	AXIN2, AKT1, APC, CTNNB1, FGFR3, EP300, NRAS, BUB1B, TP53, SRC
DIABETES MELLITUS, NONINSULIN-DEPENDENT	13	10.782779	4.38E-10	MAPK8IP1, GCK, AKT2, HMGA1, IRS1, KCNJ11, MTNR1B, PPARG, PTPN1, SLC2A2, HNF1A, TCF7L2, WFS1
LUNG CANCERALVEOLAR CELL CARCINOMA, INCLUDED	10	14.149348	2.36E-09	BRAF, CASP8, MAP3K8, CYP2A6, RASSF1, EGFR, ERBB2, KRAS, PRKN, PPP2R1B
PROSTATE CANCER	8	14.802395	8.97E-08	MAD1L1, AR, CDH1, KLF6, CHEK2, ZFHX3, PTEN, CD82
ALZHEIMER disease	7	15.30722	5.13E-07	A2M, APP, BLMH, ACE, PLAU, NOS3, PAXIP1
SCHIZOPHRENIA	7	12.026946	4.19E-06	MTHFR, AKT1, COMT, DRD3, RTN4R, SYN2, DISC1
LEUKEMIA, ACUTE LYMPHOBLASTIC	5	17.181,352	1.69E-05	TAL1, TAL2, GNB1, BCR, NUP214
HYPERTENSION, ESSENTIAL	6	12.026946	2.34E-05	AGTR1, ADD1, ATP1B1, GNB3, NOS2, NOS3
HEPATOCELLULAR CARCINOMA	5	13.363273	7.09E-05	ET, APC, CASP8, CTNNB1, TP53
THYROID CANCER, NONMEDULLARY, 2	4	19.243114	7.09E-05	SRGAP1, HRAS, NRAS, PTEN
LEUKOENCEPHALOPATHY WITH VANISHING WHITE MATTER	4	19.243114	7.09E-05	EIF2B2, EIF2B1, EIF2B5, EIF2B3
OBESITYLEANNESS, INCLUDED	6	9,6,215,569	8.15E-05	ADRB3, CARTPT, GHRL, ADRB2, MC4R, PPARG
MALARIA, SUSCEPTIBILITY TOMALARIA, RESISTANCE TO, INCLUDED	6	8.489609	1.49E-04	CISH, FCGR2B, HBB, NOS2, CD36, TNF
MENINGIOMA, FAMILIAL, SUSCEPTIBILITY TO	4	16.035928	1.49E-04	NF2, PDGFB, PTEN, SMARCE1
BECKWITH-WIEDEMANN SYNDROME	4	16.035928	1.49E-04	NSD1, CDKN1C, KCNQ1, IGF2
OVARIAN CANCEROVARIAN CANCER, EPITHELIAL, INCLUDED	4	16.035928	1.49E-04	AKT1, CDH1, CTNNB1, PRKN
HYPERCHOLESTEROLEMIA, FAMILIAL	4	13.745081	3.18E-04	ABCA1, APOA2, GHR, PPP1R17
MITOCHONDRIAL COMPLEX I DEFICIENCY	6	7.2161677	3.71E-04	NDUFAF3, TMEM126B, NDUFAF2, NDUFAF1, NDUFA1, NDUFB9
PARKINSON disease, LATE-ONSET	4	12.026946	5.31E-04	SNCAIP, MAPT, ATXN2, TBP
GASTRIC CANCERGASTRIC CANCER, INTESTINAL, INCLUDED	4	12.026946	5.31E-04	CASP10, APC, KLF6, ERBB2
DIABETES MELLITUS, PERMANENT NEONATAL	3	14.432335	0.001751964	GCK, INS, KCNJ11
JUVENILE MYELOMONOCYTIC LEUKEMIA	3	14.432335	0.001751964	CBL, ARHGAP26, PTPN11
STROKE, ISCHEMIC	3	14.432,335	0.001751964	ALOX5AP, F2, NOS3
LYMPHOMA, NON-HODGKIN, FAMILIAL	3	12.026946	0.003145343	CASP10, RAD54B, BCL10
WILMS TUMOR 1	3	12.026946	0.003145343	GPC4, IGF2, WT1
PHEOCHROMOCYTOMA	3	9.0202,096	0.007708062	RET, MAX, VHL
RHEUMATOID ARTHRITIS	3	9.0202,096	0.007708062	SLC22A4, CIITA, PADI4
RETINITIS PIGMENTOSA	3	7.2161677	0.015013329	PDE6G, CRX, ARL6
ASTHMA, SUSCEPTIBILITY TO	3	6.5601524	0.018768223	ADRB2, PHF11, TNF
PRADER-WILLI SYNDROME	3	6.5601524	0.018768223	NDN, MKRN3, MAGEL2
TRACHEOESOPHAGEAL FISTULA WITH OR WITHOUT ESOPHAGEAL ATRESIA	4	4.5816938	0.019809755	CHD7, FANCC, FANCF, MYCN
ENDOMETRIAL CANCER	2	9.6215569	0.028209828	CDH1, PTEN
EPIDERMOLYSIS BULLOSA, JUNCTIONAL, NON-HERLITZ TYPE	2	9.6215569	0.028209828	LAMC2, LAMA3
WOLF-HIRSCHHORN SYNDROME	2	9.6215569	0.028209828	CTBP1, NSD2
MACULAR DEGENERATION, AGE-RELATED, 1	2	9.6215569	0.028209828	PLEKHA1, APOE
DIABETES MELLITUS, INSULIN-DEPENDENT	2	8.0179641	0.040083771	ITPR3, HNF1A

expression and modulate the immune response of the host such as the *midbody*, *nuclear chromatin* and *PML* (promyelocytic leukaemia) *body* (each term with *p*-value $< 1 \times 10^{-4}$). PML bodies are nuclear sub-compartments that repress viral replication through entrapment or epigenetic silencing of the

viral genomes (Scherer and Stamminger, 2016). Components of PML bodies activate interferon-stimulated genes and cytokines, and may also be upregulated on induction of interferons (Scherer and Stamminger, 2016). Viruses have been known to target PML bodies to circumvent the anti-viral defences of the host cell

FIGURE 7 | Pathways associated with the host protein interactome: Number of genes from the interactome associated with selected pathways are shown.
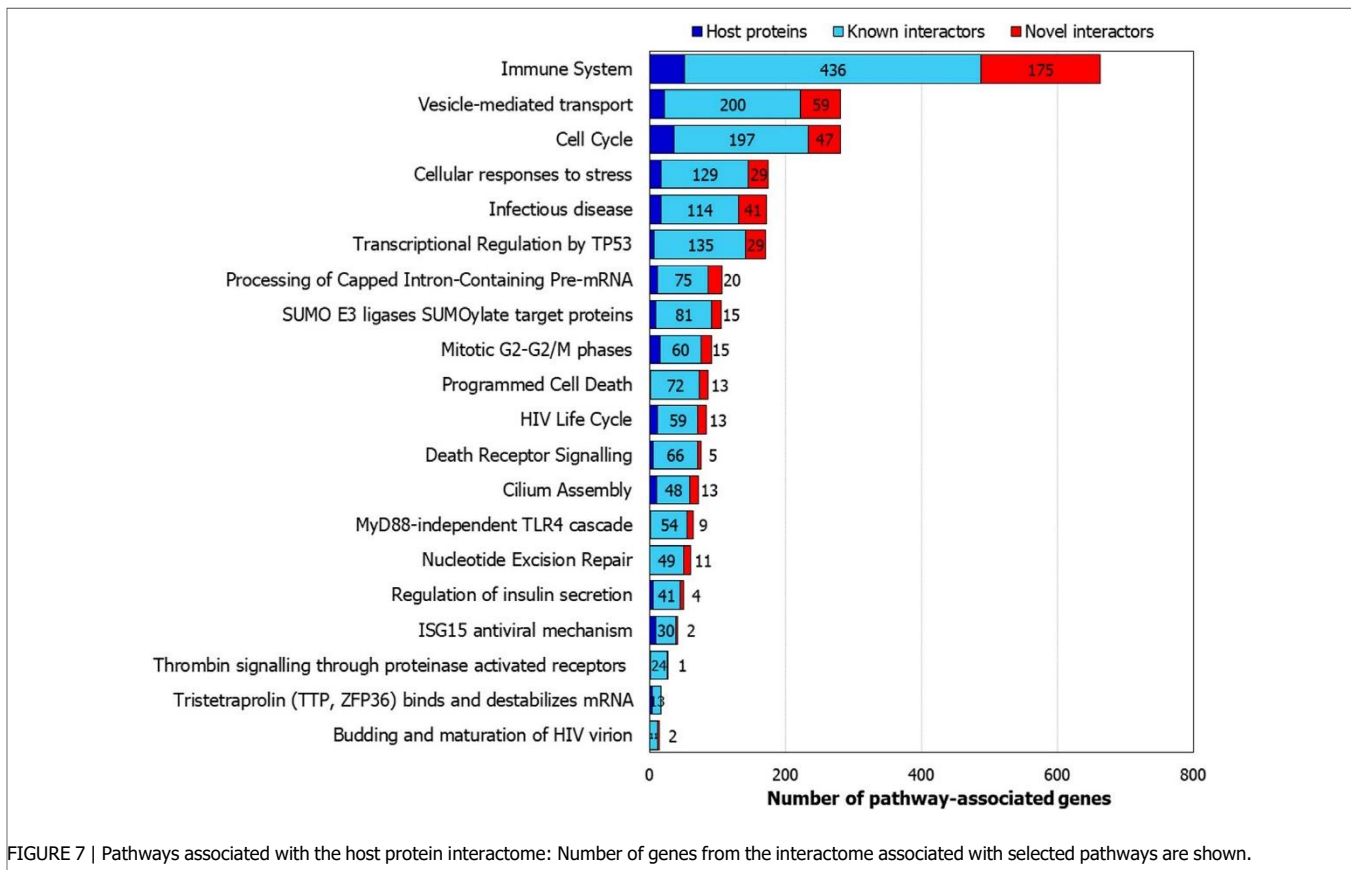
(Scherer and Stamminger, 2016). 61 proteins in the HoP interactome are PML components. These include the host protein AKAP8L, which has been known to promote retroviral gene expression, and 55 known interactors and 5 novel interactors (RNF111, SP140, ELF4, NFE2, and CIART) of other host proteins targeted by SARS-CoV-2. Our model predicted an interaction of EIF4E2 with SP140, an interferon- inducible PML component; SARS-CoV-2 may target these proteins. The midbody is a microtubule-rich structure that connects the daughter cells and marks the site of abscission during cytokinesis. Viruses have been known to recruit certain protein complexes that also localize to the midbody during cytokinesis, to the host cell membrane to promote its scission and thereby the release of viruses (Morita et al., 2010). This co-opting of proteins may explain the enrichment of midbody proteins in the HoP interactome. 83 proteins in the HoP interactome, including 11 host proteins (RHOA, CENPF, CIT, RAB8A, NUP62, SCCPDH, SPART, RDX, ARF6, CNTRL and RALA), 63 known interactors and 9 novel interactors (KIF4A, BIRC5, INCENP, ALKHB4, DNM2, DDX11, ARL2BP, ABRAXAS2 and WIS) are known to localize to the midbody.

2.8 Cellular processes

Enriched biological processes in the interactome included *G1/S and G2/M mitotic cell cycle phase transitions*, *regulation of vesicle-mediated transport*, *covalent chromatin modification* and *nuclear transport* (p-value < 1×10^{-4}). The response of

the host cell to SARS-CoV-2 infection has been shown to be significantly delayed and devoid of several anti-viralmechanisms (Blanco-Melo et al., 2020a). During early stagesof the infection, it is possible that the virus induces a G1/S phase transition to surreptitiously synergize the replication of the viral genome with that of the host genome (Fan et al.,2018). In the later stages, it may block the G2/M phase transition to maximise the levels of viral genome (Fan et al., 2018). We found novel interactions of host proteins with 34 proteins involved in cell cycle phase transition: ANAPC4, ANAPC7, ARPP19, CCNB3, CDC14B, CDC16, CDC7, CEP164, CETN2, CLSPN, CRLF3, DCTN1, DNM2, DYNC1H1, E2F6, ENSA, FBXL7, GFI1, GML, HYAL1, INHBA, JADE1, NEUROG1, NPAT, ORC2, PPM1D, RAD17, SPDYA, TAOK2, TICRR, TRIAP1, XPC, ZFP36L1 and ZNF655. Corroborating our hypothesis, a significantly large number of Vero E6 cells infected with SARS-CoV-2 were found to be in S and the G2/M phases, indicating that SARS- CoV-2 may induce cell cycle arrest between S and G2 phases to promote infection (Bouhaddou et al., 2020).

2.9 Cellular pathways

The HoP interactome showed a statistically significant enrichment of several pathways related to viral entry and infection such as *infectious disease*, *HIV life cycle*, *vesicle-mediated transport* and *membrane trafficking* (Figure 7). Several immunity-related pathways that mediate host response

162
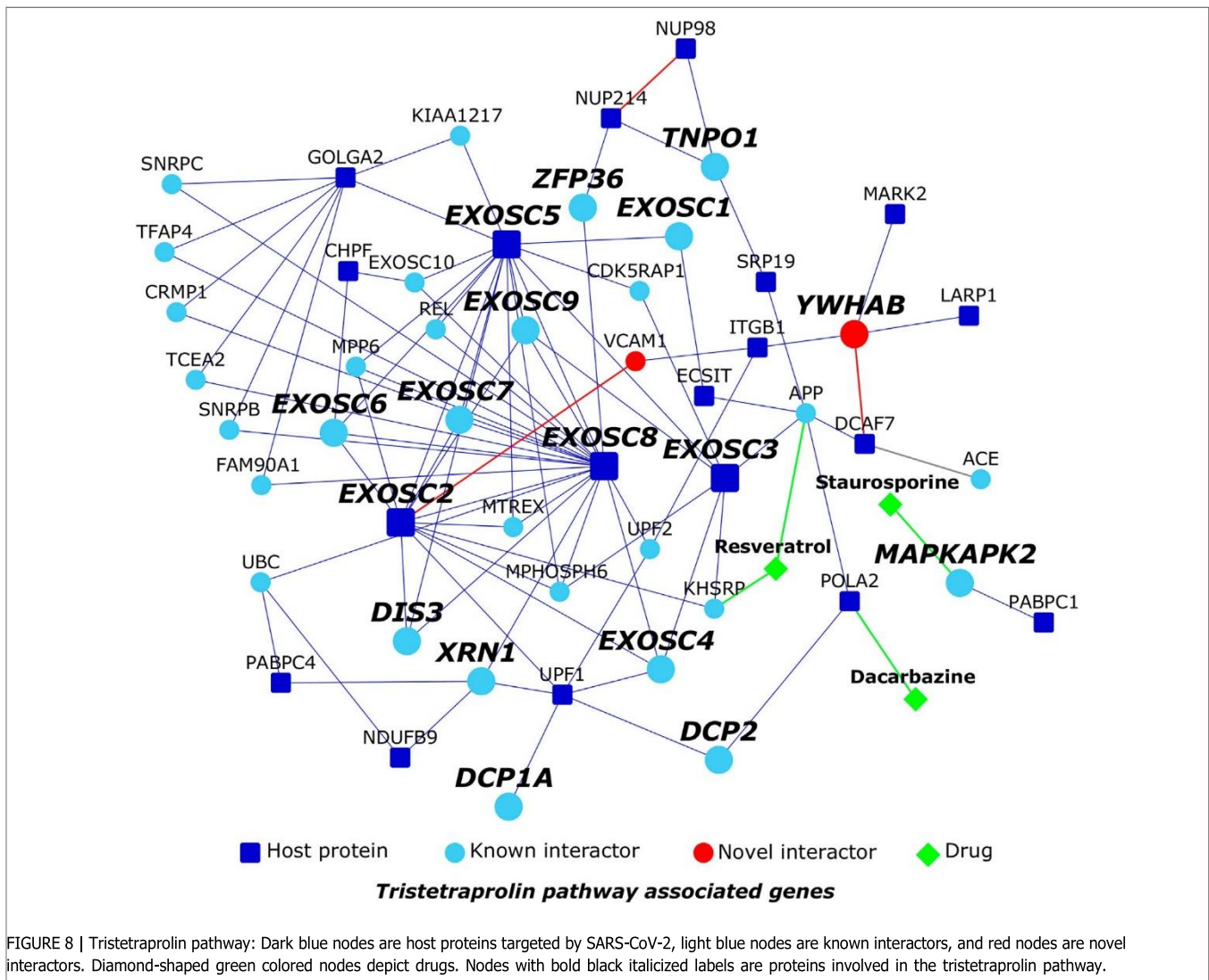
FIGURE 8 | Tristetraprolin pathway: Dark blue nodes are host proteins targeted by SARS-CoV-2, light blue nodes are known interactors, and red nodes are novel interactors. Diamond-shaped green colored nodes depict drugs. Nodes with bold black italicized labels are proteins involved in the tristetraprolin pathway.

such as *MyD88 dependent TLR4 signalling* and *ISG15 anti-viral mechanism* were also identified.

The transcriptional profile of the host cell after SARS-CoV-2 infection had revealed a remarkably limited anti-viral response compared to that elicited by seasonal influenza-A and respiratory syncytial viruses (Blanco-Melo et al., 2020a). This prompted us to inspect a post-transcriptional regulatory pathway that was enriched in the HoP interactome, namely, the Reactome pathway called *tristetraprolin (ZFP36) binds and destabilizes mRNA* (*p*-value < 1×10^{-4}). ZFP36 is an RNA-binding protein that targets AU-rich sites in the mRNA transcripts coding for immune proteins and destabilizes them by promoting deadenylation of their poly (A) tails (Moore et al., 2018; Blackshear, 2002). YWHAB increases cytoplasmic localization of ZFP36, possibly preventing destabilization of these genes and attenuation of immune response (Brook et al., 2006). We extracted the direct PPIs of the 17 genes belonging to this pathway from the HoP interactome and isolated this sub-network for further inspection (Figure 8). Our predictions

showed that the host protein DCAF7, which is known to function as a scaffold protein and a facilitator of PPIs, interacted with YWHAB (Figure 8). This raised the possibility that the virus protein Nsp9 (which interacts with DCAF7) may somehow perturb YWHAB-induced cytoplasmic localization of ZFP36 through its action on DCAF7. Nsp9 may activate or promote the sequestration of YWHAB with DCAF7, thereby reducing its capacity to form a complex with ZFP36. ZFP36- mediated destabilization of immune genes may then lead to a weakened immune response, creating an environment conducive for SARS-CoV-2 infection. We also identified 3 drugs targeting the proteins in this sub-network using Drug Bank (Wishart et al., 2008): resveratrol targeting KHSRP and APP, known interactors of the host protein EXOSC2, which is involved in the tristetraprolin (TTP) pathway, staurosporine targeting TTP- associated MAPKAPK2, which has been predicted to interact with PABPC1, and dacarbazine targeting the host protein POLA2 (Figure 8). Gene expression profiles induced by these drugs in various cell lines were found to have a negative correlation with

FIGURE 9 | Functional modules of viral budding and interferon-mediated anti-viral pathway: Host proteins are depicted as square shaped nodes. The novel predicted interaction of the host protein NUP98 with CHMP5 (red dashed line) may set off an ISG15-mediated pathway that culminates in interference of viral budding. Black colored nodes and transparent nodes indicate overexpression and underexpression in SARS-CoV-infected Calu-3 lung cells respectively. Grey colored nodes depict genes with normal expression. → indicates "*activation,*" -| indicates "*inhibition*" and—indicates "*part of the same complex/physical association*".

SARS-associated gene expression profiles, namely, that of lung fibroblast MRC5 cells infected with SARS-CoV and peripheral blood mononuclear cells of SARS patients (analysis using NextBio; https://www.nextbio.com) (Kupershmidt et al., 2010; Chattopadhyay and Ganapathiraju, 2017). Resveratrol has been proposed as a therapeutic option for SARS-CoV-2 based on its antagonistic properties against MERS-CoV (Lin et al., 2017).

2.10 Interconnections of Ciliary Proteins With Host Proteins Targeted by SARS-CoV-2

SARS coronavirus, which emerged in 2002, has been known to induce necrosis in ciliated airway epithelium of humans in a species-specific manner (Sims et al., 2005). SARS-CoV-2's host receptor ACE2 is highly expressed in ciliated respiratory cells (Sungnak et al., 2020). Cilia may serve as virus entry points and potential modulators of viral pathogenesis. This conjecture prompted us to investigate the ciliary association of the host proteins targeted by SARS-CoV-2 and their interactors in the HoP interactome. For this, we studied its overlap with an interactome of 165 ciliary proteins that we constructed in a similar manner (Karunakaran et al., 2020). The ciliary protein interactome contained 1,665 proteins. 617 of these proteins, andspecifically 30 core ciliary proteins, were also found in SARS- CoV-2's host protein interactome, and the overlap was found to

be statistically significant (*p*-value = 2.24E-10, odds ratio = 1.22). 14 novel predicted interactions connected host proteins to ciliary proteins: NUP98-CHMP5, GG3BP1-DNAH1, SEPSECS-DNAH1, NEK9-IFT43, TLE1-DNAH5, ATP6AP1-CETN2, C1orf50-ZMYND12, RAB10-IFT172, TOR1AIP1-GPR161, DNAJC19-CETN3, NLRX1-IFT46, FKBP7-TTC30B, POLA2-TMEM216 and NDUFB9-DRC7.

Pathway analysis of the 617 common proteins (i.e., commonto HoP and cilia interactomes) revealed two interesting pathways: *budding and maturation of HIV virion* (*p*-value = 1.29×10^{-6}; odds ratio = 8.8) and *anti-viral mechanism by IFN-stimulated genes* (*p*-value = 1.3×10^{-2}; odds ratio = 2.98). We predicted thatthe ciliary protein, CHMP5, involved in the former pathway, interacts with the host protein, NUP98, which is involved in thelatter pathway. This prompted us to ask whether the predicted interaction connected the functional modules of viral budding tointerferon (IFN) signaling.

2.10.1 Novel Interaction of NUP98 With CHMP5 May Activate an IFN-Stimulated Pathway That Interferes With Viral Budding

We extracted the PPIs of the 20 proteins belonging to viral budding and IFN pathways, and isolated this sub-network containing 171 proteins and 176 PPIs for further analysis. Firstly, we identified 343 functional interactions(i.e., activation, inhibition etc.) among 98 proteins in the

network. Strikingly, distinct functional modules were identified for both the pathways; CHMP5 seemed to serve as a connector from the viral budding pathway to the IFN pathway through NUP98 (Figure 9). The gene UBC was shared between the clusters.

We then checked whether the genes in these modules were differentially expressed in Calu-3 lung cells infected with SARS CoV Urbani (for 72 h) versus mock infected cells. This was done to identify the functional interactions that may remain active during viral infection. It was assumed that differential expression of the genes would directly impact the proteins encoded by them and their interactions. 20 genes including NUP98 and CHMP5 were found to be differentially expressed (Figure 9). Viruses hijack the ESCRT/VPS4 (*endosomal sorting complex required for transport*) machinery of the host cell to release viral particles through membrane scission (Pincetic et al., 2010). This machinery is normally recruited during endocytic and membrane repair processes in the host cell. The process of membrane scission is catalyzed by various ESCRT-III proteins including CHMP5 (Pincetic et al., 2010). VPS4 is an ATPase that is found in the cytoplasm in its inactive form. Activation of the VPS4 and its ATPase activity is essential for membrane budding and the release of viral particles (Pincetic et al., 2010). VPS4 is activated on membranes in the presence of its co-activator VTA (also known as LIP5). VTA is delivered to the membranes by ESCRT-III proteins such as CHMP5 (Pincetic et al., 2010). Hence, the interaction of VPS4 and VTA is facilitated by CHMP5. However, when interferons are induced in the host cell following viral infection, ISGs (*interferon stimulated genes*) such as ISG15, a dimer homologue of ubiquitin, may be activated (Pincetic et al., 2010). ISG15 may then conjugate to CHMP5 and promote its accumulation in the membrane, effectively blocking the interaction of VTA with VPS4 and preventing viral budding (Pincetic et al., 2010). The novel interaction of CHMP5 with NUP98 may serve as the critical juncture at which the IFN-stimulated anti-viral mechanism interferes with viral budding. NUP98, a protein induced on viral expression, has been shown to promote anti-viral gene expression in *drosophila* (Panda et al., 2014). Both CHMP5 and NUP98 are overexpressed following SARS-CoV Urbani infection. This interaction may serve as a signal for the initiation of ISG15-mediated interference of viral budding. ISG15 may further regulate this mechanism through feedback inhibition of NUP98. Hence, potentiation of this anti-viral mechanism through administration of recombinant interferon alfa-2b and interferon alfacon-1 may be a feasible therapeutic option for SARS-CoV-2. Both these interferons induce gene expression profiles negatively correlated with SARS-associated profiles. The machinery of ESCRT-III and VPS4 is co-opted into two subcellular structures that are intricately linked to cilia function, namely, the centrosomes and the midbody (Morita et al., 2010). It is important to study these structures as potential modulators of viral infections.

2.11 Potentially Repurposable Drugs

We followed the established approach of comparing drug-induced versus disease-associated differential expression (Sirota et al., 2011) to identify drugs for SARS-CoV-2. For

this, we used a software suite called BaseSpace Correlation Engine (previously called NextBio) (https://www.nextbio. com) (Kupershmidt et al., 2010; Chattopadhyay and Ganapathiraju, 2017). This data analysis platform was used because it allows users to study the effect of diseases and/or drugs on thousands of pre-processed publicly available gene expression datasets and has helped to identify drug candidates for diseases such as schizophrenia (Karunakaran et al., 2019b) (currently undergoing clinical trials (Vishwajit Nimgaonkar, 2022; Vishwajit Nimgaonkar, 2024)) and mesothelioma (Karunakaran et al., 2021) in the past. We compiled a list of 933 chemical compounds whose differential gene expression profiles (drug versus no drug) were negatively correlated with at least one of the four SARS differential gene expression datasets (infected versus non-infected); the 4 SARS datasets we studied were: Calu-3 epithelial cells infected for 48 h with SARS-CoV versus mock infected cells (GSE17400), Calu-3 lung cells infected for 72 h with SARS-CoV Urbani versus mock infected cells (GSE37827), lung fibroblast MRC5 cells 24 h post SARS-CoV infection (high MOI) versus mock infection (GSE56189) and PBMCs from SARS patients versus healthy subjects [GSE1739 (Reghunathan et al., 2005)]. We also compiled a list of 381 chemical compounds with gene expression profiles negatively correlated with the profile induced in human bronchial epithelial (NHBE) and lung cancer (A549) cells infected with the SARS-CoV-2 strain USA-WA1/2020 [GSE147507 (Blanco-Melo et al., 2020a)]

Although in each case, there would be some genes that are differentially expressed in the same direction for both the drug and the disease (i.e., both cause some genes to overexpress, or both cause some genes to under express), the overall effect on the entire transcriptome would be an anti-correlation. A correlation score is generated by NextBio based on the strength of the overlap between the drug and disease datasets. Statistical criteria such as correction for multiple hypothesis testing are applied and the correlated datasets are then ranked by statistical significance. A numerical score of 100 is assigned to the most significant result, and the scores of the other results are normalized with respect to this top-ranked result.

Next, we identified 1,130 drugs that target at least one protein in the HoP interactome using WebGestalt (Liao et al., 2019). We used the "redundancy reduction" feature provided by WebGestalt to prioritize drugs with highly significant overlaps with the interactome, while also capturing all the unique target gene sets. This feature used an affinity propagation algorithm, which clusters sets of genes in the interactome targeted by specific drugs using Jaccard index as the similarity metric, and identifies a "representative" for each cluster (one drug and its targets), having the most significant *p*-value among all the gene sets in that cluster. This resulted in 209 drugs for further consideration. Given a class of drugs targeting the same set of proteins, this method ensures that only those individual drugs that target a statistically significant number of proteins in the interactome are prioritized for further analysis.

Fifty-six drugs were found in common to the above two analyses, i.e., these drugs not only targeted genes in the HoP

TABLE 2 | 51 drugs with expression profiles negatively correlated with SARS-associated profiles: The correlation score is based on the strength of the overlap or enrichment between the two biosets.

Drug	Bioset 1	Bioset 2	Score (scaled negative correlations)	# Up in bioset 1 (p-val), down in bioset 2 (p-val)	# Down in bioset 1 (p-val), up in bioset 2 (p-val)
Alprenolol	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	MCF7 cells + alprenolol, 14 uM _vs._ DMSO vehicle	100	228 (9.1E-9)	74 (0.1079)
Chloramphenicol	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs._mock-infected	Liver of Crj-CD (SD)IGS rats 24 h after 28 days daily dose of 1000 mg-kg chloramphenicol _vs._ 0 mg-kg	100	75 (2.5E-9)	78 (0.0006)
Clotrimazole	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs._mock-infected	Liver of rats + CLOTRIMAZOLE at 52 mg-kg-d in corn oil by oral gavage 3 days _vs._ vehicle	100	119 (6.5E-11)	189 (0.0003)
Didanosine	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	Primary rat hepatocytes + didanosine at 50 uM in DMSO 1 day_vs._vehicle	100	30 (2.2E-6)	49 (0.0027)
Epinephrine	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	Heart of rats + EPINEPHRINE at 0375 mg-kg-d in saline by intravenous 5 days _vs._ vehicle	100	107 (2E-7)	75 (1.8E-5)
Fenofibrate	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs._mock-infected	Huvec cells treated with fenofibrate for 18 h _vs._ untreated	100	395 (4.2E-16)	230 (3.7E-15)
Fenoprofen	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs._mock-infected	Kidney of rats + FENOPROFEN at 52 mg-kg-d in corn oil by oral gavage 1 day _vs._ vehicle	100	72 (5.2E-13)	69 (0.0015)
Ifosfamide	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	Rhabdomyosarcoma xenografts F2 generation treated with ICE-T _vs._ original patient tumor untreated	100	451 (6E-20)	1733 (1.5E-7)
Irinotecan	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	MCF7 breast cancer cells treated 6 h with 5 × IC50 of topo I inhibitor SN38 _vs._ untreated	100	1,153 (4.9E-47)	500 (0.0017)
Isoniazid	PBMC from patients with SARS_vs._healthy subjects	Blood of TB patients infected with M. tuberculosis—post 2HRZE/4HR therapy _vs._ before therapy	100	334 (6E-46)	450 (4.5E-13)
Isradipine	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	HL60 cells + isradipine, 10.8 uM _vs._ DMSO vehicle	100	40 (5.7E-9)	61 (0.0229)
Nitric Oxide	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	HCT116 colon cancer cells + NO 24 h _vs._ untreated control	100	420 (7.3E-44)	231 (8.2E-9)
Paclitaxel	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	Mammary adenocarcinoma did not respond to 3 weeks carboplatin/paclitaxel treatment _vs._ untreated	100	561 (1.5E-29)	269 (1.4E-7)
Phenethyl isothiocyanate	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs._mock-infected	Primary human hepatocytes +25 uM phenethyl isothiocyanate for 48 h _vs._ vehicle	100	401 (2.8E-12)	389 (0.0062)
Riluzole	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs._mock-infected	PC3 cells + riluzole, 14.8 uM _vs._ DMSO vehicle	100	166 (6.5E-5)	258 (2.4E-6)
Sorafenib	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs.	Hodgkins lymphoma HD-MYZ cell line - 10 uM perifosine 5 uM sorafenib treated 24 h _vs._ vehicle control	100	309 (7.5E-14)	378 (5.7E-16)
Terazosin	PBMC from patients with SARS_vs._healthy subjects	Heart of rats + TERAZOSIN at 657 mg-kg-d in water by oral gavage 5 days _vs_ vehicle	100	39 (0.0228)	29 (0.031)
Tetracycline	PBMC from patients with SARS_vs._healthy subjects	Hepatocytes of female donors treated 24 h with 1 uM tetracycline _vs._ 0uM	100	50 (0.0017)	98 (8.5E-14)
Adalimumab	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	Psoriasis lesional skin of adalimumab regimen responders—wk2 _vs._ wk0	100	90 (4.6E-40)	67 (2.1E-5)
Cyclosporine	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	Lesional skins of atopic dermatitis 5 mg/kg/d CsA responders - treated 12 weeks _vs._ baseline	100	386 (2.6E-61)	1,165 (6E-19)
Infliximab	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	Ulcerative colitis colon 10 mg/kg infliximab regimen—8 w _vs._ baseline	100	170 (2.6E-54)	693 (5.7E-54)
Prednisone	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	Blood of dengue patients 2 mg/kg prednisolone treated 3 days - 1month follow up _vs._ pre-treatment	100	370 (1.4E-90)	571 (2.1E-7)
Interferon alfacon-1	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	A549 lung adenocarcinoma cells treated 24 h with 500IU infergen _vs._ untreated	100	150 (6.4E-5)	68 (2.3E-10)

TABLE 2 | (*Continued*) 51 drugs with expression profiles negatively correlated with SARS-associated profiles: The correlation score is based on the strength of the overlap or enrichment between the two biosets.

Drug	Bioset 1	Bioset 2	Score (scaled negative correlations)	# Up in bioset 1 (p-val), down in bioset 2 (p-val)	# Down in bioset 1 (p-val), up in bioset 2 (p-val)
Interferon alfa-2b	PBMC from patients with SARS _vs._ healthy subjects	Healthy whole blood - treated with IFNa-2b _vs._ not treated	100	148 (5E-12)	187 (0.012)
Dacarbazine	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	HL60 cells + dacarbazine, 22 uM _vs._ DMSO vehicle	100	127 (6.5e-16)	69 (0.0162)
Tamoxifen	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	Mammary epithelial cells 48 h with 10 uM tamoxifen _vs._ DMSO	94	343 (1.5E-26)	95 (1.9E-8)
Sumatriptan	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	Brain of rats + SUMATRIPTAN at 1100 mg-kg-d in water by oral gavage 3 days _vs._ vehicle	93	117 (4.1E-7)	87 (4.7E-6)
Nortriptyline	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	MCF7 cells + nortriptyline, 13.4 uM _vs._ DMSO vehicle	91	216 (1.2E-5)	107 (1.2E-6)
Quercetin	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	MCF7 cells + quercetin, 11.8 uM _vs._ DMSO vehicle	91	520 (1.4E-32)	104 (0.0018)
Resveratrol	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	MCF7 cells + resveratrol, 17.6 uM _vs._ DMSO vehicle	91	237 (3.9E-15)	159 (2.3E-5)
Cerivastatin	PBMC from patients with SARS _vs._ healthy subjects	Kidney of rats + cerivastatin at 7 mg-kg-d in corn oil by oral gavage 3 days _vs._ vehicle	90	46 (2.5E-5)	63 (0.0121)
Thioridazine	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	PC3 cells + thioridazine, 9.8 uM _vs._ DMSO vehicle	89	323 (1.7E-9)	105 (1.4E-5)
Mycophenolic acid	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	MCF7 cells + mycophenolic acid, 12.4 uM _vs._ DMSO vehicle	87	329 (2.4E-6)	142 (0.0142)
Granisetron	PBMC from patients with SARS _vs._ healthy subjects	Liver of rats + GRANISETRON at 175 mg-kg-d in water by oral gavage 3 days _vs._ vehicle	86	60 (0.0021)	47 (0.3425)
Ticlopidine	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	PC3 cells + ticlopidine, 13.4 uM _vs._ DMSO vehicle	85	306 (7.8E-6)	83 (0.0001)
Dobutamine	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	PC3 cells + dobutamine, 11.8 uM _vs._ DMSO vehicle	84	69 (0.0023)	42 (9.4E-9)
Permethrin	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	Neural 3D tissue constructs 16 days - treated on d14 with 2.5 uM permethrin for 2 days _vs._ untreated	81	277 (5.7E-20)	168 (1.3E-10)
Sirolimus	Calu-3 epithelial cells infected for 48 h with SARS corona virus _vs._ mock-infected	SKBR3 line (mammary adenocarcinoma overexpressing HER2) + rapamycin 24 h _vs._ vehicle	71	54 (0.46)	517 (5.7E-54)
Epirubicin	Calu-3 epithelial cells infected for 48 h with SARS corona virus _vs._ mock-infected	Breast tumors post epirubicin cyclophosphamide paclitaxel gemcitabine herceptin _vs._ baseline	68	80 (1.8E-23)	452 (1.4E-6)
Timolol	Calu-3 lung cells_SARS Cov urbani infected 72 h _vs._ mock-infected	Heart of rats + timolol at 900 mg-kg-d in water by oral gavage 5 days _vs._ vehicle	65	108 (1.7E-12)	227 (2.2E-5)
Miconazole	Calu-3 lung cells_SARS Cov urbani infected 72 h _vs._ mock-infected	HL60 cells + miconazole, 9.6 uM _vs._ DMSO vehicle	64	102 (0.0003)	108 (0.0248)
Metyrapone	Calu-3 lung cells_SARS Cov urbani infected 72 h _vs._ mock-infected	MCF7 cells + metyrapone, 17.6 uM _vs._ DMSO vehicle	62	162 (0.0004)	88 (0.0096)
Nitrazepam	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._ mock infection	Liver 310 mg per kg nitrazepam treated 3 days _vs._ vehicle control	56	117 (1.1E-8)	84 (0.033)
Perhexiline	Calu-3 lung cells_SARS Cov urbani infected 72 h _vs_ mock-infected	liver of male rat + PERHEXILINE 320 mg per kg for 5 days _vs._ vehicle	52	103 (0.0073)	223 (8E-7)
Staurosporine	PBMC from patients with SARS _vs._ healthy subjects	Primary rat hepatocytes + STAUROSPORINE at 1.3 uM in DMSO 1 day _vs._ vehicle	41	143 (0.0004)	232 (0.0285)

(Continued on following page)

TABLE 2 | (*Continued*) 51 drugs with expression profiles negatively correlated with SARS-associated profiles: The correlation score is based on the strength of the overlap or enrichment between the two biosets.

Drug	Bioset 1	Bioset 2	Score (scaled negative correlations)	# Up in bioset 1 (p-val), down in bioset 2 (p-val)	# Down in bioset 1 (p-val), up in bioset 2 (p-val)
Leflunomide	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	MCF7 cells + leflunomide, 14.8 uM _vs._ DMSO vehicle	40	70 (0.0007)	30 (0.0968)
Verapamil	PBMC from patients with SARS_vs._healthy subjects	HL60 cells + verapamil, 8.2 uM _vs._ DMSO vehicle	39	38 (0.0001)	43 (0.0299)
Hydrocortisone	Calu-3 epithelial cells infected for 48 h with SARS corona virus_vs._mock-infected	HUVECS 1uM hydrocortisone 500U/ml IL1B 2500 U/ml TNFα +1250 U/ml IFNγ 4 h_vs._vehicle	36	71 (0.1487)	324 (0.0692)
Progesterone	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	MCF7 cells + progesterone, 12.8 uM _vs._ DMSO vehicle	31	270 (2.6E-10)	116 (0.0386)
Ramipril	Lung fibroblast MRC5 cells 24 h post SARS corona virus infection high MOI Zhou et al. (2020a) _vs._mock infection	MCF7 cells + ramipril, 9.6 uM _vs._ DMSO vehicle	31	147 (1.7E-7)	53 (0.0052)
Temazepam	Calu-3 lung cells_SARS Cov urbani infected 72 h_vs._mock-infected	Cerebrocortical cells from E16.5 mice treated 8 h—0.5 uM temazepam _vs._ DMSO	27	11 (0.0534)	25 (0.0338)

Additional statistical criteria such as correction for multiple hypothesis testing are applied and the correlated biosets are then ranked by statistical significance. A numerical score of 100 is assigned to the most significant result, and the scores of the other results are normalized with respect to the top-ranked result.

interactome, but also induced gene expression profiles which are negatively correlated with that induced by SARS-CoV (Table 2) and SARS-CoV-2 (Table 3). 13 drugs showed negativecorrelation with both expression profiles. 24 of these have supporting evidence for biological relevance (see Appendix) through clinical trial data and published literature (Figure 10).

- 4 drugs showed activity against SARS-CoV-2 *in vitro* (cyclosporine, sorafenib, tamoxifen, anisomycin)
    - 1 chemical compound (nitric oxide) found here is already being tested against SARS-CoV-2 in clinical trials
    - 1 drug (ramipril) belongs to the class of receptors targeted by SARS-CoV-2
    - 5 drugs display clinical activity against SARS or MERS (resveratrol, sirolimus, mycophenolic acid, interferon alpha-2b, interferon alfacon-1)
- 3 drugs (quercetin, verapamil, progesterone) are active against influenza viruses
- 2 drugs are active against DNA viruses (leflunomide, daunorubicin), and
- 8 drugs show activity against other RNA viruses (clotrimazole, didanosine, paclitaxel, fenofibrate, cerivastatin, thioridazine, pioglitazone, miglitol)

Eight drugs from our shortlist were independently identified or prioritized by other groups, namely: leflunomide [Chen et al. (Chen et al., 2021)], sirolimus [Zhou et al. (Zhou et al., 2020a)], leflunomide, quercetin and verapamil [Gysi et al. (Gysi et al., 2020)], interferon alfa-2b, resveratrol, cyclosporine and mycophenolic acid [Li et al. (Li and De Clercq, 2020)]. Additionally, 8 out of the 24 shortlisted drugs were also found among 127 broad-spectrum antiviral drugs active against 80

viruses (https://drugvirus.info/). These are cyclosporine, leflunomide, mycophenolic acid, sirolimus, sorafenib,tamoxifen, anisomycin and verapamil. Fourteen drugs were found to induce expression profiles negatively correlated with the profiles of ICU-admitted COVID-19 patients with ARDS versus non-critical patients on oxygen [GSE172114 (Carapito et al.,2021)], namely, cerivastatin, cyclosporine, didanosine, leflunomide, miglitol, mycophenolic acid, paclitaxel, quercetin, resveratrol, sirolimus, sorafenib, tamoxifen, thioridazine and verapamil. Three drugs— didanosine, miglitol and resveratrol—induced profiles negatively correlated with that of COVID-19 patients in ICU versus healthy subjects [GSE152418 (Arunachalam et al., 2020)]. Additionally 4 drugs (sorafenib, quercetin, verapamil and cerivastatin) induced profiles negatively correlated with the profiles of ICU-admitted COVID-19 patients with ARDS versus those receiving non-intensive care [GSE157103 (Overmyer et al.,2021)] and 2 drugs (resveratrol and didanosine) with profiles of COVID-19 patients critical in ICU with ARDS versus non-criticalpatients on oxygen [GSE172114 (Carapito et al., 2021)]. These drugs could be examined for their differential clinical activity in critical versus non-critical cases.

3 DISCUSSION

In this study, to gain insight into the biological processes and pathways that may be involved in host response upon SARS-CoV-2 infection, we assembled the interactome of the host proteins targeted by the virus. The host protein (HoP) interactome has ~4,000 previously known PPIs in addition to ~2,000 PPIs that we computationally predicted. The interactome and its annotations are made available on the website that is freely

TABLE 3 | 18 drugs with expression profiles negatively correlated with COVID-associated profile: The correlation score is based on the strength of the overlap or enrichment between the two biosets.

Drug	Bioset 1	Bioset 2	Correlationscore (scaled negative correlations)	# Up in bioset 1 (p-val), down in bioset 2 (p-val)	# Downin bioset 1 (p-val), up in bioset 2 (p-val)
Didanosine	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Primary rat hepatocytes + DIDANOSINE at 500 uM in DMSO 1 day _vs._ vehicle	76	43 (3.2E-19)	20 (0.1725)
Isoniazid	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Blood of TB patients infected with M. tuberculosis - post 2HRZE/4HR therapy _vs._ before therapy	75	268 (2.4E-47)	580 (5.9E-9)
Epirubicin	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Liver 2.7 mg per kg Epirubicin treated 3 days _vs._ vehicle control	66	92 (1.4E-21)	100 (0.0001)
Paclitaxel	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Ovarian cancer OVISE cells +10X IC50 concentration of paclitaxel for 24 h _vs._ untreated	66	82 (4.2E-15)	91 (0.0358)
Daunorubicin	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Heart 3.25 mg per kg Daunorubicin treated 1 day _vs._ vehicle control	65	40 (1.3E-14)	34 (0.0004)
Rifapentine	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Kidney of rats + RIFAPENTINE at 75 mg-kg-d in corn oil by oral gavage 1 day _vs._ vehicle	64	24 (0.0117)	77 (2.2E-7)
Ticlopidine	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Liver of Crj-CD (SD)IGS rats 24h after 14 days daily dose of 300 mg-kg ticlopidine _vs_ 0 mg-kg	63	59 (2.9E-15)	59 (0.2386)
Ifosfamide	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Heart of rats + IFOSFAMIDE at 143 mg-kg-d in saline by oral gavage 5 days _vs._ vehicle	61	34 (2.6E-15)	93 (0.0091)
Quercetin	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Hep G2 hepatocarcinoma cell line cultured for 24 h with 10 nM quercetin _vs._ 0.5% DMSO	60	25 (0.0016)	233 (1.7E-12)
Resveratrol	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	AML THP-1 cells 24 h Mycobacterium tuberculosis infected - with 100 uM resveratrol _vs._ without	58	162 (2.8E-23)	293 (3.1E-5)
Tetracycline	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Primary rat hepatocytes + TETRACYCLINE at 520 uM in DMSO 0.67 days _vs._ vehicle	58	19 (0.001)	174 (6.2E-7)
Pioglitazone	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Heart of rats + PIOGLITAZONE at 1500 mg-kg-d in corn oil by oral gavage 5 days _vs._ vehicle	57	15 (0.343)	91 (7.5E-7)
Chloramphenicol	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Liver of Crj-CD (SD)IGS rats 24 h after 28 days daily dose of 1000 mg-kg chloramphenicol _vs._ 0 mg-kg	55	39 (5E-09)	23 (0.0535)
Permethrin	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Neural 3D tissue constructs 16 days - treated on d14 with 2.5 uM permethrin for 2 days _vs._ untreated	55	18 (0.4565)	157 (5.5E-5)
Miglitol	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Osteosarcoma U-2 OS cells treated 12 h with 1000 nM miglitol _vs._ DMSO	45	3 (0.0366)	31 (8.4E-6)
Nortriptyline	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Primary rat hepatocytes + NORTRIPTYLINE at 70 uM in DMSO 1 day _vs._ vehicle	42	93 (2.7E-8)	392 (0.0048)
Nitrazepam	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	Liver of rats + nitrazepam at 310 mg-kg-d in CMC by oral gavage 3 days _vs._ vehicle	41	22 (0.0471)	119 (0.0016)
Anisomycin	Bronchial epithelial NHBE and lung cancer A549 cells infected with SARS-CoV-2 strain United States-WA1/2020	PC3 cells + anisomycin, 15 uM _vs._ DMSO vehicle	41	75 (0.2238)	281 (3.5E-5)

Additional statistical criteria such as correction for multiple hypothesis testing are applied and the correlated biosets are then ranked by statistical significance. A numerical score of 100 is assigned to the most significant result, and the scores of the other results are normalized with respect to the top-ranked result.
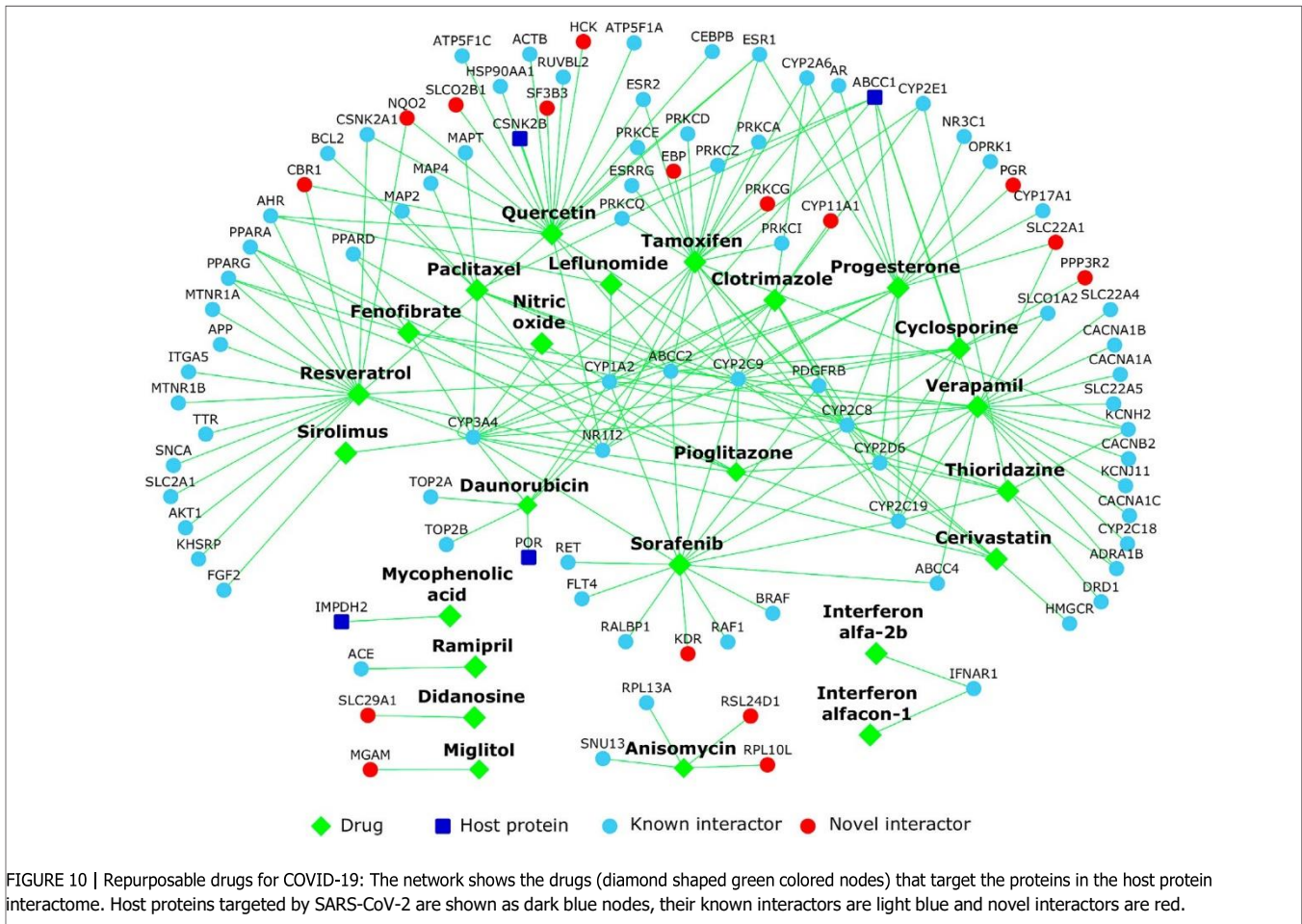
FIGURE 10 | Repurposable drugs for COVID-19: The network shows the drugs (diamond shaped green colored nodes) that target the proteins in the host protein interactome. Host proteins targeted by SARS-CoV-2 are shown as dark blue nodes, their known interactors are light blue and novel interactors are red.

accessible, *Wiki-Corona*. The HoP interactome was found to share large and statistically significant overlaps with gene expression profiles induced by SARS-CoV and SARS-CoV-2. Proteins with tissue-specific gene expression in lungs, spleen, brain and heart were also found in the interactome. Topologically connected modules in the network showed functional association to cilium organization, nuclear transport, ribonucleoprotein complex biogenesis, endosomal transport and epigenetic regulation of gene expression. The interactome was enriched for subcellular locations and host cellular processes that may be targeted by SARS-CoV-2. It also showed significant associations with several disorders including cancers, metabolic, neurological, developmental and vascular disorders. For example, the host proteins were found to directly interact with proteins associated with two co-morbidities, hypertension and diabetes, which are commonly found among COVID-19 non-survivors. Protein biomarkers showing varied expression across the different stages of COVID-19 were predicted as novel interactors of the host proteins targeted by SARS-CoV-2. The SARS-CoV-2 host proteins and ciliary proteins shared several common interactors. The role of cilia as viral entry points and modulators of viral infections should be investigated further on this premise. On further analysis of the shared interactome, we

hypothesized that the novel interaction of NUP98 with CHMP5, a ciliary (and centrosome and midbody-localizing) protein, may activate an IFN-stimulated pathway with the potential to interfere with viral budding. We shortlisted drugs potentially repurposable for COVID-19 based on the negative correlation of drug-induced versus disease-associated gene expression profiles. These included drugs with proven *in vitro* activity against SARS-CoV-2, those that were already being tested for their clinical activity against SARS-CoV-2, those with proven activity against SARS-CoV/ MERS-CoV, broad-spectrum antiviral drugs, and those identified/prioritized by other computational re-purposing studies. Our computational approach has several limitations. Drug-associated expression profiles analyzed in this study were induced in several types of cell lines (including cancer cell lines) that may not be directly relevant to COVID-19 or SARS-CoV-2 infection. The effect of the proposed repurposable drugs should be studied in human bronchial epithelial cells and/or in human lung cancer cell lines, both of which were recently used to study host transcriptional response upon SARS-CoV-2 infection (Blanco-Melo et al., 2020a). The repurposable drugs discussed in this study can simply be identified through comparative transcriptomic analysis; i.e., by comparing drug-induced expression profiles with SARS-CoV/

SARS-CoV-2-induced profiles. However, by considering the drugs targeting the proteins in the HoP interactome, we attempted to provide starting points for a mechanistic and experimentally testable basis for the negative correlation observed between the drugs and the viral infection at the transcriptomic level. These starting points are typically subnetworks (e.g., Figure 10) showing interconnections of the drug targets and the host proteins targeted by the virus in the human interactome. The sets of host proteins interacting with SARS-CoV and SARS- CoV-2, which were analyzed in our study to elucidate common pathways targeted by these viruses, were themselves identified using different protein interaction mapping techniques (AP–MS and Y2H) in different studies. These techniques differ from each other with respect to the nature of the PPIs that they detect: AP–MS identifies direct and indirect interactions among members of stable protein complexes, while Y2H may identify direct and more transient interactions between pairs of proteins. Therefore, each of these techniques may detect a different portion of the virus-host interactome. The marginal overlap observed between the sets of host proteins interacting with SARS-CoV and SARS-CoV-2 could be attributed to the differences in the interactome subspaces detected by Y2H and AP–MS, respectively. In this scenario, one may expect the neighborhood networks of these host proteins to also exhibit this discordance. However, we observed extensive interconnections between these sets of proteins, *via* direct and intermediate (known as well as novel) interactors. This shows that 1) it is the ability of the different techniques to detect different subspaces of the interactome in a complementary manner that makes them valuable, 2) machine learning methods may capture novel PPIs that other techniques fail to capture, and 3) computational methods may be employed to piece together an integrated view of the interactome, despite the limitations of the individual mapping techniques.

The novelty of this work stems from several aspects. 1) Despite an explosive increase in the generation of COVID-19 related data, knowledge on the mechanistic basis of the host cellular response to SARS-CoV-2 infection is limited. Therefore, we prioritized dataset mining and hypothesis generation over data generation by integrating and analyzing publicly available multi-omics data within the functional landscape of the protein interactome using bioinformatic tools. This approach directly contributes towards COVID-19 research prioritization, namely, selection of pathways and drugs for experimental dissection and clinical interventions. 2) Computationally predicted PPIs enhanced hypothesis generation by linking host genes across various high throughput studies in as-yet-undiscovered ways. 3) To facilitate analysis by both computational and biomedical scientists, all the results are being released in multiple data formats in open access and via an interactive webserver (see *Data Availability*). 4) The HoP interactome will facilitate several future systems biology studies derived from overlaying the interactome with data generated for research on coronaviruses, and specifically on COVID-19. In summary, the interactome will be useful for carrying out several studies in the future with rapidly emerging data to generate biologically insightful results that may be translated to biomedically actionable results.

4 METHODS

4.1 Compilation of Host Proteins and Prediction of Novel Interactions

The list of 332 host proteins identified to interact with 27 SARS-CoV-2 proteins was compiled from data files in Gordon et al. (2020a). Novel PPIs of these proteins were predicted using the HiPPIP model that we developed (Ganapathiraju et al., 2016b). Each host protein (say N1) was paired with each of the other human protein say (M1, M2, ... Mn), and each pair was evaluated with the HiPPIP model (Ganapathiraju et al., 2016b). The predicted interactions of each of the host proteins were extracted (namely, the pairs whose score is >0.5, a threshold which through computational evaluations and experimental validations was revealed to indicate interacting partners with high confidence). This resulted in 1941 newly discovered PPIs of the host proteins. The interactome figures were created using Cytoscape (Shannon et al., 2003).

The significance of the overlap of this interactome with two datasets, namely, with the ciliary protein interactome and the interactome of 120 genes differentially expressed in SARS-CoV-2-infected A549 cell line (Blanco-Melo et al., 2020a), was computed based on hypergeometric distribution.

4.2 Identification of Network Modules

Network modules among the host proteins targeted by SARS-CoV-2 and their interactors were identified using Netbox (Cerami et al., 2010). Netbox reports modularity and a scaled modularity score, as compared with the modularity observed in 1,000 random permutations of the subnetwork. Scaled modularity refers to the standard deviation difference between the observed subnetwork and the mean modularity of the random networks (Wang and Zhang, 2007).

4.3 Transcriptome and Proteome Analysis

Statistical significance of the overlaps between genes in the HoP interactome and SARS-CoV/SARS-CoV-2 induced/associated transcriptomic/proteomic datasets was computed based on hypergeometric distribution. In this method, p-value is computed from the probability of k successes in n draws (without replacement) from a finite population of size N containing exactly K objects with an interesting feature.

$$P\left(X=k\right)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

N = Total number of genes/proteins assayed.
K = Number of SARS-CoV/SARS-CoV-2-induced/associated genes/proteins.
n = Number of genes in the HoP interactome. k = K ∩ n.

4.4 Tissue-Specificity Analysis

Tissue-specificity of the genes in the HoP interactome were checked using TissueEnrich (Jain and Tuteja, 2019). The

analysis was based on genes from the GTEx database (Lonsdale et al., 2013). This included "'tissue-enriched genes" with at least 5-fold higher mRNA levels in a particular tissue compared to all the other tissues, "group-enriched genes" with at least 5-fold higher mRNA levels in a group of 2-7tissues and "tissue-enhanced genes" with at least 5-fold higher mRNA levels in a particular tissue compared to average levels in all tissues.

4.5 Functional Enrichment Analysis

Gene Ontology, Pathway and genetic disorder enrichments were computed using WebGestalt (Liao et al., 2019). WebGestalt computes the distribution of genes belongingto a particular functional category in the input list (i.e., genesin the HoP interactome/ACE2 and its interactors) and compares it with the background distribution of genes belonging to this functional category among all the genes that belongs to any functional category in the database selected by the user. Statistical significance of functional category enrichment is computed using Fisher's exact test, and corrected using the Benjamini–Hochberg method for multiple test adjustment. Annotations with FDR-corrected p-value < 0.05 were considered significant. ReactomeFIViz, a Cytoscape plugin, was used to extract known functional interactions among genes in the HoP interactome that were involved in viral budding and interferon signaling pathways (Wu et al., 2014).

4.6 Potentially Repurposable Drugs

The list of chemical compounds whose gene expression profiles correlated negatively with four SARS datatsets andone COVID-19 dataset were compiled using the BaseSpace correlation software (https://www.nextbio.com) (List 1). The datasets considered were human bronchial epithelial (NHBE) and lung cancer (A549) cells infected with the SARS-CoV-2 strain USA-WA1/2020 [GSE147507 (Blanco-Melo et al., 2020a)], Calu-3 epithelial cells infected for 48 h with SARS- CoV versus mock infected cells (GSE17400), Calu-3 lung cells infected for 72 h with SARS-CoV Urbani versus mock infected cells (GSE37827), lung fibroblast MRC5 cells 24 h post SARS- CoV infection (high MOI) versus mock infection (GSE56189) and peripheral blood mononuclear cells (PBMCs) from patients with SARS versus healthy subjects [GSE1739 (Reghunathan et al., 2005)]. Next, we identified drugs that targeted at least one gene in the HoP interactome using WebGestalt (Liao et al., 2019). After employing the "redundancy reduction" feature in WebGestalt to reduce the search space of drugs, we were left with a fewernumber of drugs (List 2). In this feature, an affinity propagation algorithm clusters gene sets in the interactome targeted by specific drugs using Jaccard index as the similarity metric, and identifies a "representative" for each cluster (one drug and its targets), having the most significant p-value among all the gene sets in that cluster. We then comparedlist 1 and list 2 to identify the drugs that not only target

proteins in the interactome but are also negatively correlated with SARS/COVID-19.

List of drugs validated to be effective against SARS-CoV-2 incell-based assays were obtained from the COVID-19 Gene and Drug Set Library (https://amp.pharm.mssm.edu/covid19/) (Kuleshov et al., 2020).

The drug-protein interactome figure was created using Cytoscape (Shannon et al., 2003).

DATA AVAILABILITY STATEMENT

The interactome, consisting of both known and novel computationally predicted protein-protein interactions, is being released via a webserver (http://severus.dbmi.pitt.edu/ corona/) and as Supplementary Table S1. The membership of the proteins in the additional data sources referred to in this workcan be found in Supplementary Table S2. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MKG carried out interactome construction, web-based dissemination, and other algorithmic analyses. KBK carried out bioinformatic analysis over the interactome with multiple omics data, including its study design and interpretation, with supervision from NB and MKG. Manuscript has been drafted by KBK and edited and approved by all authors.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsysb.2022.815237/full#supplementary-material

Supplementary Figure S1 | Network proximity of cellular entry proteins and host proteins targeted by SARS-CoV-2: Dark blue nodes are host proteins targeted by SARS-CoV-2, light blue nodes are known interactors and red nodes are novel interactors. Brown-colored nodes with bold black italicized labels are

the "cellular entry proteins" that facilitate the entry of SARS-CoV-2 into host cells.

Supplementary Table S1 | List of proteins and protein-protein interactions in the host protein interactome, and the integrated interactome with virus- host PPIs and PPIs in the neighborhood network of host proteins: Computationally predicted interactors and interactions are indicated as "novel interactors" and "novel PPIs" respectively, whereas previously known interactors and interactions are shown as "known interactors" and "known PPIs".

Supplementary Table S2 | Complete list of SARS/COVID-related biological evidences of genes in the host protein interactome: A tick mark indicates the presence of a particular evidence for a given gene.

REFERENCES

Amaral, L., Viveiros, M., and Molnar, J. (2004). Antimicrobial Activity of Phenothiazines. *In Vivo* 18 (6), 725–731.

Arunachalam, P. S., Wimmers, F., Mok, C. K. P., Perera, R. A. P. M., Scott, M., Hagan, T., et al. (2020). Systems Biological Assessment of Immunity to Mildversus Severe COVID-19 Infection in Humans. *Science* 369 (6508), 1210–1220. doi:10.1126/science.abc6261

Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network Medicine: a Network-Based Approach to Human Disease. *Nat. Rev. Genet.* 12 (1), 56–68. doi:10.1038/nrg2918

Blackshear, P. J. (2002). Tristetraprolin and Other CCCH Tandem Zinc-finger Proteins in the Regulation of mRNA Turnover. *Biochem. Soc. Trans.* 30 (Pt 6),945–952. doi:10.1042/bst0300945

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W.-C., Møller, R., Panis, M., Sachs, D., et al. (2020). *SARS-CoV-2 Launches a Unique Transcriptional Signature from* in Vitro, Ex Vivo, *and* in Vivo *Systems*. Cambridge, Massachusetts: Cell.

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Møller, R., et al. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* 181, 1036–e9. doi:10.1016/j.cell.2020. 04.026

Blasche, S., and Koegl, M. (2013). Analysis of Protein-Protein Interactions Using LUMIER Assays. *Virus-Host Interactions: Methods Protoc.* 1, 17–27. doi:10.1007/978-1-62703-601-6_2

Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., and Ciesek, S. (2020). SARS-CoV-2 Infected Host Cell Proteomics Reveal Potential Therapy Targets. *Nature* 583 (7816), 469–472.

Bouhaddou, M., Memon, D., Meyer, B., White, K. M., Rezelj, V. V., Correa Marrero, M., et al. (2020). The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell.* 182 (3), 685–712. e19. doi:10.1016/j.cell.2020.06.034

Brook, M., Tchen, C. R., Santalucia, T., McIlrath, J., Arthur, J. S. C., Saklatvala, J.,et al. (2006). Posttranslational Regulation of Tristetraprolin Subcellular Localization and Protein Stability by P38 Mitogen-Activated Protein Kinase and Extracellular Signal-Regulated Kinase Pathways. *Mol. Cell Biol.* 26 (6), 2408–2418. doi:10.1128/mcb.26.6.2408-2418.2006

Burrell, L. M., Johnston, C. I., Tikellis, C., and Cooper, M. E. (2004). ACE2, a New Regulator of the Renin-Angiotensin System. *Trends Endocrinol. Metab.* 15 (4), 166–169. doi:10.1016/j.tem.2004.03.001

Cantuti-Castelvetri, L., Ojha, R., Pedro, L. D., Djannatian, M., Franz, J., Kuivanen, S., et al. (2020). Neuropilin-1 Facilitates SARS-CoV-2 Cell Entry and Infectivity.*Science* 370 (6518), 856–860. doi:10.1126/science.abd2985

Carapito, R., Li, R., Helms, J., Carapito, C., Gujja, S., Rolli, V., et al. (2021). Identification of Driver Genes for Critical Forms of COVID-19 in a Deeply Phenotyped Young Patient Cohort. *Sci. Translational Med.* 1, eabj7521.

Cerami, E., Demir, E., Schultz, N., Taylor, B. S., and Sander, C. (2010). Automated Network Analysis Identifies Core Pathways in Glioblastoma. *PloS one* 5, e8918. doi:10.1371/journal.pone.0008918

Cham, L. B., Friedrich, S.-K., Adomati, T., Bhat, H., Schiller, M., Bergerhausen, M.,et al. (2019). Tamoxifen Protects from Vesicular Stomatitis Virus Infection. *Pharmaceuticals* 12 (4), 142. doi:10.3390/ph12040142

Chang, J., Warren, T. K., Zhao, X., Gill, T., Guo, F., Wang, L., et al. (2013). Small Molecule Inhibitors of ER α-glucosidases Are Active against Multiple

Hemorrhagic Fever Viruses. *Antivir. Res.* 98 (3), 432–440. doi:10.1016/j.antiviral.2013.03.023

Chattopadhyay, A., and Ganapathiraju, M. (2017). Demonstration Study: A Protocol to Combine Online Tools and Databases for Identifying Potentially Repurposable Drugs. *Data* 2 (2), 15. doi:10.3390/data2020015

Chen, F., Shi, Q., Pei, F., Vogt, A., Porritt, R. A., Garcia, G., Jr, et al. (2021). A Systems-Level Study Reveals Host-Targeted Repurposable Drugs against SARS-CoV-2 Infection. *Mol. Syst. Biol.* 17 (8), e10239. doi:10.15252/msb.202110239

Cheng, K.-W., Cheng, S.-C., Chen, W.-Y., Lin, M.-H., Chuang, S.-J., Cheng, I.-H.,et al. (2015). Thiopurine Analogs and Mycophenolic Acid Synergistically Inhibit the Papain-like Protease of Middle East Respiratory Syndrome Coronavirus. *Antivir. Res.* 115, 9–16. doi:10.1016/j.antiviral.2014.12.011

Cho, H., Kim, J., Kim, S., Kyaw, Y., Win, A., and Cheong, J. (2015). Sorafenib Suppresses Hepatitis B Virus Gene Expression via Inhibiting JNK Pathway. *Hepatoma Res.* 1, 97.

Cui, C., Huang, C., Zhou, W., Ji, X., Zhang, F., Wang, L., et al. (2020). *AGTR2, One Possible Novel Key Gene for the Entry of SARS-CoV-2 into Human Cells*. IEEE/ ACM transactions on computational biology and bioinformatics.

Deng, M., Zhang, K., Mehta, S., Chen, T., and Sun, F. (2003). Prediction of Protein Function Using Protein-Protein Interaction Data. *J. Comput. Biol.* 10 (6), 947–960. doi:10.1089/106652703322756168

Dunham, B., and Ganapathiraju, M. K. (2022). Benchmark Evaluation of Protein–Protein Interaction Prediction Algorithms. *Molecules* 27 (1), 41.

Ellinger, B., and Zaliani, A. (2020). Identification of Inhibitors of SARS-CoV-2 *In-Vitro* Cellular Toxicity in Human (Caco-2) Cells Using a Large Scale Drug Repurposing Collection. *Res. Square* 1, 10.

Emamjomeh, A., Goliaei, B., Torkamani, A., Ebrahimpour, R., Mohammadi, N., and Parsian, A. (2014). Protein-protein Interaction Prediction by Combined Analysis of Genomic and Conservation Information. *Genes Genet. Syst.* 89 (6), 259–272. doi:10.1266/ggs.89.259

Españo, E., Nam, J.-H., Song, E.-J., Song, D., Lee, C.-K., and Kim, J.-K. (2019). Lipophilic Statins Inhibit Zika Virus Production in Vero Cells. *Sci. Rep.* 9 (1),11461. doi:10.1038/s41598-019-47956-1

Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al. (2014). Analysis of the Human Tissue-specific Expression by Genome- wide Integration of Transcriptomics and Antibody-Based Proteomics. *Mol. Cell Proteomics* 13 (2), 397–406. doi:10.1074/mcp.m113.035600

Falzarano, D., De Wit, E., Rasmussen, A. L., Feldmann, F., Okumura, A., Scott, D. P., et al. (2013). Treatment with Interferon-A2b and Ribavirin ImprovesOutcome in MERS-CoV-Infected Rhesus Macaques. *Nat. Med.* 19 (10), 1313–1317. doi:10.1038/nm.3362

Fan, Y., Sanyal, S., and Bruzzone, R. (2018). Breaking Bad: How VirusesSubvert the Cell Cycle. *Front. Cel. Infect. Microbiol.* 8, 396. doi:10.3389/ fcimb.2018.00396

Fang, L., Karakiulakis, G., and Roth, M. (2020). Are Patients with Hypertension and Diabetes Mellitus at Increased Risk for COVID-19 Infection? *Lancet Respir. Med.* 8, e21. doi:10.1016/S2213-2600(20)30116-8

Fang, L., Karakiulakis, G., and Roth, M. (2020). Are Patients with Hypertension and Diabetes Mellitus at Increased Risk for COVID-19 Infection? *Lancet Respir. Med.* 8 (4), e21. doi:10.1016/s2213-2600(20)30116-8

Ganapathiraju, M. K., Karunakaran, K. B., and Correa-Menendez, J. (2016). Predicted Protein Interactions of IFITMs Which Inhibit Zika Virus Infection. *F1000Res.* 5, 1919. doi:10.12688/f1000research.9364.1

Ganapathiraju, M. K., Thahir, M., Handen, A., Sarkar, S. N., Sweet, R. A., Nimgaonkar, V. L., et al. (2016). Schizophrenia Interactome with 504 Novel

Protein-Protein Interactions. *npj Schizophr* 2, 16012. doi:10.1038/npjschz. 2016.12

Garzón, J. I., Deng, L., Murray, D., Shapira, S., Petrey, D., and Honig, B. (2016). A Computational Interactome and Functional Annotation for the Human Proteome. *Elife* 5, e18715. doi:10.7554/eLife.18715

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., O'Meara, M. J.,et al. (2020). *A SARS-CoV-2-Human Protein-Protein Interaction Map RevealsDrug Targets and Potential Drug-Repurposing*. London, United kingdom: Nature.

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* 583, 459–468. doi:10.1038/s41586-020-2286-9

Gralinski, L. E., and Baric, R. S. (2015). Molecular Pathology of Emerging Coronavirus Infections. *J. Pathol.* 235 (2), 185–195. doi:10.1002/path.4454

Gysi, D. M., Valle, Í. D., Zitnik, M., Ameli, A., Gan, X., Varol, O., et al. (2020). *Network Medicine Framework for Identifying Drug Repurposing Opportunities for Covid-19*. Proceedings of the National Academy of Sciences 118 (19).

Hafirassou, M. L., Meertens, L., Umaña-Diaz, C., Labeau, A., Dejarnac, O., Bonnet-Madin, L., et al. (2017). A Global Interactome Map of the Dengue Virus NS1 Identifies Virus Restriction and Dependency Host Factors. *Cell Rep.* 21 (13), 3900–3913. doi:10.1016/j.celrep.2017.11.094

Hall, O. J., Limjunyawong, N., Vermillion, M. S., Robinson, D. P., Wohlgemuth, N., Pekosz, A., et al. (2016). Progesterone-based Therapy Protects against Influenza by Promoting Lung Repair and Recovery in Females. *Plos Pathog.* 12, e1005840. doi:10.1371/journal.ppat.1005840

Handen, A., and Ganapathiraju, M. K. (2015). LENS: Web-Based Lens for Enrichment and Network Studies of Human Proteins. *BMC Med. Genomics*8 (S4), S2. doi:10.1186/1755-8794-8-s4-s2

Hart, B. J., Dyall, J., Postnikova, E., Zhou, H., Kindrachuk, J., Johnson, R. F., et al. (2014). Interferon-β and Mycophenolic Acid Are Potent Inhibitors of MiddleEast Respiratory Syndrome Coronavirus in Cell-Based Assays. *J. Gen. Virol.* 95 (Pt 3), 571–577. doi:10.1099/vir.0.061911-0

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T.,Erichsen,S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271–e8. doi:10.1016/j.cell.2020.02.052

Hoischen, C., Monajembashi, S., Weisshart, K., and Hemmerich, P. (2018). Multimodal Light Microscopy Approaches to Reveal Structural and Functional Properties of Promyelocytic Leukemia Nuclear Bodies. *Front. Oncol.* 8, 125. doi:10.3389/fonc.2018.00125

Hopf, T. A., Schärfe, C. P., Rodrigues, J. P., Green, A. G., Kohlbacher, O., Sander, C.,et al. (2014). Sequence Co-evolution Gives 3D Contacts and Structures of Protein Complexes. *Elife* 3, e03430. doi:10.7554/eLife.03430

Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., Baltier, K., Gebreab, F., et al. (2020). Dual Proteome-Scale Networks Reveal Cell-specific Remodeling of the Human Interactome. *Cell* 184 (11), 3022–40.e28.

Imai, H., Dansako, H., Ueda, Y., Satoh, S., and Kato, N. (2018). Daunorubicin, a Topoisomerase II Poison, Suppresses Viral Production of Hepatitis B Virus by Inducing cGAS-dependent Innate Immune Response. *Biochem. biophysical Res. Commun.* 504 (4), 672–678. doi:10.1016/j.bbrc.2018.08.195

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S.C., et al. (2012). Global Landscape of HIV-Human Protein Complexes. *Nature* 481 (7381), 365–370. doi:10.1038/nature10719

Jain, A., and Tuteja, G. (2019). TissueEnrich: Tissue-specific Gene Enrichment Analysis. *Bioinformatics* 35 (11), 1966–1967. doi:10.1093/bioinformatics/bty890

Jeon, S., Ko, M., Lee, J., Choi, I., Byun, S. Y., Park, S., et al. (2020). *Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs*. Washington DC: Antimicrobial Agents and Chemotherapy.

Jia, J., Liu, Z., Xiao, X., Liu, B., and Chou, K.-C. (2015). iPPI-Esml: an Ensemble Classifier for Identifying the Interactions of Proteins by Incorporating Their Physicochemical Properties and Wavelet Transforms into PseAAC. *J. Theor. Biol.* 377, 47–56. doi:10.1016/j.jtbi.2015.04.011

Karunakaran, K. B., Chaparala, S., and Ganapathiraju, M. K. (2019). Potentially Repurposable Drugs for Schizophrenia Identified from its Interactome. *Sci. Rep.*9 (1), 12682–12714. Available at: https://www.nature.com/articles/s41598-019- 48307-w. doi:10.1038/s41598-019-48307-w

Karunakaran, K. B., Chaparala, S., Lo, C. W., and Ganapathiraju, M. K. (2020). Cilia Interactome with Predicted Protein-Protein Interactions Reveals Connections to Alzheimer's Disease, Aging and Other Neuropsychiatric Processes. *Sci. Rep.* 10 (1), 15629–15716. Available at: https://www.nature. com/articles/s41598-020-72024-4. doi:10.1038/s41598-020-72024-4

Karunakaran, K. B., Chaparala, S., and Ganapathiraju, M. K. (2019). Potentially Repurposable Drugs for Schizophrenia Identified from its Interactome. *Sci. Rep.* 9 (1), 12682. doi:10.1038/s41598-019-48307-w

Karunakaran, K. B., Yanamala, N., Boyce, G., Becich, M. J., and Ganapathiraju, M. K. (2021). Malignant Pleural Mesothelioma Interactome with 364 NovelProtein-Protein Interactions. *Cancers* 13 (7), 1660. doi:10.3390/ cancers13071660

Kerslake, R., Hall, M., Randeva, H., Spandidos, D., Chatha, K., Kyrou, I., et al. (2020). Co-expression of P-eripheral O-lfactory R-eceptors with SARS-CoV-2 I-nfection M-ediators: Potential I-mplications beyond L-oss of S-mell as a COVID-19 S-ymptom. *Int. J. Mol. Med.* 46 (3), 949–956. doi:10.3892/ijmm. 2020.4646

Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., et al. (2008). Human Protein Reference Database-2009 Update. *Nucleic Acids Res.* 37 (Suppl. l_1), D767–D772. doi:10.1093/nar/ gkn892

Keskin, O., Tuncbag, N., and Gursoy, A. (2016). Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem. Rev.* 116 (8), 4884–4909. doi:10.1021/acs.chemrev.5b00683

Khan, M., Syed, G. H., Kim, S.-J., and Siddiqui, A. (2015). Mitochondrial Dynamics and Viral Infections: a Close Nexus. *Biochim. Biophys. Acta (Bba) - Mol. CellRes.* 1853 (10), 2822–2833. doi:10.1016/j.bbamcr.2014.12.040

Kotlyar, M., Pastrello, C., Pivetta, F., Lo Sardo, A., Cumbaa, C., Li, H., et al. (2015).In Silico prediction of Physical Protein Interactions and Characterization of Interactome Orphans. *Nat. Methods* 12 (1), 79–84. doi:10.1038/nmeth.3178

Kuleshov, M. V., Stein, D. J., Clarke, D. J. B., Kropiwnicki, E., Jagodnik, K. M., Bartal, A., et al. (2020). The COVID-19 Drug and Gene Set Library. *Patterns (N Y)* 1, 100090. doi:10.1016/j.patter.2020.100090

Kumaki, Y., Day, C. W., Wandersee, M. K., Schow, B. P., Madsen, J. S., Grant, D., et al. (2008). Interferon Alfacon 1 Inhibits SARS-CoV Infection in Human Bronchial Epithelial Calu-3 Cells. *Biochem. biophysical Res. Commun.* 371 (1), 110–113. doi:10.1016/j.bbrc.2008.04.006

Kumar, N., Mishra, B., Mehmood, A., Mohammad Athar, M., and M Shahid Mukhtar, M. S. (2020). Integrative Network Biology Framework Elucidates Molecular Mechanisms of Sars-Cov-2 Pathogenesis. *iScience* 23 (9), 101526. doi:10.1016/j.isci.2020.101526

Kupershmidt, I., Su, Q. J., Grewal, A., Sundaresh, S., Halperin, I., Flynn, J., et al. (2010). Ontology-based Meta-Analysis of Global Collections of High- Throughput Public Data. *PLoS One* 5 (9), 1. doi:10.1371/journal.pone. 0013066

Lambert, D. W., Yarski, M., Warner, F. J., Thornhill, P., Parkin, E. T., Smith, A. I., et al. (2005). Tumor Necrosis Factor-α Convertase (ADAM17) Mediates Regulated Ectodomain Shedding of the Severe-Acute Respiratory Syndrome-Coronavirus (SARS-CoV) Receptor, Angiotensin-Converting Enzyme-2 (ACE2). *J. Biol. Chem.* 280 (34), 30113–30119. doi:10.1074/jbc.m505111200

Li, G., and De Clercq, E. (2020). Therapeutic Options for the 2019 Novel Coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* 19, 149–150. doi:10.1038/ d41573-020-00016-0

Li, H. S., Kuok, D. I. T., Cheung, M. C., Ng, M. M. T., Ng, K. C., Hui, K. P. Y., et al. (2018). Effect of Interferon Alpha and Cyclosporine Treatment Separately andin Combination on Middle East Respiratory Syndrome Coronavirus (Mers-cov) Replication in a Human *In-Vitro* and *Ex-Vivo* Culture Model. *Antivir. Res.* 155,89–96. doi:10.1016/j.antiviral.2018.05.007

Li, Y., Klena, N. T., Gabriel, G. C., Liu, X., Kim, A. J., Lemke, K., et al. (2015). Global Genetic Analysis in Mice Unveils central Role for Cilia in Congenital Heart Disease. *Nature* 521 (7553), 520–524. doi:10.1038/nature14269

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: Gene Set Analysis Toolkit with Revamped UIs and APIs. *Nucleic Acids Res.* 9, 1. doi:10.1093/nar/gkz401

Lin, S.-C., Ho, C.-T., Chuo, W.-H., Li, S., Wang, T. T., and Lin, C.-C. (2017). Effective Inhibition of MERS-CoV Infection by Resveratrol. *BMC Infect. Dis.* 17(1), 144. doi:10.1186/s12879-017-2253-8

Liu, X., Yagi, H., Saeed, S., Bais, A. S., Gabriel, G. C., Chen, Z., et al. (2017). The Complex Genetics of Hypoplastic Left Heart Syndrome. *Nat. Genet.* 49 (7), 1152–1159. doi:10.1038/ng.3870

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* 45 (6), 580–585.

Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., et al. (2020). A Reference Map of the Human Binary Protein Interactome. *Nature* 580, 402–408. doi:10.1038/s41586-020-2188-x

Malavia, T., Chaparala, S., Wood, J., Chowdari, K., Prasad, K., McClain, L., et al. (2017). *Generating Testable Hypotheses for Schizophrenia and Rheumatoid Arthritis Pathogenesis by Integrating Epidemiological, Genomic and Protein Interaction Data Npj Schizophrenia*. London, United Kingdom: Nature.

Malavia, T. A., Chaparala, S., Wood, J., Chowdari, K., Prasad, K. M., McClain, L., et al. (2017). Generating Testable Hypotheses for Schizophrenia and Rheumatoid Arthritis Pathogenesis by Integrating Epidemiological, Genomic, and Protein Interaction Data. *npj Schizophr* 3 (1), 11. doi:10.1038/s41537-017-0010-z

Mehta, D. R., Ashkar, A. A., and Mossman, K. L. (2012). The Nitric Oxide Pathway Provides Innate Antiviral protection in Conjunction with the Type I Interferon Pathway in Fibroblasts. *PLoS One* 7, e31688. doi:10.1371/journal. pone.0031688

Messner, C. B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., et al. (2020). Ultra-high-throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst.* 11 (1), 11–24. e4. doi:10.1016/j.cels.2020.05.012

Moore, M. J., Blachere, N. E., Fak, J. J., Park, C. Y., Sawicka, K., Parveen, S., et al. (2018). ZFP36 RNA-Binding Proteins Restrain T Cell Activation and Anti-viral Immunity. *Elife* 7, 7. doi:10.7554/eLife.33057

Morgan, D. O. (2007). *Figure 5-2 Stages of Late M Phase in a Vertebrate Cell*. London, United Kingdom: New Science Press, 1.

Morita, E., Colf, L. A., Karren, M. A., Sandrin, V., Rodesch, C. K., and Sundquist, W. I. (2010). Human ESCRT-III and VPS4 Proteins Are Required for Centrosome and Spindle Maintenance. *Proc. Natl. Acad. Sci.* 107 (29), 12889–12894. doi:10.1073/pnas.1005938107

Nugent, K. M., and Shanley, J. D. (1984). Verapamil Inhibits Influenza A Virus Replication. *Arch. Virol.* 81 (1-2), 163–170. doi:10.1007/BF01309305

Omeragic, A., Kara-Yacoubian, N., Kelschenbach, J., Sahin, C., Cummins, C. L., Volsky, D. J., et al. (2019). Peroxisome Proliferator-Activated Receptor-Gamma Agonists Exhibit Anti-inflammatory and Antiviral Effects in an EcoHIV Mouse Model. *Sci. Rep.* 9 (1). Available at: https://www.nature.com/articles/s41598-019-45878-6. doi:10.1038/s41598-019-45878-6

Orii, N., and Ganapathiraju, M. K. (2012). Wiki-pi: a Web-Server of Annotated Human Protein-Protein Interactions to Aid in Discovery of Protein Function. *PLoS One* 7 (11), e49029. doi:10.1371/journal.pone.0049029

Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., et al. (2021). Large-scale Multi-Omic Analysis of COVID-19 Severity. *Cell Syst.* 12 (1), 23–40. e7. doi:10.1016/j.cels.2020.10.003

Panda, D., Pascual-Garcia, P., Dunagin, M., Tudor, M., Hopkins, K. C., Xu, J., et al. (2014). Nup98 Promotes Antiviral Gene Expression to Restrict RNA Viral Infection in Drosophila. *Proc. Natl. Acad. Sci. USA* 111 (37), E3890–E3899. doi:10.1073/pnas.1410087111

Perry, C. M., and Noble, S. (1999). Didanosine. *Drugs* 58 (6), 1099–1135. doi:10.2165/00003495-199958060-00009

Pfefferle, S., Schöpf, J., Kögl, M., Friedel, C. C., Müller, M. A., Carbajo-Lozoya, J., et al. (2011). The SARS-Coronavirus-Host Interactome: Identification of Cyclophilins as Target for Pan-Coronavirus Inhibitors. *Plos Pathog.* 7 (10), e1002331. doi:10.1371/journal.ppat.1002331

Pincetic, A., Kuang, Z., Seo, E. J., and Leis, J. (2010). The Interferon-Induced Gene ISG15 Blocks Retrovirus Release from Cells Late in the Budding Process. *J. Virol.* 84 (9), 4725–4736. doi:10.1128/jvi.02478-09

Quintana, V. M., Selisko, B., Brunetti, J. E., Eydoux, C., Guillemot, J. C., Canard, B., et al. (2020). Antiviral Activity of the Natural Alkaloid Anisomycin against Dengue and Zika Viruses. *Antivir. Res.* 176, 104749. doi:10.1016/j.antiviral. 2020.104749

Raja, K., Subramani, S., and Natarajan, J. (2013). PPInterFinder-a Mining Tool for Extracting Causal Relations on Human Proteins from Literature. *Database (Oxford)* 2013, bas052. doi:10.1093/database/bas052

Reghunathan, R., Jayapal, M., Hsu, L.-Y., Chng, H.-H., Tai, D., Leung, B. P., et al. (2005). Expression Profile of Immune Response Genes in Patients with Severe Acute Respiratory Syndrome. *BMC Immunol.* 6 (1), 2. doi:10.1186/1471-2172-6-2

Rothan, H. A., and Byrareddy, S. N. (2020). The Epidemiology and Pathogenesis of Coronavirus Disease (COVID-19) Outbreak. *J. Autoimmun.* 109, 102433. doi:10.1016/j.jaut.2020.102433

Ryang, J., Yan, Y., Song, Y., Liu, F., and Ng, T. B. (2019). Anti-HIV, Antitumor and Immunomodulatory Activities of Paclitaxel from Fermentation Broth Using Molecular Imprinting Technique. *AMB Expr.* 9 (1), 194. doi:10.1186/s13568-019-0915-1

Scherer, M., and Stamminger, T. (2016). Emerging Role of PML Nuclear Bodies in Innate Immune Signaling. *J. Virol.* 90 (13), 5850–5854. doi:10.1128/jvi.01979-15

Sehgal, N., Kumawat, K. L., Basu, A., and Ravindranath, V. (2012). Fenofibrate Reduces Mortality and Precludes Neurological Deficits in Survivors in Murine Model of Japanese Encephalitis Viral Infection. *PloS one* 7 (4), e35427. doi:10.1371/journal.pone.0035427

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303

Shendi, A., Hung, R. Y., Caplin, B., Griffiths, P., and Harber, M. (2019). The Use of Sirolimus in Patients with Recurrent Cytomegalovirus Infection after Kidney Transplantation: A Retrospective Case Series Analysis. *Saudi J. Kidney Dis. Transpl.* 30 (3), 606–614. doi:10.4103/1319-2442.261333

Sidaway, P. (2020). COVID-19 and Cancer: what We Know So Far. *Nat. Rev. Clin. Oncol.* 17 (6), 336. doi:10.1038/s41571-020-0366-2

Sims, A. C., Baric, R. S., Yount, B., Burkett, S. E., Collins, P. L., and Pickles, R. J. (2005). Severe Acute Respiratory Syndrome Coronavirus Infection of Human Ciliated Airway Epithelia: Role of Ciliated Cells in Viral Spread in the Conducting Airways of the Lungs. *J. Virol.* 79 (24), 15511–15524. doi:10.1128/jvi.79.24.15511-15524.2005

Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., et al. (2011). Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data. *Sci. Transl Med.* 3 (96), 96ra77. doi:10.1126/scitranslmed.3001318

Spirin, V., and Mirny, L. A. (2003). Protein Complexes and Functional Modules in Molecular Networks. *Proc. Natl. Acad. Sci.* 100 (21), 12123–12128. doi:10.1073/pnas.2032324100

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a General Repository for Interaction Datasets. *Nucleic Acids Res.* 34 (Suppl. 1_1), D535–D539. doi:10.1093/nar/gkj109

Sudarsanam, T. D., Sahni, R. D., and John, G. T. (2006). Leflunomide: a Possible Alternative for Gangciclovir Sensitive and Resistant Cytomegalovirus Infections. *Postgrad. Med. J.* 82 (967), 313–314. doi:10.1136/pgmj.2005.038521 Sungnak, W., Huang, N., Bécavin, C., Berg, M., and Network, H. C. A. (2020). *SARS-CoV-2 Entry Genes Are Most Highly Expressed in Nasal Goblet and Ciliated Cells within Human Airways*. ArXiv: Cornell University.

Torriani, G., Trofimenko, E., Mayor, J., Fedeli, C., Moreno, H., Michel, S., et al. (2019). Identification of Clotrimazole Derivatives as Specific Inhibitors of Arenavirus Fusion. *J. Virol.* 93 (6), e01744–18. doi:10.1128/JVI.01744-18

Touret, F., Gilles, M., Barral, K., Nougairède, A., Decroly, E., de Lamballerie, X., et al. (2020). In Vitro *Screening of a FDA Approved Chemical Library Reveals Potential Inhibitors of SARS-CoV-2 Replication*. BioRxiv: Cold Spring Harbor Laboratory.

Trepte, P., Buntru, A., Klockmeier, K., Willmore, L., Arumughan, A., Secker, C., et al. (2015). DULIP: a Dual Luminescence-Based Co-immunoprecipitation Assay for Interactome Mapping in Mammalian Cells. *J. Mol. Biol.* 427 (21), 3375–3388. doi:10.1016/j.jmb.2015.08.003

Vishwajit Nimgaonkar, M. D. P. (2024). *Stanley Medical Research I, University of P. Acetazolamide for Treatment Resistant Schizophrenia*. Stanley Medical Research Institute. Available at: https://clinicaltrials.gov/ct2/show/NCT04887792

Vishwajit Nimgaonkar, M. D. P. (2022). *Stanley Medical Research I, University of P. Cromoglycate Adjunctive Therapy for Outpatients with Schizophrenia*. Stanley Medical Research Institute.. Available at: https://clinicaltrials.gov/ct2/show/results/NCT03794076

Wang, Z., and Zhang, J. (2007). In Search of the Biological Significance of Modular Structures in Protein Networks. *Plos Comput. Biol.* 3 (6), e107. doi:10.1371/ journal.pcbi.0030107

Warner, F. J., Lew, R. A., Smith, A. I., Lambert, D. W., Hooper, N. M., and Turner, A. J. (2005). Angiotensin-converting Enzyme 2 (ACE2), but Not ACE, Is Preferentially Localized to the Apical Surface of Polarized Kidney Cells. *J. Biol. Chem.* 280 (47), 39353–39362. doi:10.1074/jbc.m508914200

Weston, S., Coleman, C. M., Haupt, R., Logue, J., Matthews, K., Li, Y., et al. (2020). Broad Anti-coronaviral Activity of FDA Approved Drugs against SARS-CoV-2 *in Vitro* and SARS-CoV *in Vivo*. *Journal of virology* 94 (21), e01218–e01220.

Wilson, K. S. (2015). Figure 1.4 – Diagram of Flagellum Origin and Cross-Section. *Probing Mech. Forces Flagella by Manipulation Media Viscosity Axonemal Struct.* 9, 1.

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., et al. (2008). DrugBank: a Knowledgebase for Drugs, Drug Actions and Drug Targets. *NucleicAcids Res.* 36 (Database issue), D901–D906. doi:10.1093/nar/gkm958

Wu, G., Dawson, E., Duong, A., Haw, R., and Stein, L. (2014). ReactomeFIViz: aCytoscape App for Pathway and Network-Based Data Analysis. *F1000Res* 3, 146. doi:10.12688/f1000research.4431.2

Wu, W., Li, R., Li, X., He, J., Jiang, S., Liu, S., et al. (2016). Quercetin as an Antiviral Agent Inhibits Influenza A Virus (IAV) Entry. *Viruses* 8 (1), 6. doi:10.3390/v8010006 Yang, S., Fu, C., Lian, X., Dong, X., and Zhang, Z.

(2019).

Understanding Human-Virus Protein-Protein Interactions Using a Human Protein Complex-Based Analysis Framework. *MSystems* 4, 1. doi:10.1128/mSystems.00303-18

You, Z. H., Lei, Y. K., Zhu, L., Xia, J., and Wang, B. (2013). Prediction of Protein-Protein Interactions from Amino Acid Sequences with Ensemble Extreme Learning Machines and Principal Component Analysis. *BMC bioinformatics* 14 (8), S10. doi:10.1186/1471-2105-14-S8-S10

Zhao, Y., Ren, J., Fry, E. E., Xiao, J., Townsend, A. R., and Stuart, D. I. (2018). Structures of Ebola Virus Glycoprotein Complexes with Tricyclic Antidepressant and Antipsychotic Drugs. *J. Med. Chem.* 61 (11), 4938–4945. doi:10.1021/acs.jmedchem.8b00350

Zhou, Y., Hou, Y., Shen, J., Mehra, R., Kallianpur, A., Culver, D. A., et al. (2020). ANetwork Medicine Approach to Investigation and Population-Based Validation of Disease Manifestations and Drug Repurposing for COVID-19. *PLoS Biology* 18 (11), e3000970.

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020). Network- based Drug Repurposing for Novel Coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* 6 (1), 14. doi:10.1038/s41421-020-0153-3

Zhu, J., Zhang, Y., Ghosh, A., Cuevas, R. A., Forero, A., Dhar, J., et al. (2014). Antiviral Activity of Human OASL Protein Is Mediated by Enhancing Signaling of the RIG-I RNA Sensor. *Immunity* 40 (6), 936–948. doi:10.1016/j.immuni. 2014.05.007

Interactome-based framework to translate disease genetic data into biological and clinical insights

7. GPX4-associated Sedaghatian Type Spondylometaphyseal Dysplasia: A Protein Interactome Perspective

The experimental chapter is based on the following pre-print publication:
Karunakaran, Kalyani B., N. Balakrishnan, and Madhavi K. Ganapathiraju. GPX4-associated Sedaghatian type spondylometaphyseal dysplasia: A protein interactome perspective. bioRxiv (2022).

Summary of this chapter

In this chapter, I demonstrate how the interactomic framework was used to gain biological and clinically translatable insights into spondylometaphyseal dysplasia, Sedaghatian type (SMDS), a rare and lethal skeletal dysplasia inherited in an autosomal recessive manner and caused by mutations in the gene GPX4. My aim was to expand the functional landscape of this poorly studied disorder. For this, I constructed three phenotype (i.e., SMDS)-centric interactomes and one GPX4-centric interactome. They contained experimentally determined and computationally predicted PPIs. I extensively characterised these interactomes using gene-phenotype association data, functional module analysis and transcriptomic analysis. I conducted extensive literature review to demonstrate the potential biological relevance of seven proteins predicted as novel interactors of GPX4 to SMDS. Lastly, I performed a comparative analysis of the differential expression profiles of dysplasia patients and the profiles induced by drugs targeting the neighbourhood network of GPX4 to identify 11 repurposable drugs for SMDS. Further, I found another potential drug candidate based on the proximity of its targets to GPX4. In summary, this study provided a functional landscape of SMDS, allowing biologists to prioritize genes, functional modules and drugs for therapeutic interventions in SMDS.

Contribution to this chapter (85%)

- Designed the study and developed the methodology of the project, which included interactome construction, functional characterisation, drug repurposing and network proximity analysis
- Curated all the datasets, performed all the analyses and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

GPX4-associated Sedaghatian Type Spondylometaphyseal Dysplasia: A Protein Interactome Perspective

Kalyani B. Karunakaran[1], N. Balakrishnan[1] and Madhavi K. Ganapathiraju[2,3*]

[1]Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, 560012

[2]Department of Biomedical Informatics, School of Medicine,

and [3]Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, 15213

*Corresponding author: madhavi@pitt.edu

Abstract

Spondylo**m**etaphyseal **d**ysplasia, **S**edaghatian type (SMDS) is a rare and lethal skeletal dysplasia inherited in an autosomal recessive manner and caused by mutations in GPX4. In order to expand the functional landscape of this poorly studied disorder and accelerate the discovery of biologically insightful and clinically actionable targets, we constructed SMDS-centric and GPX4-centric protein-protein interaction (PPI) networks, augmented with novel protein interactors predicted by our HiPPIP algorithm. The SMDS-centric networks included those that showed the interconnections of GPX4 with other putative SMDS-associated genes and genes associated with other skeletal dysplasias. The GPX4-centric network showed the interconnections of GPX4 with genes whose perturbation has been known to affect GPX4 expression. We discovered that these networks either contained or were enriched with genes associated with specific SMDS pathophenotypes, tissue-naïve/fetus-specific functional modules and genes showing elevated expression in brain and/or testis similar to GPX4. We identified 7 proteins as novel interactors of GPX4 (APBA3, EGR4, FUT5, GAMT, GTF2F1, MATK and ZNF197) and showed their potential biological relevance to GPX4 or SMDS. Comparative transcriptome analysis of expression profiles associated with chondroplasia and immune-osseous dysplasia versus drug-induced profiles revealed 11 drugs that targeted the neighborhood network of GPX4 and other putative SMDS-associated genes. Additionally, resveratrol, which is currently being tested against a skeletal dysplasia in a clinical trial, was identified as another potential candidate based on the proximity of its targets to GPX4.

Introduction

Spondylometaphyseal **d**ysplasia**s** (SDs) constitute a heterogeneous group of skeletal dysplasias characterized by abnormal development of the spine and the metaphyses of tubularbones, progressive growth and mobility impairment. **S**pondylo**m**etaphyseal **d**ysplasia, **S**edaghatian type (SMDS) is a rare and lethal SD characterized by cupping (or the inward bulging) of the metaphyseal region in long bones (i.e. metaphyseal cupping), flattening of thevertebrae (platyspondyly), abnormal shoulder bone (scapula) morphology, short upper limbs (rhizomelia), abnormal heart rate (arrhythmia), impaired electric conduction from the atrial to the ventricular chamber of the heart (atrioventricular block), cardiorespiratory arrest and several pathologic features in the brain such as the absence/underdevelopment (i.e. agenesis/hypogenesis) of the corpus callosum and underdevelopment of the cerebellum (cerebellar hypoplasia).[1-8] Respiratory distress is the primary cause of death for infants born with this congenital condition.[1,7,8] Four patients living with this debilitating disorder exhibit delayed cognitive development, severe hypotonia characterized by a lack of neck muscle control and an inability to sit or walk unsupported, and intractable epileptic seizures.[9] SMDS shows an autosomal recessive pattern of inheritance, and has been attributed to at least 3 mutations in the gene GPX4 (located in the chromosomal region 19p13.3), which codes for the protein *glutathione peroxidase*, namely, c.381C>A (p.Tyr90Ter/Y90*) (single nucleotide variant in exon 3 that introduces a translation stop before the indicated amino acid position in the protein), c.587+5G>A (single nucleotide variant in intron 4 causing splicing out of a part of exon 4, indicated here with reference to the nearest coding sequence) and c.588-8_588- 4del (5 bp deletion in intron 4 causing skipping of exon 5, indicated here with reference to the nearest coding sequence).[7,8,10] GPX4 is a multifunctional ('moonlighting') protein that not only catalyzes the reduction of lipoproteins and membrane phospholipids to prevent oxidative damage resulting from lipid peroxidation,[11] but also serves as a structural protein in the midpiece of mature sperm.[12] GPX4 activity and synthesis are influenced by Selenium levels, as its catalytic site contains a selenocysteine residue.[13] Therapeutic options for SMDS are dismally limited to controlling oxidative damage using vitamin E, n-acetyl-cysteine and coenzyme Q10 supplements. Except for some early findings on the clinical presentation of SMDS[1-6] and the characterization of the GPX4 mutations,[7,8] the etiology of this severe lethal dysplasia remains largely unexplored. Hence, novel approaches are needed to elucidate the broad themes underlying this disorder.

Protein-protein interactions (PPIs) drive the cellular machinery by facilitating a variety of biological processes including signal transduction, formation of cellular structures and enzymatic complexes. Disease-associated variants are enriched in protein cores and protein interaction interfaces.[14] Variants localized to the protein core may disrupt its tertiary structure and abolish all chances of the protein interacting with any of its interaction partners (node removal in the interactome).[14] Variants localized to interaction interfaces may perturb specific interactions (edge perturbation in the interactome).[14] Over two-thirds of disease- associated variants alter binding affinity or even establish novel interactions.[15] Hence, geneticmutations may perturb proteins and this effect may spread in the PPI network (or the 'interactome') affecting other proteins, posing deeper implications for disease development,[16]for example multiple pathophenotypes in a single disease despite the disease being associatedwith a single genotype.[17] Several studies, led by our own group and others, have successfully traced shared genetics and symptomatology among different diseases back to this complex network of PPIs.[18-21] Therefore, for functional interpretation of genetic variants associated with complex disorders, it is imperative that we place variants in the complex web of PPIs.

Mapping disease-associated variants onto PPI networks will pull in more disease-associated genes into the network, offering an opportunity to analyze communities of proteins involved in mechanisms relevant to disease etiology.

In this study, we adopted an interactome-based approach to construct an integrated functional landscape for GPX4-associated SMDS in relation with other SDs, phenotypically similar dysplasias, genes other than GPX4 speculated to be associated with SMDS and genes whose perturbation affects GPX4 expression. We identified functional modules and pathophenotypes enriched in this landscape, and used human transcriptomic data to identify groups of genes clustering with GPX4, which showed elevated expression in specific tissues.Finally, we adopted two approaches to propose a few repurposable drugs, namely, comparative analysis of drug-induced and disease-associated transcriptomes, and examining the network proximity of GPX4 to the protein targets of drugs that are being used or being tested against other skeletal dysplasias.

Results

We assembled the known PPIs of GPX4 from HPRD[22] (Human Protein Reference Database) and BioGRID[23] (Biological General Repository for Interaction Datasets), and predicted its

novel PPIs by applying the HiPPIP algorithm described in our earlier work.[20] HiPPIP computes features of protein pairs such as cellular localization, molecular function, biological process membership, genomic location of the gene, and gene expression in microarray experiments, and classifies the pairwise features as *interacting* or *non-interacting* based on a random forest model.[20] HiPPIP was shown to have high precision in its original evaluation[20] and to be superior to state-of-the-art algorithms in a recent evaluation.[24] Eighteen novel PPIs predicted by HiPPIP were experimentally tested and all eighteen were found validated to be true PPIs by collaborators as summarized in our recent work.[25] GPX4 has two known interactions, with MAPK13 (*mitogen-activated protein kinase 13*) and PRDX6 (*peroxiredoxin 6*). Additionally, we predicted seven novel interactions with APBA3 (*amyloidbeta precursor protein binding family A member 3*), EGR4 (*early growth response 4*), FUT5 (*fucosyltransferase 5*), GAMT (*guanidinoacetate N-methyltransferase*), GTF2F1 (*general transcription factor IIF subunit 1*), MATK (*megakaryocyte-associated tyrosine kinase*) and ZNF197 (*zinc finger protein 197*). **Table 1** lists the evidence supporting the biological relevance of these novel interactors to GPX4 or SMDS, identified from our analyses.

Protein interactome analysis is a useful tool for elucidating biologically relevant relationships existing at a higher level among genes, which may not be apparent by examining individual genes. Therefore, for further mechanistic characterization of GPX4, we inspected its interconnections in the human interactome with (a) other genes speculated to be associated with SMDS in the DisGeNET[26] database, (b) genes associated with other types of SDs, (c) genes associated with disorders that shared phenotypic similarity with SMDS and (d) genes whose perturbation was known to cause significant overexpression or underexpression of GPX4. Following this, we used the HumanBase toolkit[27] to identify functional modules in each of these GPX4-associated interactomes. HumanBase employs shared k-nearest-neighbors and the Louvain community-finding algorithm to cluster the genes sharing the same network neighborhoods and similar GO biological processes into functional modules. Additionally, we compiled the genes involved in 12 major pathophenotypes associated with SMDS from the MONARCH[28] database and computed the statistical significance of the enrichment of these genes in each of the interactomes, against a background of 15451 genes with phenotype associations. The selected phenotypes were atrial septal defect (290 genes), atrioventricular block (54 genes), cardiorespiratory arrest (10 genes), cerebellar hypoplasia (219 genes), myocarditis (7 genes), agenesis of corpus callosum (240 genes), arrhythmia (355

1   **Table 1: Biological relevance of novel interactors of GPX4.** The table lists different pieces of evidence gathered from our study that support the biological validity of the
2   computationally predicted novel interactors of GPX4 to SMDS or GPX4 itself. Note that in the table, "phenotypic similarity" refers to "phenotypic similarity with SMDS"
3   (unless specified otherwise). SMDS: SpondyloMetaphyseal Dysplasia, Sedaghatian type, SD: Spondylometaphyseal Dysplasia.

Gene	Potential biological relevance to GPX4/SMDS
APBA3	a) An intermediate interactor of GPX4 and IARS2, an arrhythmia-and X-linked spondyloepimetaphyseal dysplasia-associated gene (the latter is a bone dysplasia phenotypically similar to SMDS) b) A known interactor of BCR1, mutations in which leads to underexpression of KIAA0586, a gene associated with Joubert syndrome (which shares several phenotypes with SMDS) c) An intermediate interactor of GPX4 and APP, which is targeted by resveratrol, a drug currently in clinical trials for skeletal dysplasias d) Co-occurred with GPX4 and its other novel interactors (GAMT and GTF2F1), and with potential SMDS-associated genes (RPS19) and/or their novel interactors (NEDD8, PSMD8, XRCC1 and DXO) in a cellular oxidant detoxification module e) Co-occurred along with novel interactors of GPX4 (GTF2F1) and novel interactors of SD-associated genes (RXRB, SLC48A1 and SF3A2) in an epithelial morphogenesis and fetus-specific protein catabolic process f) Co-occurred with other novel interactors of GPX4 (GTF2F1) and novel interactors of genes associated with phenotypically similar bone dysplasias (DNASE1L1) in a toxic substance response module
EGR4	a) A transcription factor that regulates hind brain development in *Xenopus* b) Connected via DNMT3L to a novel interactor of the (potentially) SMDS-associated ARTN called LDB1, mutations in which lead to patterning defects c) Showed expression-based clustering with GPX4 due to elevated expression in brain tissues d) EGR4 mice mutants shared 6 phenotypes with GPX4 mice mutants: male infertility, oligozoospermia, small testis, abnormal sperm head morphology, decreased testis weight and kinked sperm flagellum e) Co-occurred in an adenylate cyclase-activating GPCR signaling module along with other novel interactors of GPX4 (MATK and ZNF197) and with potential SMDS-associated genes (ARTN and AGRP) and/or their novel interactors (CHRNB4, PSG9, TONSL, CAMK4, FCGBP, DIO1 and DHODH) f) Co-occurred in extracellular organization/cartilage development modules with other novel interactors of GPX4 (FUT5 and ZNF197) and with SD-associated genes (COL2A1 and TRPV4) and novel interactors of SD-associated genes (SPDYA, MST1, DRD2, KMT2D, GP5, LY6H, SERPINA4, CLSPN, IKZF2 and TM6SF2) g) Co-occurred in autophosphorylation and neurogenesis modules with other novel interactors of GPX4 (FUT5 and ZNF197) and genes whose perturbation leads to differential expression of GPX4 (NTRK1 and DNMT3L)
FUT5	a) Showed expression-based clustering with GPX4 due to elevated expression in the testis b) Co-occurred in extracellular organization/cartilage development modules with other novel interactors of GPX4 (EGR4 and ZNF197) and with SD-associated genes (COL2A1 and TRPV4) and novel interactors of SD-associated genes (SPDYA, MST1, DRD2, KMT2D, GP5, LY6H, SERPINA4, CLSPN, IKZF2 and TM6SF2) c) Co-occurred in autophosphorylation and neurogenesis modules with other novel interactors of GPX4 (EGR4 and ZNF197) and genes whose perturbation leads to differential expression of GPX4 (NTRK1 and DNMT3L)

GAMT	a) An intermediate interactor connecting GPX4 with the ciliary protein TERF1 b) GAMT mice mutants shared 4 phenotypes with GPX4 mice mutants decreased body weight: decreased body weight, oligozoospermia, reduced male fertility and postnatal lethality c) Co-occurred with GPX4 and its other novel interactors (APBA3 and GTF2F1), and with potential SMDS-associated genes (RPS19) and/or their novel interactors (NEDD8, PSMD8, XRCC1 and DXO) in a cellular oxidant detoxification module
GTF2F1	a) An intermediate interactor of GPX4 and PAM16, a Megarbane type SMD gene associated with metaphyseal cupping and platyspondyly b) A known interactor of ABL1, mutations in which leads to underexpression of KIAA0586, a gene associated with Joubert syndrome (which shares several phenotypes with SMDS) c) An intermediate interactor of GPX4 and two targets (AHR and CSNK2A1) of resveratrol, a drug currently in clinical trials for skeletal dysplasias d) Connected via CTDSP1 to a novel interactor of the (potentially) SMDS-associated ARTN called LDB1, mutations in which lead to patterning defects e) Co-occurred with GPX4 and its other novel interactors (APBA3 and GAMT), and with potential SMDS-associated genes (RPS19) and/or their novel interactors (NEDD8, PSMD8, XRCC1 and DXO) in a cellular oxidant detoxification module f) Co-occurred along with other novel interactors of GPX4 (GTF2F1) and novel interactors of SD-associated genes (RXRB, SLC48A1 and SF3A2) in an epithelial morphogenesis module g) Co-occurred with other novel interactors of GPX4 (APBA3) and novel interactors of genes associated with phenotypically similar bone dysplasias (DNASE1L1) in a toxic substance response module
MATK	a) Showed expression-based clustering with GPX4 due to elevated expression in brain tissues b) Co-occurred in an adenylate cyclase-activating GPCR signaling module along with other novel interactors of GPX4 (EGR4 and ZNF197) and with potential SMDS-associated genes (ARTN and AGRP) and/or their novel interactors (CHRNB4, PSG9, TONSL, CAMK4, FCGBP, DIO1 and DHODH) c) Co-occurred in a fetus-specific extracellular organization/cartilage development modules with other novel interactors of GPX4 (FUT5 and ZNF197) and with SD-associated genes (COL2A1 and TRPV4) and novel interactors of SD-associated genes (SPDYA, MST1, DRD2, KMT2D, GP5, LY6H,SERPINA4, CLSPN, IKZF2, TM6SF2, CD3G, FICD, NPTX1 and RASGRP4)
ZNF197	a) An intermediate interactor of GPX4 and the myocarditis-associated VHL b) An intermediate interactor of GPX4 and IARS2, an arrhythmia-and X-linked spondyloepimetaphyseal dysplasia-associated gene (the latter is a bone dysplasia phenotypically similar to SMDS) c) A known interactor of TRIM28, mutations in which leads to underexpression of IFT140, a gene associated with Joubert syndrome (which shares several phenotypes with SMDS) d) ZNF197 mice mutants shared the phenotype, reduced male fertility, with GPX4 mice mutants e) Recruited by VHL to inhibit HIF1A transcriptional activity. <u>Speculation</u>: GPX4 exerts its inhibitory activity on VEGFA, a critical component of the VEGF signaling pathway that is known to play a role in bone development, via its direct interaction with the ZNF197-VHL complex. Perturbed GPX4 activity in SMDS may remove this inhibitory effect and promote abnormal VEFGA protein expression and bone development f) Co-occurred in an adenylate cyclase-activating GPCR signaling module along with other novel interactors of GPX4 (MATK and EGR4) and with potential SMDS-associated genes (ARTN and AGRP) and/or their novel interactors (CHRNB4, PSG9, TONSL, CAMK4, FCGBP, DIO1 and DHODH)

	g) Co-occurred in extracellular organization/cartilage development modules with other novel interactors of GPX4 (EGR4 and ZNF197) and with SD-associated genes (COL2A1 and TRPV4) and novel interactors of SD-associated genes (SPDYA, MST1, DRD2, KMT2D, GP5, LY6H, SERPINA4, CLSPN, IKZF2 and TM6SF2)
	h) Co-occurred in a humoral immune response module along genes associated with bone dysplasia that are phenotypically similar to SMDS (IHH) and novel interactors of bone dysplasia genes (TEX28, HHIPL2, SUSD4 and SLC30A10)
	i) Co-occurred in a fetus-specific lipid storage/reactive oxygen species biosynthetic process module with novel interactors of bone dysplasia genes (CCL20, AMDHD2, HHIPL2 and SLC30A10)
	j) Co-occurred in autophosphorylation and neurogenesis modules with other novel interactors of GPX4 (EGR4 and FUT5) and genes whose perturbation leads to differential expression of GPX4 (NTRK1 and DNMT3L)

4

genes), pachygyria (133 genes; cerebral cortex malformation characterized by a fewer number of abnormally wide gyri), metaphyseal cupping (18 genes), platyspondyly (109 genes), abnormal scapula morphology (4 genes) and abnormality of the ribs (85 genes).

Interactome of GPX4 and other genes potentially associated with SMDS

The human sperm contains 648 short exon-sized sequences called sperm RNA elements corresponding to a range of unique coding and non-coding transcripts, the presence of which may increase the likelihood of live birth through natural conception in idiopathic infertile couples.[29] A sperm RNA element that was mapped to GPX4 contained an SMDS-associated variant.[29] The DisGeNET[26] database lists 6 additional SMDS-associated genes extracted from this work by the BeFree[30] text mining system (i.e. other than GPX4 which had a gene-disease association or GDA score of 0.72), namely, AGRP (agouti related neuropeptide), ARNTL (aryl hydrocarbon receptor nuclear translocator like), ARTN (artemin), LOH19CR1 (loss of heterozygosity, 19, chromosomal region 1), PSD4 (pleckstrin and Sec7 domain containing 4) and RPS19 (ribosomal protein S19) (GDA score = 0.01). BeFree employs a kernel-based approach based on morphosyntactic and dependency information to identify gene-disease associations.[30] We extracted the PPI network that connects these 6 genes (candidates) to GPX4 (target) through shortest paths as well as their own interactors (**Fig. 1**), and found that GPX4 connects to AGRP, ARNTL, ARTN, PSD4 and RPS19 through 22 intermediate interactors including 10 novel interactors (i.e. those revealed by computationally predicted PPIs in this work). These novel interactors include 6 novel interactors of GPX4 (APBA3, EGR4, GAMT, GTF2F1, MATK and ZNF197), a direct interactor of ARTN (LDB1), a directinteractor of AGRP (UPP1) and 2 novel intermediate interactors connecting GPX4 with AGRP (VHL and GHRL).

This interactome was enriched in genes associated with myocarditis (P-value = 5.33E-04, odds ratio = 55.8, genes: VHL and GPX4), atrial septal defect (P-value = 0.016, odds ratio = 3.37, genes: USP9X, SMAD4, GPX4, RPS19 and EP300) and platyspondyly (P-value = 0.018, odds ratio = 5.38, genes: GPX4, SMAD4 and TONSL). Myocarditis-associated VHL was connected to GPX4 via ZNF197. Platypondyly-associated TONSL was predicted to be a novel interactor of RPS19.

Using the HumanBase toolkit,[27] we isolated the functional modules enriched in the interactome containing the intermediate interactors connecting GPX4 to the additional 6 potential SMDS-associated genes, as well as the direct known and novel interactors of these 6 genes. We extracted tissue-specific functional modules containing genes specific to the human fetal tissue and tissue-naive (or 'global') modules containing genes playing identical roles across the tissues. Fetus-specific modules were examined based on the assumption that modules putatively active in the fetal tissue may be relevant to the congenital and neonatal underpinning of SMDS. A *cellular oxidant detoxification* module was identified in both
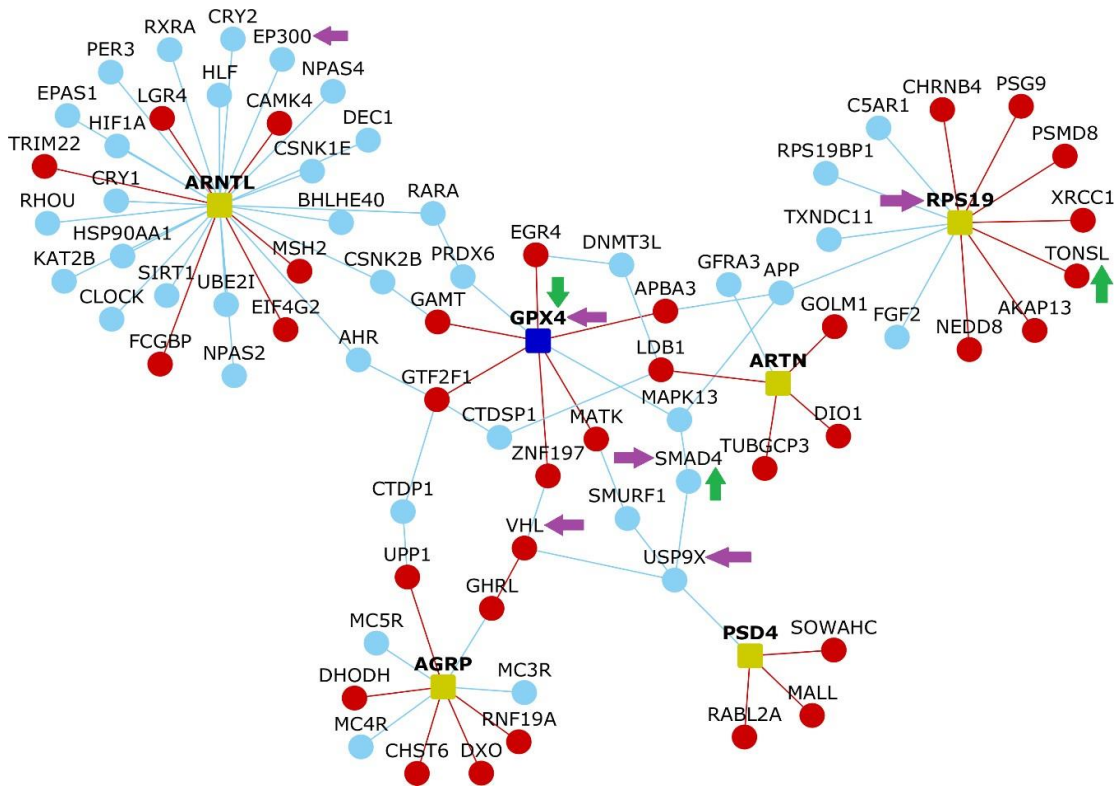
Figure 1: Interactions of GPX4 with other genes speculated to be associated with Sedaghatian type spondylometaphyseal dysplasia. This network diagram shows the shortest paths connecting the genes from the DisGeNET database (dull green colored nodes) identified through text mining to GPX4. Nodes depict proteins and edges depict PPIs. Red and light blue colored nodes denote novel and known interactors respectively. Red and light blue colored edges denote novel and known PPIs respectively. Purple colored arrows indicategenes associated with cardiac defects, whereas green colored arrows indicate those associatedwith skeletal defects.

global (M4: *Q*-value = 2.15E-03) and fetus-specific (M3: *Q*-value = 2.77E-03) contexts. GPX4 and 3 of its novel interactors (APBA3, GAMT and GTF2F1), RPS19 and 3 of its novel interactors (NEDD8, PSMD8 and XRCC1) and a novel interactor of AGRP called DXO were

detected in the global functional module. The fetus-specific module did not contain DXO, but instead contained a novel interactor of ARTN called LDB1. The *adenylate cyclase-activating G-protein coupled receptor signaling pathway* was detected both as a global (M3: *Q*-value = 4.25E-04) and fetus-specific (M2: *Q*-value = 3.4E-04) module. 3 novel interactors of GPX4 (EGR4, MATK and ZNF197), 3 novel interactors of RPS19 (CHRNB4, PSG9 and TONSL), 2 novel interactors of ARNTL (CAMK4 and FCGBP), ARTN and its novel interactor DIO1, and AGRP and its novel interactor DHODH were detected in the global module. The fetus-specific module did not contain AGRP and DHODH, but instead contained VHL, a novel intermediate interactor between GPX4 and AGRP. *Cell growth* (M1: *Q*-value = 2.38E-04) and *response to redox state* (M2: *Q*-value = 2.38E-04) were detected only as global modules. The *cell growth module* contained 2 novel interactors of ARNTL (EIF4G2 and MSH2) and a novel interactor of ARTN (TUBGCP3). The *redox state response* module contained a novel interactor of AGRP called UPP1. *N-terminal peptidyl–lysine acetylation* (M1: *Q*-value = 1.21E-04)/*histone acetylation* (M1: *Q*-value = 2.84E-04) and *regulation of DNA metabolic process* (M4: *Q*-value = 7.84E-04) were detected only as tissue-specific modules. A novel interactor of ARNTL (TRIM22) and a novel interactor of AGRP (RNF19A) belonged to the *acetylation* module, whereas MSH2 and TUBGCP3 that were earlier detected in the *cell growth* module, seemed to also belong to the *DNA metabolic process* module.

Interactome of GPX4 and genes associated with other spondylometaphyseal dysplasias

We curated genes associated with other types of SDs from a comprehensive review on genetic skeletal disorders[31] in order to map their connections to GPX4 in the human interactome. Specifically, this included a set of 9 genes associated with Kozlowski type SMD(TRPV4), spondyloenchondrodysplasia (ACP5), odontochondrodysplasia (TRIP11), Sutcliffe/corner fractures type SMD, (FN1 and COL2A1), SMD with severe genu valgam/Schmidt type SMD (COL2A1), SMD with cone-rod dystrophy (PCYT1A), SMD with retinal degeneration/axial SMD (CFAP410), dysspondyloenchondromatosis (COL2A1), achondrogenesis type 1A (TRIP11), schneckenbecken dysplasia (SLC35D1 and INPPL1) andopsismodysplasia (INPPL1). We extracted the shortest paths connecting these 9 genes (candidates) to GPX4 (target). GPX4 shared 93 intermediate interactors with these 9 genes, including 26 novel interactors (**Fig. 2**). These novel interactors included 4 direct interactors of GPX4 (APBA3, GTF2F1, MATK and ZNF197), 2 of CFAP410/C21orf2 (COL6A1 and

COL6A2), 3 of TRIP11 (ASB2, EPS8 and SERPINA4), 2 of ACP5 (EPOR and SF3A2), 2 of SLC35D1 (MACF1 and VCAM1), 2 of PCYT1A (PAK2 and RUBCN), 1 each of FN1 (CREB1), TRPV4 (RXRB) and COL2A1 (DCN) and 8 intermediate interactors (RPS6KA2, PELI2, THRA, STAT5A, SKIC, TNFSF10, PLAT and PPP2CB). GPX4 was more closely connected with the genes associated with Kozlowski type SMD (TRPV4), odontochondrodysplasia, achondrogenesis type 1A (TRIP11), Sutcliffe type SMD (FN1), SMD with cone-rod dystrophy (PCYT1A), schneckenbecken dysplasia and opsismodysplasia (INPPL1).
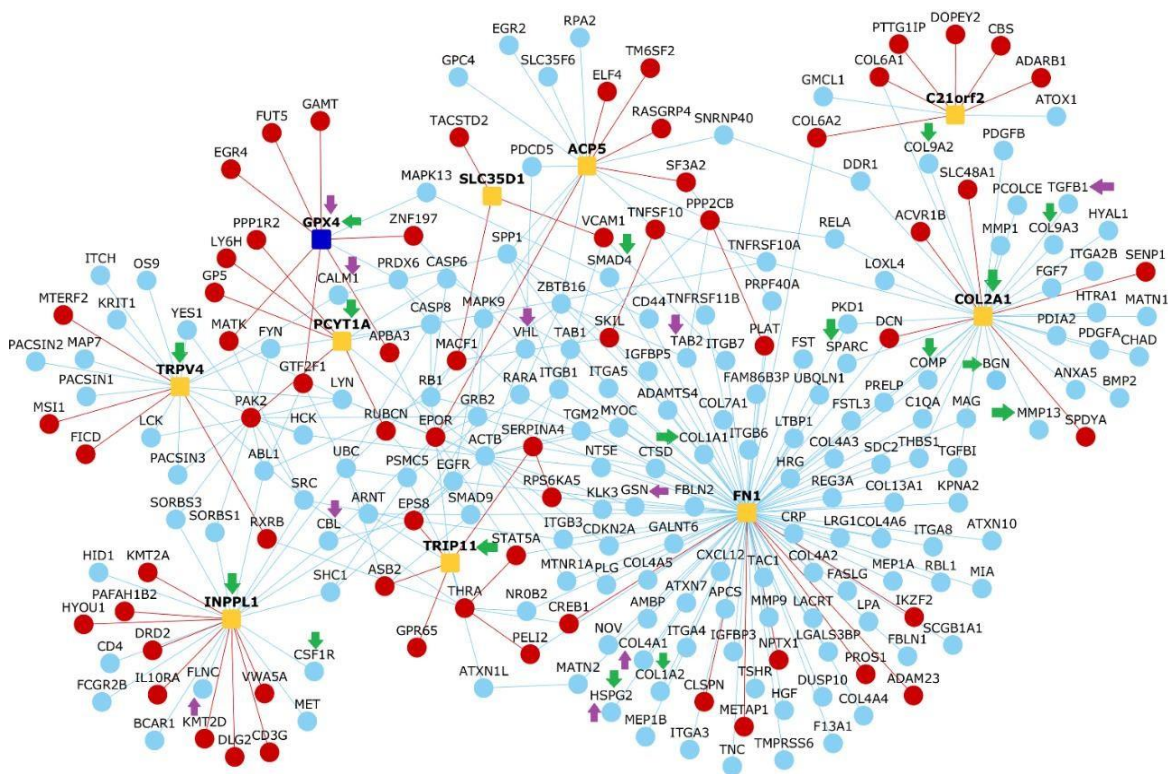
Figure 2: Interactions of GPX4 with genes associated with other spondylometaphyseal dysplasias. This network diagram shows the shortest paths connecting GPX4 with the genes associated with Kozlowski type SMD, spondyloenchondrodysplasia, odontochondrodysplasia, Sutcliffe/corner fractures type SMD, SMD with severe genu valgam/Schmidt type SMD, SMD with cone-rod dystrophy, SMD with retinal degeneration/axial SMD, dysspondyloenchondromatosis, achondrogenesis type 1A, schneckenbecken dysplasia and opsismodysplasia (orange colored nodes). Nodes depict proteins and edges depict PPIs. Red and light blue colored nodes denote novel and known interactors respectively. Red and light blue colored edges denote novel and known PPIs respectively. Purple and green arrows indicate genesassociated with cardiac and skeletal defects respectively.

This interactome was enriched in genes associated with arrhythmia (P-value = 0.041, odds ratio = 1.88, genes: TAB2, COL4A1, HSPG2, FLNC, CALM1, GPX4, CBL, VHL, GSN and TGFB1), cardiorespiratory arrest (P-value = 9.25E-03, odds ratio = 13.38, genes: COL2A1

and GPX4), myocarditis (*P*-value = 4.64E-03, odds ratio = 19.11, genes: GPX4 and VHL), metaphyseal cupping (*P*-value = 5.9E-03, odds ratio = 3.68, genes: GPX4, PCYT1A, INPPL1, TRIP11, MMP13 and COL2A1), platyspondyly (*P*-value = 4.98E-13, odds ratio = 10.43, genes: GPX4, PCYT1A, INPPL1, TRIP11, MMP13, COL2A1, HSPG2, SPARC, BGN, COMP, COL1A1, COL9A2, COL9A3, CSF1R, TRPV4, COL1A2 and SMAD4) and abnormal ribs (*P*-value = 8.9E-03, odds ratio = 3.93, genes: GPX4, PCYT1A, HSPG2, TRPV4 and SMAD4).

We examined the functional modules enriched in this interactome, which contained the intermediate interactors connecting GPX4 to the genes associated with other SDs, and the known and novel interactors of these genes. Three pairs of closely related modules were detected in global and fetal contexts: *blood vessel development* (M1: *Q*-value = 9.6E-06)/*positive regulation of cell migration* (M1: *Q*-value = 2.77E-06), *extracellular organization* (M3: *Q*-value = 3.53E-04)/*cartilage development* (M3: *Q*-value = 2.72E-03) and *regulation of cysteine-type endopeptidase activity involved in apoptotic process* (M6: *Q*- value = 6.68E-03)/*execution phase of apoptosis* (M4: *Q*-value = 6.75E-04). Three unique global modules were identified: *leukocyte tethering or rolling* (M2: *Q*-value = 3.7E-05), *response to bacterium* (M4: *Q*-value = 6.4E-03) and *morphogenesis of an epithelium* (M5: *Q*-value = 6.57E-03). Four unique fetus-specific modules were identified: *cellular component maintenance* (M2: *Q*-value = 3.39E-04), *actin filament organization* (M5: *Q*-value = 1.36E-03), *negative regulation of intrinsic apoptotic signaling pathway* (M6: *Q*-value = 2.51E-03) and *regulation of protein catabolic process* (M7: *Q*-value = 1.05E-02). 3 novel interactors of GPX4 (EGR4, FUT5 and ZNF197) co-occurred with 10 novel interactors of genes associated with other SD genes (SPDYA, MST1, DRD2, KMT2D, GP5, LY6H, SERPINA4, CLSPN, IKZF2 and TM6SF2) and 2 SD genes themselves (COL2A1 and TRPV4) in the *extracellular organization*/*cartilage development* modules in both global and fetus-specific contexts. The global module additionally contained the novel interactors THRA and ADAM23, and the fetus-specific module contained MATK (a novel interactor of GPX4), CD3G, FICD, NPTX1 and RASGRP4. 2 novel interactors of GPX4 (APBA3 and GTF2F1) co-occurred with 3 novel interactors of SD-associated genes (RXRB, SLC48A1 and SF3A2) in the *epithelial morphogenesis* (global) module. APBA3, a novel interactor of GPX4, co-occurred with 2 novel interactors of SD-associated genes (RXRB and SF3A2) in the *protein catabolic process* (fetus-specific) module.

Interactome of GPX4 and genes associated with disorders sharing phenotypic similarity with SMDS

We employed Phenogrid from the MONARCH toolkit[28] to identify the diseases that were most phenotypically similar to SMDS. The phenogrid algorithm identifies the common phenotypes between two given diseases. It then assesses the information content in each of these phenotypes (gene and disease associations) to assign a specific strength to the similarity observed between the diseases. Seven bone dysplasias appeared to be sharing more than 70% phenotypic similarity with SMDS, namely, X-linked spondyloepimetaphyseal dysplasia (77%), acrocapitofemoral dysplasia (75%), A4 type SMD (75%), Dyggve-Melchior-Clausen disease (74%), autosomal recessive Megarbane type SMD (73%), metaphyseal acroscyphodysplasia (72%) and X-linked Dyggve-Melchior-Clausen disease (71%). We checked whether genetic associations were available for these 7 dysplasias in DisGeNET.[26] 4 of them appeared to have been either causally linked or correlated with a few genes, namely, X-linked spondyloepimetaphyseal dysplasia with BGN and IARS2, acrocapitofemoral dysplasia with IHH, Dyggve-Melchior-Clausen disease with DYM and autosomal recessive Megarbane type SMD with PAM16. Next, we extracted the shortest paths connecting these 5 genes (candidates) to GPX4 (target). GPX4 was found to be connected to these 5 genes through 13 intermediate interactors including 5 novel interactors, namely, 3 novel interactors of GPX4 (APBA3, GTF2F1 and ZNF197) and 2 novel interactors of DYM (EPG5 and MAPK4) (**Fig. 3**).

This interactome was enriched in genes associated with arrhythmia (P-value = 0.042, odds ratio = 3.05, genes: TGFB1, IARS2, GPX4 and ELN), cardiorespiratory arrest (P-value = 5.9E-04, odds ratio = 54.2, genes: COL2A1 and GPX4), cerebellar hypoplasia (P-value = 8.63E-03, odds ratio = 4.95, genes: BMP4, EPG5, GPX4 and DAG1), agenesis of corpus callosum (P-value = 1.88E-03, odds ratio = 5.65, genes: BMP4, EPG5, GPX4, PTCH1 and DAG1), metaphyseal cupping (P-value = 3.73E-05, odds ratio = 45.18, genes: GPX4, COL2A1 and PAM16), platyspondyly (P-value = 5.9E-09, odds ratio = 19.9, genes: GPX4, COL2A1, PAM16, BGN, DYM, SMAD4, COL1A1 and COL1A2) and abnormal ribs (P-value = 3.8E-03, odds ratio = 9.57, genes: GPX4, SMAD4 and PTCH1). Arrhythmia-associated IARS2 is associated with X-linked spondyloepimetaphyseal dysplasia and separated from GPX4 only by 3 edges and 5 intermediate interactors (including 2 novel

interactors of GPX4, namely, APBA3 and ZNF197). EPG5, the gene associated with cerebellar hypoplasia and corpus callosum agenesis, is a novel interactor of DYM, a Dyggve-Melchior-Clausen disease gene, and is an intermediate interactor connecting DYM to GPX4. PAM16, associated with metaphyseal cupping and platyspondyly, is a Megarbane type SMD gene, and is separated from GPX4 only by 3 edges and 2 intermediate interactors (including the novel interactor of GPX4, GTF2F1). Both BGN associated with X-linked spondyloepimetaphyseal dysplasia and DYM associated with Dyggve-Melchior-Clausen disease were separated from GPX4 by 3 edges, and appeared to be linked to platyspondyly.

Figure 3: Interactions of GPX4 with genes associated with bone dysplasias sharing phenotypic similarity with Sedaghatian type spondylometaphyseal dysplasia. This network diagram shows the shortest paths connecting GPX4 with the genes associated with X-linked spondyloepimetaphyseal dysplasia, acrocapitofemoral dysplasia, Dyggve-Melchior-Clausen disease and autosomal recessive Megarbane type SMD (pink colored nodes). Nodes depict proteins and edges depict PPIs. Red and light blue colored nodes denote novel and known interactors respectively. Red and light blue colored edges denote novel and known PPIs respectively. Purple, green and orange arrows indicate genes associated with cardiac, skeletal and brain defects respectively.

Five modules were detected in both global and fetus-specific contexts from the interactome: *skeletal system development* (global M1: *Q*-value = 2.76E-06)/*embryo development*/*gastrulation* (fetus-specific M1: *Q*-value = 2.28E-04), *lipid storage* (global M2:

Figure 4: Interactions of GPX4 with genes whose perturbation is known to affect GPX4 expression. This network diagram shows the shortest paths connecting GPX4 with the genes collected from the Knockdown Atlas (black colored nodes), whose perturbation (DNMT3L knockdown/AHR knockout/JUN mutation/NTRK1 overexpression) is known to cause underexpression/overexpression of GPX4. Note that only the PPIs of the 4 perturbed genes that are most closely connected to GPX4 are shown here. Nodes depict proteins and edges depict PPIs. Red and light blue colored nodes denote novel and known interactors respectively.Red and light blue colored edges denote novel and known PPIs respectively. Purple arrows indicate genes associated with heart defects.

Q-value = 1.69E-04)/*reactive oxygen species biosynthetic process* (fetus-specific M2: *Q*-value = 9.17E-04), *humoral response* (global M3: *Q*-value = 2E-03 and fetus-specific M3: *Q*-value = 1.28E-03), *cellular response to toxic substance* (global M4: *Q*-value = 2.59E-03 and fetus-specific M4: *Q*-value = 3.76E-03) and *membrane organization* (global M5: *Q*-value = 8.47E-03 and fetus-specific M6: *Q*-value = 6.29E-03). We noted another *embryo development* module that was fetus-specific and enriched for *neurogenesis* (M5: *Q*-value = 0.01). GPX4 and 2 of its novel interactors (APBA3 and GTF2F1) co-occurred with a novel interactor of a gene associated with a bone dysplasia (DNASE1L1) in the *toxic substance response* module in both global and fetus-specific contexts. 1 novel interactor (AMDHD2) and 2 novel interactors (CDIP1 and C1orf115) uniquely occurred in this module in global andfetus-specific contexts respectively. ZNF197, a novel interactor of GPX4, co-occurred with 4novel interactors of bone dysplasia genes (TEX28, HHIPL2, SUSD4 and SLC30A10) and a bone dysplasia gene itself (IHH) in the global *humoral immune response* module. ZNF197

also co-occurred with 4 novel interactors of bone dysplasia genes (CCL20, AMDHD2, HHIPL2 and SLC30A10) in the fetus-specific *lipid storage/reactive oxygen species biosynthetic process* module.

Interactome of GPX4 and genes whose perturbation affects GPX4 expression

Using the Knockdown Atlas from the BaseSpace Correlation Engine software suite,[32] we compiled a list of 136 genes whose perturbation, in the form of gene knockout/knockdown/mutation/overexpression, significantly affects GPX4 expression. We extracted the shortest paths connecting these 136 genes (candidates) to GPX4 (target). Out of the 136 genes, 4 were closely connected to GPX4, namely, AHR, DNMT3L, JUN and NTRK1, through 33 intermediate interactors including 9 novel interactors (**Fig. 4**). These novel interactors included all the 7 novel interactors of GPX4 (APBA3, EGR4, FUT5, GAMT, GTF2F1, MATK and ZNF197) and 2 intermediate interactors (MAPK13 and SUMO1). Mutations in JUN lead to GPX4 underexpression, whereas DNMT3L knockdown, AHR knockout and NTRK1 overexpression lead to GPX4 overexpression. This interactome was enriched in genes associated with cardiorespiratory arrest (P-value = 3.2E-03, odds ratio = 23.06, genes: GPX4 and KIT) and myocarditis (P-value = 1.52E-03, odds ratio = 32.9, genes: GPX4 and VHL).

Telomere maintenance/organization was detected both as global (M1: Q-value = 1.16E-04) and fetus-specific (M1: Q-value = 1.84E-06) modules. *Positive regulation of neurogenesis* was also detected in global (M7: Q-value = 5.88E-03) and fetal (M5: Q-value = 5.34E-03) contexts. 6 additional global modules were detected: *cellular response to retinoic acid* (M2: Q-value = 1.73E-03), *detoxification* (M3: Q-value = 1.73E-03), *peptidyl-threonine phosphorylation* (M4: Q-value = 1.73E-03), *regulation of cellular protein localization* (M5: Q-value = 3.5E-03), *DNA replication* (M6: Q-value = 5.35E-03) and *protein autophosphorylation* (M8: Q-value = 1.14E-02). 3 additional fetal modules were detected: *negative regulation of myeloid cell differentiation* (M2: Q-value = 1.33E-04), *transforming growth factor beta receptor signaling pathway* (M3: Q-value = 3.25E-04) and *positive regulation of endopeptidase activity* (M4: Q-value = 7.27E-04). 3 novel interactors of GPX4 (EGR4, FUT5 and ZNF197) co-occurred with 2 perturbed genes (NTRK1 and DNMT3L) in the *protein autophosphorylation* global module and in the *neurogenesis* module that was detected in both global and fetus-specific contexts.

Tissue-elevated gene clusters in the integrated GPX4 landscape

We merged the interactomes described in the previous sections to construct an integrated GPX4 functional landscape containing 342 genes and 461 edges. Then, we sought to isolate clusters of genes showing high expression in specific tissues from this integrated interactome. For this, we generated a data matrix of genes (columns) versus 53 tissues (rows) extracted from GTEx; the cells in the matrix contained log transformed median TPM (transcripts per million) values of gene expression. Principal component analysis (PCA) was used to capture systematic variations underlying this matrix. Using Clustvis,[33] single value decomposition (SVD) with imputation was applied to this matrix to extract principal components that explain the variance in gene expression observed across the tissues. Unit variance scaling wasapplied across the matrix. PC1 and PC2 explained 75.1% and 9.6% of the total variance. The log-transformed TPM values were then converted to normalized Z-scores. Z-scores indicate the number of standard deviations that separate a given TPM value from the mean. This matrix of Z-scores was then subjected to hierarchical clustering based on Pearson's correlation coefficients and the average linkage method.

GPX4 showed high expression in brain tissues and testis. Guo et al. has reported that the human brain and testis exhibit the highest similarity in gene expression patterns among a group of 17 tissues.[34] This has been attributed to (a) shared biochemical pathways mediated by exocytotic processes and similar receptors between brain and testis tissues, and (b) the involvement of these tissues in human speciation, as a result of which the same set of genes may have been recruited and their expression patterns maintained in both the tissues by evolutionary mechanisms.[35] We detected a group of 59 genes that clustered with GPX4 and showed high expression in brain tissues and/or testis with/without elevated expression in EBV-transformed lymphocytes and spleen (**Fig. 5**). This cluster included 2 novel interactors of GPX4 that showed high expression in brain tissues, namely, ***EGR4*** and ***MATK***, and ***FUT5*** that showed high expression in testis. Additionally, 5 genes showed elevated expression in both brain and testis tissues (***HHIPL2***, ***MSI1***, COL2A1, MATN1 and MTNR1A (novel interactors shown in bold italics and disease-associated genes shown in bold), 4 genes in only testis (***CHRNB4***, ***SPDYA***, MAGEC1 and RBL1.1), 8 genes in testis, lymphocytes and spleen (**CDKN2A**, ***CLSPN***, ***MSH2***, BLOC1S2, DUSP10, KPNA2, PARP1 and UCHL5) and 39 genes in only brain tissues (***ADAM23***, ***CAMK4***, ***CBS***, ***CDIP1***, ***CHST6***, ***DLG2***, ***DRD2***,
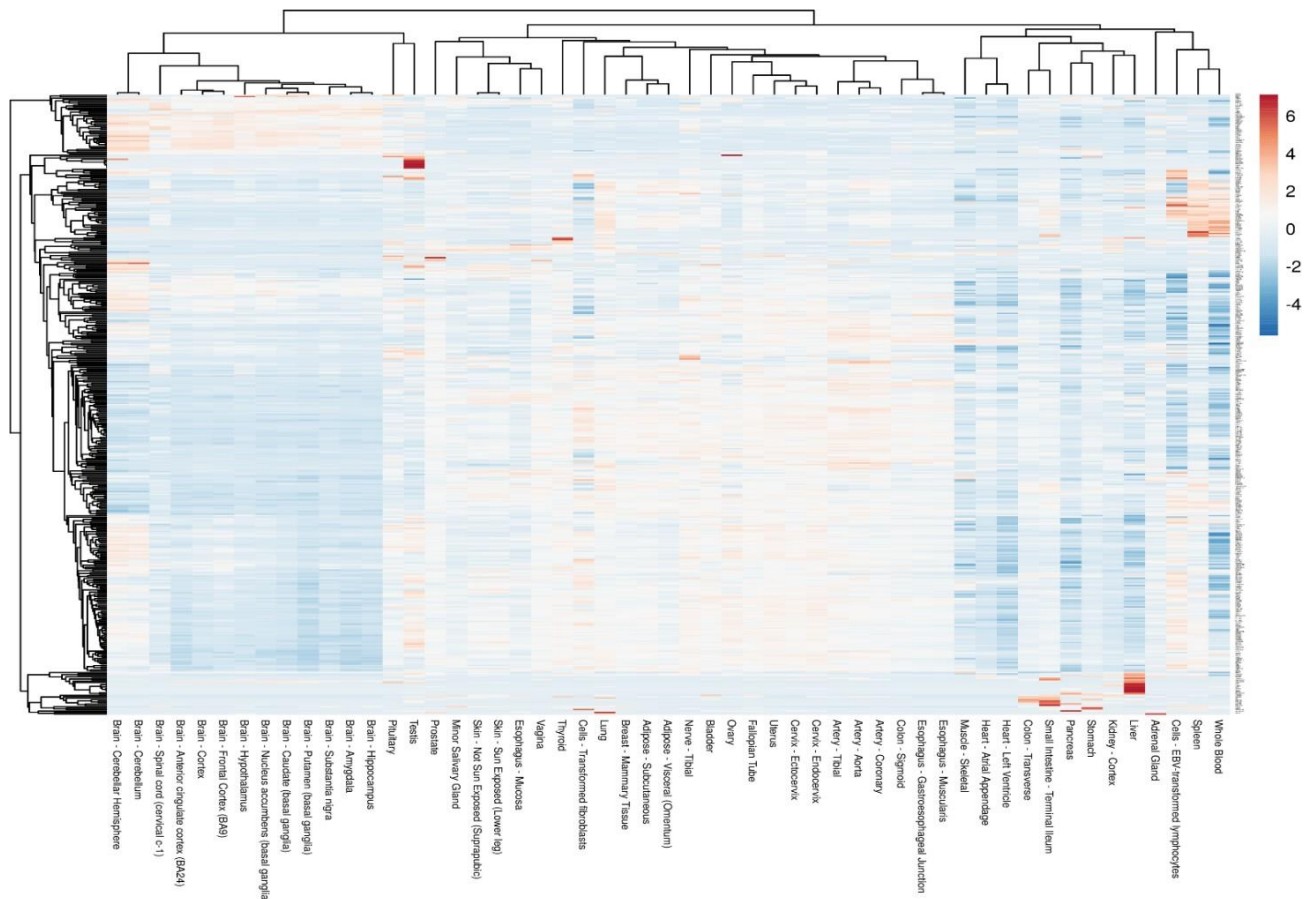
Figure 5: Clustering analysis of genes in the integrated GPX4 interactome. Variations in the expression values of the genes in the integrated GPX4 interactome across 53 tissues have been represented here in the form of a heatmap. The integrated GPX4 interactome contains the shortest paths connecting GPX4 to other genes speculated to be associated with Sedaghatian type spondylometaphyseal dysplasia, genes associated with other spondylometaphyseal dysplasias and phenotypically similar bone dysplasias. Normalized Z-scores were computed based on the $-\log_{10}$ transformed TPM (transcripts per million) values. Z-scores are computed based on the number of standard deviations that separate a given logTPM value from the mean.Clustering was performed using the hierarchical clustering method with average linkage. The dendrograms were derived from the clustering analysis based on computation of Pearson correlation coefficients between the data points. Four gene clusters were detected, namely, a cluster with elevated expression in brain tissues and/or testis with/without elevated expression in EBV-transformed lymphocytes and spleen, another cluster showing elevated expression in whole blood and/or spleen with/without elevated expression in EBV-transformed lymphocytes,and separate liver-elevated and small intestine-elevated clusters.

GOLM1, *LY6H*, *MAPK4*, *NPTX1*, *SUSD4*, ARL6IP1, CACNA1A, CALM1, CHAD, COL9A2, COL9A3, DDR1, FEZ1, FGF12, FRS3, GPM6A, HHIP, HID1, HSP90AA1, JPH3, MAG, MAP7, MAPK9, MAPT, MBP, MC4R, PACSIN1, PDIA2, SPP1, TAC1, TERF2IP and YWHAE).

GAMT, a novel interactor of GPX4, showed high expression in the liver, together with a cluster of 15 other genes (*DHODH*, *SERPINA4*, *SLC30A10*, *TM6SF2*, AMBP, APCS, APOB, CRP, HRG, LPA, MST1, NR0B2, ORM2, PLG and TMPRSS6) (**Fig. 5**). 8 genes showed elevated expression in the small intestine (**IHH**, *CCL20*, *FCGBP*, HNF4G, MEP1A, MEP1B, POU5F1 and REG3A) (**Fig. 5**). A cluster of 38 genes showed elevated expression in whole blood and/or spleen with/without elevated expression in EBV-transformed lymphocytes (**PSD4**, *GPR65*, *IL10RA*, FCGR2B, GCA, ITGA4, ITGB7, LCK, LRG1, LYN,MAP2K6, MX1, NCF2, PDE6G, PTK2B, SRGN, TNF, *VCAM1*, *CD3G*, *RASGRP4*, *UPP1*, C1QA, C5AR1, CD4, CSFR1, CTSD, CXCR6, DISC1, ELANE, FASLG, GZMM, HCK, HYAL1, ITGA2B, KLRK1, MMP9, NCR1 and TYROBP) (**Fig. 5**).

Enrichment of the brain-elevated genes in axon development, regulation of trans-synaptic signaling, regulation of ion transmembrane transport, protein localization to mitochondrion and neural nucleus development was statistically significant before applying multiple hypothesis correction. The same was true in the case of DNA recombination, aging, regulation of cyclin-dependent protein kinase activity, regulation of DNA metabolic process, mitotic cell cycle phase transition and extrinsic apoptotic signaling pathway for testis-elevated genes, and cell fate commitment, in utero embryonic development, epithelial cell proliferation, mRNA transcription, toxin transport and mesenchymal cell proliferation for small intestine-elevated genes. Enrichment of liver-elevated genes in platelet degranulation and regulation of response to wounding, and lymphocyte/whole blood/spleen-elevated genes in leukocyte differentiation, leukocyte migration, stress-activated protein kinase signaling cascade, adaptive immune response, B cell activation, regulation of inflammatory response and phagocytosis remained statistically significant after multiple hypothesis correction.

We separately inspected the interactome containing the intermediate interactors connecting the 4 genes (AHR, DNMT3L, JUN and NTRK1) whose perturbation affects GPX4 expression, to GPX4, and the known and novel interactors of these 4 genes. We identified a group of 9 testis-elevated genes that clustered with GPX4 (novel interactors shown in bold italics and perturbed genes shown in bold): *FUT5*, MX1, PDE6G, NCR1, MAGEC1, SUMO4, **NTRK1**, MSX2 and *FKBP4*.

Interconnections of GPX4 and genes associated with Joubert syndrome with Jeune asphyxiating thoracic dystrophy

From our Phenogrid analysis, we had noted that Joubert syndrome with Jeune asphyxiating thoracic dystrophy (JATD) showed 64% phenotypic similarity with SMDS. The developmental disorders shared pathophenotypes such as agenesis of corpus callosum, generalized hypotonia, cerebellar hypoplasia and atrial septal defects. Joubert syndrome with JATD shows an amalgamation of several key traits associated with Joubert syndrome such as ataxia-inducing brain stem malformations, hypotonia and cognitive impairment, and skeletal traits characteristic of JATD such as a narrow thorax that leads to respiratory failure, and rib, limb and metaphyseal dysplasia.[36] We sought to identify whether GPX4 was closely connected to the genes associated with Joubert syndrome with JATD. 3 genes, namely, IFT140, KIAA0586 and CSSP1, which were known to be linked to Joubert syndrome with JATD were extracted from DisGeNET[26], and the shortest paths connecting these genes (candidates) to GPX4 (target) were identified. IFT140, KIAA0586 and CSSP1 appeared to share 48 intermediate interactors including 14 novel interactors with GPX4 (**Fig. 6a**). The novel interactors included 4 novel interactors of IFT140 (TRAP1, TELO2, IL32 and DNAJA3), 1 novel interactor of KIAA0586 (DACT1), 5 novel interactors of GPX4 (APBA3, GAMT, GTF2F1, MATK and ZNF197) and 4 intermediate interactors (KPNA5, SP100, SNRPC and MAPK13). Additionally, we identified 2 genes from this interactome whose perturbation led to underexpression of Joubert syndrome-associated genes (**Fig. 6a**). It was interesting to note that these genes (TRIM28, ABL1 and BCR1) were known interactors of the novel interactors (ZNF197, GTF2F1 and APBA3) that we had predicted for GPX4. Knockout of TRIM28 led to the underexpression of IFT140, whereas mutations in ABL1 and BCR1 led to the underexpression of KIAA0586.

Since GPX4 showed close interconnections with the Jourbert syndrome-associated gene IFT140, which is also a ciliary protein, we checked whether GPX4 showed similar interconnections with other ciliary proteins. For this, we examined the shortest paths connecting 165 ciliary proteins[37] to GPX4. The ciliary protein that showed the closest interconnection with GPX4 was TERF1 (**Fig. 6b**). GAMT, a novel interactor of GPX4, and PRDX6, a known interactor of GPX4, acted as intermediate interactors between GPX4 and TERF1. Overall, GPX4 appeared to be connected to TERF1 through 13 intermediate

interactors, including 4 novel interactors of GPX4 (APBA3, GAMT, GTF2F1 and MATK), 1 novel interactor of TERF1 (LYN) and 2 intermediate interactors (TRIP10 and SP100).
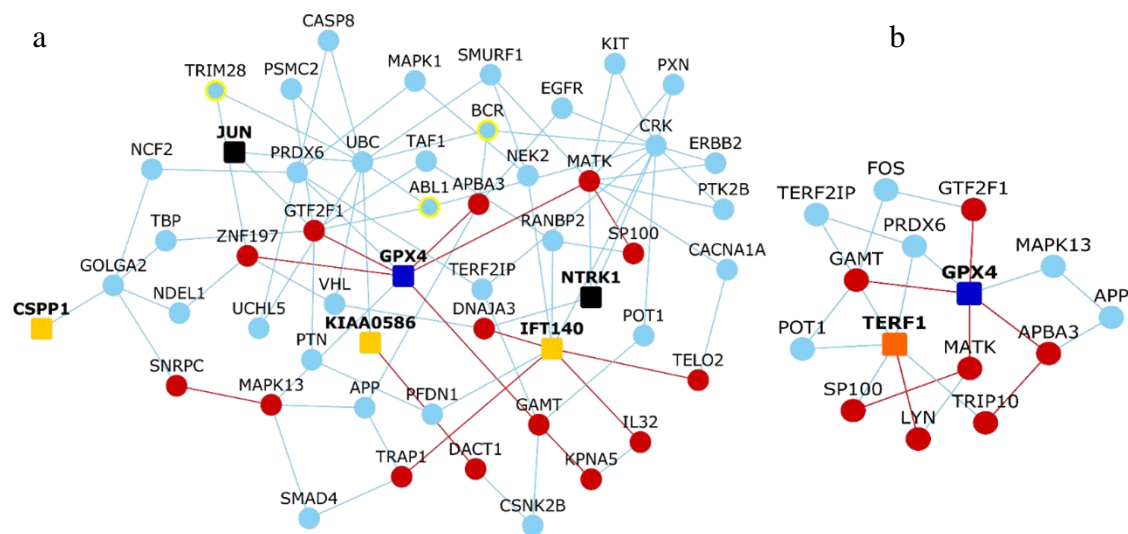
Figure 6: Interactions of GPX4 with ciliary proteins and genes associated with Jourbert syndrome. (a) shows the shortest paths connecting 3 genes associated with Joubert syndrome with Jeune asphyxiating thoracic dystrophy collected from the DisGeNETdatabase (light orange colored nodes) to GPX4; two genes whose perturbation affects GPX4expression (black colored nodes) and 3 genes (nodes with yellow colored borders) whose perturbation affects 2 Joubert syndrome-associated genes, IFT140 and KIAA0586, can also be found in this network. (b) shows the shortest paths connecting the ciliary protein TERF1 to GPX4. Nodes depict proteins and edges depict PPIs. Red and light blue colored nodes denote novel and known interactors respectively. Red and light blue colored edges denote novel and known PPIs respectively.

Detoxification was detected both as a global (M4: *Q*-value = 1.26E-03) and fetus-specific (M4: *Q*-value = 1.68E-03) module. Other enriched global modules included *positive regulation of cell development* (M1: *Q*-value = 1.18E-03), *positive regulation of actin filament bundle assembly* (M2: *Q*-value = 1.18E-03), *regulation of G1/S transition of mitotic cell cycle* (M3: Q-value = 1.18E-03), *regulation of JAK-STAT cascade* (M5: *Q*-value = 1.26E-03), *response to growth factor* (M6: *Q*-value = 1.32E-03), *regulation of proteasomal protein catabolic process* (M7: *Q*-value = 1.84E-03) and *establishment of protein localization of organelle* (M8: *Q*-value = 1.94E-03). Other enriched fetus-specific modules included *regulation of binding* (M1: *Q*-value = 4.62E-05), *cellular response to organonitrogen compound* (M2: *Q*-value = 8.91E-04) and *telomere capping* (M3: *Q*-value = 1.05E-03).

Repurposable drugs for SMDS

We adopted two approaches to identify drugs that may be tested in SMDS from the interactome of SMDS-associated genes compiled from the DisGeNET[26] database (i.e. GPX4, AGRP, ARNTL, ARTN, LOH19CR1, PSD4 and RPS19). In our first approach, we followed the established methodology of comparing drug-induced versus disease-associated differential expression profiles.[38] For this, we used a software suite called BaseSpace Correlation Engine (https://www.nextbio.com).[39,40] This data analysis platform was used because it allows users to study the effect of diseases and/or drugs on thousands of pre- processed publicly available gene expression datasets and has helped to identify drug candidates for diseases such as schizophrenia[41] (currently undergoing clinical trials[42,43]) and mesothelioma[44] in the past. We constructed the SMDS drug-protein interactome that showed the drugs that target any protein in the SMDS interactome. 36 drugs targeted 16 proteins in the interactome, including 5 novel interactors (PSMD8, XRCC1, DHODH, GHRL and VHL).We selected 4 gene expression datasets, namely, tibial growth plate hypertrophic zone - Cog mice (chondroplasia) versus wildtype littermates, tibial growth plate hypertrophic zone - Schmid mice (chondroplasia) versus wildtype littermates (GSE30628[45]), skin fibroblasts - Schimke immuno-osseous dysplasia cell line SD60 versus healthy control and skin fibroblasts - Schimke immuno-osseous dysplasia cell line SD8 versus healthy control (GSE35551[46]). Then, we compiled a list of chemical compounds whose differential gene expression profiles (drug versus no drug) were negatively correlated with at least one of the 4 dysplasia-associated differential gene expression datasets (disease versus control). Following this methodology, we identified 7 drugs negatively correlated with chondroplasia and 10 drugs with immune-osseous dysplasia. Although in each case some genes were differentially expressed in the same direction for both the drug and disorder, the overall effect on the entire transcriptome had an anti-correlation. Altogether, we identified 11 drugs as potential candidates that may be tested against SMDS in clinical trials (**Fig. 7a**), namely, anakinra, colchicine, dactinomycin, dexamethasone, fluorouracil, gemcitabine, imatinib, sirolimus, sorafenib, tretinoin and vincristine. Anakinra targets GHRL, a novel intermediate interactor between GPX4 and AGRP. Dexamethasone targets 2 potential SMDS-associated genes, RPS19 and LOH19CR1, and RARA, a known interactor of another SMDS-associated gene called ARNTL. Sorafenib targets a known interactor of GPX4 (MAPK13), a known interactor of ARNTL (HIF1A) and a novel intermediate interactor connecting GPX4 to

AGRP (VHL). Both fluorouracil and gemcitabine target SMAD4, a known intermediate interactor connecting GPXE4 to PSD4, and XRCC1, a novel interactor of the SMDS-associated RPS19.
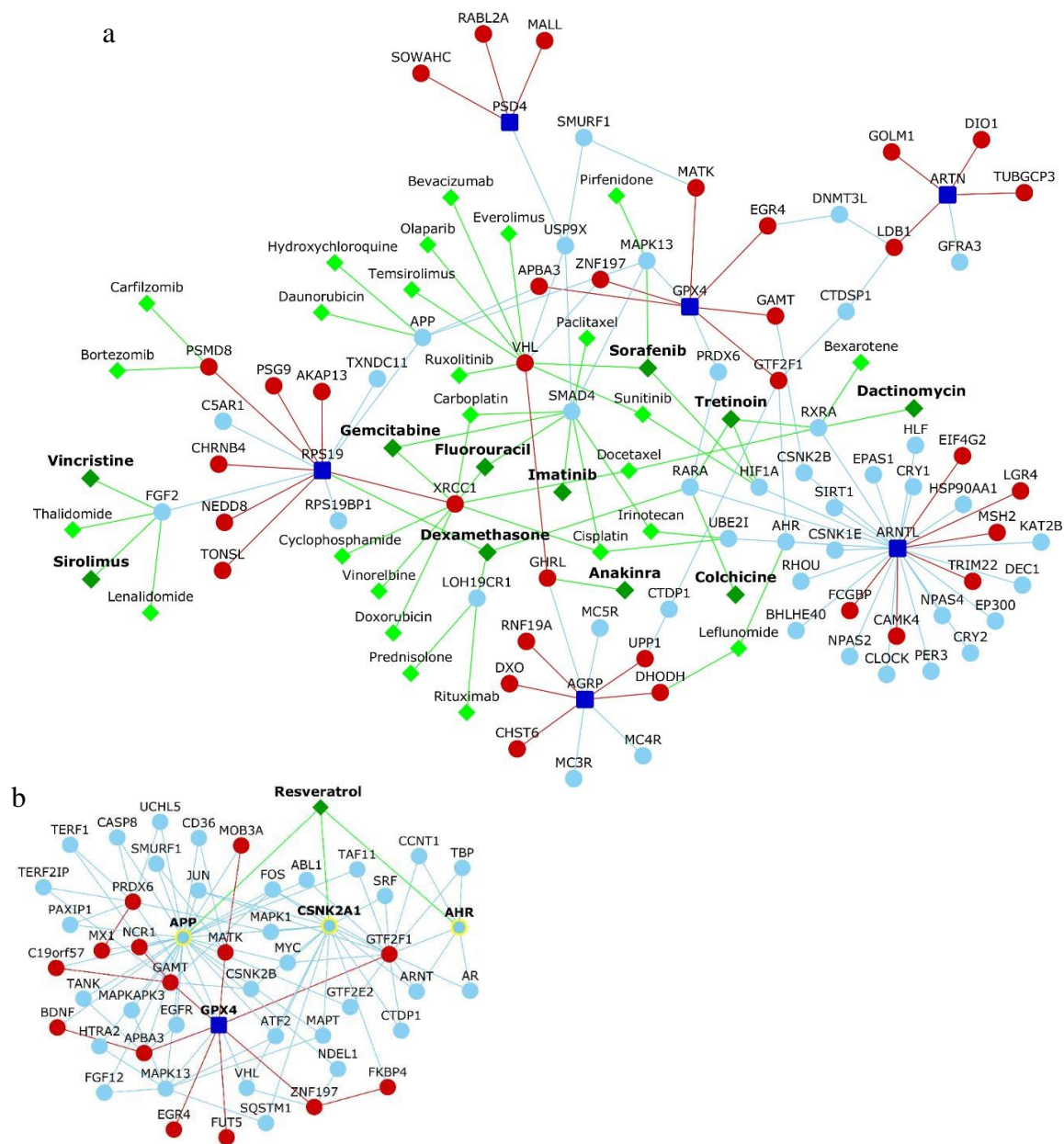
Figure 7: Drug-protein interactome of GPX4. (a) shows the drugs targeting the proteins in the interactome connecting GPX4 to other genes putatively associated with Sedaghatian type spondylometaphyseal dysplasia. Green colored nodes and edges depict drugs and drug-protein interactions respectively. Repurposable drugs identified through comparative transcriptomic analysis have been shown as dark green colored nodeswith bold labels. (b) shows the interconnections of GPX4 with 3 targets (nodes with yellow colored borders) of resveratrol. Red and light blue colored nodes denote novel and known interactors respectively. Red and light blue colored edgesdenote novel and known PPIs respectively.

In our second approach, we compiled a list of drugs that are currently labelled for or are in phase I/II/III clinical trials for different forms of skeletal dysplasias,[47] and identified their protein targets from Drug Bank.[48] This yielded a list of 56 proteins targeted by 8 drugs, namely, etidronic acid, risedronic acid, denosumab (osteogenesis imperfecta), resveratrol (pseudoachondroplasia), carbamazepine (Schmid type metaphyseal dysplasia), asfotase alfa (hypophosphatasia), burosumab (X-linked hypophosphatemia) and N-acetylcysteine (diastrophic dysplasia). We examined the shortest paths connecting each of these protein targets to GPX4 and isolated the targets that showed the closest interconnections with GPX4. 3 targets of resveratrol – AHR, CSNK2A1 and APP – appeared to be connected to GPX4 via single intermediate interactors (**Fig. 7b**). Specifically, both AHR and CSNK2A1 were found to be connected to GPX4 via GTF2F1, a novel interactor of GPX4, whereas APP was connected via a known (MAPK13) and a novel (APBA3) interactor of GPX4. We also found that resveratrol induced an expression profile that is negatively correlated with the profile of the Schimke immuno-osseous dysplasia cell line SD60 mentioned before.

Discussion

SMDS is a severely under characterized skeletal dysplasia driven by GPX4 mutations. In order to expand the functional landscape of this rare and lethal disorder, and expedite the formulation of intervention strategies, we constructed both disease-centric and gene-centric neighborhood networks of GPX4, augmented with novel interactors predicted by the HiPPIP algorithm. Three disease-centric networks were constructed for GPX4, namely, in relation with other putative SMDS-associated genes, SD-associated genes and genes associated with phenotypically similar disorders (**Fig. 1-3**). The GPX4-centric network was constructed to show the interconnections of GPX4 with genes whose perturbation has been known to affect GPX4 expression (**Fig. 4**). Our key findings from these networks were that they were enriched with/contained genes (a) linked to several SMDS pathophenotypes, (b) belonging to tissue-naïve and fetus-specific functional modules, and (c) showing elevated expression in brain and/or testis similar to GPX4. Additionally, we identified 12 drugs that target the neighborhood network of GPX4 and induce gene expression profiles negatively correlated with those associated with chondroplasia and immune-osseous dysplasia.

We used 'co-membership with GPX4 or its novel interactors in an enriched functional module' as a criterion to filter genes for experimental dissection of SMDS. Firstly, an

integrated network containing the repurposable drugs and the shortest paths of putative SMDS-associated genes, SD-associated genes, genes associated with phenotypically similar disorders and GPX4-affecting genes to GPX4 (collectively referred to as 'core genes' henceforth) was constructed. Next, subnetworks containing intermediate PPIs of the genes occurring (along with GPX4 or its novel interactors) in functional modules and the core genes were isolated (**Fig. 8**). Shared phenotypes may reflect interactome-level relationships or similarities in gene function.[49] In order to facilitate phenotypic-guided investigations of GPX4, we identified 26 genes from these subnetworks that shared at least one phenotype associated with GPX4 (as per data from Mammalian Phenotype Ontology[50]; see **Supplementary File 1**): ADAM23, AGRP, CAMK4, CLSPN, COL2A1, DHODH, DIO1, DNMT3L, DRD2, IHH, IKZF2, KMT2D, LDB1, NTRK1, RASGRP4, RPS19, RXRB, SLC30A10, SPDYA, THRA, TONSL, VHL, XRCC1 and three novel interactors of GPX4 (EGR4, GAMT and ZNF197). LDB1, CAMK4, XRCC1, EGR4 and VHL shared the most number of phenotypes with GPX4.

EGR4, which was predicted as a novel interactor of GPX4, is a transcription factor that has been shown to regulate hind brain development in *Xenopus*.[51] Mice with conditionally deleted XRCC1 (a novel interactor of RPS19) exhibited cerebellar ataxia characterized by reduction in the number of cerebellar neurons and abnormal spike activity in Purkinje cells.[52] LDB1 is a novel interactor of ARTN (a putative SMDS-associated gene). It interacts with thenovel interactors of GPX4 (EGR4 and GTF2F1) via two intermediate interactors (DNMT3L and CTDSP1). LDB1 is an adaptor protein that serves as a critical component of transcription complexes and is involved in the differentiation of various cell types (e.g. hematopoietic cells).[53] Mutant mice lacking this gene exhibit a range of developmental defects including nterior-posterior patterning, cardiac and foregut defects.[53] LDB1 has also been shown to influence another gene, LMX1B, mutations in which have been linked to a skeletal dysplasia called the Nail-Patella syndrome.[54] Accumulation of oxidized lipids leads to a process called 'ferroptosis', an iron-dependent/caspase-independent form of apoptosis. Since GPX4 inhibits the accumulation of oxidized lipids by catalyzing their reduction, GPX4 inactivation is critical for ferroptosis. ATF2, a transcription factor that can be found in the *adenylate-cyclase activating GPCR signaling pathway* module (**Fig. 8a**), has been known to inhibit ferroptosis when activated by the JNK1/2 pathway.[55] Both MAPK13 (a known interactor of GPX4) and CAMK4 (a novel interactor of ARNTL, a putative SMDS-associated gene) have been known
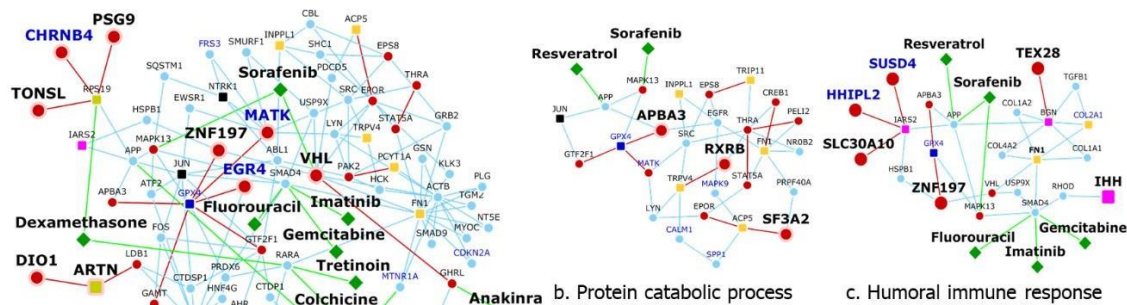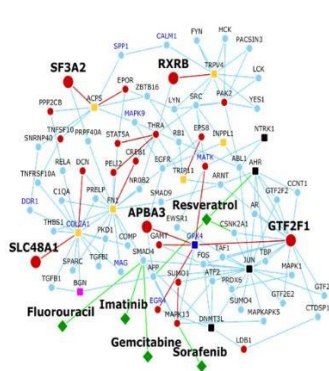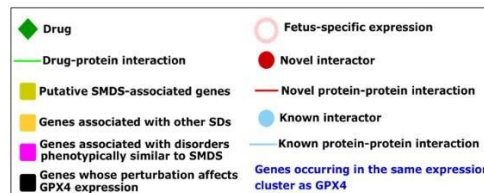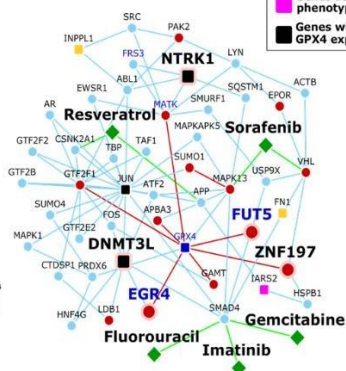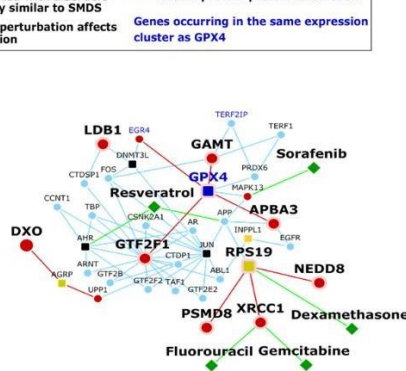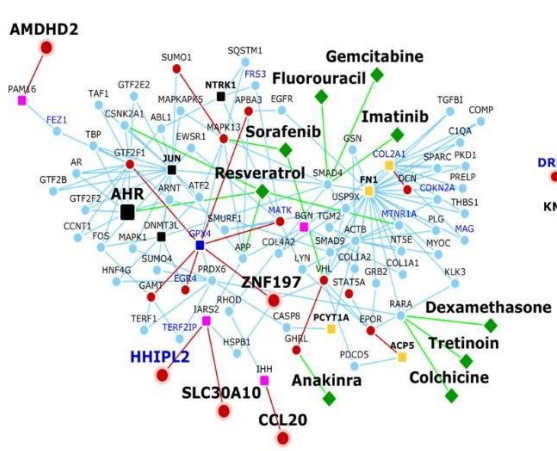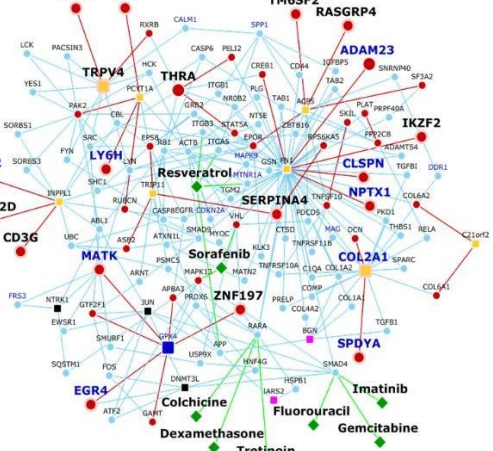
Figure 8: Functional modules with GPX4. The figure shows network diagrams for eight functional modules containing GPX4 or its novel interactors. The network nodes have been annotated as per the legend shown on the right side.

to

activate ATF2.[56,57] The exact nature of the functional interaction (e.g. activation/ inhibition) between GPX4 and MAPK13 is unclear. However, if one assumes that GPX4 may inhibit ferroptosis through some indirect action on ATF2, it may be speculated that MAPK13 needs to be activated by GPX4 before it activates ATF2 and influences the inhibition of ferroptosis;

note that ATF2 may be independently activated by CAMK4.[58] Perturbation of GPX4 activity may remove its activational effect on MAPK13, and facilitate ferroptosis. VHL is a known interactor of ZNF197, a novel interactor of GPX4 (**Fig. 8c**). GPX4 has been shown to have an inhibitory effect on elevated VEGFA,[59] which is generally activated by HIF1A. VHL recruits ZNF197 to inhibit HIF1A transcriptional activity.[60] Hence, it may be speculated that GPX4 exerts its inhibitory activity on VEGFA, via its direct interaction with the ZNF197-VHL complex. Perturbed GPX4 activity may remove this inhibitory effect on ZNF197-VHL and promote abnormal VEFGA protein expression. Since VEGF signaling is a critical regulator of bone development,[61] abnormal VEGFA expression may influence the development of skeletal pathophenotypes. Drugs exploiting these potential functional associations of GPX4 on ferroptosis (via MAPK13) and bone development (via ZNF197) could be examined in the context of SMDS. Several other genes in the functional modules have been associated with other dysplasias. Mutations in RPS19 (**Fig. 8f**) and TONSL (**Fig. 8a**) have been associated with hip dysplasia and Sponastrime dysplasia respectively,[62,63] and KMT2D (**Fig. 8h**) with Kabuki syndrome.[64] RXRB (**Fig. 8b**) has a regulatory effect on COL2A1, which has been linked to chondrodysplasia, Stickler syndrome and otospondylomegaepiphyseal dysplasia.[65]

We identified 23 genes from the network interconnecting Joubert syndrome-associated genes to GPX4 (**Fig. 6a**), which shared at least one phenotype with GPX4 (**Supplementary File 2**). Most phenotypes were shared by KIT, TRIM28, SNRPC, VHL, CACNA1A, JUN, APP, ABL1 and a novel interactor of GPX4 (GAMT). Additionally, we identified more than 50 genes that showed similar tissue expression patterns as GPX4 and connected GPX4 to other SMDS-associated genes, SD-associated genes and genes whose perturbation affected GPX4 expression. Experimental studies on the mechanistic connections of such genes to GPX4 may provide insights into SMDS etiology.

Comparative transcriptome analysis revealed 11 drugs that induced expression profiles negatively correlated with profiles associated with chondroplasia and immune-osseous dysplasia (**Fig. 7a**). Limited availability of relevant transcriptomic datasets prompted us to use the datasets of these two dysplasias, although they may exhibit etiological distinction from SMDS. We additionally shortlisted resveratrol as a potential drug that may be tested in SMDS due to the proximity of its targets to GPX4 (**Fig. 7b**); resveratrol is currently in phase

II trials for pseudoachondroplasia (ClinicalTrials.gov identifier: NCT03866200). The effect of these proposed drugs should be examined in appropriate models.

In summary, our study provides a GPX4-centric network-level view of SMDS, a functional landscape that will allow biologists to prioritize genes, functional modules and drugs for therapeutic interventions in SMDS.

Methods

Compilation of gene lists and prediction of novel interactions

SMDS-associated genes, genes associated with X-linked spondyloepimetaphyseal dysplasia, acrocapitofemoral dysplasia, Dyggve-Melchior-Clausen disease, autosomal recessive Megarbane type SMD and genes associated with Joubert syndrome (with JATD) were compiled from the DisGeNET[26] database. Genes associated with Kozlowski type SMD, spondyloenchondrodysplasia, odontochondrodysplasia, Sutcliffe/corner fractures type SMD, SMD with severe genu valgam/Schmidt type SMD, SMD with cone-rod dystrophy, SMD with retinal degeneration/axial SMD, dysspondyloenchondromatosis, achondrogenesis type 1A, schneckenbecken dysplasia and opsismodysplasia were curated from Warman et al.[31] Genes whose perturbation (gene knockout/knockdown/mutation/overexpression) affects GPX4 were collected from Knockdown Atlas (BaseSpace Correlation Engine software suite[32]). Novel PPIs of the proteins encoded by these genes were predicted using the HiPPIP model that we developed.[66] Each protein (say N1) was paired with each of the other human proteins say, (M1, M2,…Mn), and each pair was evaluated with the HiPPIP model.[66] The predicted interactions of each of the proteins were extracted (namely, the pairs whose score is >0.5, a threshold which through computational evaluations and experimental validations was revealed to indicate interacting partners with high confidence). Previously known PPIs were collected from HPRD (Human Protein Reference Database,[22] version 9) and BioGRID (Biological General Repository for Interaction Datasets,[23] version 3.4.159). The interactome figures were created using Cytoscape.[67]

Network analysis using LENS

LENS (Lens for Enrichment and Network Studies of human proteins) was used to extract the shortest paths in the human interactome connecting the various sets of genes compiled in this

study to GPX4. LENS is a web-based tool that may be used to identify pathways and diseases that are significantly enriched among the genes submitted by users.[33] The LENS algorithm finds the nearest neighbor of each gene in the interactome and includes the intermediate interactions that connect them. LENS then computes the statistical significance of the overlap of genes in the network and genes with annotations pertaining to pathways, diseases, drugs and GWASs, and reports a *P*-value computed from Fisher's exact test.

Identification of functional modules

Functional gene modules were extracted using the HumanBase toolkit[27] (https://hb.flatironinstitute.org/). HumanBase uses shared k-nearest-neighbors and the Louvain community-finding algorithm to cluster the genes sharing the same network neighborhoods and similar GO biological processes into functional modules. The *P*-values of the terms enriched in the modules are calculated using Fisher's exact test and Benjamini–Hochberg method.

Gene expression analysis

Gene expression profiles of 53 postnatal human tissues were extracted from GTEx.[68] Principal component analysis (PCA) and hierarchical clustering were used to capture relationships among the genes in the various networks constructed in our study (which connect GPX4 to other SMDS-associated genes, SD-associated genes and genes whose perturbation affects GPX4 expression). Log-transformed transcripts per million (TPM) valueswere assembled into a data matrix containing tissues as rows and genes as columns. PCA wasperformed with a web-based tool called ClustVis (https://biit.cs.ut.ee/clustvis/).[33] The data matrix was pre-processed such that 70% missing values were allowed across the rows and columns. The log(TPM) values in the matrix were centered using the unit variance scaling method, in which the values are divided by standard deviation so that each row or column hasa variance of one; this ensures that they assume equal importance while finding the components. The method called singular value decomposition (SVD) with imputation was used to extract principal components. In this method, missing values are predicted and iteratively filled using neighbouring values during SVD computation, until the estimates of missing values converge. The data matrix of tissues (rows) and genes (columns) was subjected to hierarchical clustering using the tool called Heatmapper

(http://www.heatmapper.ca/)[69] to identify tissue-based grouping patterns of genes. Pairwise distances in the data matrix were calculated using Pearson correlation and closely linked clusters were identified using the average linkage method. Dendrograms were generated by merging tissues with the smallest distance first, and those with larger distances later. In the average linkage method, the average distance of all possible pairs is considered while clustering.

Identification of repurposable drugs

The list of chemical compounds whose gene expression profiles correlated negatively with 4 dysplasia expression datasets were compiled using the BaseSpace correlation software (https://www.nextbio.com) (List 1), namely, tibial growth plate hypertrophic zone - Cog mice (chondroplasia) versus wildtype littermates, tibial growth plate hypertrophic zone - Schmid mice (chondroplasia) versus wildtype littermates (GSE30628[45]), skin fibroblasts - Schimke immuno-osseous dysplasia cell line SD60 versus healthy control and skin fibroblasts - Schimke immuno-osseous dysplasia cell line SD8 versus healthy control (GSE35551[46]). Next, we identified drugs that targeted at least one gene in the interactome of SMDS-associated genes (GPX4, AGRP, ARNTL, ARTN, LOH19CR1, PSD4 and RPS19) using Drug Bank (list 2).*[48]* We then compared list 1 and list 2 to identify the drugs that not only target proteins in the interactome but are also negatively correlated with the selected gene expression profiles.

Author contributions

MKG formulated the project idea to put the interactome-based analysis together motivated by the call "Let's Find a Cure Together" (https://www.curegpx4.org/). MKG and her students developed PPI prediction and interactome analysis algorithms and tools in prior work. KBK carried out the analysis of the interactomes, presented literature-based evidence, identified repurposable drugs and drafted the manuscript.

Acknowledgements

collaborate or provide any additional information in our capacity to researchers who wish to take this work forward, e.g., through experimental validations of the novel PPIs, or testing of the potentially repurposable FDA-approved drugs. MKG's effort is supported by Department of Biomedical Informatics, School of Medicine, University of Pittsburgh.

Competing interests

None

Data and materials availability

Cytoscape files containing the interactomes shown in Figs. 1-4, Fig. 6, Fig. 7, and Fig. 8 have been made available as Supplementary Files 3-6 respectively.

Supplementary files

Supplementary File 1: Functional modules enriched in the networks of putative SMDS-associated genes, SD-associated genes, genes associated with phenotypically similar disorders and GPX4-affecting genes, and the mouse phenotypes shared by genes belonging tothese modules.

Supplementary File 2: Functional modules enriched in the network of genes associated with Joubert syndrome (with Jeune asphyxiating thoracic dystrophy), and the mouse phenotypes shared by genes belonging to these modules.

Supplementary File 3: Cytoscape files containing the networks shown in Figs. 1-4. The networks contain mappings of known (light blue edges) and novel PPIs (red edges) and of known (light blue nodes) and novel interactors (red nodes). The legend for the other nodes are the same as that given in the corresponding figures.

Supplementary File 4: Cytoscape files containing the networks shown in Fig. 6. The networks contain mappings of known (light blue edges) and novel PPIs (red edges) and of known (light blue nodes) and novel interactors (red nodes). The legend for the other nodes are the same as that given in the corresponding sub-figures.

Supplementary File 5: Cytoscape files containing the networks shown in Fig. 7. Green colored nodes and edges depict drugs and drug-protein interactions respectively. Red and

light blue colored nodes denote novel and known interactors respectively. Red and light blue colored edges denote novel and known PPIs respectively.

Supplementary File 6: Cytoscape files containing the networks shown in Fig. 8. The legend for the nodes and the edges are the same as that given in the corresponding sub-figures.

References

1   Sedaghatian, M. & Opitz, J. M. Congenital lethal metaphyseal chondrodysplasia: a newly recognized complex autosomal recessive disorder. *American journal of medical genetics* **6**, 269-274 (1980).

2   Opitz, J. M. *et al.* Sedaghatian congenital lethal metaphyseal chondrodysplasia—observations in a second Iranian family and histopathological studies. *American journal of medical genetics* **26**, 583-590 (1987).

3   Peeden Jr, J. *et al.* Spondylometaphyseal dysplasia, Sedaghatian type. *American journal of medical genetics* **44**, 651-656 (1992).

4   Elçioglu, N. & Hall, C. M. Spondylometaphyseal dysplasia–sedaghatian type. *American journal of medical genetics* **76**, 410-414 (1998).

5   Koutouby, A., Habibullah, J. & Moinuddin, F. A. Spondylometaphyseal dysplasia: Sedaghatian type. *American journal of medical genetics* **90**, 199-202 (2000).

6   English, S. J., Gayatri, N., Arthur, R. & Crow, Y. J. Sedaghatian spondylometaphyseal dysplasia with pachygyria and absence of the corpus callosum. *American Journal of Medical Genetics Part A* **140**, 1854-1858 (2006).

7   Smith, A. C. *et al.* Mutations in the enzyme glutathione peroxidase 4 cause Sedaghatian-type spondylometaphyseal dysplasia. *Journal of medical genetics* **51**, 470-474 (2014).

8   Aygun, C. *et al.* Simplified gyral pattern with cerebellar hypoplasia in Sedaghatian type spondylometaphyseal dysplasia: a clinical report and review of the literature. *American Journal of Medical Genetics Part A* **158**, 1400-1405 (2012).

9   *SSMD DISEASE: Ultra Rare Genetic Disorder*, <https://www.curegpx4.org/about-ssmd> (2018).

10   Fedida, A. *et al.* Sedaghatian-type spondylometaphyseal dysplasia: Whole exome sequencing in neonatal dry blood spots enabled identification of a novel variant in GPX4. *Eur J Med Genet* **63**, 104020, doi:10.1016/j.ejmg.2020.104020 (2020).

11   Maiorino, M., Thomas, J. P., Girotti, A. W. & Ursini, F. Reactivity of phospholipid hydroperoxide glutathione peroxidase with membrane and lipoprotein lipid hydroperoxides. *Free radical research communications* **12**, 131-135 (1991).

12   Imai, H. *et al.* Depletion of selenoprotein GPx4 in spermatocytes causes male infertility in mice. *Journal of Biological Chemistry* **284**, 32522-32532 (2009).

13   Flohé, L., Wingender, E. & Brigelius-Flohé, R. in *Oxidative stress and signal transduction* 415-440 (Springer, 1997).

14   Laddach, A., Ng, J. C.-F., Chung, S. S. & Fraternali, F. Genetic variants and protein–protein interactions: a multidimensional network-centric view. *Current opinion in structural biology* **50**, 82-90 (2018).

15   Sahni, N. *et al.* Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647-660 (2015).

16   Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature reviews genetics* **12**, 56 (2011).

17   Loscalzo, J., Kohane, I. & Barabasi, A. L. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Molecular systems biology* **3** (2007).

18   Sakai, Y. *et al.* Protein interactome reveals converging molecular pathways among autism disorders. *Science translational medicine* **3**, 86ra49-86ra49 (2011).

19   Lim, J. *et al.* A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* **125**, 801-814 (2006).

20   Ganapathiraju, M. K. *et al.* Schizophrenia interactome with 504 novel protein–protein interactions. *npj Schizophrenia* **2**, 16012 (2016).

21   Malavia, T. *et al.* Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epidemiological, genomic and protein interaction data *npj Schizophrenia* **in Press** (2017).

22   Keshava Prasad, T. *et al.* Human protein reference database—2009 update. *Nucleic acids research* **37**, D767-D772 (2008).

23   Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic acids research* **34**, D535-D539 (2006).

24   Dunham, B. & Ganapathiraju, M. K. Benchmark Evaluation of Protein–Protein Interaction Prediction Algorithms. *Molecules* **27**, 41 (2022).

25   Karunakaran, K. B., Balakrishnan, N. & Ganapathiraju, M. K. Interactome of SARS-CoV-2 modulated host proteins with computationally predicted PPIs: Insights from illustrative translational systems biology studies. *Frontiers in Systems Biology* **in Press**, doi:doi: 10.3389/fsysb.2022.815237 (2022).

26   Piñero, J. *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*, gkw943 (2016).

27   Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nature neuroscience* **19**, 1454-1462 (2016).

28   Cacheiro, P., Haendel, M. A., Smedley, D., Consortium, I. M. P. & Initiative, M. New models for human disease from the International Mouse Phenotyping Consortium. *Mammalian Genome* **30**, 143-150 (2019).

29   Burl, R. B., Clough, S., Sendler, E., Estill, M. & Krawetz, S. A. Sperm RNA elements as markers of health. *Systems biology in reproductive medicine* **64**, 25-38 (2018).

30   Bravo, À., Piñero, J., Queralt, N., Rautschka, M. & Furlong, L. I. BeFree: a text mining system to extract relations between genes, diseases and drugs for translational research. *SMBM 2014*, 79.

31   Warman, M. L. *et al.* Nosology and classification of genetic skeletal disorders: 2010 revision. *American journal of medical genetics Part A* **155**, 943-968 (2011).

32   Kupershmidt, I. *et al.* Ontology-based meta-analysis of global collections of high-throughput public data. *PloS one* **5**, e13066 (2010).

33   Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic acids research* **43**, W566-W570 (2015).

34   Guo, J. *et al.* In silico analysis indicates a similar gene expression pattern between human brain and testis. *Cytogenetic and genome research* **103**, 58-62 (2003).

35   Matos, B., Publicover, S. J., Castro, L. F. C., Esteves, P. J. & Fardilha, M. Brain and testis: more alike than previously thought? *Open Biology* **11**, 200322 (2021).

36   Lehman, A. *et al.* Co-occurrence of Joubert syndrome and Jeune asphyxiating thoracic dystrophy. *American Journal of Medical Genetics Part A* **152**, 1411-1419 (2010).

37   Karunakaran, K. B., Chaparala, S., Lo, C. W. & Ganapathiraju, M. K. cilia interactome with predicted protein–protein interactions reveals connections to Alzheimer's disease, aging and other neuropsychiatric processes. *Scientific Reports* **10**, 1-16 (2020).

38   Sirota, M. *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science translational medicine* **3**, 96ra77-96ra77 (2011).

39   Kupershmidt, I. *et al.* Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS One* **5**, doi:10.1371/journal.pone.0013066 (2010).

40   Chattopadhyay, A. & Ganapathiraju, M. K. Demonstration Study: A Protocol to Combine Online Tools and Databases for Identifying Potentially Repurposable Drugs. *Data* **2**, 15 (2017).

41   Karunakaran, K. B., Chaparala, S. & Ganapathiraju, M. K. Potentially repurposable drugs for schizophrenia identified from its interactome. *Scientific reports* **9**, 1-14 (2019).

42   Vishwajit Nimgaonkar, M. D. P., Stanley Medical Research, I. & University of, P. (2024).

43   Vishwajit Nimgaonkar, M. D. P., Stanley Medical Research, I. & University of, P. (2022).

44   Karunakaran, K. B., Yanamala, N., Boyce, G. & Ganapathiraju, M. K. Mesothelioma Interactome with 367 Novel Protein-Protein Interactions. *bioRxiv*, 459065 (2018).

45   Cameron, T. L. *et al.* Transcriptional profiling of chondrodysplasia growth plate cartilage reveals adaptive ER-stress networks that allow survival but disrupt hypertrophy. *PloS one* **6**, e24600 (2011).

46   Baradaran-Heravi, A. *et al.* Penetrance of biallelic SMARCAL1 mutations is associated with environmental and genetic disturbances of gene expression. *Human molecular genetics* **21**, 2572-2587 (2012).

47   Marzin, P. & Cormier-Daire, V. New perspectives on the treatment of skeletal dysplasia. *Therapeutic advances in endocrinology and metabolism* **11**, 2042018820904016 (2020).

48   Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* **34**, D668-D672 (2006).

49   Zhang, S. *et al.* Calculating phenotypic similarity between genes using hierarchical structure data based on semantic similarity. *Gene* **497**, 58-65 (2012).

50   Smith, C. L. & Eppig, J. T. The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mammalian genome* **23**, 653-668 (2012).

51   Bae, C.-J., Jeong, J. & Saint-Jeannet, J.-P. A novel function for Egr4 in posterior hindbrain development. *Scientific reports* **5**, 1-10 (2015).

52 Hoch, N. C. *et al.* XRCC1 mutation is associated with PARP1 hyperactivation and cerebellar ataxia. *Nature* **541**, 87-91 (2017).

53 Mukhopadhyay, M. *et al.* Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* **130**, 495-505 (2003).

54 Dreyer, S. D. *et al.* LMX1B transactivation and expression in nail–patella syndrome. *Human molecular genetics* **9**, 1067-1074 (2000).

55 Wang, L. *et al.* ATF2 inhibits ani-tumor effects of BET inhibitor in a negative feedback manner by attenuating ferroptosis. *Biochemical and Biophysical Research Communications* (2020).

56 Gupta, S., Campbell, D., Derijard, B. & Davis, R. J. Transcription factor ATF2 regulation by the JNK signal transduction pathway. *Science* **267**, 389-393 (1995).

57 Van Dam, H. *et al.* ATF-2 is preferentially activated by stress-activated protein kinases to mediate c-jun induction in response to genotoxic agents. *The EMBO journal* **14**, 1798-1811 (1995).

58 Enslen, H., Tokumitsu, H., Stork, P., Davis, R. J. & Soderling, T. R. Regulation of mitogen-activated protein kinases by a calcium/calmodulin-dependent protein kinase cascade. *Proceedings of the National Academy of Sciences* **93**, 10803-10808 (1996).

59 Roggia, M. F., Imai, H., Shiraya, T., Noda, Y. & Ueta, T. Protective role of glutathione peroxidase 4 in laser-induced choroidal neovascularization in mice. *PLoS One* **9**, e98864 (2014).

60 Li, Z. *et al.* The VHL protein recruits a novel KRAB-A domain protein to repress HIF-1α transcriptional activity. *The EMBO journal* **22**, 1857-1867 (2003).

61 Hu, K. & Olsen, B. R. Vascular endothelial growth factor control mechanisms in skeletal growth and repair. *Developmental Dynamics* **246**, 227-234 (2017).

62 Moniz, H. *et al.* Primary hematopoietic cells from DBA patients with mutations in RPL11 and RPS19 genes exhibit distinct erythroid phenotype in vitro. *Cell death & disease* **3**, e356-e356 (2012).

63 Chang, H. R. *et al.* Hypomorphic mutations in TONSL cause SPONASTRIME dysplasia. *The American Journal of Human Genetics* **104**, 439-453 (2019).

64 Shpargel, K. B., Mangini, C. L., Xie, G., Ge, K. & Magnuson, T. The KMT2D Kabuki syndrome histone methylase controls neural crest cell differentiation and facial morphology. *Development* **147** (2020).

65   Murai, J., Ikegami, D., Okamoto, M., Yoshikawa, H. & Tsumaki, N. Insulation of the ubiquitous Rxrb promoter from the cartilage-specific adjacent gene, Col11a2. *Journal ofBiological Chemistry* **283**, 27677-27687 (2008).

66   Ganapathiraju, M. K., Karunakaran, K. B. & Correa-Menendez, J. Predicted protein interactions of IFITMs may shed light on mechanisms of Zika virus-induced microcephaly and host invasion. *F1000Res* **5**, 1919, doi:10.12688/f1000research.9364.2(2016).

67   Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).

68   Consortium, G. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).

69   Babicki, S. *et al.* Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* **44**, W147-W153 (2016).

Interactome-based framework to translate disease genetic data into biological and clinical insights

8. Potentially repurposable drugs for schizophrenia identified from its interactome

The experimental chapter is based on the following peer-reviewed publication:

Karunakaran, Kalyani B., Srilakshmi Chaparala, and Madhavi K. Ganapathiraju. Potentially repurposable drugs for schizophrenia identified from its interactome. Scientific reports 9, no. 1 (2019): 1-14.

Summary of this chapter

In this chapter, I describe the pipeline for interactome-based drug repurposing pipeline using schizophrenia (SCZ) as an example. I extracted the drugs from the SCZ drug-protein interactome. I then performed comparative transcriptome analysis of drug-induced versus SCZ-associated gene expression profiles using the BaseSpace Correlation Engine software suite, a data analysis platform used to study the effect of diseases and drugs on publicly available gene expression data. I subsequently examined the twelve drugs shortlisted in this manner for their biological validity using a series of functional enrichment tests and correlation with clinical data. This analysis resulted in evidence supporting the biological validity of 9 out of the 12 drugs. This work led to the testing of cromoglycate as adjunctive therapy for SCZ and acetazolamide for treatment-resistant SCZ (ClinicalTrials.gov NCT04887792, NCT03794076), which demonstrated the potential translational value of the drugs identified using the integrated computational approach adopted in this study.

Contribution to this chapter (85%)

- Designed the study and developed the methodology of the project, including the correlation analysis of drugs against disease and further bioinformatics analyses
- Curated all the datasets, performed all the analyses, and derived the conclusions (note that the first version of the shortlisted drugs was produced by the second author, Srilakshmi Chaparala, after which I updated this list by re-performing the correlation analysis)
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

OPEN

Potentially repurposable drugs for schizophrenia identified from its interactome

Kalyani B. Karunakaran[1], Srilakshmi Chaparala[2] & Madhavi K. Ganapathiraju[2,3]

We previously presented the protein-protein interaction network of schizophrenia associated genes, and from it, the drug-protein interactome which showed the drugs that target any of the proteins in the interactome. Here, we studied these drugs further to identify whether any of them may potentially be repurposable for schizophrenia. In schizophrenia, gene expression has been described as a measurable aspect of the disease reflecting the action of risk genes. We studied each of the drugs from the interactome using the BaseSpace Correlation Engine, and shortlisted those that had a negative correlation with differential gene expression of schizophrenia. This analysis resulted in 12 drugs whose differential gene expression (drug versus normal) had an anti-correlation with differential expression for schizophrenia (disorder versus normal). Some of these drugs were already being tested for their clinical activity in schizophrenia and other neuropsychiatric disorders. Several proteins in the protein interactome of the targets of several of these drugs were associated with various neuropsychiatric disorders. The network of genes with opposite drug-induced versus schizophrenia-associated expression profiles were significantly enriched in pathways relevant to schizophrenia etiology and GWAS genes associated with traits or diseases that had a pathophysiological overlap with schizophrenia. Drugs that targeted the same genes as the shortlisted drugs, have also demonstrated clinical activity in schizophrenia and other related disorders. This integrated computational analysis will help translate insights from the schizophrenia drug-protein interactome to clinical research - an important step, especially in the field of psychiatric drug development which faces a high failure rate.

Schizophrenia is a complex disorder with a cumulative impact of variable genetic effects coupled with environmental factors[1]. The Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC) had identified 108 genetic loci that likely confer risk for schizophrenia. Prior to this, around 25 genes were being studied for their association with the disorder[2]. While the role of genetics has been clearly validated by the genome-wide association studies (GWAS), the functional impact of the risk variants is not well understood. Several of the schizophrenia genes, especially those implicated by the GWAS have unknown functions and/or pathways. To discover the functional role of these genes, and promote discovery of novel therapeutics, we had carried out a computational analysis of the protein-protein interactions (PPI) network, or the interactome, of schizophrenia associated genes[3]. The schizophrenia interactome, comprising 101 schizophrenia genes and about 1,900 PPIs, provided valuable results highlighting the functions and pathways tied to schizophrenia genes through their protein interactome[3]. A valuable result from this study was the drug-target interactome that showed a total of 524 drugs targeting 53 proteins in the schizophrenia interactome. Many of these drugs were labeled for therapeutic value for nervous system as expected, but there were several drugs that were labeled for other anatomical systems in the human body.

As drug approvals for psychiatric indications have been facing a high failure rate in the last few years[4], it would be beneficial to study whether these drugs that target proteins from the schizophrenia interactome could be repurposed for treatment of schizophrenia. Finding alternate uses for approved drugs would be optimal, and such uses are being found in recent years[5–7].

Diseases are often considered to be driven by an abnormal or perturbed expression of a multitude of genes which together constitute unique differential (gene) expression signatures (DES)[8–12]. Drugs administered to treat

[1]Supercomputer Education and Research Centre, Indian Institute of Science, Indian Institute of Science, Bengaluru, India. [2]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, USA. [3]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, USA. Correspondence and requests for materials should be addressed to M.K.G. (email: madhavi@pitt.edu)

these diseases often revert the expression of these genes to their normal levels[13,14]. DES for disease versus normal are quantified using gene expression analysis based on microarrays and RNA sequencing methods, and are deposited in online repositories, which make the data freely available for integrated computational analyses[15]. Similarly, DES for drug-treated versus untreated is made available through Connectivity Map (CMAP)[16]. In order to analyze the suitability of these drugs for repurposing, we build over the results from our previous work on schizophrenia interactome discovery and analysis[3], utilizing large transcriptomic databases such as CMAP and Gene Expression Omnibus (GEO), and employing a bioinformatics data analysis software suite named BaseSpace Correlation Engine[17]. The approach of repurposing drugs based on the negative correlation of drug-induced versus disease-associated gene expression profiles has resulted in some valuable results in the past. Topiramate, an anti-convulsant drug used in the treatment of epilepsy, was identified to be potentially repurposable for inflammatory bowel disease (IBD), based on the negative correlation of drug-induced profiles extracted from CMAP and disease-associated profile from GEO[18]. They further validated the efficacy of this drug in a rodent model of IBD[18].

Many genes harboring variants associated with schizophrenia, such as DTNBP1, DAOA, NRG1 and RGS4, show differential gene expression in post-mortem brain samples obtained from schizophrenia patients compared with normal controls[19]. In schizophrenia, it has been pointed out that the effect of genetic variants may, in fact, be reflected on gene expression rather than on the structure of the proteins coded by these genes[20]. Gene expression has been described as a 'psychiatric endophenotype' in schizophrenia[19]. A psychiatric endophenotype may broadly be defined as a measurable phenotype, namely, any neuroanatomical, physiological, psychological, biochemical or molecular aspect of brain function, having some definitive disease-associated genetic component, and contributing to a larger behavioral trait such as 'cognitive dysfunction' or 'psychosis' underlying a complex disorder such as schizophrenia[19]. The 'definitive genetic component', in this case, could be a set of disease susceptibility genes harboring sequence variants affecting the expression of the susceptibility genes themselves, or a set of genes differentially expressed in patients compared with healthy subjects. These genes may uncover novel pathways underlying some behavioral trait contributing to disease etiology. For example, it was recently shown that expression of genes associated with immunological processes vary with cognitive performance in familial schizophrenia[21]. So, our method to identify repurposable drugs may be tested on schizophrenia, in which differential gene expression plays a critical role.

Results

In our prior work[3], we presented 524 drugs that target any of the proteins in the Schizophrenia Interactome[3]. We pruned this large list of drugs by comparing differential expression profiles induced by drug to profiles associated with schizophrenia, using our *in silico* protocol, and shortlisted drugs that had a negative correlation between these expression profiles[22]. We carried out bioinformatics analysis on the shortlist of drugs identified thus, to answer the following questions on their biological validity to schizophrenia (see Fig. 1): Have any of these drugs been considered for clinical trials? Are the genes targeted by these drugs associated with neuropsychiatric disorders? Are the genes with opposite expression in drug versus schizophrenia associated with morphological or physiological phenotypes of the mammalian nervous system? Do other drugs targeting the same genes as the shortlisted drugs show clinical activity in neuropsychiatric disorders? Are any genes in the PPI network of the genes targeted by the shortlisted drugs associated with neuropsychiatric disorders? Are any genes in the PPI network of genes with opposite expression in drug versus schizophrenia involved in pathways relevant to schizophrenia? Are they also GWAS genes associated with traits or diseases having a pathophysiological overlap with schizophrenia? These questions were based on the fact that genes associated with traits related to the nervous system and genes linked to neuropsychiatric disorders have been shown to converge in specific co-expression modules, indicating shared genetic basis and disease mechanisms[23]. Drugs used for treatment of a neuropsychiatric disorder may be repurposable for schizophrenia by virtue of shared genes and mechanisms. Each of these sources of information is assessed separately in parallel, highlighting which of the drugs have multiple sources of supporting information.

We followed an established approach to identify drugs that have opposite differential expression to the differential expression of schizophrenia (i.e., genes over-expressed in schizophrenia are under-expressed by drug treatment and vice versa)[8]. We identified such drugs using the BaseSpace Correlation Engine software suite, a data analysis platform used to study the effect of diseases and/or drugs on publicly available gene expression data[17]. This analysis resulted in 12 drugs. Although in each case, there are some genes that are differentially expressed in the same direction for both the drug and disorder, the overall effect on the entire transcriptome has an anti-correlation, leading to 12 drugs as potential candidates for further studies (Table 1 and Fig. 2). The top 5 drugs by the score of anti-correlation are cromoglicic acid, bepridil, acetazolamide, dimenhydrate, cinnarizine, of which bepridil and dimenhydrate may be excluded due to their side-effects related to nervousness and hallucinations (see Table 1), thus leaving cromoglicic acid, acetazolamide and cinnarizine as top candidates. There were 30 drugs indicated for schizophrenia in DrugBank[24]. 23 out of these occur in the schizophrenia drug-protein interactome (77%). We checked the overlap of drugs indicated for other diseases to infer the specificity of this result, namely, coronary heart disease (25%), lung cancer (50%), diabetes (33%), chronic kidney disease (0%), post-traumatic stress disorder (75%) and bipolar disorder (66%). As expected, there was a larger overlap with neurological disorders compared to other unrelated disorders. 50% overlap with lung cancer drugs may be explained by the large number of drug targets implicated in cancers, and their vital role in numerous basic cellular functions. Eleven of these did not have relevant datasets in BaseSpace, or even though a negative correlation was found, the p-value was insignificant for schizophrenia gene expressions studies. Of 23 *known* schizophrenia drugs – six of them, namely, clozapine, haloperidol, molindone, perphenazine, amitriptyline and nortriptyline, had negative correlation with schizophrenia and 6 others had a positive correlation with schizophrenia. Sources of datasets in which differential expression is observed is listed in Data File 1.
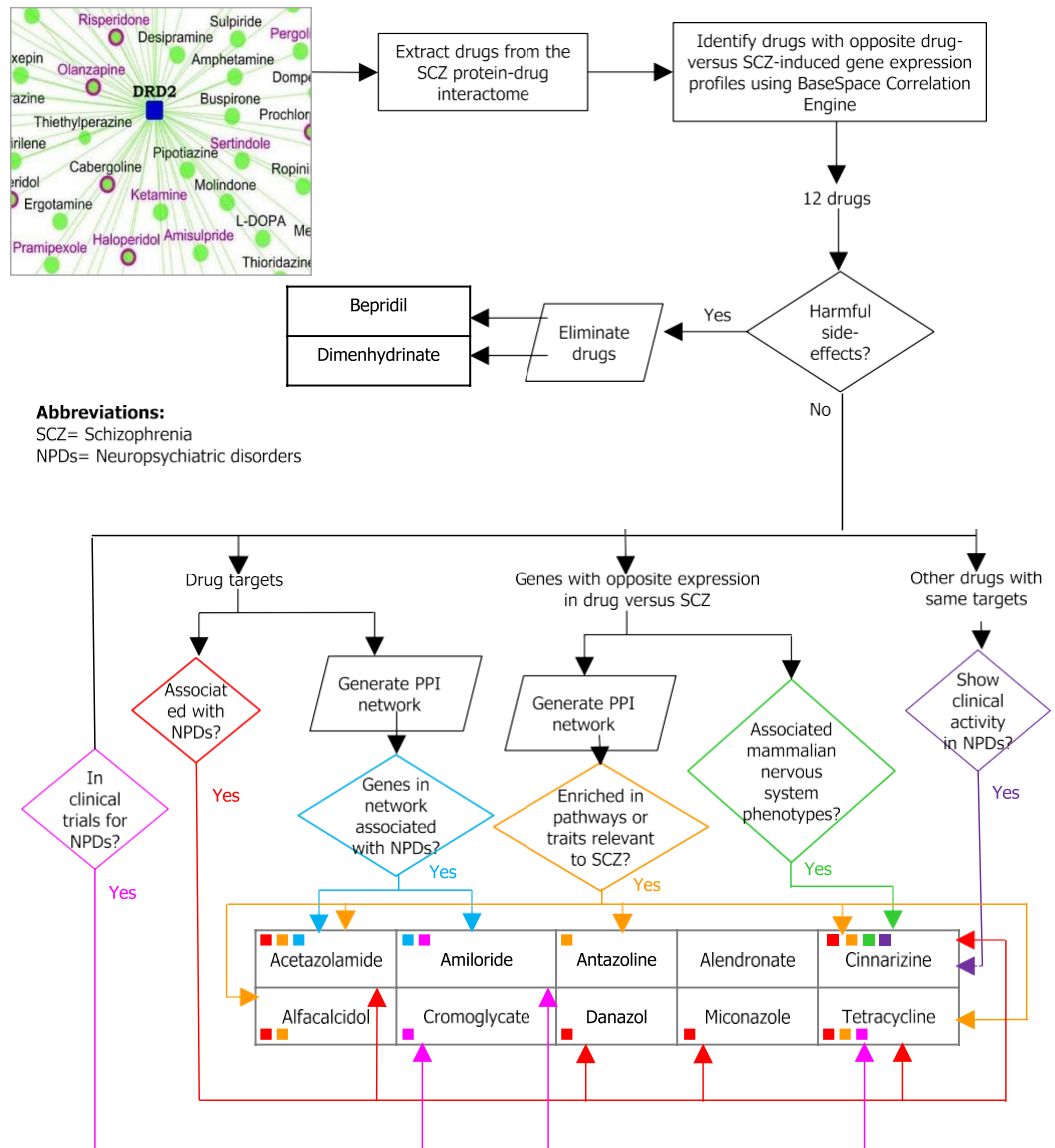
Figure 1. Graphical abstract depicting the steps taken in this study to assess the biological validity of the shortlisted drugs. Drugs were extracted from the schizophrenia drug-protein interactome and screened for negative correlation of drug-induced versus disease-associated gene expression profiles. Drugs shortlisted in this manner were further checked for their toxicity, and eliminated if they were found to have harmful side effects. The targets of the remaining drugs and their network of protein-protein interactions were checked for their association with schizophrenia (SCZ)/other neuropsychiatric disorders (NPDs) using DisGeNET. Genes with opposite expression in drug-induced versus disease-associated profile were analyzed for their association with nervous system phenotypes (Mammalian Phenotype Ontology). Their networks were analyzed for enrichment of SCZ-associated pathways/GWAS traits. Apart from this, it was checked whether the shortlisted drugs are already being tested against NPDs (NIH Clinical Trials), and whether other drugs with the same targets show clinical activity in NPDs. Different sources of supporting information are shown by lines of different colors. Each of the drugs is also tagged with little squares of colors of corresponding supporting information. For example, amiloride is supported by "genes in network associated with neuropsychiatric disorders" (blue) and "in clinical trials for neuropsychiatric disorders" (bright pink). Acetazolamide, cinnarizine and tetracycline each are supported by 3 sources of supporting information.

Acetazolamide. The protein targets of acetazolamide are carbonic anhydrases (CA*) and aquaporin (AQP1). We collected known and computationally predicted PPIs of these targets of acetazolamide and queried the DisGeNet[25] database whether any of the proteins in this interaction network are associated with neuropsychiatric disorders. While Fig. 2 shows the protein targets only from schizophrenia interactome, Fig. 3 shows the network of all protein targets (orange colored nodes) of acetazolamide and their PPIs. Nineteen genes within this network are associated with various neuropsychiatric disorders (nodes with green border in Fig. 3; Data File 2): AQP1 and CA2, which are acetazolamide targets, DAXX, EPHB2, HSPD1, SLC4A3, SLC9A1, SRC, TCF4, TNK2,

218

Drug	Drug class	Original therapeutic purpose(s)	Pharmacokinetic details: dosage form, delivery route, half-life	Toxicity	Correlation with all data types, Overall correlation score	Correlation with SCZ gene expression study, Correlation score	Bs1	Bs2	Bs1 & Bs2 up	Bs1 & Bs2 down	Bs1 up & Bs2 down	Bs1 down & Bs1 up
Amitriptyline	Dibenzo-cycloheptenes	Major depressive disorder and anxiety disorders, treatment of secondary depression in schizophrenia	Tablet, oral, 25 hours	Abnormally low blood pressure, confusion, convulsions, dilated pupils and other eye problems on overdosing, and withdrawal symptoms including gastrointestinal disturbances, anxiety, and insomnia	Negative, 76	Negative, 100	HL60 cells + amitriptyline, 12.8 uM _vs_ DMSO vehicle	Hippocampus tissues from schizophrenia patients _vs_ normals	9	3	19	31
Haloperidol	Alkyl-phenyl-ketone	Schizophrenia and other psychoses, delusional disorders, ballism, and Tourette syndrome, adjunctive therapy in mental retardation, chorea associated with Huntington's disease	Solution/tablet, oral, 24 hours	Cardiovascular effects, extrapyramidal symptoms, tardive dyskinesia, neuroleptic malignant syndrome, hematologic effects	Negative, 70	Negative, 68	HL60 cells + haloperidol, 10 μM vs. DMSO vehicle	Prefrontal cortex Brodmann area 46 of schizophreniacs with short DOI vs. helathy controls	3	6	30	1
Molindone	Indoles and derivatives	Schizophrenia, other psychoses and aggressive type of undersocialized conduct disorder	Tablet, oral, not available	Not available	Negative, 76	Negative, 57	MCF7 + molindone, 12.8 μM vs. DMSO vehicle	Hippocampus tissues from schizophrenia patients _vs_ normals	22	23	174	21
Clozapine	Dibenzo-diazepines	Atypical antipsychotic drug used in schizophrenia	Tablet, oral, 4 to 12 hours	Black-box warning for agranulocytosis	Negative, 59	Negative, 100	HL60 cells + clozapine, 10 μM vs. DMSO vehicle	Prefrontal cortex Brodmann area 46 of schizophreniacs with short DOI vs. helathy controls	10	23	101	2
Nortriptyline	Dibenzo-cycloheptenes	Clinical depression, treatment of depressive symptoms in schizophrenia (dose adjustments are necessary to safely use the drug in schizophrenia, as it has been shown to exacerbate psychosis)	Capsule, oral, 26 hours	Cardiac dysrhythmias, severe hypotension, shock, congestive heart failure, pulmonary edema, convulsions, and CNS depression, including coma on overdosing, and withdrawal symptoms include gastrointestinal disturbances, anxiety, and insomnia	Negative, 50	Negative, 89	HL60 cells + nortriptyline, 13.4 uM _vs_ DMSO vehicle	Hippocampus tissues from schizophrenia patients _vs_ normals	6	6	26	3
Perphenazine	Phenothiazines	Schizophrenia and the manic phases of bipolar disorder	Tablet, oral, 8 to 12 hours	Stupor or coma, convulsive seizures in children	Negative, 80	Negative, 100	HL60 cells + perphenazine, 10 μM vs. DMSO vehicle	Prefrontal cortex Brodmann area 46 of schizophreniacs with short DOI vs. helathy controls	4	7	78	7
Acetazo-lamide	Thiadiazole sulfonamides	Glaucoma, mountain sickness	Tablet, oral, 3 to 9 hours	Not available	Negative, 76	Negative, 100	MCF7 cells + acetazolamide, 18 uM _vs_ DMSO vehicle	Whole blood from schizophrenic patients _vs_ healthy controls_ GPL6947	67	38	95	119
Alendronate	Bisphos-phonates	Osteoporosis	Tablet, oral, 10 years	Damage of oesophagus	Negative, 68	Negative, 57	Heart of rats + ALENDRONIC ACID at 138 mg-kg-d in CMC by oral gavage 5d _vs_ vehicle	Associative striatum tissues from schizophrenia patients _vs_ normals	1	14	12	8
Continued												

Drug	Drug class	Original therapeutic purpose(s)	Pharmacokinetic details: dosage form, delivery route, half-life	Toxicity	Correlation with all data types, Overall correlation score	Correlation with SCZ gene expression study, Correlation score	Bs1	Bs2	Bs1 & Bs2 up	Bs1 & Bs2 down	Bs1 up & Bs2 down	Bs1 down & Bs1 up
Alfacalcidol	Vitamin D and derivatives	Vitamin D supplement	Capsule, oral, not available	Hypercalcemia	Negative, 46	Negative, 100	Liver of rats + ALFACALCIDOL at 043 mg-kg-d in CMC by oral gavage 1d _vs_ vehicle	Prefrontal cortex Brodmann area 46 - schizophrenics with short DOI _vs_ healthy controls	22	120	125	25
Amiloride	Aminop-yrazines	Hypertension, heart failure, edema	Tablet, oral, 6 to 9 hours	Dehydration and electrolyte imbalance	Negative, 58	Positive, 66	HL60 cells + amiloride, 13.2 uM _vs_ DMSO vehicle	Whole blood from schizophrenic patients _vs_ healthy controls_ GPL6947	36	26	14	18
Antazoline	Phenylben-zamines	Nasal congestion, allergic conjunctivitis	Liquid, opthalmic, not available	Not available	Negative, 60	Negative, 91	HL60 cells + antazoline, 13.2 uM _vs_ DMSO vehicle	Whole blood from schizophrenic patients _vs_ healthy controls_ GPL6947	5	10	10	24
Bepridil	Phenylben-zamines	Angina	Tablet, oral, 24 to 50 hours	Gastrointestinal problems, dizziness, asthenia, nervousness	Negative, 77	Negative, 40	HL60 cells + bepridil, 10 uM _vs_ DMSO vehicle	Neural progenitors derived from donor stably expressing GFP - schizophrenia _vs_ normal	36	51	73	68
Cinnarizine	Diphenyl-methanes	Motion sickness, vertigo	Tablet, oral, 3 to 4 hours	Drowsiness, skin problems, lethargy, movement problems	Negative, 64	Negative, 100	HL60 cells + cinnarizine, 10.8 uM _vs_ DMSO vehicle	Whole blood from schizophrenic patients _vs_ healthy controls_ GPL6947	20	15	21	57
Cromoglicic acid	Chromones	Asthma prophylaxis, aerosol	Solution, oral, 1.3 hours	Cough, nasal congestion, nausea, sneezing and wheezing	Negative, 84	Negative, 64	MCF7 cells + cromoglicic acid, 7.8 uM _vs_ DMSO vehicle	Neural progenitors derived from donor stably expressing GFP - schizophrenia _vs_ normal	15	18	36	13
Danazol	Estrane steroids	Endometriosis, fibrocystic breast disease, hereditary angioedema	Capsule, oral, 24 hours	Not available	Negative, 61	Negative, 100	HL60 cells + danazol, 11.8 uM _vs_ DMSO vehicle	Whole blood from schizophrenic patients _vs_ healthy controls_ GPL6947	173	335	460	1264
Dimenhy-drinate	Diphenyl-methanes	Motion sickness, nausea	Solution, intramuscular or intravenous, 1 to 4 hours	Delerium, hallucinations, excitement	Negative, 64	Negative, 50	HL60 cells + dimenhydrinate, 8.6 uM _vs_ DMSO vehicle	Whole blood from schizophrenic patients _vs_ healthy controls_ GPL6947	22	10	35	6
Miconazole	Benzylethers	Antifungal medication used in vaginal infections	Tablet, buccal, not available	Oral toxicity in mice at LD50 = 3800 mg/kg	Negative, 60	Negative, 50	HL60 cells + miconazole, 9.6 uM _vs_ DMSO vehicle	Hippocampus tissues from schizophrenia patients _vs_ normals	11	19	66	20
Tetracycline	Tetracyclines	Antibiotic used in acne, cholera, brucellosis, plague, malaria, and syphilis	Capsule, oral, 6 to 12 hours	Oral toxicity in mice at LD50 = 808 mg/kg	Negative, 49	Negative, 100	HL60 cells + tetracycline, 8.4 uM _vs_ DMSO vehicle	Whole blood from schizophrenic patients _vs_ healthy controls_ GPL6947	26	20	44	149

Table 1. Details of known schizophrenia drugs and drugs identified as potentially repurposable for schizophrenia: Pharmacokinetic information is collected from DrugBank (www.drugbank.ca). Known schizophrenia drugs are shown in italics.

....

TRAF1, TRAF2, MTUS2, PICK1, GRM3, OLR1, TBP, PML and FOS, giving credence to the consideration that it has a potential application to schizophrenia. Acetazolamide has been shown to have high inhibitory activity against human CA2 (hCA II), the ubiquitous cytosolic enzyme (inhibition constant, $K_i = 12$ nM) and human CA7 ($K_i = 2.5$ nM), the brain-specific form of the enzyme[26]. Human CA2 was found to be catalytically highly active (defined in terms of K_{cat}/K_m for CO_2 hydration described by two ionizations at pKa 6.2 and 7.5, with a maximum
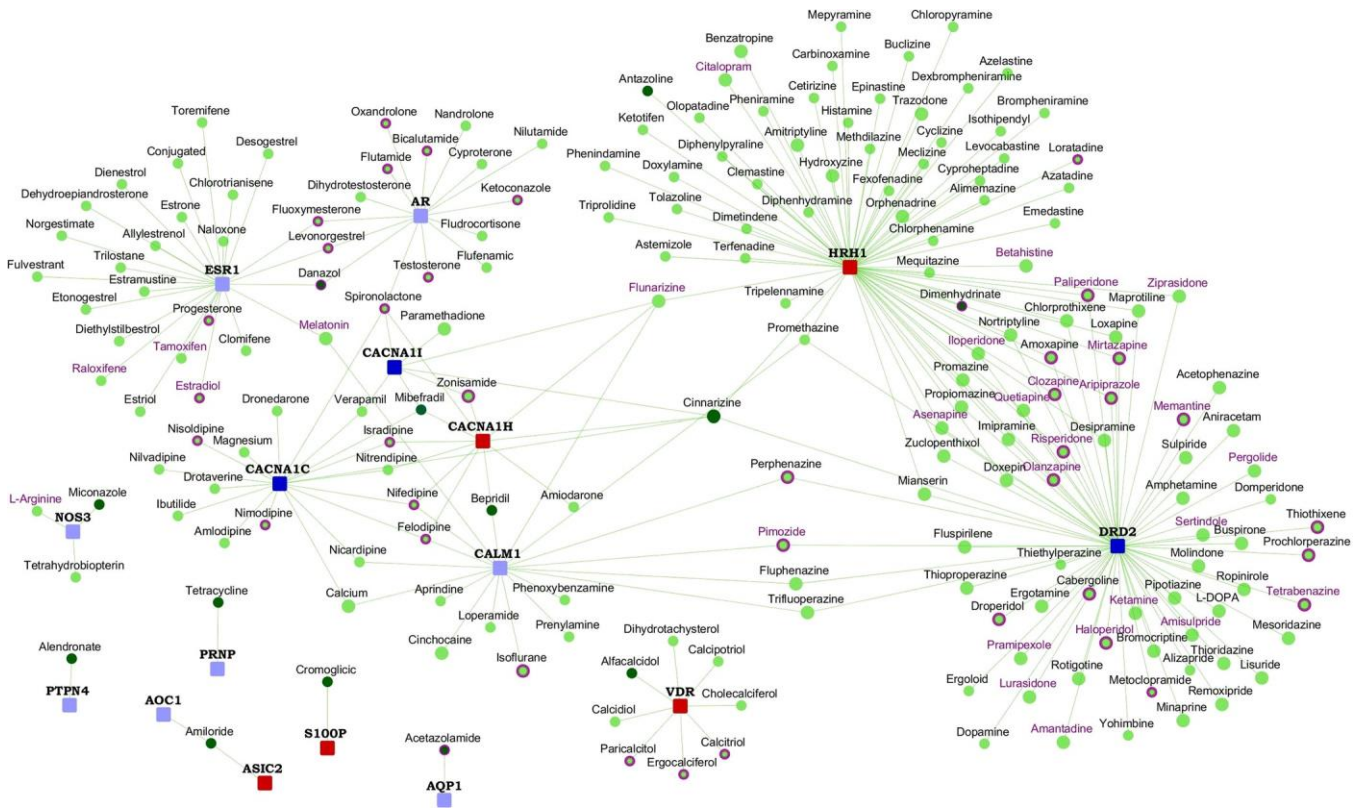
Figure 2. Drugs potentially repurposable for schizophrenia: The network highlights the shortlisted drugs that may be potentially repurposed for schizophrenia. The shortlisted drugs are shown as round nodes colored in dark green, and other drugs are shown as light green nodes. FDA approved drugs are shown with purple borders. Drugs with purple labels are in clinical trials for schizophrenia. Schizophrenia genes are square nodes colored in dark blue, known interactors are colored in light blue and novel interactors in red.

approaching $8 \times 10^7 \, M^{-1} \, s^{-1}$)[27]. K_{cat}/K_m for human CA2 is 1.5×10^{8} [27]. The increase in extracellular pH which accompanies neural activity is generated by the exchange of external H^+ for cytosolic Ca^{2+}. This process, and its impact on the glutamate receptors, NMDARs, has been shown to be regulated by CA14 in the synaptic microenvironment[28]. On these lines, it is interesting to note that CA3 has been predicted to be a novel interactor of the glutamate receptor, GRM3, mutations in which have been associated with schizophrenia[29].

We assembled the network of PPIs of genes that are differentially expressed by each of the shortlisted drugs and carried out network and enrichment analysis using a tool called LENS[30]. The networks of genes found to be differentially expressed in acetazolamide, antazoline and cinnarizine, having an anti-correlation in schizophrenia, were shown to be enriched in ubiquitination and proteasome degradation pathways (Data File 3). The ubiquitin proteasome system has been identified as an important pathway in several genetic studies of neuropsychiatric disorders including Alzheimer's disease, Parkinson's disease, psychosis and bipolar disorder[31]. Many gene expression studies performed on blood collected from schizophrenia patients, and on post-mortem samples of hippocampus, prefrontal cortex and temporal cortex of patients have pointed at abnormalities in the ubiquitin proteasome pathway, which targets protein for degradation in the cell[31]. Moreover, reduced protein ubiquitination, reduced levels of ubiquitin and ubiquitin-like activases and ligases, were identified in a region of the brain called the left superior temporal gyrus in schizophrenia patients[31]. Left superior temporal gyrus, the volume of which has been shown to decrease in schizophrenia patients, is involved in the development of auditory hallucinations and thought process abnormalities seen in schizophrenia[31]. Interestingly, acetazolamide which has been shown to mediate diuretic effects through its action on AQP1, induces AQP1 ubiquitination, and a proteasome inhibitor reversed its downregulatory action on AQP1[32]. RAD51AP1 and AQR are novel interactors of the calcium channel CACNA1C and the nicotinic receptor CHRNA7 respectively in the schizophrenia interactome, found to have an anti-correlated expression in schizophrenia and acetazolamide treatment. It has been shown that UAF1, an interaction partner of USP1 deubiquitinating enzyme, associates with RAD51AP1, which interacts with RAD51 to mediate homologous recombination repair[33]. NEDD4-1, an ubiquitin ligase, has been shown to promote the sorting of newly synthesized calcium voltage gated channels for proteasomal degradation[34]. Suppression of AQR in HepG2, a liver cancer line, has been shown to inhibit protein ubiquitination[35]. It has been shown that the expression of nicotinic receptors on the cell surface is regulated by the ubiquitin proteasomal system[36]. The networks of genes found to be differentially expressed in alfacalcidol and tetracycline, having an anti-correlation in schizophrenia, were shown to be enriched in the neutrophil degranulation pathway (Data File 3). Degranulating activity of neutrophils has been attributed to dysfunctional permeability of the blood-brain barrier in schizophrenia[37].

Figure 3. Network of PPIs among targets of acetazolamide: The network shows protein-protein interactions that connect the targets of acetazolamide, which are shown as orange colored nodes. Nodes that connect these target genes are shown as grey colored nodes. Nodes with light green borders are genes associated with neuropsychiatric disorders. Novel interactions are shown as red edges and known interactions as blue edges.

Network of genes which were differentially expressed in acetazolamide and had an anti-correlation with schizophrenia were found to be significantly enriched for association to rheumatoid arthritis (Data File 3). Recently, the reduced prevalence of rheumatoid arthritis observed in schizophrenia patients was attributed to SNPs (single nucleotide polymorphisms) in the HLA region that conferred differential risk for schizophrenia and rheumatoid arthritis[38]. The interactomes of schizophrenia and rheumatoid arthritis genes also showed a significant overlap even outside of HLA genes, and shared common pathways[38].

Alfacalcidol. Alfacalcidol targets the protein VDR which was found to be overexpressed in whole blood obtained from schizophrenic patients compared to healthy controls (fold change (FC) = 2.21, p-value = 0.0037)[39]. The network of genes differentially expressed in alfacalcidol was enriched in GWAS genes associated with inflammatory bowel disease (Data File 3). The incidence of schizophrenia has been shown to be high in patients with immune-mediated inflammatory diseases such as inflammatory bowel disease, rheumatoid arthritis and multiple sclerosis[40].

Amiloride. With our focus on candidate drugs for repurposing (i.e. those that exhibited a negative correlation to schizophrenia but are not currently labeled for this use), we queried the ClinicalTrials.gov database (https://clinicaltrials.gov/) and found that amiloride is being tested in clinical trials for its efficacy in attention deficit hyperactivity disorder.

We analyzed the PPI network of proteins targeted by the drug amiloride (Fig. 4), despite its positive correlation with schizophrenia gene expression because its overall correlation with a range of schizophrenia datasets was negative, and because of the biological characteristics of its targets. The protein targets of amiloride are ASIC1, ASIC2, AOC1, SLC9A1, PLAU, SCNN1A, SCNN1B, SCNN1G and SCNN1D (orange nodes in Fig. 4). The network of PPIs among these targets of amiloride shows that 12 genes, including ASIC2, AOC1 and PLAU, which are amiloride targets, NEDD4, STX1A, MAPK1, HECW1, DAO, CSNK2A1, LASP1, SMG6 and PICK1 are associated with various neuropsychiatric disorders (nodes with green border in Fig. 4; Data File 2). ASIC2 was a computationally predicted interactor of the gene SMG6, structural variants in which have been associated with schizophrenia or bipolar disorder in a Spanish population[41,42]. SMG6 is located in the chromosomal region 17p13.3, linked to lissencephaly, a neuronal migration disorder arising from incomplete neuronal migration to the cerebral cortex during gestation, and characterized by an absence of normal convolutions in the cerebral cortex and an abnormally small head (or microcephaly)[42,43]. ASICs (*acid-sensing ion channels*) are members of the epithelial Na^+ channel (ENaC) family of ion channels, expressed in the nervous system[44]. It was shown in a study that ASIC2 is not expressed at the cell surface of high grade glioma (brain tumor) cells and this may be responsible for the constitutively activated inward Na^+ current, which promotes increased cell growth and
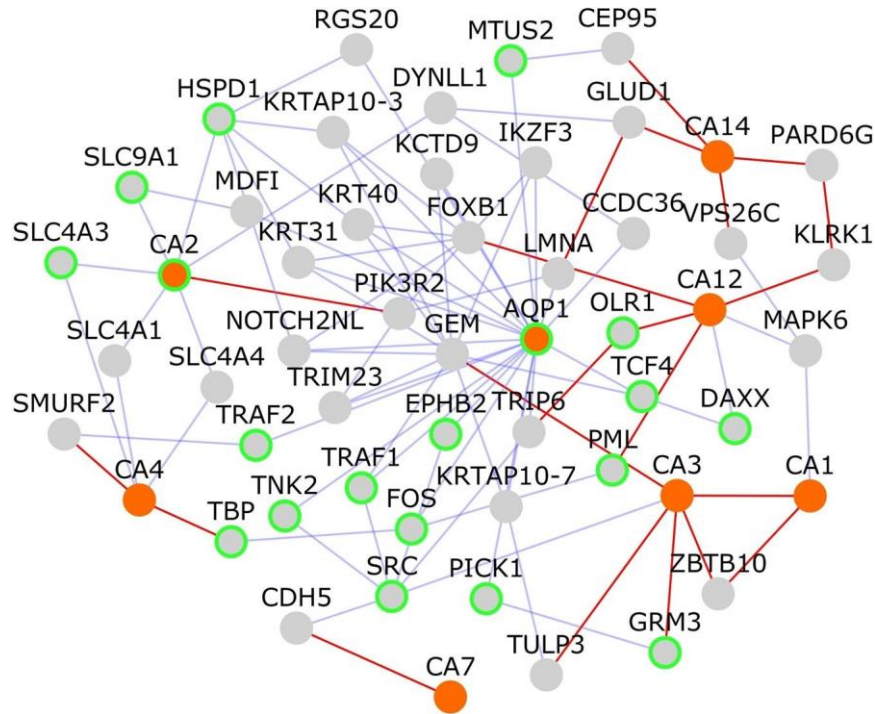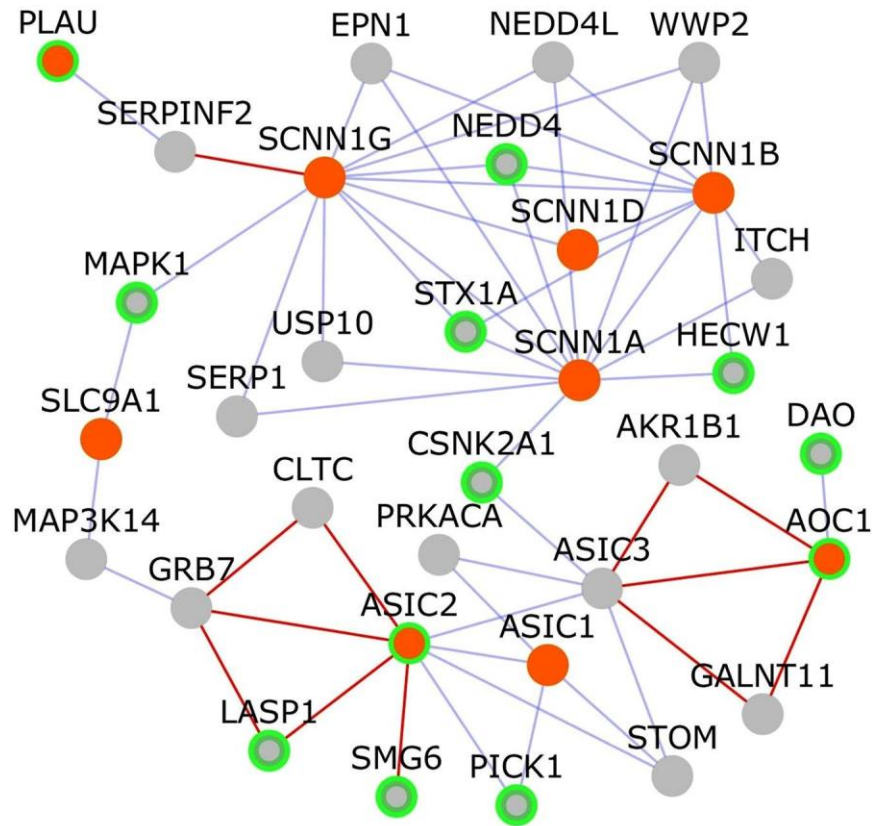
Figure 4. Network of PPIs among targets of amiloride: The network shows protein-protein interactions that connect the targets of amiloride, which are shown as orange colored nodes. Nodes that connect these target genes are shown as grey colored nodes. Nodes with light green borders are genes associated with neuropsychiatric disorders. Novel interactions are shown as red edges and known interactions as blue edges.

migration in these cells[44]. In such glioma cells, compounds such as glycerol and the transcriptional regulator, sodium 4-phenylbutyrate, were shown to inhibit the constitutively activated inward Na$^+$ current and reduce cell growth and migration[44]. These compounds were shown to induce the movement of ASIC2 to the plasma membrane, and prevent the active inward current through negative regulatory mechanisms, reducing the ability of glioma cells to proliferate and migrate[44].

Antazoline. The networks of genes found to be differentially expressed in antazoline having an anti-correlation in schizophrenia, were shown to be enriched in ubiquitination and proteasome degradation path- ways (Data File 3). The network of genes differentially expressed in antazoline, and with an opposite expression in schizophrenia, was significantly enriched in GWAS genes associated with brain connectivity (Data File 3). Abnormal interactions between brain networks have been pointed out to be an important contributing factor in schizophrenia etiology[45].

Cinnarizine. The networks of genes found to be differentially expressed in cinnarizine having an anti-correlation in schizophrenia, were shown to be enriched in ubiquitination and proteasome degradation pathways (Data File 3). We checked whether any of the genes having anti-correlated expression on cinnarizinetreatment and in schizophrenia were associated with mammalian phenotype ontology (MPO) terms related to various morphological or physiological aspects of the nervous system (http://www.informatics.jax.org/)[46]. It wasfound that mutations in 13 genes were associated with relevant MPO terms, namely, AHI1, ENTPD1, IFNGR1, NAP1L1, NPTN, PIK3CA, PKN2, PRKDC, PTGS2, RBM12, SEC. 23 A, SS18L1 and UBE3A. Two of these genes, IFNGR1 and AHI1, both linked to 'abnormal depression-related behavior', are predicted to have a novel interaction between them. Depressive symptoms have been observed in schizophrenia patients[47]. IFNGR1 has been found to be necessary for the induction of IDO, the tryptophan synthesizing enzyme, which plays a role in depressive behavior, induced by inflammation[47]. AHI1 is associated with susceptibility to schizophrenia and autism[47]. Mice lacking neuronal expression of AHI1 had reduced levels of tyrosine kinase receptor B and a depressive phenotype, which was alleviated by antidepressants and overexpression of TRKB[47]. BDNF/TRKB signaling has been shown to play a key role in depression. Altered BDNF/TRKB signaling in the prefrontal cortex, hippocampus and nucleus accumbens has been shown to give rise to depressive phenotype induced by inflammation[48]. Another gene, UBE3A, was associated with increased dopamine and serotonin levels, abnormal brain wave pattern, cerebral cortex morphology, dendrite morphology, GABA-mediated receptor currents, long term potentiation and
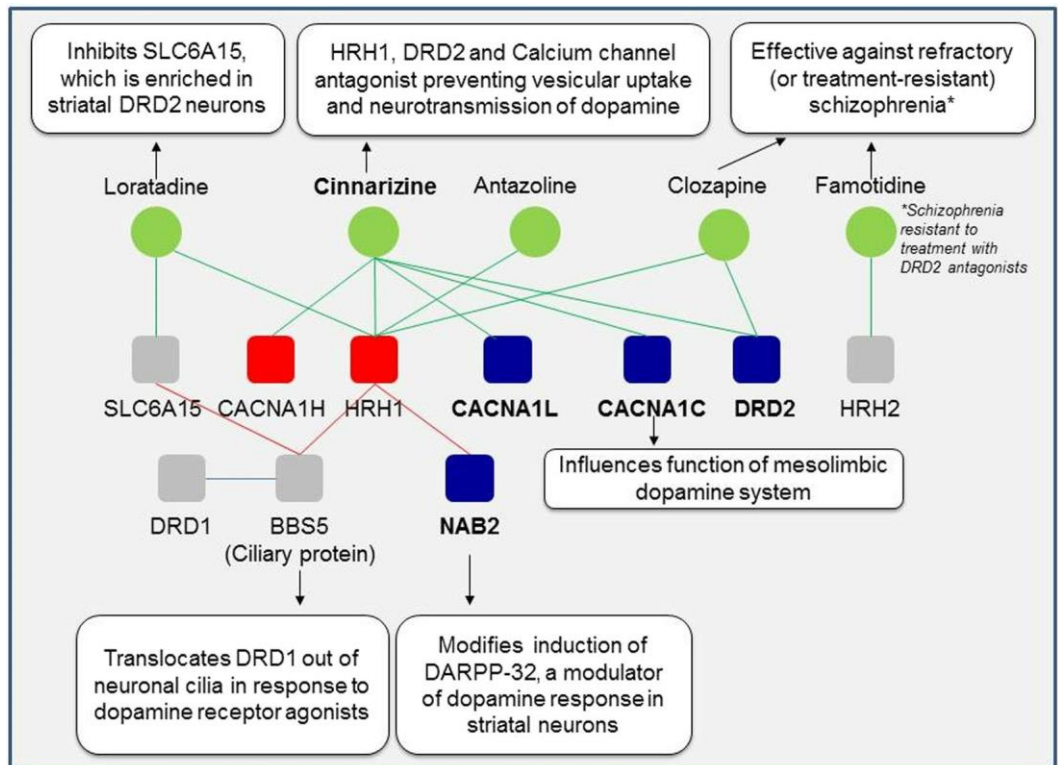
Figure 5. Cinnarizine and its targets in the schizophrenia interactome: The drug cinnarizine is shown here with the proteins it targets from the schizophrenia interactome. 4 additional proteins (BBS5, DRD1, HRH2 and SLC6A15) and 3 additional drugs (loratadine, clozapine and famotidine) that are relevant to the hypothesis are also shown. Cinnarizine, targets 3 schizophrenia genes and 2 novel interactors which constitute calcium channels, and histamine & dopamine receptors. Since histamine antagonists are known to reduce dopamine levels through their action on dopamine receptors, and calcium channel antagonists are known to reduce dopamine neurotransmission, the HRH1, DRD2 and calcium channel antagonist, cinnarizine, may be repurposable for schizophrenia. Another shortlisted drug, antazoline, is not part of the reasoning presented here even though it is an HRH1 antagonist. Schizophrenia genes are shown as dark blue colored nodes, novel interactors are red colored nodes and genes relevant to the hypothesis, which are not in the schizophrenia interactome, are shown as grey colored nodes.

- -

nervous system electrophysiology. Yet another gene, NPTN, was linked to abnormal synaptic transmission in the central nervous system and abnormal dendritic spine morphology.

The network of genes differentially expressed in cinnarizine was enriched in GWAS genes associated with inflammatory bowel disease (Data File 3).

We queried Drug Bank[24] to find drugs that targeted the same genes as the shortlisted drugs, and checked whether they demonstrated any clinical activity in schizophrenia or other neuropsychiatric disorders. Risperidone, nimodipine, nilvadipine, flunarizine, nifedipine, cannabidiol and clozapine target the same genes as cinnarizine. Flunarizine (targeting CALM1, CACNA1H) showed good efficacy and tolerability for the treatment of schizophrenia[49]. Nifedipine (which targets CALM1, CACNA1H) enhanced learning and memory in schizophrenic patients with tardive dyskinesia[50]. Cannabidiol (which targets CACNA1H) shows beneficial effects as an adjunctive drug along with existing anti-psychotic medication in schizophrenia[51]. Risperidone (targeting DRD2) is used to treat schizophrenia, bipolar disorder, and irritability in autistic patients[52–54]. Nimodipine (CACNA1C) has been found effective for treating resistant bipolar mood disorder[55]. Nilvadipine (CACNA1C) was found to be effective in treatment of schizophrenia[56]. Clozapine (targeting HRH1) is effective in treatment-resistant schizophrenia[57].

Cinnarizine targets CACNA1H which is found to be overexpressed in neural progenitor cells differentiated for 2 days from induced pluripotent stem cells of schizophrenia patients versus healthy subjects (FC = 3.1227, p-value = 4.10E-20)[58]. Cinnarizine targets HRH1, which has been linked to schizophrenia etiology. It also targets CACNA1C, associated with bipolar disorder, schizophrenia and depressive disorder, and CACNA1H, associated with epilepsy and autism. It targets DRD2, linked to bipolar disorder, schizophrenia, depressive disorder, Parkinson's disease and attention deficit hyperactivity disorder.

Cromoglicic acid. Cromoglicic acid is being tested in clinical trials for its efficacy in Alzheimer's disease. It has been reported that cromoglicic acid in combination with ibuprofen reduces the levels of amyloid-beta protein levels, a pathological biomarker in Alzheimer's disease, and promotes a neuroprotective state by activating

microglia and inducing phagocytosis of amyloid-beta proteins[59]. Based on this work, cromoglicic acid has been considered for further study by our clinical collaborators and is currently in clinical trials (ClinicalTrials.gov Identifier: NCT03794076).

Danazol and miconazole. Danazol and miconazole target ESR1 and NOS3, both associated with Alzheimer's disease. NOS3 was also identified as a potential target for schizophrenia based on its druggability, membership in schizophrenia-related biological pathways and differential expression in schizophrenia[60].

Tetracycline. Minocycline, a broad spectrum tetracycline antibiotic (where tetracycline is one of the short-listed drugs), has been shown to be effective as an adjunctive drug, improving the effect of antipsychotic drugs in schizophrenia[61]. Tetracycline targets PRNP, linked to depressive disorder, Huntington disease-like 1 and Alzheimer's disease. The network of genes differentially expressed in tetracycline was enriched in GWAS genes associated with inflammatory bowel disease (Data File 3).

In summary, clinical trial data, network-based analyses and literature review support the biological validity of 9 out of the 12 drugs proposed to be repurposable for schizophrenia, namely, acetazolamide, alfacalcidol, amiloride, antazoline, cinnarizine, cromoglicic acid, danazol, miconazole and tetracycline.

Discussion

In this section, we discuss cinnarizine and alfacalcidol further due to abundant evidences in literature pointing at their potential utility as repurposable drugs for schizophrenia.

Cinnarizine, an HRH1 (*histamine receptor H1*), DRD2 (*dopamine receptor D2*) and calcium channel antagonist commonly used to treat motion sickness, may be re-purposed to treat symptoms of schizophrenia (see Fig. 5)[62]. Histamine receptors are highly expressed in brain regions associated with the higher cognitive func- tions disturbed in schizophrenia[63]. Leu49Ser mutation in HRH1 was associated with susceptibility to schizophrenia[64]. Schizophrenia patients have elevated levels of n-tele-methylhistamine, a histamine metabolite, in their cerebrospinal fluid and reduced HRH1 binding in their frontal cortex and cingulate gyrus[65]. According to the revised dopamine hypothesis of schizophrenia, hyperactive dopamine transmission in the mesolimbic areas such as the ventral tegmental area and ventral striatum including nucleus accumbens contribute to disease etiology[66]. Many studies have demonstrated a crosstalk between the dopaminergic and the histamine neuron systems. Compounds acting at histamine receptors have been shown to modulate extracellular striatal dopamine levels[67]. Enhanced release of neuronal histamine was observed on DRD2 activation and in methamphetamine or phencyclidine-induced animal models of schizophrenia[68,69]. Histamine antagonists inhibit behavioral sensitization arising from increased levels of extracellular dopamine[69–72]. The fact that refractory schizophrenia may be treated with clozapine, an HRH1 antagonist, indicates that extra-dopaminergic systems, namely, the histamine neuron system, contribute to schizophrenia etiology[57,69]. Clozapine also exhibits strong affinity to dopaminergic receptors and decreases hyperactivity of the mesolimbic dopaminergic pathway by blocking 5-HT2A (*5-hydroxytryptamine receptor 2A*)[66]. Famotidine, an HRH2 antagonist, significantly reduced psychotic symptoms in schizophrenia patients[73]. The examples of clopazine and famotidine indicate that a drug such as cinnarizine acting as a DRD2 and HRH1 antagonist may serve to alleviate psychotic symptoms arising from the interplay of dopaminergic and histamine neuron systems. Cinnarizine prevents vesicular uptake of dopamine[74]. It shows antagonistic activity at the calcium channel, CACNA1C, whose reduced levels attenuate the function of the mesolimbic dopaminergic pathway and impair behavioral responses to dopamine stimulants[75]. Calcium channel antagonists reduce neurotransmission of dopamine[76]. Even though our computational analysis supports the repurposing of cinnarizine to treat schizophrenia symptoms, its clinical utility can only be validated after experiments in pre-clinical models such as cell lines or animal models, and in clinical trials. Being an anti-histamine, cinnarizine causes drowsiness and its anti-dopaminergic activity may induce Parkinsonism and depression[77]. HRH1, targeted by cinnarizine, was predicted to interact with the schizophrenia gene NAB2. NAB2 modifies the induction of DARPP-32, which modulates the response to dopamine in striatal neurons[78]. HRH1 has also been predicted to interact with BBS5, a ciliary protein. BBS5 interacts with DRD1 and is involved in translocating DRD1 out of the cilia in response to dopamine receptor agonists, thereby implicating neuronal cilia in dopamine signaling[79]. BBS5 was predicted to interact with SLC6A15, which is enriched in striatal DRD2 neurons and inhibited by loratadine, an HRH1 antagonist[80,81].

The drug alfacalcidol, an analog of vitamin D, commonly used as a vitamin D supplement, or to treat conditions involving imbalance in calcium metabolism such as hypercalcemia and imbalance in bone metabolism such as osteoporosis, may be potentially re-purposed to treat dopaminergic symptoms in schizophrenia, possibly in combination with dopamine receptor antagonists such as clozapine[82,83]. Deficiency of vitamin D exerting its effects through VDR (*vitamin D receptor*) has been observed in schizophrenia patients[84]. Dopaminergic aspects of schizophrenia etiology as proposed by the dopamine hypothesis of schizophrenia may, at least in part, be treated by vitamin D supplementation[66]. In a study based on 9,114 subjects from the Northern Finland 1966 birth cohort, vitamin D supplementation in the first year of life was associated with reduced risk of schizophrenia in males[85]. Several studies have noted an interplay between vitamin D and dopaminergic systems[86]. VDR is highly expressed in brain regions associated with schizophrenia, namely, the hippocampus, prefrontal cortex and dopaminergic neurons in substantia nigra of rats and humans[87]. During early stages of development, VDR is expressed in the mesencephalon precisely at the time when monoamine cells differentiate to dopaminergic cells and dopaminergic systems are innervated[86]. Mice pups with vitamin D deficiency have reduced levels of the enzyme COMT (*catechol-O-methyltransferase*), which converts the dopamine metabolite DOPAC (*3,4-Dihydroxyphenylacetic acid*) into HVA (*homovanillic acid*) and affects the dopamine turnover[86]. In rats with vitamin D deficiency, the effect of MK-801, an NMDA (*N-methyl-D-aspartate*) receptor antagonist which indirectly activates dopaminergic activity and also induces hyperlocomotion in animals, was found to be attenuated with the use of haloperidol, a

DRD2 (*Dopamine Receptor D2*) anatgonist[88]. In SH-SY5Y cells routinely used to model neural functions, VDR overexpression resulted in increased dopamine levels, overexpression of TH (*tyrosine hydroxylase*) which is an enzyme involved in the production of the precursor of dopamine called L-DOPA and overexpression of DRD2 whose increased activity has been noted in schizophrenia models, among other regulatory effects on genes associated with the dopaminergic system[89,90]. On treatment of these SH-SY5Y cells with calcitriol, a biologically active form of vitamin D, increased levels of dopamine metabolites such as HVA, increased COMT levels and reduced DRD2 expression were observed[90,91]. Even though there are several studies supporting the efficacy of vitamin D supplementation in treating schizophrenia symptoms[85,92], several groups have argued that these studies were irreplicable and that randomized controlled trials in larger cohorts would be necessary to ascertain its clinical utility, if any[93].

In this study, we shortlisted several drugs potentially repurposable for schizophrenia based on the negative correlation of drug-induced versus disease-associated gene expression profiles. Even though this approach has resulted in some valuable results in the past, it has several limitations. The gene expression profiles analyzed in this study were induced by drugs in cancer cell lines[16], and not in cell lines relevant to schizophrenia. The biological validity of our study will be strengthened if we perform our analysis with gene expression profiles induced by drugs in neuronal cell lines such as SH-SY5Y, in patient-derived induced pluripotent stem cells or in animal models of schizophrenia. However it is to be noted that such data has been shown to be valuable for repurposing drugs even for non-cancer diseases. Specific examples include repurposing of topiramate, an anti-epileptic drug, for inflammatory bowel disease[18], repurposing of drugs for schizophrenia[94] and repurposing of drugs for bipolar disorder[95]. These studies show that the drug-induced profiles generated in non-neural cells and deposited in CMAP are amenable to analysis involving neuropsychiatric disorders. Our future analysis will also focus on interrogating gene expression datasets of larger sample sizes. In summary, we showed that the drugs repurposable for schizophrenia may be identified from the schizophrenia drug-protein interactome based on gene expression profiles induced by the drug versus associated with the disease, and augmented our findings with clinical trial data, network-based analyses, and literature review. Through this study, we disseminate this list of drugs potentially repurposable drugs for schizophrenia to the scientific community so as to enable clinical translation of these results.

Methods

Identification of potentially repurposable drugs using BaseSpace correlation engine. In an earlier work, we constructed the protein-protein interaction network of schizophrenia genes, and then identified the drugs that target any of the proteins in this interactome[3]. Several of these drugs were known to have therapeutic value for nervous system, but there were several drugs that targeted other anatomical systems in the human body[3]. In this work, as a mechanism of shortlisting drugs for further analysis, we selected those that targeted more than two proteins in the schizophrenia interactome or those that target proteins that are also targeted by many drugs. While the first criterion helps in selecting drugs with the capacity to exert several pharmacological actions, a feature that is critical to targeting a disease as multifactorial as schizophrenia, the second criterion may point in the direction of highly druggable targets. For identifying repurposable drugs, it is essential that we tap into undiscovered regions of the PPI network. So, we also included drugs targeting novel proteins predicted to interact with known schizophrenia-associated genes[96]. Next step involved identifying the drugs that have opposite differential expression to the differential expression of schizophrenia (i.e., genes over-expressed in schizophrenia are under-expressed by drug treatment and vice versa). We studied each of these drugs in comparison to gene expression profiles of schizophrenia by using the software suite called BaseSpace (http://www.nextbio.com/b/nextbio.nb). BaseSpace Correlation Engine is used to study the effect of diseases and/or drugs on publicly available gene expression data[17]. Bioset 1 ('BS1') or a particular cell line, tissue or blood sample in which differential expression by drug has been studied was compared with a bioset 2 ('BS2'), another cell line, tissue or blood sample in which differential expression in schizophrenia patients was studied. A correlation score is generated by the tool based on the strength of the overlap or enrichment, between the two biosets. Additional statistical criteria such as correction for multiple hypothesis testing are applied and the correlated biosets are then ranked by statistical significance. A numerical score of 100 is assigned to the most significant result, and the scores of the other results are normalized with respect to the top-ranked result. We excluded drugs with unacceptable levels of toxicity or undesirable pharmacokinetics.

Network analysis using LENS. LENS (Lens for Enrichment and Network Studies of human proteins) is a web-based tool which may be used to identify pathways and diseases that are significantly enriched among the genes submitted by users[30]. The LENS algorithm finds the nearest neighbor of each gene in the interactome and includes the intermediate interactions that connect them. LENS then computes the statistical significance of the overlap of genes in the network and genes with annotations pertaining to pathways, diseases, drugs and GWASs, and reports a p-value computed from Fisher's exact test.

Shortlisted drugs which are being tested in clinical trials against various neuropsychiatric disorders were identified from NIH Clinical Trials (https://clinicaltrials.gov/).

Differential expression of the novel interactor VDR in whole blood obtained from schizophrenia patients was identified from GSE38485[39], and that of CACNA1H in induced pluripotent stem cells of schizophrenia patients was identified from GSE92874[58].

Association of the various genes in the network of PPIs among targets of the shortlisted drugs was identified from DisGeNET, a database that integrates human gene-disease associations from expert curated databases and text-mining derived associations[25].

Drugs that targeted the same genes as the shortlisted drugs were identified from DrugBank (https://www.drugbank.ca/)[24].

Preprint publication. An earlier version of this article was deposited into preprint server bioRxiv, where it appeared online on October 13, 2018[97].

Data Availability

Data sharing is not applicable to this article as no datasets were generated during the current study. Source of data that was analyzed in this study has been described in Methods and Data File 1.

References

1. Corvin, A. P. Two patients walk into a clinic…a genomics perspective on the future of schizophrenia. *BMC biology* **9**, 77, https://doi.org/10.1186/1741-7007-9-771741-7007-9-77 (2011).
2. Farrell, M. S. *et al*. Evaluating historical candidate genes for schizophrenia. *Molecular psychiatry*. https://doi.org/10.1038/mp.2015.16 (2015).
3. Ganapathiraju, M. K. *et al*. Schizophrenia interactome with 504 novel protein-protein interactions. *NPJ Schizophr* **2**, 16012, https://doi.org/10.1038/npjschz.2016.12 (2016).
4. Thomas, D. W. *et al*. Clinical development success rates 2006–2015. San Diego: Biomedtracker/Washington, DC: BIO/Bend: Amplion (2016).
5. Athauda, D. *et al*. Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. *The Lancet* **390**, 1664–1675 (2017).
6. Kinnings, S. L. *et al*. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS computational biology* **5**, e1000423 (2009).
7. Evans, J. M., Donnelly, L. A., Emslie-Smith, A. M., Alessi, D. R. & Morris, A. D. Metformin and reduced risk of cancer in diabetic patients. *Bmj* **330**, 1304–1305 (2005).
8. Sirota, M. *et al*. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science translational medicine* **3**, 96ra77–96ra77 (2011).
9. Maertzdorf, J. *et al*. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proceedings of the National Academy of Sciences* **109**, 7853–7858 (2012).
10. Chaussabel, D. *et al*. Analysis of significance patterns identifies ubiquitous and disease-specific gene-expression signatures in patient peripheral blood leukocytes. *Annals of the New York Academy of Sciences* **1062**, 146–154 (2005).
11. Kumar, A. *et al*. (Am Soc Hematology, 2014).
12. Chiu, I. M. *et al*. A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. *Cell reports* **4**, 385–401 (2013).
13. Duran-Frigola, M., Mateo, L. & Aloy, P. Drug repositioning beyond the low-hanging fruits. Current Opinion in. *Systems Biology* **3**, 95–102 (2017).
14. Pushpakom, S. *et al*. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery* **18**, 41 (2019).
15. Barrett, T. *et al*. NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research* **41**, D991–D995 (2012).
16. Lamb, J. *et al*. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *science* **313**, 1929–1935 (2006).
17. Kupershmidt, I. *et al*. Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS one* **5**, e13066 (2010).
18. Dudley, J. T. *et al*. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Science translational medicine* **3**, 96ra76–96ra76 (2011).
19. Bray, N. J. Gene expression in the etiology of schizophrenia. *Schizophrenia bulletin* **34**, 412–418 (2008).
20. Huo, Y., Li, S., Liu, J., Li, X. & Luo, X.-J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nature communications* **10**, 670 (2019).
21. Ukkola-Vuoti, L. *et al*. Gene expression changes related to immune processes associate with cognitive endophenotypes of schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **88**, 159–167 (2019).
22. Chattopadhyay, A. & Ganapathiraju, M. K. Demonstration Study: A Protocol to Combine Online Tools and Databases for Identifying Potentially Repurposable Drugs. *Data* **2**, 15 (2017).
23. Li, M. *et al*. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
24. Wishart, D. S. *et al*. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research* **34**, D668–D672 (2006).
25. Piñero, J. *et al*. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015** (2015).
26. Carta, F. *et al*. Carbonic anhydrase inhibitors: inhibition of cytosolic carbonic anhydrase isozymes II and VII with simple aromatic sulfonamides and some azo dyes. *Chemical biology & drug design* **74**, 196–202 (2009).
27. Vullo, D. *et al*. Carbonic anhydrase inhibitors. Inhibition of the human cytosolic isozyme VII with aromatic and heterocyclic sulfonamides. *Bioorganic & medicinal chemistry letters* **15**, 971–976 (2005).
28. Makani, S. *et al*. NMDA receptor-dependent afterdepolarizations are curtailed by carbonic anhydrase 14: regulation of a short-term postsynaptic potentiation. *Journal of Neuroscience* **32**, 16754–16762 (2012).
29. Egan, M. F. *et al*. Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proceedings of the National Academy of Sciences* **101**, 12604–12609 (2004).
30. Handen, A. & Ganapathiraju, M. K. LENS: web-based lens for enrichment and network studies of human proteins. *BMC medical genomics* **8**, S2 (2015).
31. Rubio, M. D., Wood, K., Haroutunian, V. & Meador-Woodruff, J. H. Dysfunction of the ubiquitin proteasome and ubiquitin-like systems in schizophrenia. *Neuropsychopharmacology* **38**, 1910 (2013).
32. Zhang, J. *et al*. Aquaporin-1 translocation and degradation mediates the water transportation mechanism of acetazolamide. *Plos one* **7**, e45976 (2012).
33. Cukras, S. *et al*. The USP1-UAF1 complex interacts with RAD51AP1 to promote homologous recombination repair. *Cell Cycle* **15**, 2636–2646 (2016).
34. Rougier, J.-S., Albesa, M., Abriel, H. & Viard, P. Neuronal precursor cell-expressed developmentally down-regulated 4-1 (NEDD4-1) controls the sorting of newly synthesized CaV1. 2 calcium channels. *Journal of biological chemistry* **286**, 8829–8838 (2011).
35. Song, C. *et al*. AQR is a novel type 2 diabetes-associated gene that regulates signaling pathways critical for glucose metabolism. *Journal of Genetics and Genomics* **45**, 111–120 (2018).
36. Christianson, J. C. & Green, W. N. Regulation of nicotinic receptor expression by the ubiquitin–proteasome system. *The EMBO journal* **23**, 4156–4165 (2004).
37. Shcherbakova, I. *et al*. Activation of kallikrein-kinin system, degranulating activity of neutrophils and blood-brain barrier in schizophrenia. *Zhurnal nevrologii i psikhiatrii imeni SS Korsakova* **98**, 38–41 (1998).
38. Malavia, T. A. *et al*. Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epidemiological, genomic, and protein interaction data. *NPJ schizophrenia* **3**, 11 (2017).

39. de Jong, S. *et al*. A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. *PloS one* **7**, e39498 (2012).

40. Marrie, R. A. *et al*. Increased incidence of psychiatric disorders in immune-mediated inflammatory disease. *Journal of psychosomatic research* **101**, 17–23 (2017).

41. Yu, H. *et al*. Association study of suppressor with morphogenetic effect on genitalia protein 6 (SMG6) polymorphisms and schizophrenia symptoms in the Han Chinese population. *Neuropsychiatry (London)* **6**, 223–228 (2016).

42. Tabares-Seisdedos, R. *et al*. Evidence for association between structural variants in lissencephaly-related genes and executive deficits in schizophrenia or bipolar patients from a Spanish isolate population. *Psychiatric genetics* **18**, 313–317 (2008).

43. Dobyns, W. & Truwit, C. Lissencephaly and other malformations of cortical development: 1995 update. *Neuropediatrics* **26**, 132–147 (1995).

44. Vila-Carriles, W. H. *et al*. Surface expression of ASIC2 inhibits the amiloride-sensitive current and migration of glioma cells. *Journal of Biological Chemistry* **281**, 19220–19232 (2006).

45. Yu, Q. *et al*. Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging. *Current topics in medicinal chemistry* **12**, 2415–2425 (2012).

46. Smith, C. L. & Eppig, J. T. The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mammalian genome* **23**, 653–668 (2012).

47. Xu, X. *et al*. Neuronal Abelson helper integration site-1 (Ahi1) deficiency in mice alters TrkB signaling with a depressive phenotype. *Proceedings of the National Academy of Sciences* **107**, 19126–19131 (2010).

48. Zhang, J.-c, Yao, W. & Hashimoto, K. Brain-derived neurotrophic factor (BDNF)-TrkB signaling in inflammation-related depression and potential therapeutic targets. *Current neuropharmacology* **14**, 721–731 (2016).

49. Bisol, L. W. *et al*. Is flunarizine a long-acting oral atypical antipsychotic? A randomized clinical trial versus haloperidol for the treatment of schizophrenia. *The Journal of clinical psychiatry* **69**, 1572–1579 (2008).

50. Schwartz, B. L., Fay-McCarthy, M., Kendrick, K., Rosse, R. B. & Deutsch, S. I. Effects of nifedipine, a calcium channel antagonist, on cognitive function in schizophrenic patients with tardive dyskinesia. *Clinical neuropharmacology* **20**, 364–370 (1997).

51. McGuire, P. *et al*. Cannabidiol (CBD) as an adjunctive therapy in schizophrenia: a multicenter randomized controlled trial. *American Journal of Psychiatry* **175**, 225–231 (2017).

52. Scott, L. J. & Dhillon, S. Spotlight on risperidone in irritability associated with autistic disorder in children and adolescents. *CNS drugs* **22**, 259–262 (2008).

53. Marder, S. R. & Meibach, R. C. Risperidone in the treatment of schizophrenia. *The American Journal of Psychiatry* **151**, 825 (1994).

54. Sajatovic, M., Subramoniam, M. & Fuller, M. A. Risperidone in the treatment of bipolar mania. *Neuropsychiatric disease and treatment* **2**, 127 (2006).

55. Goodnick, P. J. The use of nimodipine in the treatment of mood disorders. *Bipolar disorders* **2**, 165–173 (2000).

56. YAMADA, K. *et al*. Effectiveness of nilvadipine in two cases of chronic schizophrenia. *Psychiatry and clinical neurosciences* **49**, 237–238 (1995).

57. Humbert-Claude, M., Davenas, E., Gbahou, F., Vincent, L. & Arrang, J.-M. Involvement of histamine receptors in the atypical antipsychotic profile of clozapine: a reassessment *in vitro* and *in vivo*. *Psychopharmacology* **220**, 225–241 (2012).

58. Narla, S. *et al*. Common developmental genome deprogramming in schizophrenia—Role of Integrative Nuclear FGFR1 Signaling (INFS). *Schizophrenia research* **185**, 17–32 (2017).

59. Zhang, C. *et al*. Cromolyn Reduces Levels of the Alzheimer's Disease-Associated Amyloid β-Protein by Promoting Microglial Phagocytosis. *Scientific reports* **8**, 1144 (2018).

60. Chellappa, S. A. *et al*. Meta-analysis of genomic variants and gene expression data in schizophrenia suggests the potential need for adjunctive therapeutic interventions for neuropsychiatric disorders. *Journal of Genetics* **98**, 60 (2019).

61. Liu, F. *et al*. Minocycline supplementation for treatment of negative symptoms in early-phase schizophrenia: a double blind, randomized, controlled trial. *Schizophrenia research* **153**, 169–176 (2014).

62. Wilder-Smith, C., Schimke, J., Osterwalder, B. & Senn, H. Cinnarizine for prevention of nausea and vomiting during platin chemotherapy. *Acta Oncologica* **30**, 731–734 (1991).

63. Jin, C. & Panula, P. The laminar histamine receptor system in human prefrontal cortex suggests multiple levels of histaminergic regulation. *Neuroscience* **132**, 137–149 (2005).

64. García-Martín, E., Ayuso, P., Luengo, A., Martínez, C. & Agúndez, J. A. Genetic variability of histamine receptors in patients with Parkinson's disease. *BMC medical genetics* **9**, 15 (2008).

65. Mahmood, D. Histamine H3 receptors and its antagonism as a novel mechanism for antipsychotic effect: a current preclinical & clinical perspective. *International journal of health sciences* **10**, 564 (2016).

66. Brisch, R. *et al*. The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Frontiers in psychiatry* **5** (2014).

67. Hussain, N., Flumerfelt, B. & Rajakumar, N. Muscarinic, adenosine A 2 and histamine H 3 receptor modulation of haloperidol-induced c-fos expression in the striatum and nucleus accumbens. *Neuroscience* **112**, 427–438 (2002).

68. Benveniste, E. N. Cytokine actions in the central nervous system. *Cytokine & growth factor reviews* **9**, 259–275 (1998).

69. Ito, C. The role of the central histaminergic system on schizophrenia. *Drug News Perspect* **17**, 383–387 (2004).

70. Baucum, A. J., Rau, K. S., Riddle, E. L., Hanson, G. R. & Fleckenstein, A. E. Methamphetamine increases dopamine transporter higher molecular weight complex formation via a dopamine-and hyperthermia-associated mechanism. *Journal of Neuroscience* **24**, 3436–3443 (2004).

71. Hernandez, L., Auerbach, S. & Hoebel, B. G. Phencyclidine (PCP) injected in the nucleus accumbens increases extracellular dopamine and serotonin as measured by microdialysis. *Life sciences* **42**, 1713–1723 (1988).

72. Galosi, R. *et al*. Dopaminergic effects of histamine administration in the nucleus accumbens and the impact of H1-receptor blockade. *Neuropharmacology* **40**, 624–633 (2001).

73. Meskanen, K. *et al*. A randomized clinical trial of histamine 2 receptor antagonism in treatment-resistant schizophrenia. *Journal of clinical psychopharmacology* **33**, 472–478 (2013).

74. Terland, O. & Flatmark, T. Drug-induced parkinsonism: cinnarizine and flunarizine are potent uncouplers of the vacuolar H+-ATPase in catecholamine storage vesicles. *Neuropharmacology* **38**, 879–882 (1999).

75. Terrillion, C. E. *et al*. Reduced levels of Cacna1c attenuate mesolimbic dopamine system function. *Genes, Brain and Behavior* (2017).

76. Mena, M. A. *et al*. Effects of calcium antagonists on the dopamine system. *Clinical neuropharmacology* **18**, 410–426 (1995).

77. Teive, H. A., Troiano, A. R., Germiniani, F. M. & Werneck, L. C. Flunarizine and cinnarizine-induced parkinsonism: a historical and clinical analysis. *Parkinsonism & related disorders* **10**, 243–245 (2004).

78. Chandwani, S. *et al*. Induction of DARPP-32 by brain-derived neurotrophic factor in striatal neurons *in vitro* is modified by histone deacetylase inhibitors and Nab2. *PloS one* **8**, e76842 (2013).

79. Domire, J. S. *et al*. Dopamine receptor 1 localizes to neuronal cilia in a dynamic process that requires the Bardet-Biedl syndrome proteins. *Cellular and Molecular Life Sciences* **68**, 2951–2960 (2011).

80. Cuboni, S. *et al*. Loratadine and Analogues: Discovery and Preliminary Structure–Activity Relationship of Inhibitors of the Amino Acid Transporter B0AT2. *Journal of medicinal chemistry* **57**, 9473–9479 (2014).

81. Chandra, R. *et al*. Reduced Slc6a15 in nucleus accumbens D2-neurons underlies stress susceptibility. *Journal of Neuroscience* **37**, 6527–6538 (2017).

82. Shiraishi, A. *et al*. The advantage of alfacalcidol over vitamin D in the treatment of osteoporosis. *Calcified tissue international* **65**, 311–316 (1999).
83. Meltzer, H. Y. Clinical studies on the mechanism of action of clozapine: the dopamine-serotonin hypothesis of schizophrenia. *Psychopharmacology* **99**, S18–S27 (1989).
84. Kočovská, E., Gaughran, F., Krivoy, A. & Meier, U.-C. Vitamin-D Deficiency As a Potential Environmental Risk Factor in Multiple Sclerosis, Schizophrenia, and Autism. *Frontiers in psychiatry* **8** (2017).
85. McGrath, J. *et al*. Vitamin D supplementation during the first year of life and risk of schizophrenia: a Finnish birth cohort study. Schizophrenia research 67, 237-245 (2004).
86. Kesby, J. P., Cui, X., Burne, T. H. & Eyles, D. W. Altered dopamine ontogeny in the developmentally vitamin D deficient rat and its relevance to schizophrenia. *Frontiers in cellular neuroscience* 7 (2013).
87. Eyles, D. W., Smith, S., Kinobe, R., Hewison, M. & McGrath, J. J. Distribution of the vitamin D receptor and 1α-hydroxylase in human brain. *Journal of chemical neuroanatomy* **29**, 21–30 (2005).
88. Kesby, J. P., Burne, T. H., McGrath, J. J. & Eyles, D. W. Developmental vitamin D deficiency alters MK 801-induced hyperlocomotion in the adult rat: An animal model of schizophrenia. *Biological psychiatry* **60**, 591–596 (2006).
89. Krabbe, S. *et al*. Increased dopamine D2 receptor activity in the striatum alters the firing pattern of dopamine neurons in the ventral tegmental area. *Proceedings of the National Academy of Sciences* **112**, E1498–E1506 (2015).
90. Pertile, R. A., Cui, X. & Eyles, D. W. Vitamin D signaling and the differentiation of developing dopamine systems. *Neuroscience* **333**, 193–203 (2016).
91. Steddon, S. J., Schroeder, N. J. & Cunningham, J. Vitamin D analogues: how do they differ and what is their clinical role? *Nephrology Dialysis Transplantation* **16**, 1965–1967 (2001).
92. Eyles, D. W. *et al*. The association between neonatal vitamin D status and risk of schizophrenia. *Scientific reports* **8**, 17692 (2018).
93. Brown, H. E. & Roffman, J. L. Vitamin supplementation in the treatment of schizophrenia. *CNS drugs* **28**, 611–622 (2014).
94. Zhao, K. & So, H.-C. Drug repositioning for schizophrenia and depression/anxiety disorders: A machine learning approach leveraging expression data. *IEEE journal of biomedical and health informatics* (2018).
95. Kidnapillai, S. *et al*. The use of a gene expression signature and connectivity map to repurpose drugs for bipolar disorder. *The World Journal of Biological Psychiatry*, 1–9 (2018).
96. Kondej, M., Stępnicki, P. & Kaczor, A. Multi-target approach for drug discovery against schizophrenia. *International journal of molecular sciences* **19**, 3105 (2018).
97. Karunakaran, K. B., Chaparala, S. & Ganapathiraju, M. K. Potentially repurposable drugs for schizophrenia identified from its interactome. *bioRxiv*, 442640 (2018).

Acknowledgements

Author Contributions

The study has been designed by M.K.G. and K.B.K. S.C. carried out correlation analysis of drugs against diseases. K.B.K. carried out literature study and further bioinformatics analysis of the shortlisted drugs. Manuscript has been prepared by K.B.K. and edited by M.K.G.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-48307-w.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Interactome-based framework to translate disease genetic data into biological and clinical insights

9. Generalized and social anxiety disorder interactomes show distinctive overlaps with striosome and matrix interactomes

The experimental chapter is based on the following peer-reviewed publication:
Karunakaran, Kalyani B., Satoko Amemori, N. Balakrishnan, Madhavi K. Ganapathiraju, and Ken-ichi Amemori. Generalized and social anxiety disorder interactomes show distinctive overlaps with striosome and matrix interactomes. Scientific reports 11, no. 1 (2021): 1-25.

Summary of this chapter

In this chapter, I demonstrate the methodology to examine the unifying and differentiating biological themes underlying related disorders, using five anxiety disorder subtypes (generalized anxiety disorder, social anxiety disorder, obsessive-compulsive disorder, specific phobia, and panic disorder). I assembled anxiety disorder interactomes containing experimentally determined protein-protein interactions (PPIs). I found that the genes co-occurring across these interactomes were enriched for expression in the striatum, suggesting a potential relationship between anxiety disorders and striatal gene dysfunction. The disorder interactomes were refined based on their intersection with two striatal compartments, striosome, and matrix. The striosome and matrix interactomes were constructed by mapping genes differentially expressed in these compartments in various species to their human orthologues. Specific anxiety disorder interactomes showed significant and distinctive overlaps with the striosome and matrix interactomes. Principal component and hierarchical clustering analyses provided insights into the aetiological differentiation of the disorders. Systematic gene expression analysis with the interactomes constrained to contain only those genes shared with striatal compartment interactomes revealed a bifurcation of the anxiety disorders. Enrichment patterns of the genes in specific brain regions and signalling pathways influenced this bifurcation. The study suggested that the functionally distinct striatal systems constituted by the striosome and matrix may contribute to the development of anxiety disorders.

Contribution to this chapter (75%)

- Designed the study and developed the methodology of the project, including interactome construction and all the components of comparative interactome analysis, namely, transcriptomic analysis, signalling pathway analysis, principal component and hierarchical clustering analysis
- Curated all the datasets, performed all the analyses, and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

scientific reports

OPEN

Generalized and social anxiety disorder interactomes show distinctive overlaps with striosome and matrix interactomes

Kalyani B. Karunakaran[1], Satoko Amemori[2], N. Balakrishnan[1], Madhavi K. Ganapathiraju[3,4] & Ken-ichi Amemori[2]

Mechanisms underlying anxiety disorders remain elusive despite the discovery of several associated genes. We constructed the protein–protein interaction networks (interactomes) of six anxiety disorders and noted enrichment for striatal expression among common genes in the interactomes. Five of these interactomes shared distinctive overlaps with the interactomes of genes that were differentially expressed in two striatal compartments (striosomes and matrix). Generalized anxiety disorder and social anxiety disorder interactomes showed exclusive and statistically significant overlaps with the striosome and matrix interactomes, respectively. Systematic gene expression analysis with the anxiety disorder interactomes constrained to contain only those genes that were shared with striatal compartment interactomes revealed a bifurcation among the disorders, which was influenced by the anterior cingulate cortex, nucleus accumbens, amygdala and hippocampus, and the dopaminergic signaling pathway. Our results indicate that the functionally distinct striatal pathways constituted by the striosome and the matrix may influence the etiological differentiation of various anxiety disorders.

Anxiety is a mental state evoked in anticipation of a potential threat. Individuals may either exhibit acute levels of anxiety in response to an immediate threat or persistent levels of heightened anxiety (trait anxiety) in non-threatening situations as part of a 'neurotic' personality trait[1]. While both these non-pathological forms of anxiety may have evolved to protect the individual from potential dangers, the latter predisposes the individual to a range of anxiety disorders (ADs), depression, or both[2]. ADs, including generalized anxiety disorder (GAD), social anxiety disorder (SAD), specific phobia, post-traumatic stress disorder (PTSD) and obsessive–compulsive disorder (OCD), affect 284 million people (63% females, 2.5–7% variation by country), and are among the most prevalent mental health and neurodevelopmental disorders (WHO and IHME, 2017)[3]. About 31% of U.S. adults experience at least one AD during their lifetime[1,4]. ADs exhibit substantial familial aggregation with 30–50% heritability[5,6], and about 50% comorbidity of various AD types[7]. Despite the discovery of several genes associated with these disorders through linkage[8,9] and genome-wide association studies[10–14], the neurobiological implications of their genetic architectures remain elusive.

Several factors may reflect the distinct etiologies of various ADs, including different diagnostic definitions, regionally-specific neural activity and region-specific gene expression in the brain. By definition, SAD symptoms are conditional and 'externally' provoked by exposure to social situations[15]. In contrast, GAD is 'internally' provoked in the absence of any apparent anxiety-inducing event[15]. Regionally-specific neural activity has been associated with these disorders. For example, pregenual anterior cingulate cortex (ACC) of GAD patients showed hyperactivity correlated with their treatment responses[16], whereas bilateral amygdala showed hyperactivity in response to emotional stimuli in SAD patients[17]. Neural activity patterns within specific brain regions and among anatomically/functionally connected regions may underlie cognitive and emotional states in anxiety and can be correlated with transcriptional profiles[18–24]. It is thus possible that psychiatric morbidities such as ADs that are strongly driven by specific brain regions or networks would exhibit abnormalities in region-specific

[1]Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, India. [2]Institute for the Advanced Study of Human Biology, Kyoto University, Kyoto, Japan. [3]Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, USA. [4]Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, USA. ✉email: madhavi@pitt.edu; amemori.kenichi.7s@ kyoto-u.ac.jp

transcriptional signatures. Although anxiety-linked regional gene expression has been examined in post-mortem human brain tissues, blood samples and pharmacogenomic animal models[25–29], the functional consequences of such regional specificities remain unclear. In the current study, we examined ADs within the mechanistic framework of the protein–protein interaction (PPI) network (or the 'interactome'), which has revealed higher- order relationships in the genetic structures of complex disorders[30–33].

Protein–protein interactions (PPIs) drive the cellular machinery by facilitating a variety of biological processes, including signal transduction, the formation of cellular structures and enzymatic complexes. The effect of genetic mutations and abnormal gene expression may affect proteins and PPIs, posing deeper implications for disease development, such as multiple pathophenotypes that cannot be attributed to a single genotype in a disease[34]. Such effects can be explained through the analysis of the interactome, which allows examination of shared genetics, biological pathways and symptomatology[30–33].

We constructed the interactomes for six types of ADs (GAD, SAD, OCD, specific phobia, panic disorder and PTSD interactomes) using genes associated with each AD in biomedical literature as starting points. We identified the transcriptional profiles that characterized each of the anxiety disorder interactomes (ADIs) by computing the enrichment of the genes contained in these ADIs for high/moderate expression in specific human brain regions (Fig. 1a-c). For example, enrichment of a specific ADI for ACC-expressed genes was determined by comparing the distribution of ACC-expressed genes in that ADI against the background distribution of ACC-expressed genes among all the genes expressed in the brain. We then performed principal component and hierarchical clustering analyses to characterize region-wise expression patterns of ADIs and delineate AD groups (Fig. 1d). We observed that the genes that commonly co-occur in all the ADIs were strikingly enriched for expression in the striatum. We conducted a detailed interactome-based analysis of two striatal subdivisions called striosomes and matrix. The striatum is the primary input of the basal ganglia and is critical to motor control and motivated behaviors[35]. The striatum itself is histologically and neurochemically segregated into the striosome (or patch) and matrix compartments, which have differential gene expression signatures, anatomical connections and developmental patterns[36]. Striosomes are labyrinthine structures found embedded within the extra-striosomal matrix[37]. Striosomes express elevated levels of the μ-opioid receptors (MORs) in rodents[38] and Kv4 potassium channel subunit (KChIP1) in primates[37], whereas matrix expresses elevated levels of calbindin[39], somatostatin[40], encephalin[41] and acetylcholinesterase[42]. Striosomes are preferentially innervated by cortical areas implicated in limbic and evaluative processes such as the caudal orbitofrontal cortex (cOFC), pregenual anterior cingulate cortex (pACC)[43–45] in primates, and prelimbic cortex (PL) in rodents[46]. The medium spiny neurons (MSNs) in the striosome and matrix send projections to the substantia nigra pars reticulata and the external and internal segments of the globus pallidus, but only striosome MSNs have projections to the dopamine neurons in substantia nigra pars compacta[47,48].

It is noteworthy that both the striatal subdivisions are preferentially innervated by regions that may govern various aspects of anxiety. Both striosomes and the matrix arise from progenitor cell populations constituting the lateral ganglionic eminence[47]. However, striosomal neurons migrate out into the striatum from the lateral ganglionic eminence earlier than the matrix[47]. The functions of striosomes and matrix are yet to be fully eluci- dated. Nevertheless, studies have implicated them in reward-guided decision making and motivational conflict during cost–benefit decision making[44,46,49–51], and demonstrated their differential involvement in Huntington's disease[52], Parkinson's disease[53], motor stereotype[54], and drug addiction[47]. Interestingly, the striosome/matrix interactomes (SMIs), assembled using genes differentially expressed in the striosome and the matrix, showed preferential overlap with specific ADIs. Further, the genes shared between specific ADIs and SMIs showed discrete expression patterns, which allowed us to cluster the various ADs. Our findings implicate striatum as one of the focal points of etiological differentiation of ADs by showing that region-specific expression patterns underlying these disorders emerge only when the ADIs are constrained to include those genes that are shared with the SMIs.

Results

Expression of ADI genes in brain regions. Genes associated with six types of ADs (from ref[55] Fig. 1. Suggested scheme for exploring a suspected anxiety disorder), namely, PTSD, OCD, GAD, SAD, specific phobia and panic disorder, were extracted from DisGeNET[56] (Supplementary Table S1). Note that (a) DisGeNET cata-logs gene-disease associations described in animal models such as rats and mice, in addition to those described in human studies, and (b) many of the genes cataloged in DisGeNET may not share a causal relationship with the disease, and may instead only be associated with disease susceptibility or endophenotypes. Using RNA- sequencing data of 13 postnatal human brain regions obtained from GTEx[57], we attempted to identify whether these genes were enriched for expression in a specific brain region in a statistically significant manner. Genes with high/medium expression (transcripts per million (TPM) ≥ 9) in these 13 brain regions that were not house-keeping genes (from Human Protein Atlas[58]) were considered. For the enrichment analysis, we computed the distribution of genes expressed in a specific brain region among AD-associated genes and compared it with the background distribution of genes expressed in this particular brain region among all the genes that were assayed for expression in any brain regions. Statistically significant expression in a particular brain region was computed using a hypergeometric test (see Methods). No significant enrichment was found for any brain regions among AD-associated genes. This led us to examine these genes from the perspective of anxiety disorder interactomes (ADIs). This framework allowed us to include a larger number of genes in the enrichment analysis and examine AD-associated regions in the context of other mechanistically linked genes.

To assemble the network of PPIs (i.e., interactome) for each type of AD, we collected known PPIs from Human Protein Reference Database (HPRD)[59] and the Biological General Repository for Interaction Datasets (BioGRID)[60]. The GAD, SAD, PTSD, specific phobia, OCD, and panic disorder interactomes (Supplementary Data S1) were constructed in this manner (see Methods). The ADIs showed significant enrichment in several of
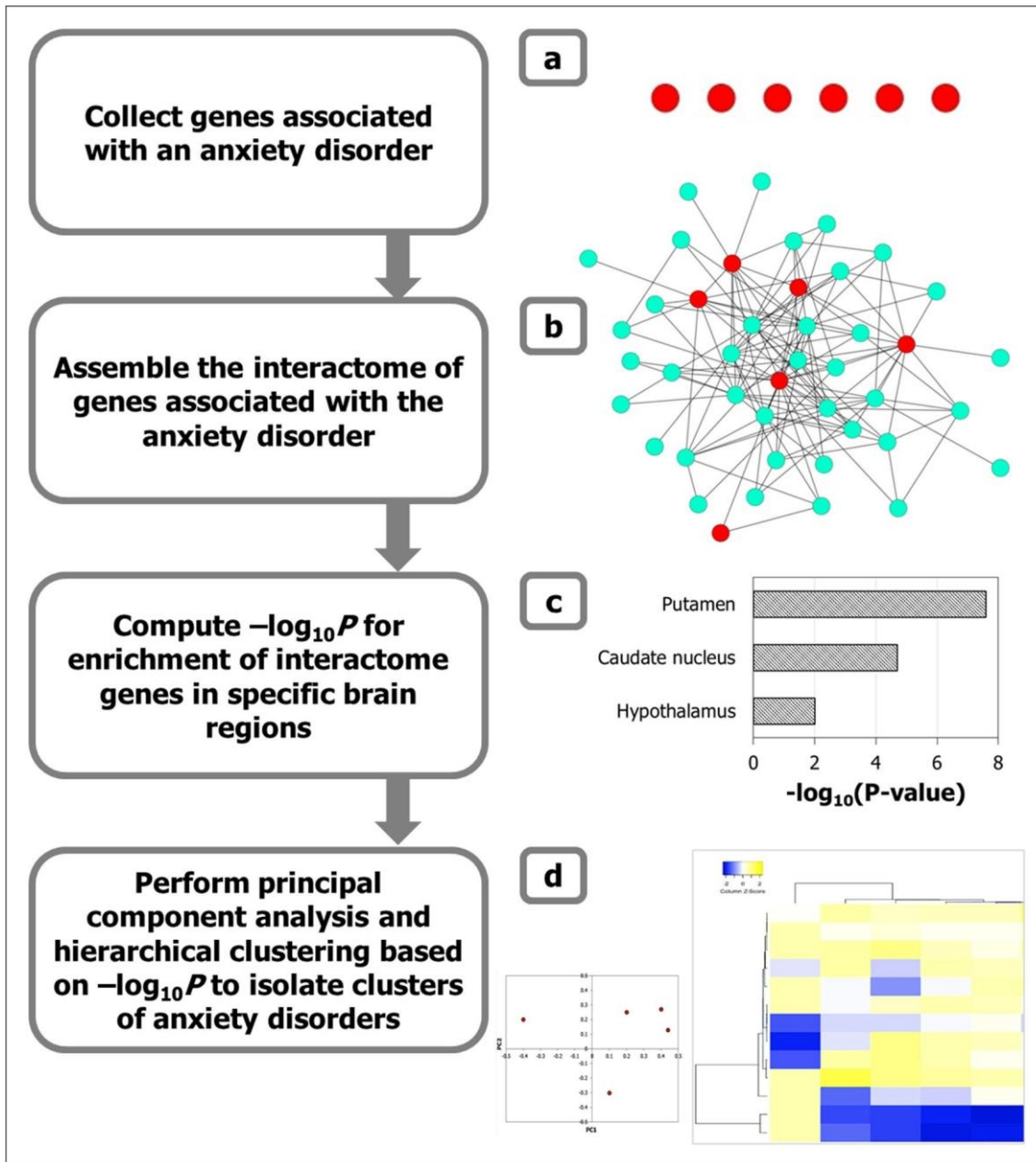
Figure 1. Methodology to identify regional specificity of ADs in the brain. (**a**) Genes associated with six types of ADs were compiled (depicted as red nodes), and (**b**) their protein interactomes assembled by curating the interactions of the proteins encoded by them with other proteins (i.e. interactors depicted as cyan colored nodesin the network diagram). These protein–protein interactions or PPIs are depicted as edges in the network. (**c**) The enrichment of the individual ADIs with genes showing medium/high expression (TPM ≥ 9) in specific brain regions were computed, and the statistical significance of these enrichments were calculated as negative logarithm of p values (i.e. $-\log_{10}P$). TPM = transcripts per million. (**d**) Principal component analysis (PCA) was performed to identify specific grouping patterns from the data matrix of $-\log_{10}P$ of enrichment of each ADI inspecific brain regions. Principal components which explain a large percentage of the variance observed across this data matrix were identified, and the component loadings denoting the correlation of the original variables($-\log_{10}P$ of specific brain regions) with the principal components were examined to interpret the observed patterns. The data matrix was also subjected to hierarchical clustering to delineate closely related groups of ADs. In the heat map, regions were colored according to the z-scores indicating their mean enrichment in the ADI. The z-scores indicate the number of standard deviations that separate a given p value from the mean. Highz-scores correspond to high enrichment for the specific region.

233
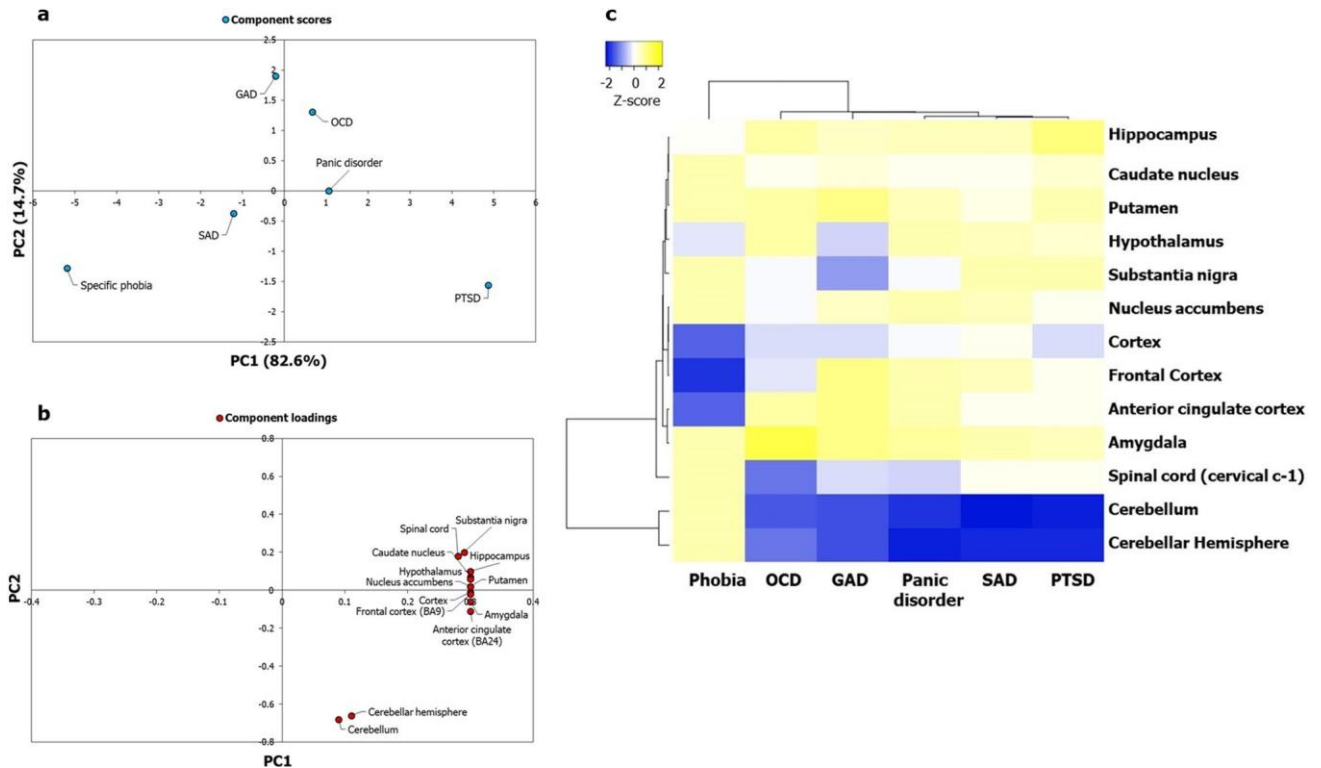
Figure 2. Principal component analysis and clustering analyses of the ADIs based on enrichment patterns in the brain failed to capture putative regional specificities and clusters of the ADs. (**a**) PCA was performed with the *p* values of enrichment of the ADI genes in 13 brain regions compiled from GTEx. *p* values were transformed to $-\log_{10}P$ values and a data matrix with brain regions (rows) and ADs (columns) (represented as a heatmap in (**c**)) was constructed out of these log-transformed values. Unit variance scaling was applied across this matrix. SVD with imputation was used to extract the principal components (PCs). Component scores (n = 6) corresponding to PC1 and PC2 explaining 82.6% and 14.7% of the total variance were plotted along X and Y axes respectively. (**b**) Component loadings of 13 dimensions, i.e. brain regions, contributing to PC1 and PC2 shown in (**a**) were plotted along X and Y axes respectively. Relatively equal and moderate contribution of all the brain regions, except the cerebellum and the cerebellar hemisphere, shows that ~97% of the variance captured by PC1 and PC2 plotted in (**a**) may not have reflected regional specificities of the ADs. (**c**) Variations in region-wise enrichment of ADI genes (computed based on GTEx data) are represented in the form of a heatmap. *p* values indicating statistical significance of enrichment were converted into $-\log_{10}P$ values. Each cell in the heatmap depicts a normalized z-score derived from a $-\log_{10}P$ value corresponding to a brain region. Z-scores indicate relative enrichment of specific brain regions in an ADI and are computed based on the number of standard deviations that separate a given *p* value from the mean. Clustering was performed using the hierarchical clustering method with average linkage. The dendrograms were derived from the clustering analysis is based on computation of Pearson correlation coefficients between the data points. The region-wise enrichment profile of specific phobia seems to be distinct from that of OCD, GAD, panic disorder, SAD and PTSD, which were all identified to be closely related. Clusters among the latter five disorders were not distinctive. The clustered heatmap was created using Heatmapper (http://www.heatmapper.ca/).

the 13 brain regions extracted from GTEx[57]. First, we performed enrichment analyses to determine whether genes in ADIs tend to be

overrepresented in a specific brain region. *P* values indicating the statistical significance of the overlap between the interactomes and genes expressed in specific brain regions were computed based on the hypergeometric test (see Methods). *P* value < 0.05 after multiple test adjustments using the Benjamini–Hochberg method was considered to be significant.

We then sought to understand whether the specific values indicating the significance of enrichment of the ADIs in the brain regions revealed any underlying clustering patterns in terms of regional specificities among the ADs themselves. For this, we generated a data matrix of ADIs (columns) versus brain regions (rows); each cell contained the negative of log-transformed *p* values. Single value decomposition (SVD) with imputation was applied to this matrix to extract principal components that explain the variance observed with region-wise enrichment of gene expression across the ADs. Principal component analysis (PCA) is used to capture systematic variations underlying datasets. Unit variance scaling was applied across the matrix. Six principal components were extracted from the matrix, out of which PC1 and PC2 explained 82.6% and 14.7% of the total variance (Fig. 2a). PTSD and specific phobia seemed to be separated from GAD, OCD, SAD and panic disorders (Fig. 2a). The latter four had low component scores, and clusters among them were not apparent (Fig. 2a). The log-transformed *p* values of enrichment in each brain region were then converted to normalized z-scores. Z-scores

234

indicate the number of standard deviations that separate a given p value from the mean. This matrix of z-scores was then subjected to hierarchical clustering based on Pearson's correlation coefficients and the average linkage method (Fig. 2c). OCD, GAD, panic disorder, SAD and PTSD were depicted as being closely related compared with specific phobia (Fig. 2c). Two factors seemed to differentiate the five ADs from specific phobia: (a) theirlower enrichment in the cerebellum and the spinal cord, and (b) their higher enrichment in cortical regions such as the frontal cortex and the ACC (compared with specific phobia).

Next, we sought to identify specific brain regions that were relatively more influential than others in delineating the specific pattern of clustering observed among the ADs. To this end, we examined component loadings that have contributed to PC1 and PC2. Component loadings are values depicting the correlation of the originalvariables in our data matrix—negative log of p value of enrichment for specific brain regions—with each of the extracted principal components. On plotting the component loadings of the brain regions for PC1 and PC2 across X and Y axes, we noticed that almost all of the brain regions were moderately correlated with PC1 and PC2and contributed relatively equally to them (Fig. 2b). Hence, although the PC plot of ADs explained ~ 97% of the variance observed in region-wise enrichment (Fig. 2b), it may not have captured region specificities underlying the ADs, except for the ability of cerebellar structures, spinal cord and cortical regions to differentiate specificphobia from OCD, GAD, panic disorder, SAD and PTSD. The same pattern of moderate correlation of most of the component loadings with PC1 and PC2 and their equivalent contribution to the principal components was observed with a larger set of 26 brain regions extracted from Allen Brain Atlas (genes that are not housekeeping genes and have logRPKM > 2, where RPKM is Reads Per Kilobase per Million mapped reads, were considered) (Supplementary Fig. S1).

Potential striatal association of ADIs. We checked the overlap among the six ADIs. Thirty-six genes were found to be shared among all the interactomes (Supplementary Table S2). These genes were found to be enriched only in striatal genes, i.e., genes with high/medium expression in the caudate nucleus, putamen and the nucleus accumbens (all at p value = 0.0187) (based on data extracted from GTEx), suggesting a potential relationship between ADs and striatal gene dysfunction.

Early developmental mechanisms controlled by genetic susceptibility factors and gene-environment inter- actions may modulate the response pattern of an individual to threat stimuli[5]. Genes involved in establishing neuronal connectivity in the adult brain are often regulated by genes that specify the neuronal identity and brain regionalization during the early stages of brain development. We reasoned that the striatum-enriched genes that co-occur in all the ADs could be closely connected with genes that set up the cellular and molecular architecture of the striatum during early developmental stages, namely, transcriptional regulators and signaling transduction molecules controlling neuronal development and neurotransmission[61]. Specifically, these genes are DLX1-6, GSX2, EBF1, ISL1, FOXP1/2, DRD1/2, GNAL, ADCY5, PPP1R1B, STEP and RASGRP2[61]. We indeed observed that these genes involved in striatal development were closely interconnected with our striatum-enriched genes (Supplementary Fig. S2). Several anxiety-associated genes had direct interactions with striatal development genes, such as PRKCA with ADCY5 and DLX3, and ESR1 with ISL1 (Supplementary Fig. S2). We found this network to be enriched in dorsal thalamus genes (p value = 0.0143) in addition to striatal genes (p value = 0.0016) (based on data extracted from Allen Brain Atlas). These results imply that the striatum is critical for ADs fromthe perspective of their interactomes.

Distinctive overlaps of ADIs and SMIs. Striatum is composed of two neurochemically segregated compartments called the striosomes and the matrix, which are characterized by their distinct gene expression profiles[47]. We analyzed these structures from an interactome perspective. We compiled the list of genes differentially expressed in the striosomes and the matrix compartments (Supplementary Table S3) of various species (rat,mouse, ferret, cat, monkey and human)[47], and mapped them to their human orthologs. From enrichment analysis using Gene Ontology terms, we noted that some genes differentially expressed in the striosomes (HTR2A, HTR2C, CHRM1-5 and DRD4) were involved in the serotonergic signaling pathway (p value = 2.12E−10); human striosomes are known to be enriched in serotonin receptors[62]. Some genes differentially expressed in thematrix neurons (CDH8, CDK5, CNR1, HTR2A and SLC17A6) were found to be involved in the glutamatergicsynaptic transmission (p value = 1.93E−5). We examined whether the interactome of the genes differentially expressed in either of these striatal compartments were significantly enriched in each of the ADIs.

Computation of overlaps among the ADIs revealed that they themselves do not segregate into any groups (Fig. 3a). However, specific ADIs showed significant and distinctive overlaps (Fig. 3b and Table 1) with the striosome/matrix interactomes (SMIs) (Supplementary Data S2). The striosome interactome (SI) shared 28% of its constituent genes with the GAD interactome (227 genes out of a total of 810 SI genes) (p value = 1.31E−4) (i.e. out of the total number of 810 genes present in SI, 227 genes were also found in the GAD interactome) and 53% (427/810) of genes with the OCD interactome (p value = 5.21E−8). These genes shared with the SI account for 38% (227/595) and 25% (427/1718) of the genes in the GAD and OCD interactomes, respectively. Out of the 1718 genes in the OCD interactome, 639 were not shared with any other ADI. 17.5% of these 639 (112/639) genes were shared with the SI (p value = 2.27E−7) (overlap between OCD and SI is illustrated in the form of anetwork diagram in Supplementary Fig. S3). Genes shared between OCD and SI showed high enrichment for genes associated with human motor and behavioral stereotypes (Human Phenotype Ontology[63]) (Supplementary Note S1). This observation is in line with the findings of a study that demonstrated the ability of excessive activation in the striosomes (compared with the matrix) to predict the degree of drug-induced motor stereotypy in rats[54]; both the activation ratio and drug-induced stereotypy have been shown to be under the regulation of cholinergic interneurons in the striatum[64–66]. The matrix interactome (MI) shared 14.35% (30/209) of its constituent genes with the SAD interactome (p value = 8.23E−3). The MI shared 23% (48/209) of its constituent genes
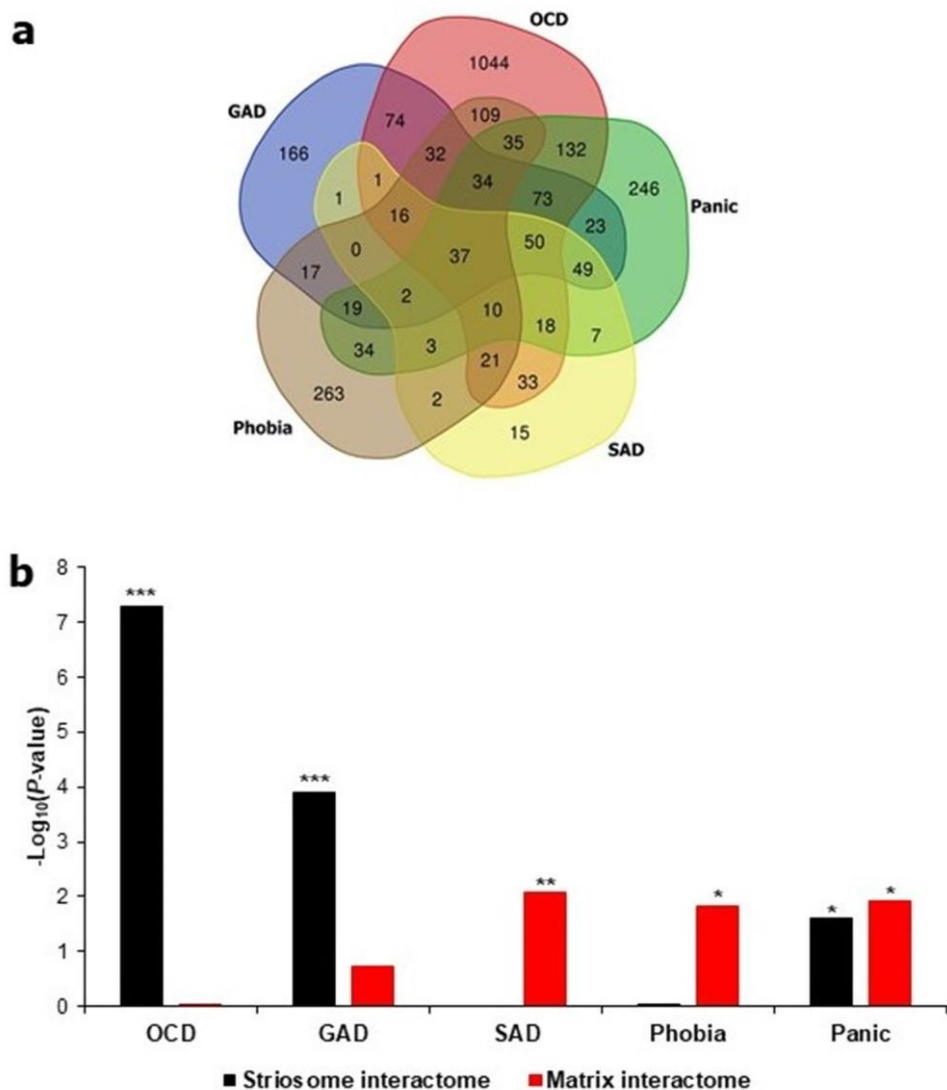
Figure 3. ADIs showed preferential and statistically significant overlaps with either SI or MI. (**a**) The number of proteins shared among ADIs is shown as a Venn diagram. (**b**) The number of common proteins shared between a particular ADI and SMIs assembled from the genes differentially expressed in striosome/matrix was computed. $-\log_{10}P$ values derived from this analysis are shown as bars in the figure. Overlaps with p value<0.05(i.e. $-\log_{10}P>1.3$) after correction for multiple hypotheses were considered to be statistically significant. *, ** and *** corresponds to p value < 0.05, p value < 0.01 and p value < 0.001 respectively. Four out of the 5 ADIs shown in the figure shared statistically significant and exclusive overlaps with either the SI (OCD and GAD) or the MI (SAD and phobia). While the panic disorder interactome shared statistically significant overlaps with both of the striatal compartments, the overlap shared with the MI was more statistically significant than that shared with the SI. From (**a**), it is clear that the ADIs themselves do not segregate into any groups. However, (**b**) shows that they exhibit preferential overlap with SMIs. For example, (**a**) shows that the OCD interactome shares 186 genes with the SAD interactome, whereas (**b**) clarifies that the OCD and SAD interactomes exhibitpreferential overlaps with the striosome and MIs respectively. The Venn diagram tool provided as part of the Bioinformatics & Evolutionary Genomics toolkit was used to create the Venn diagram (http://bioinformatics. psb.ugent.be/webtools/Venn/).

with the phobia interactome (p value$=0.0154$). The genes shared with the MI account for 11% (30/266) and 7.5% (48/634) of the genes in the SAD and phobia interactomes. 32.5% (68/209, p value $= 0.012$) and 26% (212/810, p value$=0.025$) of genes found in MI and SI respectively were also found in the panic disorder interactome. In addition to striatum-expressed genes, thalamus-expressed genes were enriched in the network of striatal developmental regulators and anxiety-associated genes (Supplementary Fig. S2). Unlike striatal sub-compartments, molecularly distinct subdivisions of the thalamus did not show any preferential association with any ADI (Supplementary Methods and Supplementary Note S2).

Overlap between Interactomes and the *P* value of significance	Strisome Interactome (810 genes)	Matrix Interactome (209 genes)
Generalized anxiety disorder interactome (595 genes)	227 (1.31E−04)	n.s
Obsessive compulsive disorder interactome (1718 genes)	427 (5.21E−08)	n.s
Panic disorder interactome (773 genes)	212 (0.025)	68 (0.012)
Social anxiety disorder interactome (266 genes)	n.s	30 (8.23E−03)
Specific phobia interactome (634 genes)	n.s	48 (0.0154)

Table 1. Overlap of the anxiety disorder and striatal subcompartment interactomes. The table shows the statistics of the overlaps shared between the five anxiety disorder interactomes and the interactomes of the striatal subdivisions.

In summary, each ADI shared a statistically significant overlap with SMIs, i.e., the GAD interactome shared an overlap with the SI, OCD with SI, phobia with MI and SAD with MI. Although panic disorder showed significant overlaps with SI and MI, the statistical significance of the overlap with MI was higher than that of the overlap with SI. These results raised the possibility that gene dysfunction occurring in striosomes could underlie the symptoms of GAD and OCD, whereas matrix dysfunction could underlie SAD and phobia. Hence, five groups of shared genes contributing to statistically significant overlaps of ADIs with SMIs (referred to as 'AD-SMIs' henceforth) were delineated from this analysis, namely, GAD-striosome, OCD-striosome, SAD-matrix, phobia-matrix andpanic-matrix. We focused on these five AD-SMIs speculating that they would allow us to probe the etiologicaldifferentiation of these ADs in terms of their affiliation to one of the two striatal compartments.

Expression of AD-SMI genes in brain regions. Studies have noted a substantial overlap between functional connectivity and gene co-expression patterns within and between cortical and striatal networks[18–23]. Functional connectivity can be defined as a temporal correlation between brain regions, often derived from co-activated fMRI signals at resting state[67]. Co-expression patterns are derived from Pearson and Spearman correlations that assess the transcriptional similarity of genes. Based on this, we speculated that testing the regionalexpression patterns of AD-SMIs may reveal brain regions showing expression of these shared genes and perhaps,exhibiting functional connectivity with the striatal compartments and governing key anxiety traits. However, it is important to note that (a) the validity of this speculation is supported only for a limited number of brain networks, and (b) the degree and the nature of the interaction between transcriptional similarity and functionalconnectivity are yet to be fully elucidated[23]. We thus examined the expression patterns of AD-SMIs (e.g., genes shared between GAD interactome and SI) in the brain.

Lists of genes expressed in 13 postnatal human brain regions were extracted from GTEx[57], and their enrichment in each of the five AD-SMIs was systematically computed using the hypergeometric test (see Methods). Multiple brain regions showed significant enrichment in each of these groups. Following this, we used the negative of log-transformed *p* values denoting the significance levels of enrichment as input data for PCA. Five principal components were extracted, and from this, we selected PC1 and PC2, explaining 58.6% and 27.5% of the variance observed with region-wise enrichment in gene expression to interpret the AD groupings. Notably,we observed that OCD, GAD, SAD, and panic disorder exhibited clearer patterns of clustering when the overlap with striatal compartments was taken into consideration (Fig. 4a) compared to when this overlap was not taken into consideration (Fig. 2a). Next, we generated a heat map of z-scores derived from the log-transformed *p* values and employed Pearson correlation and average linkage method to identify clusters of ADs. Firstly, we observed two main clusters (Fig. 4c): the first cluster included SAD-matrix and panic-matrix, and the second cluster included GAD-striosome and a sub-cluster consisting of OCD-striosome and phobia-matrix. The occurrence of GAD-striosome and OCD-striosome in the same cluster (Fig. 4c) is supported by previous observations link-ing striatal stimulation and striatal beta oscillation to a key feature in OCD called cognitive inflexibility, which manifested as a repetitive pattern of a major symptom observed in GAD called pessimistic valuation[50]. Secondly, with the introduction of striatal subdivisions in our analysis with AD-SMIs (Fig. 4c), we were able to delineate the clustering of caudate nucleus and putamen, and nucleus accumbens and ACC, a pattern that was not clear with ADIs (Fig. 2c). This demonstrates the biological validity of our approach. Additionally, it is notable that thespinal cord did not influence the clustering of the AD-SMIs (Fig. 4c).

Further, we plotted the correlation of the component loadings with PC1 and PC2 to assess whether the observed pattern of clustering reflected regional specificities (Fig. 4b). It was observed that some brain regions had an uneven influence over the clustering pattern when overlap with striatal compartments was taken into consideration (Fig. 4b), namely, ACC, amygdala, hippocampus, nucleus accumbens, putamen and caudate, compared to when the striatal overlap was not taken into consideration (Fig. 2b). Heightened reactivity in the first three regions has been associated with clinical anxiety[68]. We identified two groups of brain regions that may more or less act as functional units to influence the etiology of these two purported 'types' of ADs (Fig. 4c): (1) amygdala and hippocampus and (2) ACC and nucleus accumbens. They are referred to here as 'functional units' by virtue of them being tight clusters in the dendrogram, and only in terms of their potential contribution towards anxiety etiology (see Supplementary Discussion). Higher enrichment of SAD-matrix andpanic-matrix in the amygdala and hippocampus may have segregated them from the cluster of GAD-striosome, OCD-striosome and phobia-matrix that showed lower enrichment in these same regions (Fig. 4c). Relatively higher enrichment of GAD-striosome in ACC and nucleus accumbens may have led to its segregation from the
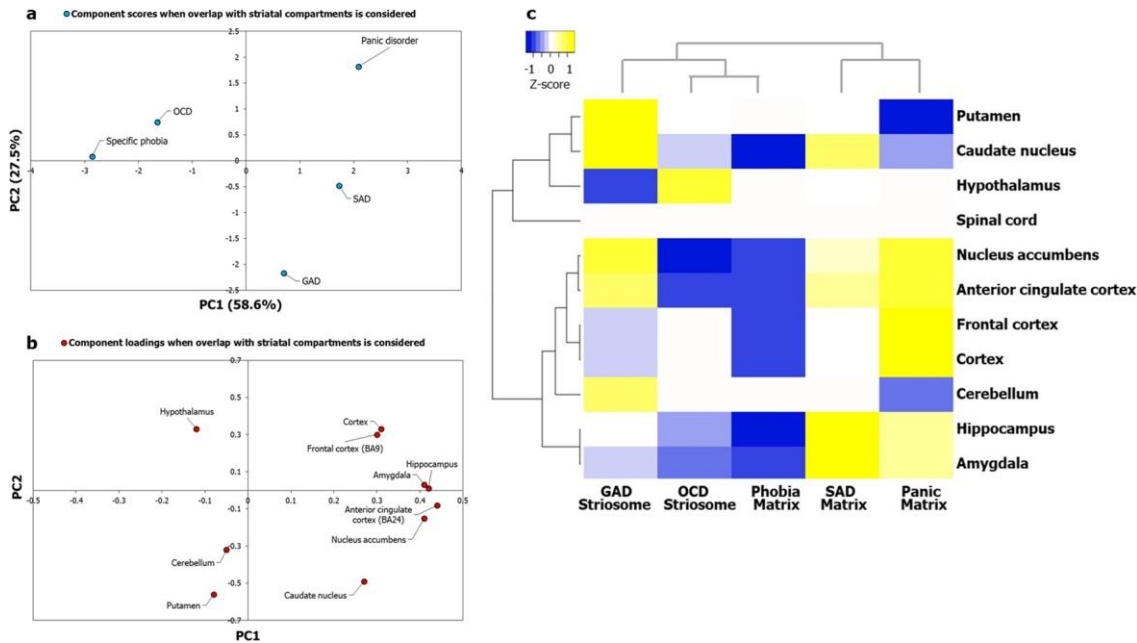
Figure 4. Principal component analysis and clustering analyses of AD-SMIs revealed uneven influence of key brain regions linked to anxiety etiology and distinct AD groups. (**a**) PCA was performed with the *p* valuesof enrichment of genes that co-occur in specific ADIs and SMIs. As observed in Fig. 5, specific ADIs share discrete and distinct statistically significant overlaps with either SI or MI, namely, GAD and striosome, OCD and striosome, phobia and matrix, SAD and matrix, and panic disorder and matrix. The enrichment of these genes among those exhibiting high/medium expression in 13 brain regions (TPM > 9) compiled from GTEx was checked. The statistical significance of region-wise enrichment was computed as *p* values. These values were transformed to − log₁₀P values, which were then assembled into a data matrix containing brain regions as rows and ADs as columns (represented as a heatmap in Fig. (**c**)). Unit variance scaling was applied across this matrix. Single value decomposition (SVD) with imputation was used to extract the principal components (PCs). Component scores of GAD-striosome, OCD-striosome, phobia-matrix, SAD-matrix and panic disorder-matrix (n = 5) corresponding to PC1 and PC2 explaining 58.6% and 27.5% of the total variance were plotted along X and Y axes respectively. (**b**) Component loadings of 10 dimensions, i.e. brain regions, contributing to PC1 and PC2 shown in (**a**) were plotted along X and Y axes respectively. In contrast with the pattern of component loadings of brain regions observed in Fig. 2b, PCA with genes shared between ADIs and SMIs appears to have captured uneven influences of several brain regions such as the ACC, amygdala, hippocampus,nucleus accumbens, putamen and caudate on the grouping patterns of ADs. (**c**) Variations in region-wise enrichment of genes (computed from GTEx data) shared between ADIs and SMIs are represented in the form of a heatmap. Specifically, normalized z-scores computed based on the –log₁₀ transformed *p* values, indicating the statistical significance of enrichment of GAD-striosome, OCD-striosome, phobia-matrix, SAD-matrix andpanic disorder-matrix, are shown in the figure. Z-scores indicate relative enrichment of specific brain regions in the gene sets and are computed based on the number of standard deviations that separate a given *p* value from the mean. Clustering was performed using the hierarchical clustering method with average linkage. The dendrograms were derived from the clustering analysis based on computation of Pearson correlation coefficientsbetween the data points. Two main clusters were detected among the ADs. SAD and panic formed one cluster.OCD and phobia formed a sub-cluster within the second main cluster. GAD was an outgroup to the sub-cluster of OCD and phobia. The clustered heatmap was created using Heatmapper (http://www.heatmapper.ca/).

sub-cluster of OCD-striosome and phobia-matrix (Fig. 4c). Based on this analysis, we constructed 'brain maps' for two categories of ADs, namely, those sharing interactome overlaps with striosome and with matrix (Fig. 5). We examined whether the AD-SMIs were enriched for expression in 23 postnatal human brain regions and three transitory fetal structures (lateral ganglionic eminence, medial ganglionic eminence and rhombic lip) available in Allen Brain Atlas[69]. PCA showed that the clustering among the ADs was clearer when overlap with the striatal compartments was taken into consideration (Fig. 6a) compared with when this overlap was not taken into consideration (Supplementary Fig. S1a). Despite using a more diverse and numerous dataset, hierarchical clustering revealed the preservation of the grouping of GAD-striosome with OCD-striosome and SAD-matrix with panic-matrix, both of which appeared as sub-clusters within the main cluster in this analysis;phobia was detected as an outgroup to this main cluster (Fig. 6c). Two groups of brain structures seemed to be highly influential in this clustering (Fig. 6c): (1) medial ganglionic eminence (MGE) and lateral ganglionic eminence (LGE) and (2) parietal neocortex. LGE is a source of striatal projection neurons and gives rise to both striosomes and matrix neurons[70]. MGE populates cortical layers and differentiates into interneurons[71]. The segregation of the sub-clusters of GAD-striosome and OCD-striosome from SAD-matrix and panic-matrix could have stemmed from the higher enrichment of LGE and MGE in the former group compared with the

238
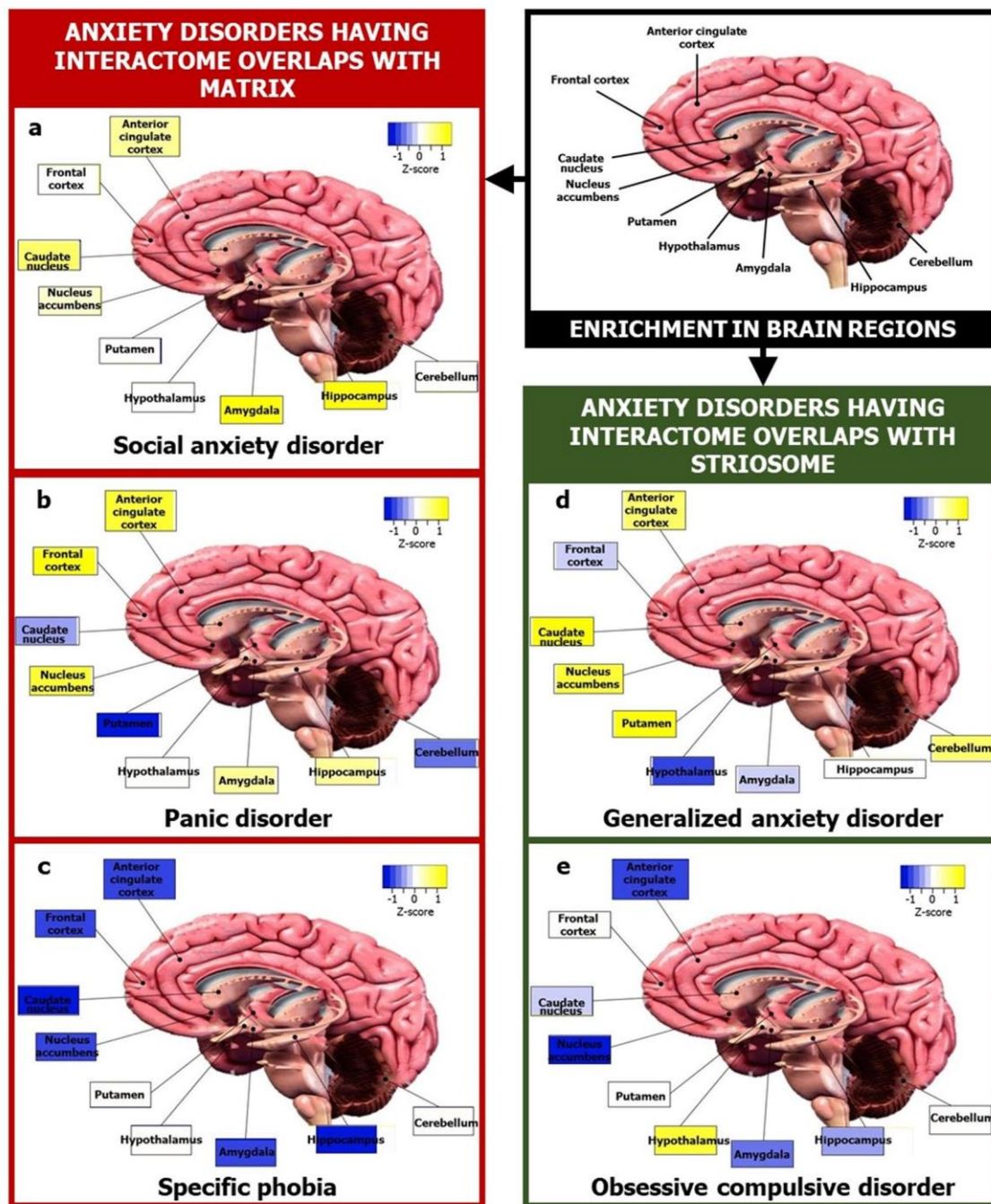
Figure 5. Map of focal brain regions in various ADs. The figure shows brain regions that may be linked to the etiology of two categories of ADs, namely, those with striosomal (green boxes) affiliation and those with matrix (red boxes) affiliation, in terms of interactome overlaps. The regions in the 'brain map' of each AD was colored according to the normalized z-score for that region indicating its mean enrichment in the ADI. This wascomputed using data from GTEx as shown in Fig. 4c. Prominent involvement of ACC in generalized anxiety disorder (**d**) and that of amygdala and hippocampus in social anxiety disorder (**a**) can be noted. The brain section image is a royalty-free stock illustration (ID: 1401181217) downloaded from Shutterstock titled 'Human Brain Anatomy Sagittal Section with Labels, 3D Rendering'.

latter group (Fig. 6c). It was interesting to note that phobia-matrix showed high enrichment for LGE and MGEsimilar to GAD-striosome and OCD-striosome (Fig. 6c), which may explain their occurrence in the same cluster in Fig. 4c and hint at shared etiology rooted in perturbations of genes expressed during the early stages of striosome-matrix compartment specification. Phobia-matrix had a lower enrichment for parietal cortex (and CGE which clustered together) compared to GAD-striosome and OCD-striosome. GAD-striosome showed exclusive enrichment for ACC (Fig. 6c). Lastly, we examined the correlation of component loadings with PC1
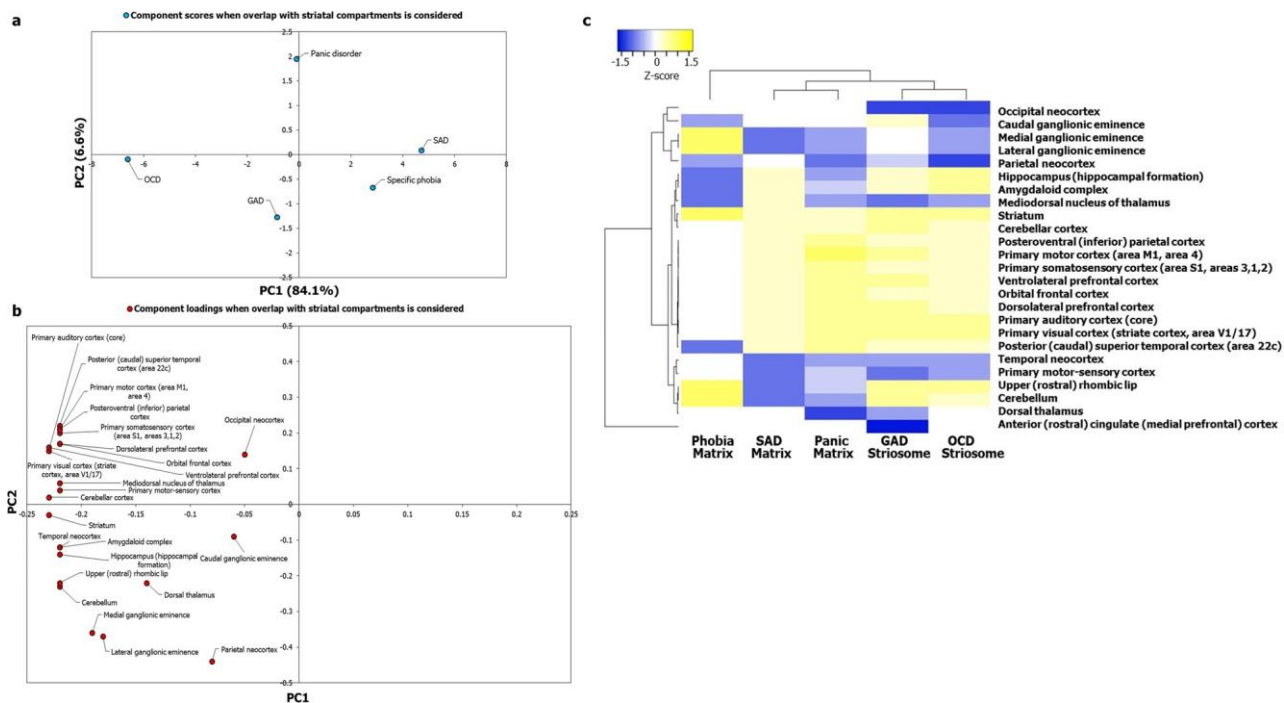
Figure 6. Principal component analysis and clustering analyses of AD-SMIs based on a larger expression dataset recapitulated the AD groups and revealed embryonic structures potentially relevant to anxiety etiology.

(**a**) PCA was performed with the *p* values of enrichment of genes that co-occur in specific ADIs and SMIs. The enrichment of these genes among those expressed across 26 brain regions compiled from Allen Brain Atlas having log(RPKM) > 2 was checked. The statistical significance of region-wise enrichment was computed as *p* values. These values were transformed to $-\log_{10}P$ values, and then assembled into a data matrix containing brainregions as rows and ADs as columns (represented as a heatmap in (**c**)). Unit variance scaling was applied

across this matrix. Single value decomposition (SVD) with imputation was used to extract the principal components (PCs). Component scores of GAD-striosome, OCD-striosome, phobia-matrix, SAD-matrix and panic disorder- matrix (n = 5) corresponding to PC1 and PC2 explaining 84.1% and 6.6% of the total variance were plotted along X and Y axes, respectively. (**b**) Component loadings of 24 dimensions, i.e., brain regions, contributing to PC1 and PC2 shown in (**a**) were plotted along X and Y axes, respectively. This figure recapitulates findings from GTEx data (Fig. 4c) with data from Allen Brain Atlas, namely, observation of clearer grouping patterns among ADs when the striatal overlap is taken into consideration and uneven contribution of multiple brain regions to these patterns such as medial ganglionic eminence, lateral ganglionic eminence and parietal cortex. (**c**) Variations in region-wise enrichment of genes between ADIs and SMIs are represented in the form of a heatmap. Specifically, normalized z–scores computed based on the –$\log_{10}$ transformed *p* values, indicating thestatistical significance of enrichment of GAD-striosome, OCD-striosome, phobia-matrix, SAD-matrix and panic disorder-matrix, are shown in the figure. Z-scores indicate relative enrichment of specific brain regions in the gene sets and are computed based on the number of standard deviations that separate a given *p* value from the mean. Clustering was performed using the hierarchical clustering method with average linkage. The dendrograms were derived from the clustering analysis based on the computation of Pearson correlation coefficients between the data points. Clustering of GAD with OCD and SAD with panic disorder seen withGTEx data in Fig. 4c is recapitulated here. Phobia was identified as a separate cluster altogether. The clusteredheatmap was created using Heatmapper (http://www.heatmapper.ca/).

and PC2, which captured 90.7% of the variance observed in region-wise enrichment of gene expression. This ascertained the fact that the observed pattern of clustering was unevenly influenced by regional specificities that become apparent when overlap with striatal compartments is taken into consideration (Fig. 6b). These regional specificities were not observed when the striatal overlap was not considered (Supplementary Fig. S1b). From this detailed analysis, we can confirm that regional specificities underlying the various ADs are revealed only when discrete subnetworks of their interactomes that contain genes differentially expressed in the striosome/ matrix and their interactors are examined. Since functional connectivity and gene expression are correlated in the brain[18–23], this raises the possibility that the etiology of specific ADs may be rooted in the discrete functionalconnections of specific striatal sub-compartments with other brain regions, such as the ACC and amygdala, that govern traits specific to these ADs.

Signaling pathways enriched in key brain regions influential in the grouping of ADs. We isolated the gene sets that were responsible for the enrichment of the AD-SMIs in the four brain regions—amygdala, hippocampus, ACC and nucleus accumbens—that were presumably more influential than the other brainregions (Fig. 4b) in producing the grouping pattern seen in Fig. 4a,c, i.e., the genes that were responsible for the
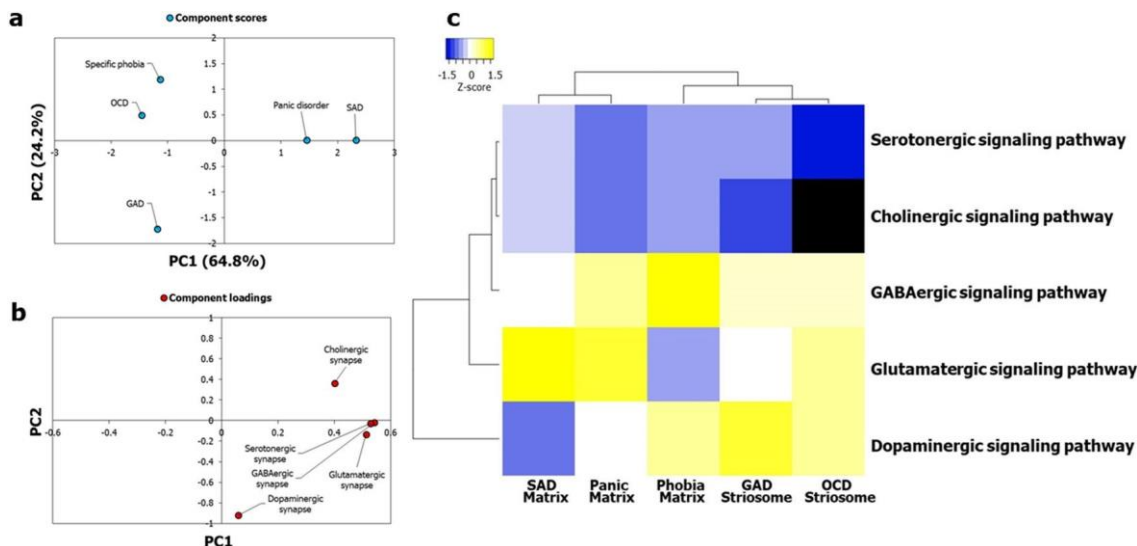
Figure 7. Neuronal synaptic signaling pathways enriched among genes expressed in influential brain regionsrecapitulated the AD groups. Gene sets in GAD-striosome, OCD-striosome, phobia-matrix, SAD-matrix andpanic-matrix showing moderate/high expression in the four key brain regions (i.e., amygdala, hippocampus, ACC and nucleus accumbens) were selected. The KEGG pathways enriched in them were then identified.
(**a**) PCA was performed with the *p* values of enrichment of five neuronal synaptic signaling pathways in each of these gene sets, namely, cholinergic, dopaminergic, GABAergic, glutamatergic and serotonergic signaling pathways. These values were transformed to –$\log_{10}$P values and were then assembled into a data matrix containing pathways as rows and the gene sets as columns (represented as a heatmap in (**c**)). Unit variance scaling was applied across this matrix. Single value decomposition (SVD) with imputation was used to extractthe principal components (PCs). Component scores of gene sets from GAD-striosome, OCD-striosome,
phobia-matrix, SAD-matrix and panic disorder-matrix (n = 5) corresponding to PC1 and PC2 explaining 64.8% and 24.2% of the total variance were plotted along X and Y axes respectively. Genes in GAD-striosome, OCD-striosome, phobia-matrix, SAD-matrix and panic-matrix that were involved in at least one of the five neuronal synaptic signaling pathways, and showed expression in at least one of the four brain regions that were influentialin producing the pattern seen in Fig. 4a,c, recapitulated the AD groups seen in Fig. 4a,c. (**b**) Component loadings of 5 dimensions, i.e. signaling pathways, contributing to PC1 and PC2 shown in (**a**) were plotted along X and Y axes respectively. The loading value of the dopaminergic signaling pathway indicated that it had highinfluence over the observed grouping pattern. GABAergic and glutamatergic signaling pathways seemed to
have a moderate influence. (**c**) Pathway enrichment of these gene sets are represented in the form of a heatmap.Specifically, normalized z-scores computed based on the –$\log_{10}$ transformed *p* values, indicating the statisticalsignificance of pathway enrichment of these gene sets, are shown in the figure. The dendrograms were derived from hierarchical clustering analysis based on the computation of Pearson correlation coefficients between the data points. The clustered heatmap was created using Heatmapper (http://www.heatmapper.ca/).

---

p values of enrichment used for generating Fig. 4a and Fig. 4c. For example, the gene set for GAD-striosome contained the genes shared between GAD-striosome and the four brain regions, i.e., the genes which co-occurred in GAD interactome and SI, and showed expression in at least one of the four brain regions. For a particular AD-SMI, the same set of genes was found to be expressed across the four brain regions, which led us to considerthem together in the corresponding gene set. The signaling pathways (KEGG[72]) that were significantly enriched(*p* value after multiple test adjustment < 0.05) in each of these gene sets were identified using WebGestalt[73]. A data matrix of the enriched pathways (rows) and disorder-striatal compartment combinations (columns) was created; each cell contained −$\log_{10}$P of enrichment of each pathway. Firstly, on PCA and clustering analysis ofthis matrix, we observed that five neuronal synaptic signaling pathways, namely, the cholinergic, dopaminergic,GABAergic, glutamatergic and serotonergic signaling pathways, could produce a grouping pattern (Fig. 7a, c)similar to that shown in Fig. 4a and Fig. 4c. Out of these, the dopaminergic signaling pathway had a very high influence on the grouping pattern, and the GABAergic and glutamatergic signaling pathways had a moderate influence (Fig. 7b). Based on this, we speculated that the conservative set of genes co-occurring in each of the AD-SMIs and expressed in the four connected limbic structures are those that influence dopamine signaling. Secondly, we noted that when all the enriched pathways were considered, phobia-matrix clustered alongside SAD-matrix and panic-matrix (Supplementary Fig. S4), and not with GAD-striosome and OCD-striosome. Collectively, the gene sets that were examined, as discussed above, contained 84 genes (Fig. 8). Out of these, 12 were known to be associated with ADs (CDH2, DLG4, DRD1, GRIK2, MAPT and NTRK2 associated with OCD, NTRK3 with OCD and panic disorder, ADORA2A with panic disorder and specific phobia, GAD1 with GAD and panic disorder, NPY with GAD, panic disorder and specific phobia, GAD2 with GAD, OCD and panic disorder, and DRD2 with GAD, OCD, SAD and specific phobia). These genes were used as starting points for interactome construction in our study. The rest of the 72 genes, including three genes that were differentially
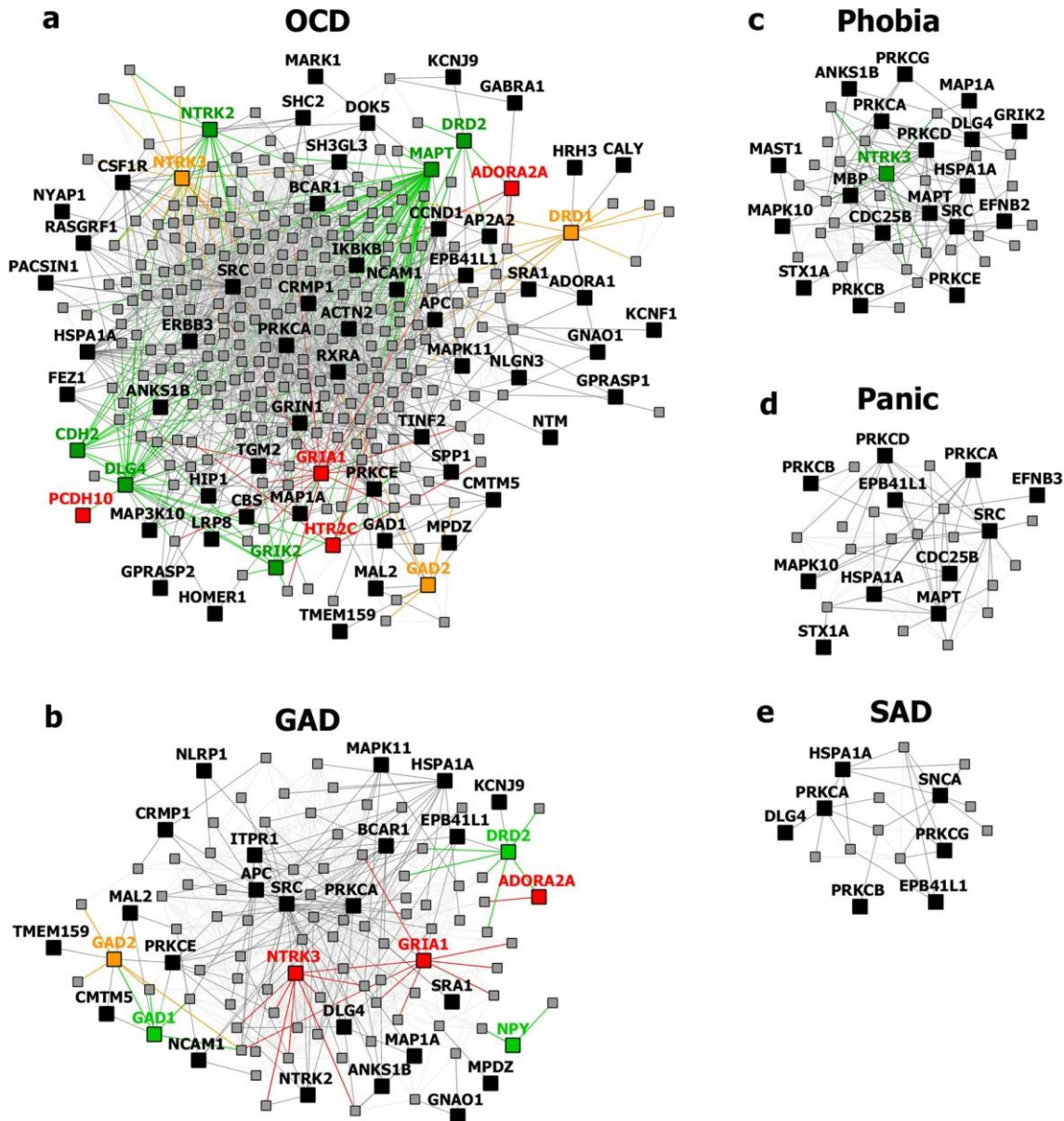
Figure 8. Network layout of the genes expressed in the influential brain regions in the AD-SMIs. The network diagrams show the genes having moderate/high expression in the four brain regions that were highly influential in identifying the disorder groups shown in Fig. 4c, namely, amygdala, hippocampus, ACC and nucleus accumbens. These genes have been highlighted in (**a**) OCD interactome, (**b**) GAD interactome, (**c**) phobia interactome, (**d**) panic disorder interactome and (**e**) SAD interactome. In each of the network diagrams, green colored nodes depict genes associated with the particular AD. Red colored nodes indicate genes differentially expressed in striosomes/matrix that shows interactome overlap with the given AD; (**a**) OCD and (**b**) GAD shows interactome overlap with striosome, whereas (**c**) phobia, (**d**) panic disorder and (**e**) SAD show overlap with matrix. Orange colored nodes indicate genes that are associated with the particular AD as well as differentially expressed in the striatal compartment. Black colored nodes are intermediary genes interconnectingthe AD genes. The majority of the red and black colored nodes shown in the diagrams, i.e. genes that are yet to be studied in the context of human ADs, have already been linked to anxiety-like/repetitive/perturbed social behavior in animal models and humans (see Table 2). The validity of these genes as candidates for further investigation of anxiety mechanisms is supported by this evidence and their proximity in the interactome to genes that have been previously associated with human ADs (i.e., green colored nodes). The network diagramswere created using Cytoscape (version 3.7.2).

expressed in striosomes/matrix (GRIA2, HTR2C and PCDH10)[47], have not been studied so far in the context of human ADs. However, 59 (82%) of these genes have already been linked to anxiety-like/linked behavior/ physiological states, repetitive behavior or perturbed social behavior through gene knockout, differential gene expression/methylation studies in animal models and genetic variant studies in humans (see evidence from

each of these studies in Table 2). This demonstrates the validity of our interactome-based method due to two reasons: we (a) identified other genes that were previously studied in animal models of anxiety using an unbiased approach from the ADIs, and (b) showed that they were closely connected to the genes associated with human ADs, thereby providing a mechanistic framework to study these putative anxiety-associated genes in the future.

Discussion

Neuroimaging, neurochemical and gene-based approaches have provided valuable insights on neurobiological and genetic themes underlying AD etiology[9,103–105]. Regional neural activity has been correlated with anxiety traits such as pessimistic valuation and negative affectivity associated with facial expression[16,17]. Since resting- state functional connectivity in the brain shares substantial overlap with correlated regional gene expression patterns[18–23], perturbed region-specific transcriptional signatures could underlie cognitive and emotional states in anxiety[24]. In order to capture the complexity of neuropsychiatric phenotypes, such transcriptional signatures need to be examined within an interactomic framework, in which the effect of a perturbed gene spreads in the protein interactome and affects other proteins and the biological processes mediated by them. We thus reasoned that perturbations in region-specific transcriptional profiles may be reflected in the interactome and could underlie region-specific activity in ADs. Several studies have identified higher-order biological relationships existing among genes from the functional landscape of the interactome[30–33]. In this study, we adopted a data- driven interactome and transcriptome-based approach to elucidate common and distinctive neurobiological themes underlying six types of ADs. Clustering of ADIs based on the region-specific expression of their constituent genes failed to reveal any clustering among OCD, GAD, SAD and panic disorder. Moreover, analysis of component loadings correlated with PC1 and PC2 revealed that this approach failed to capture regional specificities underlying the ADs (Fig. 2b and Supplementary Fig. S1b). Therefore, contrary to our speculation, the clustering of ADs was not directly evident at the level of their interactomes (Fig. 2c). It was evident only after filtering out genes that showed no overlap with the interactomes of genes specifically expressed in striatal compartments called the striosome and the matrix (Figs. 4c, 6c) and retaining only those genes that did show an overlap (Fig. 3).

In our study, striatum-expressed genes, which are closely connected to modulators of striatal development, were found to co-occur in all the ADIs (Supplementary Fig. S2). Genes active in the serotonergic synapse (KEGG[72]) were enriched (p value = 1.75E−07) among these 36 genes, namely, APP, GNAI2, GNAI3, GNB5, KCNJ9, MAPK1, MAPK3, PRKCA and SLC6A4 (serotonin transporter). The fact that selective serotonin reup- take inhibitors (SSRIs) function as broad-spectrum drugs across ADs (and the highly co-morbid major depressive disorder)[106], led us to prioritize striatum as a focal region, with potential involvement in neurodevelopmental underpinnings of AD[1]. Two factors seemed to support this suspicion: striatum is (a) anatomically and function-ally connected to brain regions that have been extensively studied in the context of anxiety, such as the amygdala, hippocampus and the prefrontal cortex[107], and (b) involved in key cognitive processes impaired in ADs, such as attention, motivation, fear conditioning and reward prediction error[107]. The main finding in this direction was that each of the ADIs, except the PTSD interactome, shared a preferential and statistically significant overlap with the striosome/matrix interactomes (i.e., SMIs) (Fig. 3).

We then attempted to cluster the ADs based on the region-specific expression of genes overlapping between the ADIs and the SMIs (i.e., AD-SMIs). This revealed clearer clustering patterns among the disorders (Fig. 4c and Fig. 6c), which seemed to be influenced by regional specificities, with some brain regions showing a strong correlation with PC1 and PC2 (Fig. 4b and Fig. 6b). The enrichment patterns of AD-SMIs in specific brain regions may reflect the related brain circuitry, which produces heightened responses to triggers and sequentially generates anxiety (see Supplementary Discussion)[108]. The same set of genes contributed towards the enrichment of each of the AD-SMIs across the four regions that were highly influential in AD-SMI clustering, namely, amygdala, hippocampus, ACC and nucleus accumbens. This may reflect the concerted roles (a) of these regions in the limbic circuit or (b) of neuronal synaptic signaling pathways across these regions (Fig. 7). Another reason could be the inclusion of any gene with high/moderate rather than tissue-specific expression in the expression profile of each region (see Supplementary Discussion).

Our results point to a scenario wherein the functionally distinct striatal pathways constituted by the striosome and the matrix, act as 'diverging points' for the etiological differentiation of various ADs. Genetic perturbations associated with a specific AD may (a) modulate the interactome in one of the two compartments or even their progenitor populations, such as LGE (Fig. 6c), (b) influence their functional connectivity with other regions and (c) govern the 'route' of development of key anxiety traits. GAD- and OCD-associated traits generated in this manner may include internally generated ruminations (involving the phenomenon of interoceptive awareness[109]), cognitive rigidity[50] and pessimistic valuation[44,49], controlled by ACC, which targets the striosome compartment[44]. SAD-associated traits may include heightened emotional responses to neutral stimuli and excessive emotional contextualization during social information processing, controlled by the amygdala that collates information from the sensorimotor cortices[110], which in turn targets the matrix compartment. The behavioral traits characterizing striosome- and matrix-associated ADs could also arise from differential responses of these striatal compartments to reinforcement contingencies[111] and differential involvement in resolving motivational conflicts[112]. Both these paradigms are intricately tied to anxiety etiology and influenced by dopamine circuits. We noted that the dopaminergic signaling pathway may be highly influential (Fig. 7b) in producing the AD-SMI clustering seen in Fig. 4a,c. Therefore, it is possible that the bifurcation in ADs may stem from differential dopamine dynamics, namely, differential electrically evoked dopamine release[113], dopamine levels[114], dopaminergic innervation[114] and modulation of state transitions in the striatal compartments[115], and the preferential striosomal projection to dopaminergic neurons in substantia nigra pars compacta[116]. GABAergic and glutamatergic signaling were noted to be moderately influential in AD-SMI clustering (Fig. 7b). The role of our three key neuronal synaptic

Gene	GAD-striosome	OCD-striosome	Phobia-matrix	Panic-matrix	SAD-matrix	Evidence
ADORA2A	✓	✓				AD-associated gene from DisGeNET
CDH2		✓				AD-associated gene from DisGeNET
DLG4	✓	✓		✓	✓	AD-associated gene from DisGeNET
DRD1		✓				AD-associated gene from DisGeNET
DRD2	✓	✓				AD-associated gene from DisGeNET
GAD1	✓	✓				AD-associated gene from DisGeNET
GAD2	✓	✓				AD-associated gene from DisGeNET
GRIK2		✓		✓		AD-associated gene from DisGeNET
MAPT		✓	✓	✓		AD-associated gene from DisGeNET
NPY	✓					AD-associated gene from DisGeNET
NTRK2	✓	✓				AD-associated gene from DisGeNET
NTRK3	✓	✓		✓		AD-associated gene from DisGeNET
ACTN2		✓				Differential gene expression (fold change = 1.5, p value = 0.005) in the amygdala of restrained C57BL/6J mice versus the amygdala of non-restrained C57BL/6J mice assessed by dark/light exploration test and elevated plus maze.[74] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
ADORA1		✓				Increased occurrence of the GG genotype of the ADORA1 rs2228079 polymorphism in Gilles de la Tourette syndrome patients having obsessive compulsive disorder/behavior (p value = 0.021) and depression (p value = 0.032) as co-morbidities.[76] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
ANKS1B	✓	✓		✓		Differential hydroxymethylation and correlated gene expression in mice region in mice exposed to early life stress with later anxiety behaviors.[77] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]

Continued

Gene						Description
AP2A2		■				Differential methylation in prenatally stressed young bull calves compared with control bull calves.[78] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
APC	■	■				APC knockout mice exhibited repetitive behaviors, reduced social interest (autistic traits),learning and memory deficits compared with wild type littermates.[79]
BCAR1	■	■				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
CABP1		■				CABP1 knockout mice tested for hippocampal-dependent spatial and fear-related memories exhibited a mild anxiety phenotype as assessed by the open-field test.[80] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
CALY		■				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
CBS		■				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
CCND1		■				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
CDC25B			■	■		Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
CMTM5	■	■				Differential gene expression after stress exposurein healthy adults with a reported history of childhood adversity compared with a matched control group without adverse childhood experiences.[81]
CRMP1	■	■				CRMP1 knockout mice tested for development of schizophrenia symptoms exhibited anxiety- related behavior as assessed by the elevated plus maze test.[82] Differential expression in the prefrontal cortex (dorsolateral) of obsessive- compulsive disorder patients versus healthycontrols.[75]
CSF1R		■				Male mice models of adult-onset leukodystrophy with axonal spheroids and pigmented glia (ALSP) with heterozygous mutation in CSF1R exhibit depression and anxiety-like behaviour.[83]

Continued

	Col1	Col2	Col3	Col4	Col5	
DDR1	▓					Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
DOK5		▓				Differentially regulated in the blood, hemibrain and spleen of mice subjected to a social-stress model of PTSD.[84] Differential expression in the prefrontal cortex (dorsolateral) of obsessive- compulsive disorder patients versus healthycontrols.[75]
EFNB2				▓		Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
EPB41L1	▓	▓	▓		▓	Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
FEZ1		▓				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
GABRA1		▓				Increased anxiety in a mouse model of Wolfram syndrome has been linked to reduced expression of GABRA1 in the frontal cortex and temporal lobe.[85]
GNAI3			▓	▓	▓	Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
GNAO1	▓	▓				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
GPRASP1		▓				Differential gene expression (fold change = − 1.6, p value = 0.006) in the amygdala of restrained C57BL/6J mice versus the amygdala of non-restrained C57BL/6J mice assessed by dark/light exploration test and elevated plus maze.[74]
GPRASP2		▓				Mice with GPRASP2 deletion displayed behaviour relevant to autism spectrum disorder in three-chamber social arena test, social dyadic test, tube test for social dominance, perturbed innate social behaviour in nest building test and increased stereotypical behaviour.[86]
GRIA1	▓	▓				Mice with deleted GRIA1 show a distinct anxiety phenotype in elevated O-maze and dark–light box tests.[87] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]

Continued

Gene						Description
GRIN1						Mice with non-synonymous mutation in GRIN1 showed abnormal social interactions and abnormal anxiety-like behaviors in light/dark transition and the elevated plus maze tests.[88] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
HOMER1						Mice with overexpressed HOMER1A in the basal and lateral amygdala exhibit impaired fear conditioning and autism-like social impairment.[89]
HRH3						Mice with deleted HRH3 exhibited reduced anxiety in the elevated plus and zero mazes involving exploratory behaviour and avoidable anxiety-provoking stimuli, and enhanced acoustic startle responses in the presence of unavoidable anxiety-provoking stimuli.[90] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
HSPA1A						Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
HTR2C						HTR2C knockout mice exhibited reduced anxiety-like behavior in elevated zero maze and open field tests.[91]
IKBKB						Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
KCNF1						Differential gene expression in the ventromedial prefrontal cortex (fold change = 1.4, p value = 0.004), amygdala (fold change = 1.8, p value = 0.007) and hippocampus (fold change = 2, p value = 0.008) of restrained C57BL/6J mice versus the amygdala of non-restrained C57BL/6J mice assessed by dark/light exploration test and elevated plus maze.[74]
LRP8						Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
MAP1A						Sex-dependent differential regulation of MAP1A expression in male mice upon prenatal exposure to stress suspected to be associated with anxiety-like behaviors.[92] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
MAP3K10						Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]

Continued

Gene						Description
MAPK10			�	�		Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
MAPK11	�	�				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
MARK1		�				Autism spectrum disorder associated genetic variants; MARK1 overexpression in the prefrontal cortex of postmortem brain samples of autistic patients.[73] Differential gene expression in the ventromedial prefrontal cortex (fold change = 1.3, p value = 0.005) of restrained C57BL/6J mice versus the amygdala of non-restrained C57BL/6J mice assessed by dark/light exploration test and elevated plus maze.[74] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
MAST1				�		Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
MPDZ	�					Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
NCAM1	�					Mice with homozygous (NCAM−/−) and heterozygous (NCAM+/−) mutations in NCAM exhibited anxiety-like behavior in a light/dark avoidance test.[94]
NLGN3		�				Male mice lacking NLGN3 were socially submissive to their wild-type littermates; this social submission correlated with increased anxiety in these mice.[95] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
NLRP1	�					Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
NTM		�				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
NYAP1		�				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
PACSIN1		�				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
PCDH10		�				Mice lacking a copy of PCDH10 (Pcdh10+/−) exhibited reduced social approach behavior, but no anxiety-like or repetitive behaviors.[96]

Continued

Gene						Evidence
PRKCA	▦	▦	▦	▦		Genetic variant mapped to PRKCA rs4790904 was found to be significantly associated with PTSD.[97] Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
PRKCD			▦	▦		Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
PRKCE	▦	▦				Mice with deleted PRKCE showed reduced anxiety-like behaviour.[98]
PRKCG	▦			▦		Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
RXRA		▦				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
SHC2		▦				Differentially regulated in the blood, hemibrain and spleen of mice subjected to a social-stress model of PTSD.[84] Differential expression in the prefrontal cortex (dorsolateral) of obsessive- compulsive disorder patients versus healthy controls.[75]
SNCA					▦	mRNA and protein abundance levels of SNCA were twice as high in the hippocampus of the anxious LEW (Lewy) rats when compared to SHR (spontaneously hypertensive) rats; anxiety was measured using open field, light/dark box and elevated plus maze tests.[99] 44% increase in protein abundance levels of SNCA was noted in the striatum of anxious LEW rats.[99]
SPP1		▦				Differential expression in mice that underwent social stress for 13 days.[100]
SRA1	▦	▦				Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
SRC	▦	▦	▦	▦		Differential expression in the prefrontal cortex (dorsolateral) of obsessive-compulsive disorder patients versus healthy controls.[75]
STX1A			▦	▦	▦	STX1A mRNA expression was significantly higher in the lymphocytes of drug-naive highfunctioning autism patients compared with controls.[101] Differential expression in the prefrontal cortex (dorsolateral) of obsessive- compulsive disorder patients versus healthy controls.[75]
TMEM159	▦	▦				Genetic variants in TMEM159 linked to brain arousal in the resting state.[102]

Table 2. Genes present in AD-SMIs that were expressed in the four key brain regions influential in producing the grouping pattern of ADs (shown in Fig. 4a,c). A colored cell indicates the presence of the particular gene in the corresponding AD-SMI. The final column presents the literature evidence supporting the potential role of the given genes in anxiety etiology.

signaling pathways is well-supported in anxiety studies (see Supplementary Discussion). Further investigations are necessary to delineate their roles in the association between ADs and the striatal compartments.

Our speculation on the potential effects of genetic perturbations on AD etiology is valid when they are examined as neurodevelopmental disorders, i.e., under the assumption that ADs develop when genetic risk factors affect the formation of neural circuits that mediate anxiety and subsequently modulate their responsiveness to anxiety-inducing events[117]. However, a converse scenario is also conceivable, wherein anxious behavior and adverse events alter gene expression via epigenetic mechanisms and modulate the neural circuits and susceptibility to ADs (e.g., the genes NR3C1[118] and FKBP5[119] undergo stress-induced epigenetic alterations)[120]. Hence, it is important to note that our results do not illustrate any cause-and-effect relationships among genetic perturbations, disrupted neural circuits/signaling pathways and ADs.

The PTSD interactome was excluded from our analysis as it did not show any distinctive overlap with either the striosome or matrix interactomes. However, we noted that its mechanism could be mostly, if not entirely, different from the other ADs. Corroborating previous studies[121], the PTSD interactome showed the highest enrichment for the hippocampus followed by the striatum, substantia nigra and amygdala, a pattern that was not shown by any other AD (Fig. 2c). This may justify its placement in a separate group (trauma and stress-related disorders) in DSM-5, unlike the case of OCD, which has been placed in a separate group (OCD and related disorders) despite its etiology being very closely related to the other ADs.

Specific phobia clustered independently from GAD, SAD, OCD, panic disorder and PTSD in the analysis with ADIs, which was performed without taking the striatal subdivisions into consideration (Fig. 2c). Lower enrichment of cortical regions, including the frontal cortex and ACC, and higher enrichment for the spinal cord and the cerebellum could differentiate specific phobia from the rest of the ADs. However, ADI clustering with the inclusion of the striosome-matrix division led to its grouping with OCD, characterized by lower enrichment for the amygdala and hippocampus compared with SAD and panic disorder (Fig. 4c). Analysis with a dataset containing a larger number of regions again isolated specific phobia from the rest of the ADs (Fig. 6c). This isolation was characterized by a high enrichment of three transitory fetal structures—LGE, MGE and upper rhombic lip—and the cerebellum and the striatum, compared with the rest of the ADs. Together, these results seem to indicate (a) the etiological distinction of specific phobia from the other ADs, possibly driven by the spinal cord and the cerebellum, and (b) a possible association with GAD and OCD during the early stages of striosome- matrix specification, as indicated by its affiliation to LGE, a focal region in this process.

We used an interactome-driven model to examine the interrelatedness of ADs, using AD-associated genes as starting points. Our study (a) demonstrates the validity of the protein interactome as a data integration platform, (b) provides evidence supporting the role of the striatum in AD etiology and (c) proposes striosome-matrix specification as a key process with the potential to explain the neurodevelopmental origins of ADs. However, our study has several limitations, and our results should be interpreted with caution. Firstly, we used cross-species data (in addition to human data) as starting points for the construction of ADIs and SMIs (after mapping it to orthologous human data). This cross-species approach is necessary to connect research in animal models to human ADs and gather a systems-level view of the multiple biological levels that affect AD etiology (gene, local and global neuronal circuitry and behavior). Nevertheless, it is not advisable to draw direct transcriptomic/proteomic/phenotypic equivalences between humans and animal models unless the biological levels are comprehensively characterized in all the species, and a clear equivalence of factors defining a condition such as 'anxious behavior' is demonstrated in both the species[122]. Secondly, the reliability of our ADIs could be limited by the variability in differential AD diagnosis across a large number of studies (compiled by DisGeNET) from which the AD-associated genes were extracted. Challenges to the differential diagnosis of ADs include the high comorbidity of ADs, their comorbidity with other psychiatric disorders such as major depressive disorder, bipolar disorder and schizophrenia, and their occurrence with a range of other conditions such as substance use disorders, asthma, thyroid disease and complex partial seizures[120]. Thirdly, our ADIs are static networks, which do not incorporate data at various levels of granularity (cell, tissue and organism) or spatiotemporal points. Therefore at this stage, the interactome model will not be able to account for the complex and dynamic events that influence AD development, ranging from genetic mutations and PPI perturbations to gene-environment interactions, and the varying developmental trajectories of anxiety symptoms. Lastly, further investigations are necessary to characterize anxiety as an emergent property driven by specific neural circuits or neural mass effects. Our study proposes striatum and its subdivisions as one of the several candidate regions that may be prioritized for anxiety research.

In summary, our study reveals distinctive interactome overlaps shared between different ADs and striatal compartments and a bifurcation among ADs that are influenced by key anxiety-associated regions and neuronal signaling pathways. Our study proposes striatum as one of the focal regions for future AD research.

Methods

Compilation of genes associated with ADs. Genes associated with six types of ADs (from ref[55] Fig. 1. Suggested scheme for exploring a suspected anxiety disorder), namely, post-traumatic stress disorder (PTSD), obsessive–compulsive disorder (OCD), generalized anxiety disorder (GAD), social anxiety disorder (SAD), specific phobia and panic disorder were extracted from DisGeNET[56]. A gene-disease association score ≥ 0.01 was chosen to ensure that at least one publication has linked the gene in question with the disease. Note that 'association' of a gene with an AD here does not imply 'causality' in most cases, and may only indicate an association with disease susceptibility or behavioral endophenotypes of diseases. We compiled a list of 33 genes for GAD, 109 genes for PTSD including chronic PTSD, 134 genes for OCD including obsessions, OCD behaviour, OCD trait and OCD personality, 22 genes for SAD, 22 genes for specific phobia including claustrophobia and toco- phobia, and 93 genes for panic disorder including panic disorder with agoraphobia (Supplementary Table S1).

Construction of interactomes. Protein–protein interactions (PPIs) in the human interactome were compiled from Human Protein Reference Database (HPRD)[59] and the Biological General Repository for Inter-action Datasets (BioGRID)[60] using the Cytoscape plugin, Bisogenet[123]. The network building options were: organism—Homo sapiens, biorelation type—protein–protein interaction, data sources— BioGRID and HPRD,method—input nodes and its neighbors upto a distance of 1. The extracted interactomes included direct interactors of genes associated with the specific AD and intermediate interactors connecting AD-associated genes. This resulted in the GAD interactome with 595 genes and 3517 PPIs, SAD interactome with 266 genes and 887 PPIs,PTSD interactome with 2119 genes and 19,687 PPIs, specific phobia interactome with 634 genes and 4799 PPIs,OCD interactome with 1718 genes and 15,359 PPIs and panic disorder interactome with 773 genes and 4261 PPIs (Supplementary Data S1). The interactomes were visualized using Cytoscape[124].

Construction of striosome and matrix interactomes. Genes differentially expressed in striosome andmatrix compartments of the striatum were compiled from Table 1 of Graybiel et al.[47] In cases where the genes were from other species, the corresponding human genes were retrieved from the Homologene database (https://www.ncbi.nlm.nih.gov/homologene), these genes were mapped to their human homologs (taxon id: 9606). Thisyielded 49 and 19 striosome and matrix genes, respectively (Supplementary Table S3). The striosome and matrix interactomes with 827 genes and 6274 PPIs, and 213 genes and 779 PPIs respectively were then assembled from the human interactome (Supplementary Data S2), as explained before in the case of ADIs.

Gene expression enrichment in brain regions. We checked the enrichment of AD-SMI genes among genes expressed in specific brain regions. RNA-Seq data from the brains of adult donors was extracted from GTEx[57]. Genes with high or medium expression (transcripts per million (TPM) ≥ 9) in 13 brain regions were included, provided that they were not housekeeping genes, i.e. genes detected in all the tissues with transcriptsper million ≥ 1, as identified in the Human Protein Atlas[58]. TPM is a metric for quantifying gene expression; itdirectly measures the relative abundance of transcripts. 9638 genes were considered as housekeeping genes. A gene matrix transpose (GMT) file was created with amygdala (1953 genes), ACC-BA24 (2269 genes), caudatenucleus (2229 genes), cerebellar hemisphere (3978 genes), cerebellum (3968 genes), cortex (2706 genes), frontalcortex-BA9 (2872 genes), hippocampus (1949 genes), hypothalamus (2374 genes), nucleus accumbens (2464 genes), putamen (1892 genes), spinal cord-cervical c-1 (2408 genes) and substantia nigra (1949 genes). For an independent analysis, a GMT file with genes having log(Reads Per Kilobase per Million mapped reads) > 2 in 26 brain regions, that are not housekeeping genes, were compiled from Allen Brain Atlas[69]. This included amygdaloid complex (4278 genes), anterior (rostral) cingulate (medial prefrontal) cortex (7022 genes), caudal ganglionic eminence (4303 genes), cerebellar cortex (4580 genes), cerebellum (4563 genes), dorsal thalamus (4271 genes), dorsolateral prefrontal cortex (4570 genes), hippocampus (hippocampal formation) (4713 genes), inferolateral temporal cortex -area TEv, area 20 (7436 genes), lateral ganglionic eminence (4448 genes), medial ganglionic eminence (4395 genes), mediodorsal nucleus of thalamus (4503 genes), occipital neocortex (4457 genes), orbital frontal cortex (4617 genes), parietal neocortex (4443 genes), posterior (caudal) superior temporal cortex-area 22c (4558 genes), posteroventral (inferior) parietal cortex (4475 genes), primary auditorycortex (core) (4540 genes), primary motor cortex-area M1, area 4 (4545 genes), primary motor-sensory cortex(4488 genes), primary somatosensory cortex-area S1, areas 3,1,2 (4520 genes), primary visual cortex-striate cor-tex, area V1/17 (4528 genes), striatum (4628 genes), temporal neocortex (4384 genes), upper (rostral) rhombic lip (4480 genes) and ventrolateral prefrontal cortex (4575 genes).

The GMT files served as inputs for a gene over-representation analysis (GSEA) based on the hypergeometric distribution. In this method, the *p* value is computed from the probability of k successes in n draws (without replacement) from a finite population of size N containing exactly k objects with an interesting feature.

$$P\,(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

N = Total number of genes expressed in any brain regions. K = number of genes expressed in a particular brain region. n = number of genes co-occurring in an ADI and SMIs. k = number of common genes between K and n (genes co-occurring in an ADI and SMIs, that are also expressed in a particular brain region). Enrichment ratio is computed as the ratio of k/n and K/N.

Signaling pathway enrichment in brain regions. WebGestalt was used to compute the distributionof genes involved in a specific signaling pathway in the gene sets that were responsible for the enrichment of the AD-SMIs in highly influential brain regions, and compare it with the background distribution of genes belonging to this pathway among all the genes associated with any pathway in the selected database (KEGG[72])[73].Statistical significance of the enrichment was computed using Fisher's exact test, and corrected using the Benja-mini–Hochberg method for multiple test adjustment.

Principal component analysis. Principal component analysis (PCA) was used to capture relationships between the ADIs, and between AD-SMIs. Negative log-transformed *p* values indicating the statistical significance of enrichment of various brain regions in ADIs/AD-SMIs were assembled into a data matrix containing brain regions as rows and ADs as columns; each cell in the matrix contained a $-\log_{10}P$ value. PCA was per- formed with a web-based tool called ClustVis (https://biit.cs.ut.ee/clustvis/)[125]. The data matrix was pre-pro-

cessed to include only those rows and columns that contained less than 70% missing values. The $-\log_{10}P$ valuesin the matrix were further centered using the unit variance scaling method, in which the values are divided by standard deviation so that each row or column has a variance of one; this ensures that they assume equal importance while finding the components. The method called singular value decomposition (SVD) with imputation was used to extract principal components. In this method, missing values are predicted and iteratively filled using neighbouring values during SVD computation, until the estimates of missing values converge. The number of principal components computed was equal to the number of column dimensions in the data matrix, i.e. the number of ADIs or gene sets shared between ADIs and SMIs, in our case. PCA essentially transformed our original variables ($-\log_{10}P$) into uncorrelated variables called principal components. These principal components were ranked in the descending order of the percentage of total variance explained by them. We extracted the firsttwo components, i.e. PC1 and PC2, and plotted the component scores of each tissue on a 2D plane to capture the angle of most variability and delineate grouping patterns based on approximated distances between the scores.Scores corresponding to PC1 and PC2 were plotted along the X and Y axes respectively.

After an initial assessment of potential clusters, we extracted factor/component loadings corresponding to all the brain regions that contributed to the selected principal components. Component loadings are correlationcoefficients between the variables in rows and the factors (i.e. PC1, PC2 etc.). The squared value of a component loading gives the percentage of the variance explained by a particular original variable, and essentially its contribution to the principal components. We plotted the loading of each brain region corresponding to PC1 (X-axis) and PC2 (Y-axis) to examine their influence on the grouping patterns observed on the PC plot.

Clustering analysis. The data matrix of brain regions (rows) and ADIs or AD-SMIs (columns) was subjected to hierarchical clustering using the tool called Heatmapper (http://www.heatmapper.ca/)[126], in order to check whether the grouping patterns observed in the PC plot are valid. Pairwise distances in the data matrix were calculated using Pearson correlation and they were 'linked' using the average linkage method. Dendro- grams were generated by merging tissue samples with the smallest distance first, and those with larger distanceslater. In the average linkage method, the average distance of all possible pairs is considered while clustering[127].

Data availability
ADIs and SMIs analyzed in this study have been made available as Supplementary Data S1 and Supplementary Data S2, respectively.

References

1. Gross, C. & Hen, R. The developmental origins of anxiety. *Nat. Rev. Neurosci.* **5**, 545–552 (2004).
2. Chambers, J. A., Power, K. G. & Durham, R. C. The relationship between trait vulnerability and anxiety and depressive diagnoses at long-term follow-up of Generalized Anxiety Disorder. *J. Anxiety Disord.* **18**, 587–607 (2004).
3. Kyu, H. H. *et al.* Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of DiseaseStudy 2017. *The Lancet* **392**, 1859–1922 (2018).
4. Kessler, R. C. *et al.* Comorbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow- up. *Psychol. Med.* **38**, 365 (2008).
5. Hettema, J. M., Neale, M. C. & Kendler, K. S. A review and meta-analysis of the genetic epidemiology of anxiety disorders. *Am. J. Psychiatry* **158**, 1568–1578 (2001).
6. Taylor, S. Etiology of obsessions and compulsions: A meta-analysis and narrative review of twin studies. *Clin. Psychol. Rev.* **31**, 1361–1372 (2011).
7. Kessler, R. C., Chiu, W. T., Demler, O. & Walters, E. E. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62**, 617–627 (2005).
8. Shimada-Sugimoto, M., Otowa, T. & Hettema, J. M. Genetics of anxiety disorders: Genetic epidemiological and molecular studies in humans. *Psychiatry Clin. Neurosci.* **69**, 388–401 (2015).
9. Smoller, J. W., Block, S. R. & Young, M. M. Genetics of anxiety disorders: The complex road from DSM to DNA. *Depress. Anxiety* **26**, 965–975 (2009).
10. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **21**, 1391–1399 (2016).
11. Levey, D. F. *et al.* Reproducible genetic risk loci for anxiety: Results from~ 200,000 participants in the Million Veteran Program. *Am. J. Psychiatry* **177**, 223–232 (2020).
12. Dunn, E. C. *et al.* Genome-wide association study of generalized anxiety symptoms in the Hispanic Community Health Study/ Study of Latinos. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **174**, 132–143 (2017).
13. Stein, M. B. *et al.* Genetic risk variants for social anxiety. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **174**, 120–131 (2017).
14. Arnold, P. D. *et al.* Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1181 (2018).
15. Edition, F. Diagnostic and statistical manual of mental disorders. *Am. Psychiatr. Assoc.* DSM-5. — 5th ed, **203**, 222–223 (2013).
16. Whalen, P. J. *et al.* A functional magnetic resonance imaging predictor of treatment response to venlafaxine in generalized anxiety disorder. *Biol. Psychiat.* **63**, 858–863 (2008).
17. Hattingh, C. J. *et al.* Functional magnetic resonance imaging during emotion recognition in social anxiety disorder: An activa- tion likelihood meta-analysis. *Front. Hum. Neurosci.* **6**, 347 (2013).
18. Anderson, K. M. *et al.* Gene expression links functional networks across cortex and striatum. *Nat. Commun.* **9**, 1–14 (2018).
19. Richiardi, J. *et al.* Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
20. Wang, G.-Z. *et al.* Correspondence between resting-state activity and brain gene expression. *Neuron* **88**, 659–666 (2015).
21. Krienen, F. M., Yeo, B. T., Ge, T., Buckner, R. L. & Sherwood, C. C. Transcriptional profiles of supragranular-enriched genesassociate with corticocortical network architecture in the human brain. *Proc. Natl. Acad. Sci.* **113**, E469–E478 (2016).
22. Patania, A. *et al.* Topological gene expression networks recapitulate brain anatomy and function. *Netw. Neurosci.* **3**, 744–762(2019).

23. Mills, B. D. *et al.* Correlated gene expression and anatomical communication support synchronized brain activity in the mouse functional connectome. *J. Neurosci.* **38**, 5774–5787 (2018).
24. Sylvester, C. M. *et al.* Functional network dysfunction in anxiety and anxiety disorders. *Trends Neurosci.* **35**, 527–535 (2012).
25. Le-Niculescu, H. *et al.* Convergent functional genomics of anxiety disorders: Translational identification of genes, biomarkers, pathways and mechanisms. *Transl. Psychiatry* **1**, e9–e9 (2011).
26. Wingo, A. P. & Gibson, G. Blood gene expression profiles suggest altered immune function associated with symptoms of general-ized anxiety disorder. *Brain Behav. Immun.* **43**, 184–191 (2015).
27. Song, Y., Liu, Y., Wu, P., Zhang, F. & Wang, G. Genome-wide mRNA expression analysis of peripheral blood from patients with obsessive-compulsive disorder. *Sci. Rep.* **8**, 1–8 (2018).
28. Bracha, H. S., Garcia-Rill, E., Mrak, R. E. & Skinner, R. Postmortem locus coeruleus neuron count in three American veterans with probable or possible war-related PTSD. *J. Neuropsychiatry Clin. Neurosci.* **17**, 503–509 (2005).
29. Su, Y. A. *et al.* Dysregulated mitochondrial genes and networks with drug targets in postmortem brain of patients with post-traumatic stress disorder (PTSD) revealed by human mitochondria-focused cDNA microarrays. *Int. J. Biol. Sci.* **4**, 223 (2008).
30. Sakai, Y. *et al.* Protein interactome reveals converging molecular pathways among autism disorders. *Science translational medicine* **3**, 86ra49 (2011).
31. Lim, J. *et al.* A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* **125**, 801–814 (2006).
32. Ganapathiraju, M. K. *et al.* Schizophrenia interactome with 504 novel protein–protein interactions. *NPJ Schizophrenia* **2**, 16012 (2016).
33. Malavia, T. *et al.* Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epide-miological, genomic and protein interaction data *npj Schizophrenia* **in Press** (2017).
34. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
35. Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **31**, 359–387 (2008).
36. Graybiel, A. M. & Ragsdale, C. W. Histochemically distinct compartments in the striatum of human, monkeys, and cat demon-strated by acetylthiocholinesterase staining. *Proc. Natl. Acad. Sci.* **75**, 5723–5726 (1978).
37. Mikula, S., Parrish, S. K., Trimmer, J. S. & Jones, E. G. Complete 3D visualization of primate striosomes by KChIP1 immunostain-ing. *J. Comp. Neurol.* **514**, 507–517 (2009).
38. Herkenham, M. & Pert, C. B. Mosaic distribution of opiate receptors, parafascicular projections and acetylcholinesterase in rat striatum. *Nature* **291**, 415–418 (1981).
39. Martin, L., Hadfield, M., Dellovade, T. & Price, D. The striatal mosaic in primates: Patterns of neuropeptide immunoreactivity differentiate the ventral striatum from the dorsal striatum. *Neuroscience* **43**, 397–417 (1991).
40. Gerfen, C. R. The neostriatal mosaic. I. Compartmental organization of projections from the striatum to the substantia nigra in the rat. *J. Comp. Neurol.* **236**, 454–476 (1985).
41. Koshimizu, Y. *et al.* Paucity of enkephalin production in neostriatal striosomal neurons: Analysis with preproenkephalin–green fluorescent protein transgenic mice. *Eur. J. Neurosci.* **28**, 2053–2064 (2008).
42. Graybiel, A. Correspondence between the dopamine islands and striosomes of the mammalian striatum. *Neuroscience* **13**, 1157–1187 (1984).
43. Eblen, F. & Graybiel, A. M. Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *J. Neurosci.* **15**, 5999–6013 (1995).
44. Amemori, S., *et al.* Microstimulation of primate neocortex targeting striosomes induces negative decision-making. *Eur. J. Neu-rosci.* **51**, 731–741 (2020).
45. Amemori, K.-I. & Graybiel, A. M. Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat. Neurosci.* **15**, 776–785 (2012).
46. Friedman, A. *et al.* A corticostriatal path targeting striosomes controls decision-making under conflict. *Cell* **161**, 1320–1333 (2015).
47. Crittenden, J. R. & Graybiel, A. M. Basal Ganglia disorders associated with imbalances in the striatal striosome and matrix compartments. *Front. Neuroanat.* **5**, 59 (2011).
48. Fujiyama, F. *et al.* Exclusive and common targets of neostriatofugal projections of rat striosome neurons: A single neuron-tracing study using a viral vector. *Eur. J. Neurosci.* **33**, 668–677 (2011).
49. Amemori, K.-I. & Graybiel, A. M. Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat. Neurosci.* **15**, 776 (2012).
50. Amemori, K.-I., Amemori, S., Gibson, D. J. & Graybiel, A. M. Striatal microstimulation induces persistent and repetitive negative decision-making predicted by striatal beta-band oscillation. *Neuron* **99**, 829-841. e826 (2018).
51. White, N. M. & Hiroi, N. Preferential localization of self-stimulation sites in striosomes/patches in the rat striatum. *Proc. Natl. Acad. Sci.* **95**, 6486–6491 (1998).
52. Lawhorn, C., Smith, D. M. & Brown, L. L. Striosome-matrix pathology and motor deficits in the YAC128 mouse model of Huntington's disease. *Neurobiol. Dis.* **32**, 471–478 (2008).
53. Koizumi, H. *et al.* Response of striosomal opioid signaling to dopamine depletion in 6-hydroxydopamine-lesioned rat model of Parkinson's disease: A potential compensatory role. *Front. Cell. Neurosci.* **7**, 74 (2013).
54. Canales, J. J. & Graybiel, A. M. A measure of striatal function predicts motor stereotypy. *Nat. Neurosci.* **3**, 377–383 (2000).
55. Goodwin, G. M. The overlap between anxiety, depression, and obsessive-compulsive disorder. *Dialog. Clin. Neurosci.* **17**, 249 (2015).
56. Piñero, J. *et al.* DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids Res.* **45**(D1), D833–D839 (2016).
57. Consortium, G. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
58. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 394–403 (2015).
59. Keshava Prasad, T. *et al.* Human protein reference database—2009 update. *Nucleic Acids Res.* **37**, D767–D772 (2009).
60. Stark, C. *et al.* BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
61. Chang, C.-W. *et al.* Identification of a developmentally regulated striatum-enriched zinc-finger gene, Nolz-1, in the mammalian brain. *Proc. Natl. Acad. Sci.* **101**, 2613–2618 (2004).
62. López-Giménez, J. F., Mengod, G., Palacios, J. M. & Vilaró, M. T. Human striosomes are enriched in 5-HT2A receptors: Autora-diographical visualization with [3H] MDL100, 907,[125I](±) DOI and [3H] ketanserin. *Eur. J. Neurosci.* **11**, 3761–3765 (1999).
63. Robinson, P. N. *et al.* The Human Phenotype Ontology: A tool for annotating and analyzing human hereditary disease. *Am. J.Hum. Genet.* **83**, 610–615 (2008).
64. Crittenden, J. R., Lacey, C. J., Lee, T., Bowden, H. A. & Graybiel, A. M. Severe drug-induced repetitive behaviors and striatal overexpression of VAChT in ChAT-ChR2-EYFP BAC transgenic mice. *Front. Neural Circuits* **8**, 57 (2014).
65. Saka, E., Iadarola, M., Fitzgerald, D. & Graybiel, A. Local circuit neurons in the striatum regulate neural and behavioral responses to dopaminergic stimulation. *Proc. Natl. Acad. Sci.* **99**, 9004–9009 (2002).

66. Aliane, V., Perez, S., Bohren, Y., Deniau, J.-M. & Kemel, M.-L. Key role of striatal cholinergic interneurons in processes leading to arrest of motor stereotypies. *Brain* **134**, 110–118 (2011).
67. Woodward, N. D. & Cascio, C. J. Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiat.* **72**, 743–744(2015).
68. Shackman, A. J. & Fox, A. S. (Am Psychiatric Assoc, 2021).
69. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391 (2012).
70. Kelly, S. M. *et al.* Radial glial lineage progression and differential intermediate progenitor amplification underlie striatal com- partments and circuit organization. *Neuron* **99**, 345-361. e344 (2018).
71. Lavdas, A. A., Grigoriou, M., Pachnis, V. & Parnavelas, J. G. The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex. *J. Neurosci.* **19**, 7881–7888 (1999).
72. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2007).
73. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
74. Mozhui, K. *et al.* Strain differences in stress responsivity are associated with divergent amygdala gene expression and glutamate-mediated neuronal excitability. *J. Neurosci.* **30**, 5357–5367 (2010).
75. Jaffe, A. *et al.* Genetic neuropathology of obsessive psychiatric syndromes. *Transl. Psychiatry* **4**, e432–e432 (2014).
76. Janik, P., Berdyński, M., Safranow, K. & Żekanowski, C. Association of ADORA1 rs2228079 and ADORA2A rs5751876 Poly- morphisms with Gilles de la Tourette Syndrome in the Polish Population. *PloS one* **10**, e0136754 (2015).
77. Papale, L. A., Madrid, A., Li, S. & Alisch, R. S. Early-life stress links 5-hydroxymethylcytosine to anxiety-related behaviors. *Epigenetics* **12**, 264–276 (2017).
78. Littlejohn, B. P. *et al.* Influence of prenatal transportation stress-induced differential DNA methylation on the physiological control of behavior and stress response in suckling Brahman bull calves. *J. Anim. Sci.* **98**, skz68 (2020).
79. Mohn, J. *et al.* Adenomatous polyposis coli protein deletion leads to cognitive and autism-like disabilities. *Mol. Psychiatry* **19**, 1133–1142 (2014).
80. Yang, T. *et al.* Ca2+-binding protein 1 regulates hippocampal-dependent memory and synaptic plasticity. *Neuroscience* **380**, 90–102 (2018).
81. Dieckmann, L., Cole, S. & Kumsta, R. Stress genomics revisited: Gene co-expression analysis identifies molecular signaturesassociated with childhood adversity. *Transl. Psychiatry* **10**, 1–11 (2020).
82. Yamashita, N. *et al.* Mice lacking collapsin response mediator protein 1 manifest hyperactivity, impaired learning and memory, and impaired prepulse inhibition. *Front. Behav. Neurosci.* **7**, 216 (2013).
83. Chitu, V. *et al.* Phenotypic characterization of a Csf1r haploinsufficient mouse model of adult-onset leukodystrophy with axonal spheroids and pigmented glia (ALSP). *Neurobiol. Dis.* **74**, 219–228 (2015).
84. Muhie, S. *et al.* Molecular indicators of stress-induced neuroinflammation in a mouse model simulating features of post- traumatic stress disorder. *Transl. Psychiatry* **7**, e1135–e1135 (2017).
85. Raud, S. *et al.* Relation between increased anxiety and reduced expression of alpha1 and alpha2 subunits of GABAA receptors in Wfs1-deficient mice. *Neurosci. Lett.* **460**, 138–142 (2009).
86. Edfawy, M. *et al.* Abnormal mGluR-mediated synaptic plasticity and autism-like behaviours in Gprasp2 mutant mice. *Nat. Commun.* **10**, 1–15 (2019).
87. Weber, T. *et al.* Adult AMPA GLUA1 receptor subunit loss in 5-HT neurons results in a specific anxiety-phenotype with evidence for dysregulation of 5-HT neuronal activity. *Neuropsychopharmacology* **40**, 1471–1484 (2015).
88. Umemori, J. *et al.* ENU-mutagenesis mice with a non-synonymous mutation in Grin1 exhibit abnormal anxiety-like behaviors, impaired fear memory, and decreased acoustic startle response. *BMC. Res. Notes* **6**, 203 (2013).
89. Banerjee, A., Luong, J. A., Ho, A., Saib, A. O. & Ploski, J. E. Overexpression of Homer1a in the basal and lateral amygdala impairsfear conditioning and induces an autism-like social impairment. *Mol. Autism* **7**, 1–15 (2016).
90. Rizk, A., Curley, J., Robertson, J. & Raber, J. Anxiety and cognition in histamine H3 receptor$^{-/-}$ mice. *Eur. J. Neurosci.* **19**, 1992–1996 (2004).
91. Heisler, L., Zhou, L., Bajwa, P., Hsu, J. & Tecott, L. Serotonin 5-HT2C receptors regulate anxiety-like behavior. *Genes Brain Behav.* **6**, 491–496 (2007).
92. Ambeskovic, M. *et al.* Ancestral stress alters lifetime mental health trajectories and cortical neuromorphology via epigenetic regulation. *Sci. Rep.* **9**, 1–14 (2019).
93. Maussion, G. *et al.* Convergent evidence identifying MAP/microtubule affinity-regulating kinase 1 (MARK1) as a susceptibility gene for autism. *Hum. Mol. Genet.* **17**, 2541–2551 (2008).
94. Stork, O. *et al.* Anxiety and increased 5-HT1A receptor response in NCAM null mutant mice. *J. Neurobiol.* **40**, 343–355 (1999).
95. Kalbassi, S., Bachmann, S. O., Cross, E., Roberton, V. H. & Baudouin, S. J. Male and female mice lacking neuroligin-3 modify the behavior of their wild-type littermates. *Eneuro* **4**, https://doi.org/10.1523/ENEURO.0145-17.2017 (2017).
96. Schoch, H. *et al.* Sociability deficits and altered amygdala circuits in mice lacking Pcdh10, an autism associated gene. *Biol. Psychiat.* **81**, 193–202 (2017).
97. Liu, Y. *et al.* Association of variant rs4790904 in protein kinase C alpha with posttraumatic stress disorder in a US Caucasianand African-American veteran sample. *J. Depress. Anxiety* **2**, S4-001 (2013).
98. Lesscher, H. M. *et al.* Amygdala protein kinase C epsilon regulates corticotropin-releasing factor and anxiety-like behavior. *Genes Brain Behav.* **7**, 323–333 (2008).
99. Chiavegatto, S. *et al.* Expression of α-synuclein is increased in the hippocampus of rats with high levels of innate anxiety. *Mol. Psychiatry* **14**, 894–905 (2009).
100. Stankiewicz, A. M., Goscik, J., Majewska, A., Swiergiel, A. H. & Juszczak, G. R. The effect of acute and chronic social stress on the hippocampal transcriptome in mice. *PLoS One* **10**, e0142195 (2015).
101. Nakamura, K. *et al.* Genetic and expression analyses reveal elevated expression of syntaxin 1A (STX1A) in high functioning autism. *Int. J. Neuropsychopharmacol.* **11**, 1073–1084 (2008).
102. Jawinski, P. *et al.* Human brain arousal in the resting state: A genome-wide association study. *Mol. Psychiatry* **24**, 1599–1609 (2019).
103. Cannistraro, P. A. & Rauch, S. L. Neural circuitry of anxiety: Evidence from structural and functional neuroimaging studies. *Psychopharmacol. Bull.* **37**, 8–25 (2003).
104. Lowry, C. A., Johnson, P. L., Hay-Schmidt, A., Mikkelsen, J. & Shekhar, A. Modulation of anxiety circuits by serotonergic systems. *Stress* **8**, 233–246 (2005).
105. Nutt, D. J. The pharmacology of human anxiety. *Pharmacol. Ther.* **47**, 233–266 (1990).
106. Cassano, G. B., Rossi, N. B. & Pini, S. Psychopharmacology of anxiety disorders. *Dialogues Clin. Neurosci.* **4**, 271 (2002).
107. Lago, T., Davis, A., Grillon, C. & Ernst, M. Striatum on the anxiety map: Small detours into adolescence. *Brain Res.* **1654**, 177–184(2017).
108. Ghashghaei, H., Hilgetag, C. C. & Barbas, H. Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage* **34**, 905–923 (2007).
109. Khalsa, S. S., Rudrauf, D., Feinstein, J. S. & Tranel, D. The pathways of interoceptive awareness. *Nat. Neurosci.* **12**, 1494–1496 (2009).

110. LeDoux, J. Rethinking the emotional brain. *Neuron* **73**, 653–676 (2012).
111. Bloem, B., Huda, R., Sur, M. & Graybiel, A. M. Two-photon imaging in mice shows striosomes and matrix have overlapping but differential reinforcement-related responses. *Elife* **6**, e32353 (2017).
112. Friedman, A. *et al.* Chronic stress alters striosome-circuit dynamics, leading to aberrant decision-making. *Cell* **171**, 1191-1205. e1128 (2017).
113. Salinas, A. G., Davis, M. I., Lovinger, D. M. & Mateo, Y. Dopamine dynamics and cocaine sensitivity differ between striosome and matrix compartments of the striatum. *Neuropharmacology* **108**, 275–283 (2016).
114. Brimblecombe, K. R. & Cragg, S. J. The striosome and matrix compartments of the striatum: A path through the labyrinth from neurochemistry toward function. *ACS Chem. Neurosci.* **8**, 235–242 (2017).
115. Prager, E. M. *et al.* Dopamine oppositely modulates state transitions in striosome and matrix direct pathway striatal spiny neurons. *Neuron* **108**, 1091–1102.e5 (2020).
116. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74**, 858–873 (2012).
117. Leonardo, E. D. & Hen, R. Anxiety as a developmental disorder. *Neuropsychopharmacology* **33**, 134–140 (2008).
118. Turecki, G. & Meaney, M. J. Effects of the social environment and stress on glucocorticoid receptor gene methylation: A sys-tematic review. *Biol. Psychiat.* **79**, 87–96 (2016).
119. Zannas, A. & Binder, E. Gene–environment interactions at the FKBP5 locus: Sensitive periods, mechanisms and pleiotropism. *Genes Brain Behav.* **13**, 25–37 (2014).
120. Craske, M. G. *et al.* Anxiety disorders. *Nat. Rev. Dis. Prim.* **3**, 17024. https://doi.org/10.1038/nrdp.2017.24 (2017).
121. Joshi, S. A., Duval, E. R., Kubat, B. & Liberzon, I. A review of hippocampal activation in post-traumatic stress disorder. *Psycho- physiology* **57**, e13357 (2020).
122. Breschi, A., Gingeras, T. R. & Guigó, R. Comparative transcriptomics in human and mouse. *Nat. Rev. Genet.* **18**, 425 (2017).
123. Martin, A. *et al.* BisoGenet: A new tool for gene network building, visualization and analysis. *BMC Bioinform.* **11**, 1–9 (2010).
124. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L. & Ideker, T. Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
125. Metsalu, T. & Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* **43**, W566–W570 (2015).
126. Babicki, S. *et al.* Heatmapper: Web-enabled heat mapping for all. *Nucleic Acids Res.* **44**, W147–W153 (2016).
127. Mukherjee, P. *et al.* Lower effective connectivity between amygdala and parietal regions in response to fearful faces in schizo- phrenia. *Schizophr. Res.* **134**, 118–124 (2012).

Author contributions

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-97418-w.

Correspondence and requests for materials should be addressed to M.K.G. or K.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps andinstitutional affiliations.

Interactome-based framework to translate disease genetic data into biological and clinical insights

10. Drug Contraindications in Comorbid Diseases: a Protein Interactome Perspective

The experimental chapter is based on the following publication under review:
Karunakaran, Kalyani B., Madhavi Ganapathiraju, Sanjeev Jain, Samir Brahmachari, and Narayanaswamy Balakrishnan. Drug Contraindications in Comorbid Diseases: a Protein Interactome Perspective. bioRxiv (2022).
Current status: Accepted for publication in Network Modeling Analysis in Health Informatics and Bioinformatics (18 December 2023)

Summary of this chapter

In this chapter, I demonstrate the methodology to examine the target networks of drugs to elucidate patterns predictive of their clinical activity. To identify the potential mechanisms of adverse drug reactions (ADRs) within comorbid diseases, I systematically examined six pairs of comorbid diseases, three pairs of non-comorbid diseases, and their FDA-approved drugs. The relative risk of comorbidity was proportional to disease network similarity measures, showing that disease networks contained comorbidity patterns. I compiled four categories of drugs for each disease pair based on their clinical activity, and constructed four corresponding drug target networks (DTNs). I then examined the DTNs for enrichment among genes involved in disease protein-protein interaction (PPI) networks, tissue-specific expression, and pathways. Using principal component analysis and downstream analyses, I identified specific disease protein sets, pathways, and tissues closely related to each of the four DTNs. I found that the target networks of disease A drugs not contraindicated in disease B were affiliated with the disease B network, associated pathways and tissues, while the target networks of disease A drugs contraindicated in disease B were affiliated with the disease A network, related pathways and tissues. These correlations indicated that the aetiological associations between the two comorbidities could play an active role in their therapeutic alleviation. The study concluded that examining DTN enrichment in pathways, tissues, and PPI networks of comorbid diseases can provide valuable insights into drug safety and efficacy in comorbid conditions.

Contribution to this chapter (75%)

- Designed the study and developed the methodology of the project, including DTN and disease interactome construction and all the components of comparative DTN analysis, namely, comparative analysis of patient comorbidity data and network similarity measures, transcriptomic analysis, pathway analysis and principal component analysis
- Curated all the datasets, performed all the analyses, and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

Drug Contraindications in Comorbid Diseases: a Protein Interactome Perspective

Kalyani B. Karunakaran[1], Madhavi K. Ganapathiraju[2,3*], Sanjeev Jain[4],Samir K. Brahmachari[5,6] and N. Balakrishnan[1*]

[1]Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, India

[2]Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, USA

[3]Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, USA

[4]National Institute of Mental Health and Neuro-Sciences (NIMHANS), Bangalore, India

[5]Academy of Scientific and Innovative Research, CSIR-4PI, Bangalore, India [6]Indraprastha

Institute of Information Technology, Delhi, India

Correspondence to: N. Balakrishnan (balki@iisc.ac.in) and Madhavi K. Ganapathiraju (madhavi@pitt.edu)

e-mail IDs of other authors: Kalyani B. Karunakaran: kalyanik@iisc.ac.in Sanjeev Jain: sjain.nimhans@gmail.com Samir K. Brahmachari: skb@igib.in

Abstract

Adverse drug reactions (ADRs) are leading causes of death and drug withdrawals and frequently co-occur with comorbidities. However, systematic studies on the effects of drugs in comorbidities are lacking. Drug interactions with the cellular protein-protein interaction (PPI) network give rise to ADRs. We selected 6 comorbid disease pairs, identified the drugs used in the treatment of the individual diseases 'A' and 'B'– 44 drugs in anxiety and depression, 128 in asthma and hypertension,48 in chronic obstructive pulmonary disease and heart failure, 58 in type 2 diabetes and obesity, 58 in Parkinson's disease and schizophrenia, and 84 in rheumatoid arthritis and osteoporosis – and categorized them based on whether they aggravate the comorbid condition. We constructed drug target networks (DTNs) and examined their enrichment among genes in disease A/B PPI networks, expressed across 53 tissues and involved in ~1000 pathways. To pinpoint the biological features characterizing the DTNs, we performed principal component analysis and computed the Euclidean distance between DTN component scores and feature loading values. DTNs of disease A drugs not

contraindicated in B were affiliated with proteins common to A/B networks or uniquely found in the B network, similarly regulated common pathways, and disease-B specific pathways and tissues. DTNsof disease A drugs contraindicated in B were affiliated with common proteins or those uniquely found in the A network, differentially regulated common pathways, and disease A-specific pathways and tissues. Hence, DTN enrichment in pathways, tissues, and PPI networks of comorbid diseases will help identify drugs contraindications in comorbidities.

Keywords: comorbidities, drugs, adverse drug reactions, drug contraindications, drug target networks

1. Introduction

Comorbidity is the phenomenon in which one or more diseases co-exist with a primary disease in patients. Comorbidities are the norm rather than exceptions among chronic conditions and pose a significant threat to the physical and psychosocial wellbeing of patients [1]. Comorbidities increase with age, and the risk of mortality increases with the number of comorbidities. A longitudinal study (1992-2006) has shown that the mortality risk increased by 25% in patients with 3-4 chronic comorbidities and by 80% in those with 5 or more comorbidities, both in comparison with individuals having no chronic conditions [2]. The prevalence of comorbidities increases from 10% in 0-19 year-olds to 78% in individuals aged 80 or more [3]. The prevalence of comorbidity in women of age groups of 18-44 years, 45-64 years, and $\geq$65 years was 68%, 95%, and 99% and in men, it was 72%, 89%, and 97% [4]. As per the US National Comorbidity Survey Replication (NCS-R) survey, 73.8-98.2% of the respondents reported having at least one comorbid condition along with a primary condition [1]. The most striking finding from this report was that the estimates of individual disease burden based on the respondents' perception of their health condition decreased substantially when adjusted for comorbidity [1]. This effect was particularly magnified for neurological disorders, chronic pain, anxiety disorders, major depressive disorder, and diabetes, all of which contribute immensely to the global disease burden [1]. For example, anxiety disorders collectively affect 284 million people (63% females, 2.5-7% variation by country) around the world, and are among the mostprevalent mental health and neurodevelopmental disorders (WHO and IHME, 2017) [5].

Disease comorbidity may increase the likelihood of experiencing adverse drug reactions [6-8]. Drugsthat are beneficial in the treatment of one disease may aggravate or even cause comorbid conditions, giving rise to adverse drug reactions, e.g. beta-blockers that treat hypertension and heart disease may aggravate asthma [6], trimethoprim and sulfamethoxazole to treat AIDS may increase the patient's susceptibility to Stevens-Johnson syndrome and toxic epidermal necrolysis [7]; malaria patients withAIDS and osteoarthritis treated with artemisinin based combination antimalarial therapy were 3 times more likely to experience adverse side effects [8]. Serious adverse drug reactions constitute the

fourth leading cause of death in the U.S. with 100,000 deaths per year, and about 2 million patients in the U.S. experiencing adverse drug reactions per year [9]. Patient fatalities have led to the withdrawal of 19 drugs from the U.S. market during 1998-2007 [9]. These aspects highlight the importance of re-examining drug design, and the need to develop drugs in light of disease mechanisms governing comorbidities.

Network medicine is an integrative framework for examining the mechanistic effects of disease-associated genes within the context of the human protein-protein interaction (PPI) network (or the 'interactome') [10]. The emerging network medicine paradigm in systems biology has prompted systematic data-driven investigations of the effects of drugs on diseases. It captures the essence of the Fourth Paradigm, i.e. Data-Intensive Scientific Discovery [11, 12]. This framework allows data capture and combines theory and computation to facilitate the translation of biological data into biologically insightful and clinically actionable results. The primary applications of this framework are uncovering disease-associated genes, identifying biomarkers that will improve disease screening, clinical diagnosis, and patient stratification, and prioritizing drug targets and pathways for therapeutic intervention [12].

Drugs that target proteins may perturb the PPI network to elicit the intended therapeutic response or an unintended adverse event or side effect [13]. The extensive interconnectivity of the network components suggests that perturbations at the genomic or proteomic level that affect PPIs may disrupt cellular functions and affect other proteins in the neighborhood network, posing deeper implications for several aspects of the disease such as comorbidity and phenotypic responses to drugs [10].

Although the side effects or adverse events precipitated by drugs in specific diseases have been investigated within the framework of the PPI network [14-19], the effects of multiple drugs and their contraindications on comorbid conditions remain largely unexplored. Some studies have provided key insights on the influence of disease-associated PPI networks, biological pathways, and tissues on drug action. Pairs of drugs used for the same disease have shown significant adverse events when the network modules of their protein targets overlap with each other or with a network of disease-associated genes ('overlapping exposure', statistical significance p-value ≤ 0.007), e.g. the anti-hypertensive drug nadroparin increased hyperkalemia, an adverse effect of spironolactone, another anti-hypertensive drug [20]. The targets of both cancer and non-cancer drugs were enriched by 1.8 folds among tissue-specific proteins (p-value = 2E-06), and this enrichment became magnified to 2.3 folds when the targets of non-cancer drugs were considered alone [21]. Drugs that are currently in the market are twice as likely to act on tissue-specific proteins than on housekeeping proteins [22].

In this study, we attempt to elucidate the mechanisms underlying drug contraindications in pairs of comorbid diseases. Our findings suggest the relationship between the PPI networks of disease-associated proteins and drug targets, and the pathway membership and tissue specificity of the drug target networks as critical biological factors influencing adverse drug reactions in comorbidities.

2. Methods

2.1 Compilation of drugs indicated for specific diseases

The Drug Bank database [23] (version 5.1.8) was used to compile the lists of drugs indicated for each of the 14 diseases. After compiling these lists, we used the TWOSIDES database [24] (version 0.1) – a publicly available database of drugs and associated adverse events – to categorize these drugs with respect to their effects on the disease pairs, specifically, (a) drugs effective in disease A and not contraindicated in disease B, (b) drugs effective in disease B and not contraindicated in disease A, (c) drugs effective in disease A and contraindicated in disease B, and (d) drugs effective in disease B and contraindicated in disease A. Drugs associated with specific adverse effects (belonging to (c) and (d)) were identified using their 'condition concept names' (descriptions of adverse events). The lists of the condition concept names used for identifying the drugs belonging to the 4 groups for each of the disease pairs can be found in **Additional File 1: Table S1**, and the drug lists can be found in **Additional File 2: Table S2.**

2.2 Construction of drug target protein-protein interaction (PPI) networks

The proteins targeted by the drugs (**Additional File 3: Table S3**) belonging to the 4 categories were retrieved from the Drug Bank database [23] using the DGIdb (drug gene interaction database) web portal [25]. The PPIs of these drug targets in the human interactome were compiled from Human Protein Reference Database (HPRD; version 9) [26] and the Biological General Repository for Interaction Datasets (BioGRID; version 4.3.194) [27] using the Cytoscape plugin, Bisogenet [28]. The network building options were: organism - *Homo sapiens*, biorelation type - *protein-protein interaction*, data sources - *BioGRID and HPRD*, method - *input nodes and its neighbors upto a distance of 1*.

2.3 Compilation of disease-associated genes

The genes associated with each of the 14 diseases in the 3 non-comorbid pairs and 6 comorbid pairswere compiled from the DisGeNET database [29] (version 7). The non-comorbid pairs were (I) Multiple sclerosis (DisGeNET ID: C0026769) – Peroxisomal disorders (C0282528), (II) Schizophrenia (C0036341) – Rheumatoid arthritis (C0003873), (III) Asthma (C0004096) –

Schizophrenia (C0036341). The comorbid pairs were (IV) Anxiety (C0003467) – Depression (C0011570), (V) Asthma (C0004096) – Hypertension (C0085580), (VI) Chronic obstructive pulmonary disorder (COPD) (C0024117) – Heart failure (C0018801), (VII) Type 2 diabetes (C0011860) – Obesity (C0028754), (VIII) Rheumatoid arthritis (C0003873) – Osteoporosis (C0029456) and (IX) Parkinson's disease (C0030567) – Schizophrenia (C0036341) (**Additional File 4: Table S4**). 100 top-ranking genes associated with each of the diseases were curated based on their gene-disease association scores (GDA). Although the range of the GDA scores among the 100 top- ranking genes varied across our selected diseases (multiple sclerosis (0.11-0.5), peroxisomal disorders(0.01-0.32), schizophrenia (0.43-0.9), rheumatoid arthritis (0.33-0.7), asthma (0.29-0.7), anxiety (0.1-0.5), mental depression (0.34-0.6), essential hypertension (0.03-0.063), chronic obstructive airway disease (0.11-0.9), heart failure (0.3-0.6), non-insulin-dependent diabetes mellitus (0.4-1), obesity (0.4-1), osteoporosis (0.13-0.9) and Parkinson's disease (0.23-0.7)), a minimum GDA of ≥ 0.01 was chosen to ensure that at least one publication has linked the gene in question with the disease. Note that 'association' of a gene with a disease here does not imply causality in most cases and may only indicate an association with disease susceptibility or an endophenotype.

1.4 Construction of disease protein-protein interaction (PPI) networks

The PPI networks of the proteins encoded by the disease-associated genes were assembled by extracting their protein interactors from the PPI repositories BioGRID [27] and HPRD [26] usingBisoGenet [28] and the network building options specified before. The input nodes for the construction of each of the disease networks were the 100 top-ranking genes compiled from the DisGeNET database.

2.5 Calculation of network similarity measures

Matching node ratio (N_M) was measured as the ratio of the total number of common nodes shared between the two PPI networks of a comorbid pair and the total number of unique nodes in the two disease networks [30].

$$N_M = \frac{A_n \cap B_n}{A_n \cup B_n}$$

(1)

A_n = Number of nodes in the PPI network of disease A

B_n = Number of nodes in the PPI network of disease B

Matching link ratio (L$_M$) was measured as the ratio of the total number of common links (i.e. edges) shared between the two PPI networks of a comorbid pair and the total number of unique links in the two disease networks [30].

$$(2)$$

$$L_M = \frac{A_l \cap B_l}{A_l \cup B_l}$$

A_l = Number of links in the PPI network of disease A

B_l = Number of links in the PPI network of disease B

The same formula shown above was also used to calculate the matching link ratio for common links of path length 2 and path length 3. Links of specific path lengths were retrieved using the Cytoscape application called NetworkAnalyzer [31, 32].

2.6 Calculation of comorbid associations

Relative risk (RR$_{AB}$) measures comorbidity by comparing the observed prevalence of a pair of comorbid diseases (A and B) in the population with the expected number, which is calculated based on the prevalence of the individual diseases A and B in the population.

$$RR_{AB} = \frac{N_{AB}N}{N_A N_B} \qquad (3)$$

N_A = Total number of patients diagnosed with disease A

N_B = Total number of patients diagnosed with disease B

N_{AB} = Total number of patients diagnosed with both disease A and disease B

N = Total number of patients in the population

For the calculation of relative risks of disease pairs, we downloaded the HuDiNe dataset (http://sbi.upf.edu/data/hudine/) containing processed hospital claims data of 13,039,018 U.S. individuals who had applied for support from the U.S. Medicare program during 1990-1993 [33]. Comorbidity data was available for five out of our six comorbid disease pairs and two out of the three non-comorbid pairs in HuDiNe. Specifically, data was not available for Anxiety – Depression and Multiple sclerosis – Peroxisomal disorders. Hence, N$_A$, N$_B$ and N$_{AB}$ were extracted for seven out of the nine disease pairs. The diseases were specified in the form of their ICD-9 codes (at three digits level): asthma (ICD-9: 493), hypertension (ICD-9: 401), type 2 diabetes (ICD-9: 250), obesity (ICD-

9: 278), chronic obstructive pulmonary disease (ICD-9: 496), heart failure (ICD-9: 428), Parkinson's disease (ICD-9: 332), schizophrenia (ICD-9: 295), rheumatoid arthritis (ICD-9: 714) and osteoporosis (ICD-9: 733). The population size N was considered to be 13,039,018, i.e. the total number of individuals represented in the HuDiNe dataset.

2.7 Pathway enrichment analysis

WebGestalt [34] was used to compute the distribution of genes involved in specific signalling pathways in the drug target networks, and compare it with the background distribution of genes belonging to this pathway among all the genes associated with any pathway in the selected database (Reactome) [35]. Statistical significance of the enrichment was computed using Fisher's exact test and corrected using the Benjamini-Hochberg method for multiple test adjustment.

2.8 Gene expression enrichment analysis

The enrichment of the drug target networks in genes expressed in specific tissues was computed using RNA-sequencing data from 53 postnatal human tissues extracted from GTEx [36] (version 8). Genes with high or medium expression (transcripts per million (TPM) ≥ 9) in 53 tissues were included, provided that they were not housekeeping genes, i.e. genes detected in all the tissues with transcripts per million ≥ 1, as identified in the Human Protein Atlas [37]. TPM is a metric for quantifying gene expression; it directly measures the relative abundance of transcripts. The GMT files served as inputs for a gene over-representation analysis (GSEA) based on hypergeometric distribution. The following GWAS datasets were selected in TSEA-DB [38] for identification of disease-specific tissues (trait IDs are given in parentheses): anxiety (4679), depression (5315), chronic obstructive pulmonary disease (571), heart failure (5333), asthma (5259), hypertension (169), type 2 diabetes (4628), obesity (1031), Parkinson's disease (4607), schizophrenia (5215), rheumatoid arthritis (4614) and osteoporosis (746).

BaseSpace Correlation Engine (https://covid-19.ce.basespace.illumina.com/c/nextbio.nb) was used to identify the correlations between the gene expression profile induced by maprotiline in PC3 cells (Broad Connectivity Map (CMAP 2.0) [39]), the expression profile associated with major depressive disorder and generalized anxiety disorder (GSE98793 [40]) and the expression profile of adrenal cortex. The software uses a non-parametric rank-based approach to compute the extent of enrichment of a particular set of genes (or 'bioset') in another set of genes [41].

2.9 Principal component analysis

Principal component analysis (PCA) was used to capture relationships between the drug target networks and the disease networks/biological pathways/tissues. For each disease pair, negative log-

transformed p-values indicating the statistical enrichment of the disease networks/biological pathways/tissues in the 4 drug target networks were assembled into a data matrix containing disease networks/biological pathways/tissues as rows and drug target networks as columns; each cell in the matrix contained a $-\log_{10}P$ value. Following the established approach [42], log transformation was performed to reduce the influence of extreme values on the obtained PCs. PCA was performed with a web-based tool called ClustVis (https://biit.cs.ut.ee/clustvis/) [43]. The data matrix was pre-processed such that 70% missing values were allowed across the rows and columns. The $-\log_{10}P$ values in the matrix were further centred using the unit variance scaling method, in which the values are divided by standard deviation so that each row or column has a variance of one; this ensures that they assume equal importance while finding the components. The method called singular value decomposition (SVD) with imputation was used to extract principal components. In this method, missing values are predicted and iteratively filled using neighbouring values during SVD computation, until the estimates of missing values converge. The factor/component loadings corresponding to the disease networks/pathways/tissues that contributed to the selected principal components were also extracted. Component loadings are correlation coefficients between the variables in rows and the factors (i.e. PC1, PC2 etc.). The squared value of a component loading gives the percentage of the variance explained by a particular original variable, and essentially its contribution to the principal components. Finally, for each of the disease pairs, the Euclidean distance between the principal component scores of each of the drug target networks were computed for all the component loading values pertaining to the particular biological modality. This resulted in a list of the specific disease protein sets/pathways/tissues that may be closely related to each of the different drug target networks.

2. Results

To identify potential mechanisms of adverse drug interactions within comorbid diseases, we systematically studied pairs of comorbid diseases ('disease A' and 'disease B') and their FDA-approved drugs. We separated the drugs into two groups, namely, disease A drugs that are (a) contraindicated and (b) not contraindicated in disease B, and disease B drugs that are (c) contraindicated and (d) not contraindicated in disease A We then constructed the interactomes of the proteins targeted by these drugs and examined these drug target interactomes in the context of three biological factors, namely, (i) proteins exclusive to interactomes of diseases A and B and those that are in their intersection, and (ii) biological pathways and (iii) tissues associated with these drug target interactomes.

Specifically, we selected three pairs of non-comorbid diseases as negative controls and six pairs of comorbid diseases for our analysis. The non-comorbid pairs were: (I) Multiple sclerosis – Peroxisomal disorders [44], (II) Schizophrenia – Rheumatoid arthritis [45-47], (III) Asthma –

Schizophrenia [48]. The comorbid pairs were (IV) Anxiety – Depression [49], (V) Asthma – Hypertension [50, 51], (VI) Chronic obstructive pulmonary disorder (COPD) – Heart failure [52,53], (VII) Type 2 diabetes – Obesity [54, 55], (VIII) Rheumatoid arthritis – Osteoporosis [56] and (IX) Parkinson's disease – Schizophrenia [57].

The drugs indicated for use in each of the diseases were retrieved from Drug Bank (version 5.1.8) [23]. For each pair, we categorized the drugs into the four groups (a-d) mentioned earlier, based on their clinical activity in the diseases, collected from the TWOSIDES database (version 0.1) [24], a compendium of drugs and their contraindications (see **Additional File 2: Table S2**). Drugs contributing to specific adverse effects were collected by manually selecting relevant 'condition concept names' (**Additional File 1: Table S1**). For example, to identify the anxiolytic drugs that maycause depression, the condition concept names, *depression*, *major depression*, *depressive symptom*, *depression suicidal*, *depression postoperative*, *postpartum depression*, *depressive delusion,* and *agitated depression*, were selected. The list of anxiolytic drugs was then compared with the list of drugs associated with these condition concept names. The matching drugs were compiled into groups 'a' and 'c', for example, "drugs effective in anxiety and contraindicated in depression". Similarly, groups 'b' and 'd' drugs were compiled. The proteins targeted by the drugs belonging to groups a andb were retrieved by querying the Drug Bank database through the DGIdb drug-genee interaction database) web portal [25] (see **Additional File 3: Table S3**). Finally, the protein-protein interaction (PPI) networks of the drug targets were assembled by extracting their protein interactors from the PPI repositories BioGRID [27] (version 4.3.194) and HPRD [26] (version 9) using a Cytoscape plugin called BisoGenet [28].

The methodology of our study is illustrated in **Fig. 1**. To characterize the 4 classes of drug target networks (DTNs), we examined 3 types of data that may reflect their biological profiles, namely (i) disease PPI networks, (ii) biological pathways and (iii,) tissue gene expression. Specifically, we conducted gene overrepresentation analyses based on hypergeometric distribution to check the enrichment of the DTNs among proteins that are unique to/shared between networks of disease A and disease B, genes showing high/moderate expression in 53 tissues across the human body, and proteinsinvolved in ~1000 biological pathways. Overlaps computed in this manner with each of the 3 types of biological data were considered to be statistically significant at p-value < 0.05 after multiple test adjustments with the Benjamini-Hochberg method.

As a first step towards identifying the specific biological data modalities (disease subnetworks/pathways/tissues) that were relatively more 'closer' to each of the different types of DTNs in terms of Euclidean distance, we generated a data matrix of the DTNs (columns) versus the
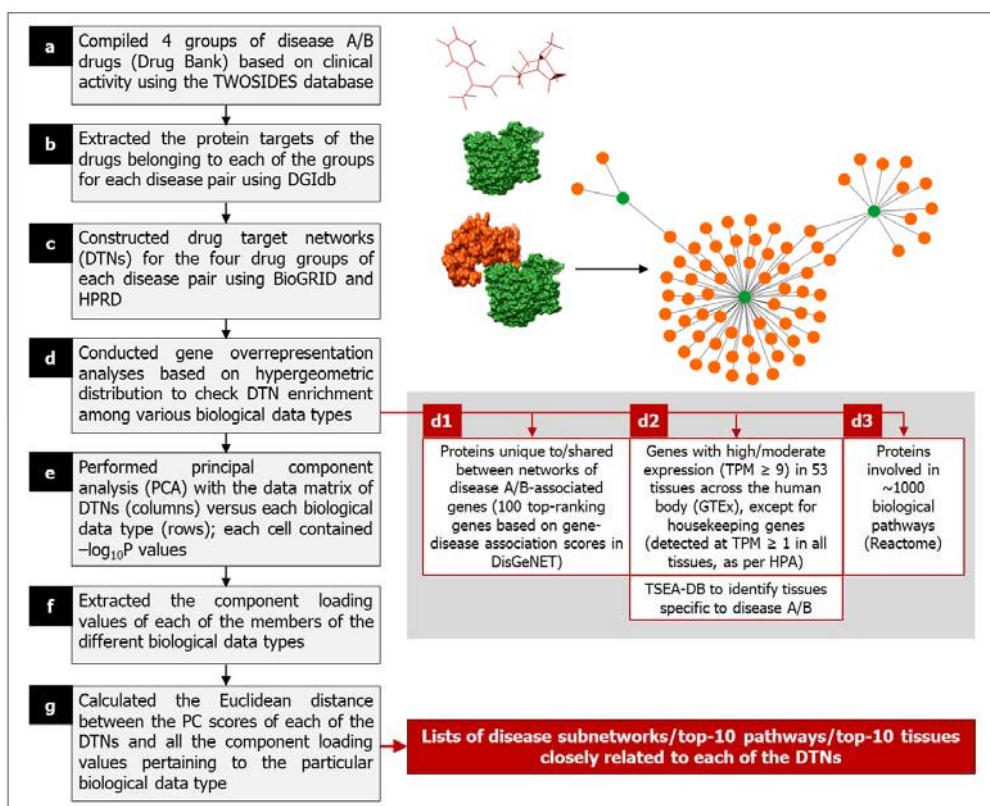
Figure 1: Framework for characterizing the drugs that target comorbid disease pairs. Our methodology to characterize drug target networks (DTNs) contained seven steps: (a) Retrieval of the drugs indicated for use against each of the diseases using Drug Bank and their categorization into four groups based on their clinical activity in the comorbid diseases, namely, disease A drugs not contraindicated in disease B, disease B drugs not contraindicated in disease A, disease A drugs contraindicated in disease B and disease B drugs contraindicated in disease A. (b) Identification of the proteins collectively targeted by the drugs in each of the groups by querying Drug Bank through DGIdb. (c) Construction of DTNs using the protein targets as input nodes and assembling their immediate neighbors in the human protein-protein interaction network up to a distance of 1, based on data from the PPI repositories BioGRID and HPRD. (d) Performing gene enrichment analysis with the four DTNs (corresponding to each of the disease pairs) in 3 biological data types: (d1) disease protein-protein interaction networks, (d2) tissue gene expression and (d3) biological pathways. (e) Generation of a data matrix containing the enriched disease protein sets/tissues/pathways as rows, DTNs as columns and log-transformed p-values in each of the cells, and using the matrix as an input for principal component analysis. (f) Extraction of component loading values of each of the enriched disease protein sets/tissues/pathways that correspond to each of the principal components. (g) Calculation of the Euclidean distance between the principal component scores of each of the DTNs and the component loading values of the disease protein sets/tissues/pathways. These steps resulted in the identification of the top disease protein sets, tissues and pathways that were closely associated with each of the DTNs. Databases: BioGRID (Biological General Repository for Interaction Datasets), DGIdb (Drug Gene Interaction database), DisGeNET (Disease Gene association NETwork), Drug Bank, GTEx (Genotype-Tissue Expression), HPRD (Human Protein Reference Database), Reactome, TSEA-DB (Tissue-Specific Enrichment Analysis DataBase) and TWOSIDES. Abbreviations: DTN – Drug Target Network, PCA – Principal Component Analysis and TPM – Transcripts Per Million.

various members of the biological data modality (rows) (for example, for the data modality 'disease subnetwork', the members would be 'common to both the networks', 'unique to disease A network'

and 'unique to disease B network' and for the data modality 'tissue', the members would be 'amygdala', 'aorta', 'lungs' etc.). Each cell contained the negative of log-transformed p-values. $-\log_{10}$ transformed p-values have been used as inputs for PCA in previous studies [58, 59]. Following the established approach [42], log transformation was performed to reduce the influence of extreme values on the obtained PCs. Single value decomposition (SVD) with imputation and unit variance scaling was applied to this matrix to extract principal components that explained the variance observed with each of the data modalities across the DTNs. Principal component analysis (PCA) has been applied to matrices containing gene-level association scores in several studies [59]. PCA is primarily used to capture systematic variations underlying datasets. All the principal components generated after this analysis were considered for our study, since they may together reveal underlying clustering patterns among the different DTNs. Following this, we extracted the component loading values of each of the members of the different data modalities, which correspond to each of the principal components representing the relationships among the DTNs. Component loadings are values depicting the correlation of the original variables in our data matrix — negative log of p-values of enrichment for specific disease subnetworks/pathways/tissues — with each of the extracted principal components. Lastly, we calculated the Euclidean distance between the principal component scores of each of the DTNs specifically in the context of each data modality and all the corresponding component loading values. This yielded a list of the specific disease subnetworks/pathways/tissues that are presumably closely related to each of the different DTNs.

3.1 Disease network similarity and comorbid associations

Relative risk is an experiential measure of comorbidity as it compares the observed prevalence of a pair of comorbid diseases in the population with the expected number, which is calculated based on the prevalence of the individual diseases in the population. We then explored whether this information was embedded in the disease networks, i.e., whether the relative risk of comorbidity of the disease pairs would be reflected in the similarity of the disease networks. For each of the comorbid pairs, we computed four established network similarity measures, namely, matching node ratio (N_M) for all the nodes shared between the two disease networks, and the matching link ratio (L_M) [30] for all the (i) shared links (i.e. edges), (ii) shared links of path length 2 (connecting two nodes via one intermediate node) and (iii) shared links of path length 3 (connecting two nodes via two intermediate nodes) between the two disease networks.

We computed the relative risk for each of the disease pairs observed in hospital claims data of 13,039,018 U.S. individuals who had filed for support from the Medicare program during the period of 1990-1993, made available as the HuDiNe dataset [33]. The ICD-9 codes corresponding to pairs of

diseases diagnosed as primary and secondary conditions, along with the number of individuals who were diagnosed with diseases A or B or both (N_A, N_B and N_{AB}, respectively) were available (see **Methods**). Comorbidity data was available for five out of our six comorbid disease pairs (i.e. exceptfor Anxiety – Depression) and two out of the three non-comorbid pairs (i.e. except for Multiple sclerosis – Peroxisomal disorders in HuDiNe.
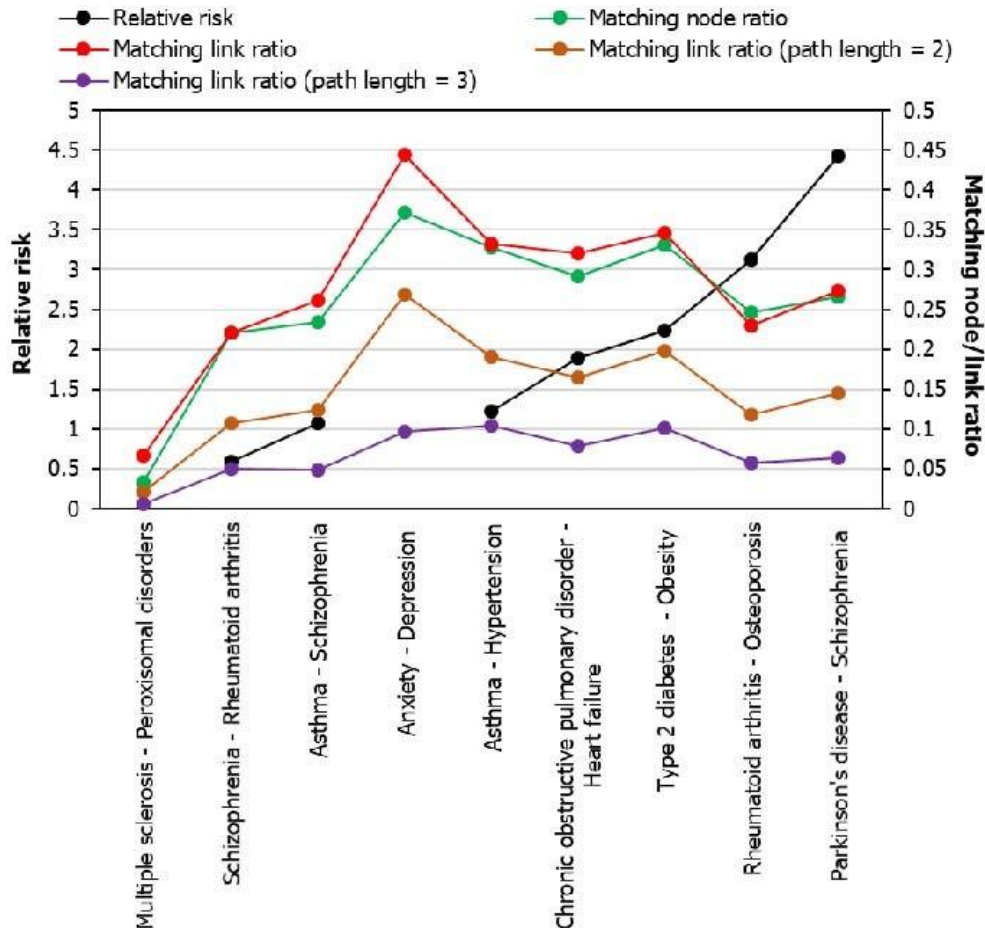
Figure 2: Comparison of disease network similarity measures and comorbid associations.
The graph shows the relationship between relative risk (black data points) andfour measures of network similarity, namely, matching node ratio (green data points), matching link ratio of all shared edges (red data points), matching link ratio of all shared

For each of the diseases considered, the top 100 genes associated with the disease were curated from the DisGeNET database (version 7) [29] based on their gene-disease association (GDA) scores (see **Additional File 4: Table S4**). The GDA score ranges from 0 to 1 and is computed for a gene based on the number of publications supporting its association with the disease, and the number and types of database sources (levels of curation (expert-curated/computationally-predicted) and the model

268

organisms in which the association was validated). The 100 top-ranking genes collected in this manner were used as starting points for the construction of disease networks. Here also, the network isassembled by extracting PPIs from BioGRID and HPRD using a Cytoscape plugin BisoGenet, similar to assembling of DTNs. Then, we systematically conducted network overlap analyses with each of the9 disease pairs and identified the proteins (a) shared between the two disease networks, (b) unique todisease A and (c) unique to disease B (**Table 1**).

Table 1: Overlap of the disease networks. The table shows the statistics of the overlaps shared between the two diseases in each of the nine disease pairs that were examined in our study.

Disease pair	# Proteins in disease A network	# Proteins in disease B network	# Shared proteins	p-value of overlap	Odds ratio of overlap	% Shared proteins in disease A network	% Shared proteins in disease B network
Multiple sclerosis (A) – Peroxisomal disorders (B)	2418	727	284	5.97E-70	2.9	12%	39%
Schizophrenia (A) – rheumatoid arthritis (B)	2662	2424	918	6.86E-208	2.56	34.5%	38%
Asthma (A) – Schizophrenia (B)	3041	2662	1084	1.36E-228	2.41	36%	41%
Anxiety (A) – Depression (B)	3342	3054	1732	1.86E-628	3.06	52%	57%
Asthma (A) – Hypertension (B)	3041	2515	1371	1.85E-500	3.23	45%	54.50%
Chronic obstructive pulmonary disease (A) –heart failure (B)	3736	2922	1505	3.12E-371	2.48	40%	51.50%
Type 2 diabetes (A) – Obesity (B)	2471	2490	1232	3.66E-503	3.6	50%	49%
Rheumatoid arthritis (A) – osteoporosis (B)	2424	3681	1206	1.30E-270	2.43	50%	33%
Parkinson's disease (A) – schizophrenia (B)	3200	2662	1232	2.88E-310	2.6	38.50%	46%

The relative risk between diseases was proportional to the matching node and link ratios (**Fig. 2**). The control disease pairs showed low relative risks and smaller disease network overlaps, whereas three out of five comorbid disease pairs showed high relative risks and larger network overlaps, namely, Asthma – Hypertension, COPD – Heart failure and Type 2 diabetes – Obesity. However, this trend was not seen in the comorbid pairs, Rheumatoid arthritis – Osteoporosis and Parkinson's disease – Schizophrenia. Specifically, their higher relative risks (compared with other comorbid pairs), were not

accompanied by a corresponding increase in the network overlaps. ~85% of the human interactome awaits experimental discovery [60]. Hence, two factors may have led to the underestimation of the network overlaps. Firstly, the inherent incompleteness of these disease networks [60]. Secondly, the tendency of incomplete networks to exhibit small overlaps [60].

3.2 Druggability of disease networks

Next, we tested the potential of each of the disease subnetworks to be acted upon by drugs or their susceptibility to pharmacological modulation (druggability), by examining their enrichment among a group of 4,463 proteins deemed to be druggable [61], similar to the approach followed in a previous study [62]. These proteins are bound with high affinity at specific binding sites by drugs that follow the Lipinski's 'rule-of-five', i.e. orally bioavailable drugs with specific molecular characteristics that influence their pharmacokinetic ability to enter systemic circulation and act on their target sites (**Table 2**) [63].

Table 2: Overlaps of the disease protein sets with druggable targets. $-\log_{10}P$ values computed for each of the nine tested disease pairs using a hypergeometric test. The $-\log_{10}P$ values indicate the statistical significance of the overlaps shared by each of the disease protein sets (top column headings) with a group of 4463 druggable proteins. *, ** and *** indicate low, medium and high levels of statistical significance. , and indicate non-significant overrepresentation, non-significant underrepresentation and significant underrepresentation respectively.

Disease pairs	Common to both the networks	Unique to disease A network	Unique to disease B network
Multiple sclerosis (**A**) – peroxisomal disorders (**B**)	7.38**	19.52***	2.09*
Schizophrenia (A) – Rheumatoid arthritis	13.26**	14.36***	2.4*
Asthma (**A**) – schizophrenia (**B**)	19.18***	9.41**	0.89
Anxiety (**A**) – Depression (**B**)	5.57**	0.001	12.05***
Asthma (**A**) – Hypertension (**B**)	31.34***	3.19*	9.59**
Chronic obstructive pulmonary disease (**A**) – heart failure (**B**)	34.73***	1.06	9.16**
Type 2 diabetes (**A**) – Obesity (**B**)	18.65***	1.47*	7.05**
Rheumatoid arthritis (**A**) – Osteoporosis (**B**)	21.96***	7.17**	1.97*
Parkinson's disease (**A**) – Schizophrenia (**B**)	19.93***	0.27	0.3

We found that the proteins shared between the two diseases were the most significantly enriched for druggable targets in 5 out of the 6 tested comorbid pairs (**Table 2**). In case of the sixth pair, namely

anxiety and depression, the proteins that are exclusive to the depression network were found to be more enriched for druggable targets. In 2 out of the 5 disease pairs that shared many common drug targets, the drug target proteins were significantly enriched in G protein-coupled receptor activity (p-value <0.05) (**Table 3**).

Table 3: Enrichment of Gene Ontology molecular functions among druggable targets and proteins unique to the depression network. The odds ratio of enrichment of two specific Gene Ontology molecular functions among druggable proteins, proteins unique to the depression network and proteins common to the anxiety and depression networks have been shown. p-values indicating the statistical significance of these enrichments have been shown in parentheses. Note that druggable proteins show higher enrichment for transmitter-gated channel activity compared to G protein-coupled peptide receptor activity, in terms of odds ratio of enrichment. The overrepresentation of a more druggable class (glutamate-gated Ca^{2+} channel activity) among proteins unique to the depression network (and not among the common proteins) would have altered the enrichment pattern for anxiety and depression in comparison with other the disease pairs (as shown in Table 2).

Protein sets / Gene Ontology Molecular Function	4463 druggable proteins – odds ratio (p-value)	Proteins common to anxiety and depression networks – odds ratio (p-value)	Proteins unique to depression network – odds ratio (p-value)
Transmitter-gated channel activity	**4.2 (< 1E-15)**	**8.65 (0.037)**	-
G protein-coupled peptide receptor activity	3.8 (< 1E-15)	3.6 (6.6E-03)	**10.46 (9.3E-07)**

Based on these observations and the finding in the previous section that relative risk varies in tandem with network similarity measures, we speculated that contraindications in comorbidities may arise from drug action on druggable proteins shared between the networks of comorbid diseases (**Table 2**). This led to two corollaries: (i) the target networks of the group 'a' and 'c' drugs (effective in disease A and contraindicated in disease B or vice versa) may show the highest enrichment for the proteins/pathways/tissues shared between the two disease networks and (ii) the target networks of the groups 'b' and 'd' drugs (effective in disease A and *not* contraindicated in disease B or vice versa) may show the highest enrichment for proteins/pathways/tissues unique to disease A (or B respectively).

3.3 Disease networks and drug target networks

To test these corollaries, we systematically computed the overlaps between three groups of disease proteins, namely, proteins that are (a) common to disease A and disease B networks, (b) unique to disease A network and (c) unique to disease B network, and four classes of DTNs, namely, target networks of drugs effective in disease A and (a) contraindicated and (b) *not* contraindicated in diseaseB, and target networks of drugs effective in disease B (c) contraindicated and (d) *not* contraindicated in disease A (**Table 4**); previous studies have examined the overlaps between the PPI networks of drug targets and disease-associated proteins [20, 64]. For each of the six disease pairs, we created a data matrix of DTNs (columns) versus disease subnetworks (rows), which contains -log(p-values) indicating the statistical significance of these enrichments. This data matrix was used as the input for PCA. In order to identify the specific disease subnetworks that were the nearest to each of the DTNs, we calculated the Euclidean distance between the PC scores of each of the DTNs across all the extracted axes and the corresponding component loading values of all the disease subnetworks across these axes (following the methodology depicted in **Fig. 1**). By counting the two disease subnetworks that were the closest to each of the different DTNs, we identified two predominant patterns.

In 10 out of the 12 cases, the DTNs of drugs used for a specific disease and not contraindicated in a comorbid condition were found to be closest/second closest to the proteins uniquely found in the network of the comorbid condition. Additionally, in 9 out of the 12 cases, they were closest/second closest to the proteins shared between the networks of both the diseases. In contrast, the DTNs of drugs used for a specific disease and contraindicated in a comorbid condition were found to be closest/second closest to the proteins uniquely found in the network of the disease for which these drugs were primarily used in 8 out of the 12 cases. Additionally, in 9 out of the 12 cases, they were closest/second closest to the proteins shared between the networks of both the diseases.

These observations led us to speculate two scenarios. Firstly, disease A drugs that are not contraindicated in disease B may target proteins unique to the disease B subnetwork involved in mechanisms that are either inconsequential/beneficial for disease B, but whose modulation is certainly beneficial for the treatment of disease A. Alternatively, they may target common mechanisms that are dysregulated in a similar manner in both the diseases and pharmacologically modulate them in a similar direction. Secondly, disease A drugs may become contraindicated in disease B when they target either (a) common mechanisms that are pharmacologically oppositely modulated in a manner that benefits disease A but aggravates disease B or (b) mechanisms unique to disease A that aggravate disease B. Additionally, we hypothesized that biological processes such as signalling pathways that function at a higher level than disease subnetworks could be regulating the action of drugs under comorbid conditions.

272

Table 4: Overlaps of the disease protein sets with the four classes of drug target networks. –log$_{10}$P values computed for each of the nine tested disease pairs using a hypergeometric test. The –log$_{10}$P values indicate the statistical significance of the overlaps shared between each of the disease protein sets (top column headings) and the target networks of the four classes of drugs (row headings). *, ** and *** indicate low, medium and high levels of statistical significance. , and indicate non-significant overrepresentation, non-significant underrepresentation and significant underrepresentation respectively.

MULTIPLE SCLEROSIS (A) AND PEROXISOMAL DISORDERS (B)	Common to multiple sclerosis (A) and peroxisomal disorders (B) networks	Unique to multiple sclerosis network (A)	Unique to peroxisomal disorders network (B)
DTN of drugs effective in multiple sclerosis (A)	113.15**	247.78***	24.52*
DTN of drugs effective in peroxisomal disorders (B)	19.56***	9.78**	1.6*
RHEUMATOID ARTHRITIS (A) AND SCHIZOPHRENIA (B)	Common to rheumatoid arthritis (A) and schizophrenia (B) networks	Unique to rheumatoid arthritis network (A)	Unique to schizophrenia network (B)
DTN of drugs effective in rheumatoid arthritis (A) and not contraindicated in schizophrenia (B)	202.91***	76.56*	82.54**
DTN of drugs effective in schizophrenia (B) and not contraindicated in rheumatoid arthritis (A)	198.8***	17.71*	120.54**
DTN of drugs effective in rheumatoid arthritis (A) and contraindicated in schizophrenia (B)	235.72***	157.11**	119.54*
DTN of drugs effective in schizophrenia (B) and contraindicated in rheumatoid arthritis (A)	257.06***	48.54*	207.08**
SCHIZOPHRENIA (A) AND ASTHMA (B)	Common to asthma (B) and schizophrenia (A) networks	Unique to asthma network (B)	Unique to schizophrenia network (A)
DTN of drugs effective in schizophrenia (A) and not contraindicated in asthma (B)	0.71**	0.75	1.5***
DTN of drugs effective in asthma (B) and not contraindicated in schizophrenia (A)	6.6***	0.99**	0.64*
DTN of drugs effective in schizophrenia (A) and contraindicated in asthma (B)	5.12**	0.32*	7.3***
DTN of drugs effective in asthma (B) and contraindicated in schizophrenia (A)	8.98***	3.2**	0.25
ANXIETY (A) AND DEPRESSION (B)	Common to anxiety (A) and depression (B) networks	Unique to anxiety network (A)	Unique to depression network (B)

DTN of drugs effective in anxiety (A) and not contradicated in depression (B)	42.51***	0.63	10.51*
DTN of drugs effective in depression (B) and not contraindicated in anxiety (A)	221.32***	0.32	24.98*
DTN of drugs effective in anxiety (A) and contraindicated in depression (B)	70.77***	1.03	20.96*
DTN of drugs effective in depression (B) and contraindicated in anxiety (A)	259.82***	29.51**	17.05*
ASTHMA (A) AND HYPERTENSION (B)	Common to asthma (A) and hypertension (B) networks	Unique to asthma network (A)	Unique to hypertension network (B)
DTN of drugs effective in asthma (A) and not contradicated in hypertension (B)	385***	21.29**	7.93*
DTN of drugs effective in hypertension (B) and not contraindicated in asthma (A)	423***	3.01	6.77*
DTN of drugs effective in asthma (A) and contraindicated in hypertension (B)	571***	30.17*	0.45
DTN of drugs effective in hypertension (B) and contraindicated in asthma (A)	351***	104.14**	58.71*
CHRONIC OBSTRUCTIVE PULMONARY DISEASE (A) AND HEART FAILURE (B)	Common to chronic obstructive pulmonary disease (A) and heart failure (B) networks	Unique to chronic obstructive pulmonary disease network (A)	Unique to heart failure network (B)
DTN of drugs effective in chronic obstructive pulmonary disease (A) and not contraindicated in heart failure (B)	279.3***	17.47*	43.59**
DTN of drugs effective in heart failure (B) and not contraindicated in chronic obstructive pulmonary disease (A)	8.03***	0.63	0.67
DTN of drugs effective in chronic obstructive pulmonary disease (A) and contraindicated in heart failure (B)	255.32***	33.88**	15.83*
DTN of drugs effective in heart failure (B) and contraindicated in chronic obstructive pulmonary disease (A)	314***	8.92*	55.19**
TYPE 2 DIABETES (A) AND OBESITY (B)	Common to type 2 diabetes (A) and obesity (B) networks	Unique to type 2 diabetes network (A)	Unique to obesity network (B)
DTN of drugs effective in diabetes (A) and not contraindicated in obesity (B)	140.81***	11.69*	1.16
DTN of drugs effective in obesity (B) and not contraindicated in diabetes (A)	18.56***	0.73	2.59*
DTN of drugs effective in diabetes	232.99***	27.39**	10.93*

(A) and contraindicated in obesity (B)			
DTN of drugs effective in obesity (B) and contraindicated in diabetes (A)	54.79***	0.32	21.27*
RHEUMATOID ARTHRITIS (A) AND OSTEOPOROSIS (B)	Common to rheumatoid arthritis (A) and osteoporosis (B) networks	Unique to rheumatoid arthritis network (A)	Unique to osteoporosis network (B)
DTN of drugs effective in rheumatoid arthritis (A) and not contradicated in osteoporosis (B)	219.6***	17.51*	30.9**
DTN of drugs effective in osteoporosis (B) and not contraindicated in rheumatoid arthritis (A)	908*	126.08	2118***
DTN of drugs effective in rheumatoid arthritis (A) and contraindicated in osteoporosis (B)	272.65***	16.29*	68.31**
DTN of drugs effective in osteoporosis (B) and contraindicated in rheumatoid arthritis (A)	255.5***	0.29	237.99*
PARKINSON'S DISEASE (A) AND SCHIZOPHRENIA (B)	Common to Parkinson's disease (A) and schizophrenia (B) networks	Unique to Parkinson's disease network (A)	Unique to schizophrenia network (B)
DTN of drugs effective in Parkinson's disease (A) and not contradicated in schizophrenia (B)	83.25***	15.5**	6.6*
DTN of drugs effective in schizophrenia (B) and not contraindicated in Parkinson's disease (A)	72.82***	15.97**	6.73*
DTN of drugs effective in Parkinson's disease (A) and contraindicated in schizophrenia (B)	25.51***	4.25**	4*
DTN of drugs effective in schizophrenia (B) and contraindicated in Parkinson's disease (A)	156.68***	103.63**	7.87*

3.4 Biological pathways and drug target networks

We identified the pathway associations of the DTNs using the gene set analysis toolkit called WebGestalt [34]. WebGestalt computes statistical significance enrichment of a functional group (e.g., a Reactome pathway) in an input gene list using Fisher's exact test using the Benjamini-Hochberg method for multiple test adjustment. For each of the 6 disease pairs, a data matrix of DTNs (columns) versus Reactome pathways (rows) containing corresponding enrichments was used as inputs for PCA, and the Euclidean distance between the PC scores of each of the DTNs across all the extracted axes and the corresponding component loading values of all the pathways across these axes were computed. For each of the disease pairs, we retrieved the top-10 pathways closest to each of the DTNs

out of all the pathways enriched in the DTNs (**Additional Files 5-10: Figures S1-S6)**. Confirming our earlier suspicions, we noted that disease A DTN classes without contraindications in disease B were nearest to pathways possibly underlying both the diseases or uniquely associated with B, which are similarly regulated, i.e. upregulated or downregulated together, in the two comorbid diseases. On the other hand, disease A DTN classes with contraindication in disease B were nearest to pathways underlying both the diseases or unique to disease A that are differentially regulated, i.e. upregulated in one disease and downregulated in the other or vice versa.
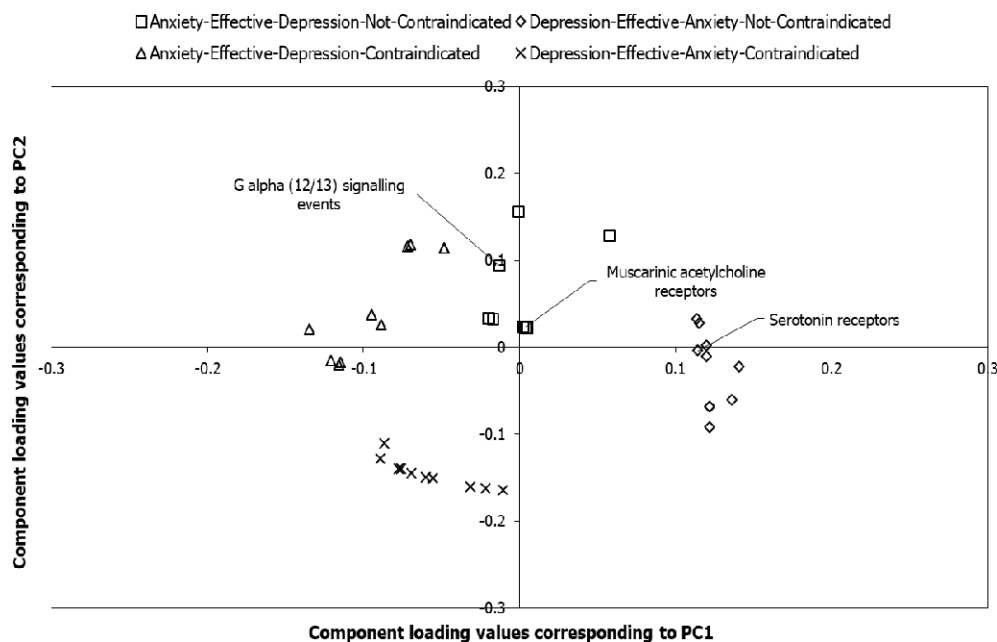
Figure 3: Pathways associated with the target networks of anxiety and depression drugs. The component loading values shown in the figure correspond to component scores of 4 drug target networks (DTNs) of anxiety and depression along PC1 and PC2, which explain 41.6% and 33.1% of the total variance respectively. The top-10 pathways that appeared to be highly related to each of the 4 DTNs, which were obtained after computing the Euclidean distance between the component loading values and the component scores, are shown as square-shaped data points for the DTN of drugs effective in anxiety and not contraindicated in depression, diamond-shaped data points for the DTN of drugs effective in depression and not contraindicated in anxiety, triangle-shaped data points for the DTN of drugs effective in anxiety and contraindicated in depression and cross-mark-shaped data points for the DTN of drugs effective in depression and contraindicated in anxiety. 'G α(12/13) signaling events' and 'muscarinic acetylcholine receptors' shown here are among the top-10 pathways associated with anti-anxiety drugs that are not contraindicated in depression. The drug maprotiline shown in **Fig. 4** corroborates this by showing antagonistic activity on adrenergic and muscarinic acetylcholine receptors. Similarly, serotonin receptors are associated with anti-depressants that are not contraindicated in anxiety; flupentixol and mirtazapine corroborate this by showing antagnostic activity on serotonin receptors.
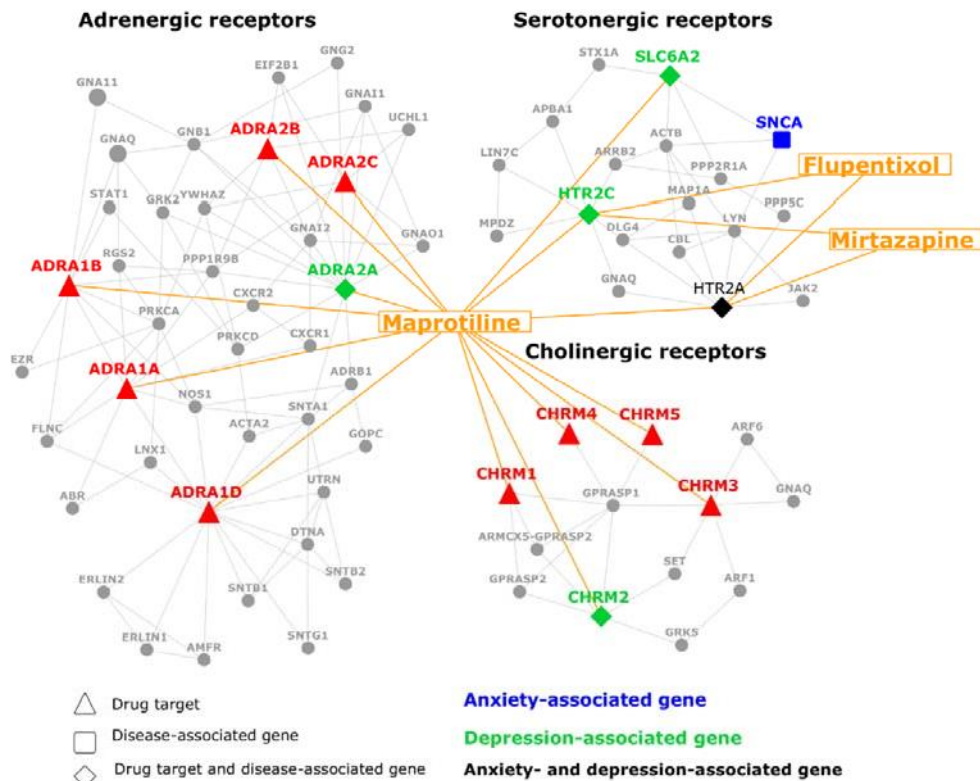
Figure 4: Network diagram showing the relationship between the targets of maprotiline, flupentixol and mirtazapine, and genes associated with anxiety and depression. The different families of receptors and transporter proteins targeted by maprotiline, flupentixol and mirtazapine and their interactions with the proteins encoded by anxiety (disease A) and/or depression (disease B) associated genes have been shown. Note that maprotiline (an anti-anxiety (disease A) drug not contraindicated in depression (disease B)) targets a higher number of proteins associated uniquely with depression in the adrenergic, serotonergic and cholinergic systems, which is in line with our observation that disease A drugs that are not contraindicated in disease B are closely associated with proteins uniquely found in the disease B network (i.e. depression in this specific example). Serotonin receptors were found to be associated in our analysis with depression drugs not contraindicated in anxiety; antagonistic activity on serotonin receptors is shown by two such drugs shown in the diagram (flupentixol and mirtazapine).

G alpha (12/13) signalling events' and 'muscarinic acetylcholine receptors' were identified among the top-10 pathways that were close to anxiety drugs not contraindicated in depression in our study (**Fig. 3**). Adrenergic receptor signalling could be regulated via G α(12/13); Gα12 and Gα13 have been shown to mediate alpha-1 adrenergic receptor-induced JNK activation in rat cardiomyocytes [65]. The drug maprotiline was among our list of anxiety drugs without contraindications in depression (**Fig. 4**). Corroborating this, clinical data suggested that the drug is effective in alleviating anxiety symptoms co-occurring with depression [66]. Maprotiline acts an inhibitor of SLC6A2 (sodium- dependent noradrenaline transporter) and inhibits noradrenaline reuptake in the brain. It also acts as an

277

antagonist to alpha-1 adrenergic receptors (ADRA1A, ADRA1B and ADRA1D) and alpha-2 adrenergic autoreceptors and heteroreceptors (ADRA2A, ADRA2B and ADRA2C), and enhances central noradrenergic and serotonergic functions, which have been linked to alleviation of anxiety and depression [67]. Maprotiline also acts as a weak antagonist to muscarinic acetylcholine receptors (CHRM1, CHRM2, CHRM3, CHRM4 and CHRM5); enhanced cholinergic signaling has been linkedto both anxiety and depression [68]. It is notable that maprotiline targets a higher number of proteins associated uniquely with depression (ADRA2A, HTR2C, SLC6A2 and CHRM2) in the adrenergic,
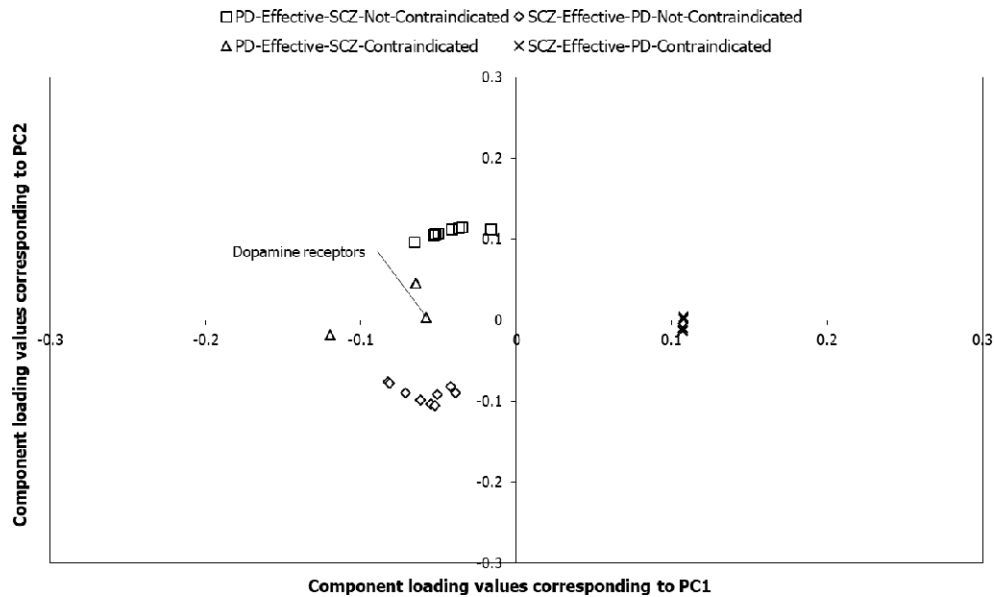
Figure 5: Pathways associated with the target networks of Parkinson's disease and schizophrenia drugs. The component loading values shown in the figure correspond to component scores of 4 drug target networks (DTNs) of Parkinson's disease (PD) and schizophrenia (SCZ) along PC1 and PC2, which explain 47.3% and 38.2% of the total variance respectively. The top-10 pathways that appeared to be highly related to each of the
4 DTNs, which were obtained after computing the Euclidean distance between the component loading values and the component scores, are shown as square-shaped data pointsfor the DTN of drugs effective in PD and not contraindicated in SCZ, diamond-shaped data points for the DTN of drugs effective in SCZ and not contraindicated in PD, triangle-shapeddata points for the DTN of drugs effective in PD and contraindicated in SCZ and cross-mark-shaped data points for the DTN of drugs effective in SCZ and contraindicated in PD. Dopamine receptors are among the top-10 pathways associated with PD drugs contraindicated in SCZ. Corroborating this, the drugs levodopa and ropinirole shown in **Fig. 6** stimulate dopaminergic receptors to alleviate Parkinsonian symptoms, but at the risk of inducing a hyperdopaminergic state conducive to the SCZ development.

serotonergic and cholinergic systems (**Fig. 4**). It targets only one receptor associated with both anxiety and depression (HTR2A), and no gene uniquely associated with anxiety (**Fig. 4**). These observations are in line with our findings with the overlap of DTNs with disease subnetworks, i.e. DTNs of disease A drugs that are not contraindicated in disease B (e.g. maprotiline) are closely associated with proteins uniquely found in the disease B network (i.e. depression in this specific example). Since

maprotiline has been discontinued from usage since 2020 in U.S. [69], note that we are citing this drug only as a demonstrative example. Drugs acting on the serotonergic system is known to be effective in both short-term and long-term treatment of patients with major depressive disorder and anxiety disorders [70]. 'Serotonin receptors' was identified among the top-10 pathways that were close to depression drugs not contraindicated in anxiety (**Fig. 3**). This may suggest the broad- spectrum efficacy of drugs acting on serotonin receptors in both the conditions. Two such drugs in ourstudy displayed antagonistic activity on serotonin receptors – flupentixol [71] and mirtazapine [72] (both acting on HTR2A and HTR2C) – and have been used to treat depression accompanied by anxiety symptoms (**Fig. 4**).
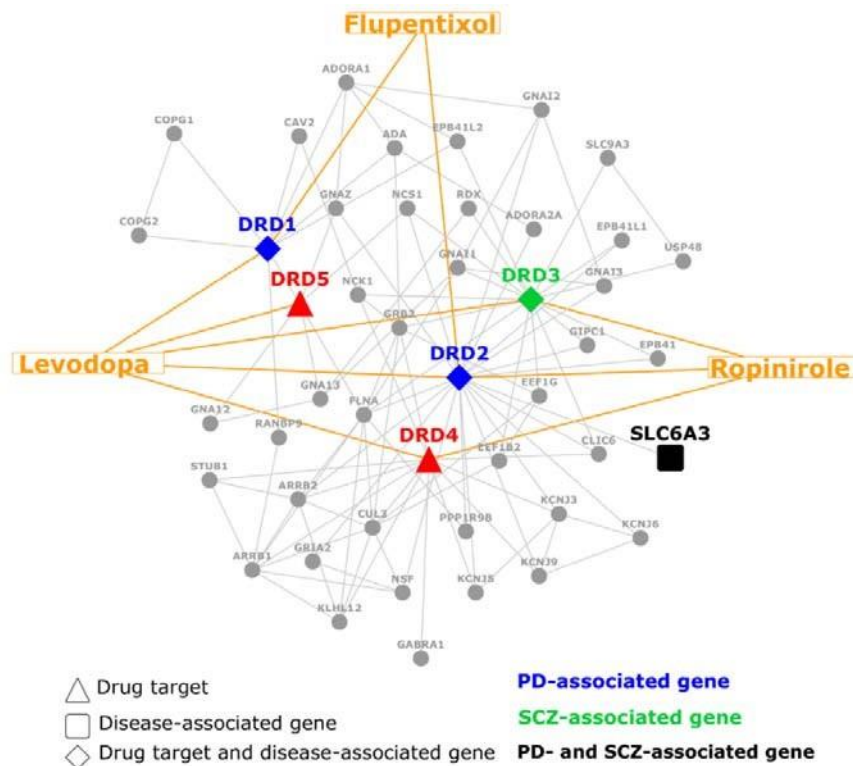
Figure 6: Network diagram showing the relationship between the targets of levodopa and ropinirole and genes associated with Parkinson's disease and schizophrenia. The specific dopamine receptors targeted by levodopa, ropinirole and flupentixol and their interactions with the proteins encoded by Parkinson's disease and/or schizophrenia associated genes have been shown. Note that levodopa and ropinirole are used in the treatment of Parkinson's disease (disease A), but contraindicated in schizophrenia(disease B), and flupentixol is used in the treatment of schizophrenia, but contraindicated in Parkinson's disease. Note that levodopa and ropinirole target a higher number of dopamine receptors associated uniquely with Parkinson's disease, which supports our finding that disease A drugs that are contraindicated in disease B are closely associated with proteins uniquely found in the disease A network (i.e. Parkinson's disease in this specific example).

The pathway 'dopamine receptors' was found to be close to PD drugs contraindicated in SCZ (**Fig. 5**), indicating that the enhancement in dopamine levels brought about by PD drugs may in fact induce

SCZ, which has been linked to a hyperdopaminergic state [57]. The dopamine agonists belonging to this group of PD drugs have been shown to induce psychosis, namely, levodopa (acting on DRD1, DRD2, DRD3, DRD4 and DRD5) and ropinirole (DRD2, DRD3 and DRD4) (**Fig. 6**) [73, 74]. It is notable that levodopa and ropinirole target a higher number of dopamine receptors associated uniquely with Parkinson's disease (DRD1 and DRD2) (**Fig. 6**). It targets only one dopamine receptor(DRD3) uniquely associated with schizophrenia (**Fig. 6**). These observations are in line with our findings with the overlap of DTNs with subnetworks, i.e. DTNs of disease A drugs that are contraindicated in disease B (e.g. levodopa and ropinirole) are closely associated with proteins uniquely found in the disease A network (i.e. Parkinson's disease in this specific example).

3.5 Tissues and drug target networks

Using RNA-sequencing data of 53 postnatal human tissues obtained from GTEx [36] (version 8), we attempted to identify whether the four DTN classes showed any tissue-specific patterns. Genes with high/medium expression (transcripts per million (TPM) ≥ 9) in these 53 tissues, which were not housekeeping genes (as per the Human Protein Atlas [37]), were considered. For DTNs of each disease pair, we computed the distribution of genes expressed in a specific tissue among the DTN genes and compared it with the background distribution of genes expressed in this tissue among all the genes that were assayed for expression in any of the 53 tissues. We generated a data matrix of DTNs (columns) versus tissues (rows) containing the negative of log-transformed p-values and performed PCA with this matrix as the input. We calculated the Euclidean distance between the PC scores of each of the DTNs and the component loading values of all the tissues. For each of the disease pairs, we retrieved the top-10 tissues that were nearest to the four DTNs (**Additional Files 11-16: Figures S7-S13**). Following this, we employed the tissue-specific enrichment analysis database (TSEA-DB)[38] to retrieve the top-3 tissues that may be preferentially affiliated with the diseases in each of the pairs. TSEA-DB is a reference database for information on disease-associated tissues, specifically, the tissues in GTEx that show significant enrichment of genes harbouring diseases- associated variants compiled from the GWAS catalog [38]. We checked whether the top-3 tissues associated with each of the diseases in a disease pair (according to TSEA-DB) appeared among the list of tissues identified to be nearest to each of the 4 DTNs pertaining to this disease pair in our analysis. Out of the 11 tissues identified to be closer to the target networks of drugs used for a specificdisease and not contraindicated in a comorbid condition, 6 were found to be associated with the comorbid condition as per TSEA-DB, whereas 3 were associated with the specific disease for which the drugs were used and 2 were associated with the disease as well as the comorbid condition.

Conversely, out of the 9 tissues identified to be closer to the target networks of drugs used for a specific disease and contraindicated in a comorbid condition, 5 were found to be associated with the

specific disease, whereas 3 were associated with the comorbid condition in which the drugs were contraindicated and one was associated with the disease as well as the comorbid condition.
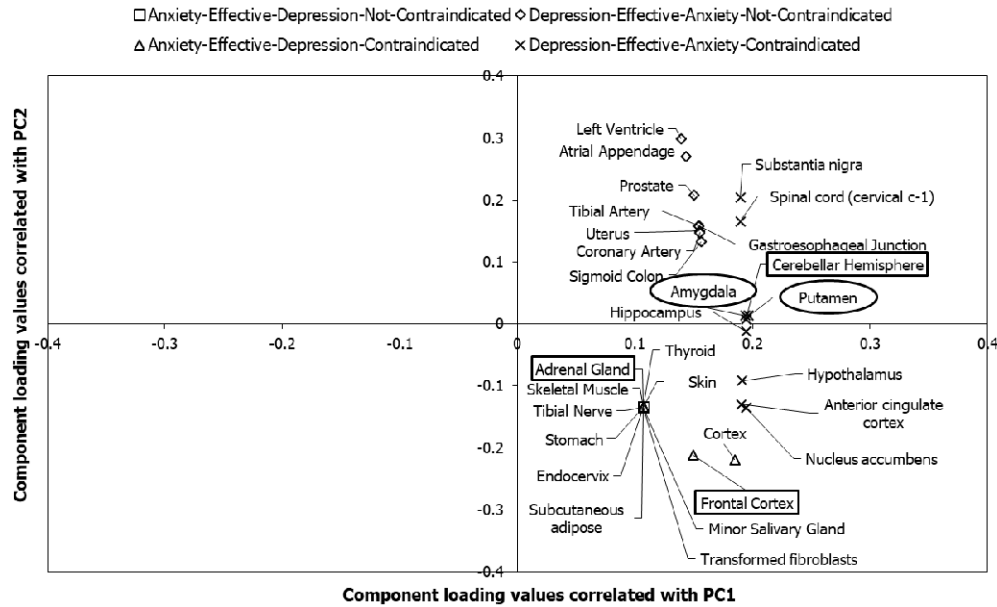
Figure 7: Tissues associated with the target networks of anxiety and depression drugs. The component loading values shown in the figure correspond to component scores of 4 DTNs of anxiety and depression along PC1 and PC2, which explain 90.7% and 6.5% of the total variance respectively. The tissues that were exclusively associated with each of the 4 DTNs among the top-ten tissues that were identified to be highly related to the DTNs, after computing the Euclidean distance between the component loading values and the componentscores, are shown as square-shaped data points for the DTN of drugs effective in anxiety andnot contraindicated in depression, diamond-shaped data points for the DTN of drugs effective in depression and not contraindicated in anxiety, triangle-shaped data points for theDTN of drugs effective in anxiety and contraindicated in depression and cross-mark-shaped data points for the DTN of drugs effective in depression and contraindicated in anxiety. Thetissues shown in circular and rectangular boxes were also identified to be highly specific to anxiety and depression respectively by TSEA-DB (due to a significant enrichment of anxiety/depression-associated variants). Note that adrenal cortex, which was identified to be associated with anti-anxiety (disease A) drugs that are not contraindicated in depression (disease B), is a tissue enriched with depression (i.e. disease B) associated variants. This corroborates our finding that disease A drugs that are not contraindicated in disease B are affiliated with disease B-specific tissues.

These percentages obtained with a low number of tissues suggest cautious interpretation. Nevertheless, these results seem to corroborate our previous findings with disease subnetworks and biological pathways. Specifically, the networks of disease A drugs that are not contraindicated for disease B seemed to be nearest to tissues preferentially affiliated with disease B. This could indicatethat these tissues could be equally important to the pathophysiology of disease A and its therapeuticalleviation (as they might be to these same aspects of disease B), despite showing a high enrichment for genes harbouring disease B-associated variants. For example, the adrenal gland was detected as a tissue highly specific to depression by TSEA-DB. In our analysis, this tissue appeared to be nearest to

the DTN of anxiety drugs that were not contraindicated in depression (**Fig. 7**), indicating that targeting of the adrenal gland may be vital to treat anxiety without aggravating comorbid depressive symptoms. The adrenal gland is an organ in the endocrine system that secretes the cortisol hormone, following the activation of the hypothalamic-pituitary-adrenal (HPA) axis by psychological stressors[75]. Several studies support the role of the adrenal gland as a focal point for depression. The adrenal gland exhibits a 70% increase in its volume in depressed individuals before successful anti-depressant treatment as well as in comparison with their matched controls [76, 77]. The cortisol hormone secreted by the adrenal gland, upon stress-induced activation of the HPA axis, has been linked to depressive symptoms in humans and monkeys. Increased cortisol levels have been positively correlated with depressive behaviour in rhesus macaques [78]. Enhanced cortisol secretion has been observed in depressive individuals [79], and has been proposed to (a) increase susceptibility to depression [80] and (b) be correlated with the stress experienced by depressed individuals [81].

Hyperactivation of the HPA axis has been noted in generalized anxiety disorder [82]. Treatment with selective serotonin reuptake inhibitors (SSRIs) has been shown to reduce HPA hyperactivity in both depressed patients and patients with generalized anxiety disorder [83-85]. Therefore, it is possible that anti-anxiety drugs that do not aggravate depressive symptoms target the adrenal gland, which produces the cortisol hormone, an effector or 'endpoint' of the HPA axis that seems to be regulated in a similar manner in depression as well as anxiety. We performed comparative transcriptome analysis of disease-associated, tissue-associated and drug-induced gene expression profiles using the BaseSpace Correlation Engine to analyse this hypothesis. BaseSpace Correlation Engine software suite is a data analysis platform that is used to study the effect of diseases and drugs on publicly available gene expression data [86].

As mentioned in the previous section, maprotiline was found among our list of anti-anxiety drugs that are not contraindicated in depression; clinical data supports its utility in the treatment of anxiety symptoms associated with depression [66]. The differential gene expression (DGE) profile induced by maprotiline (12.8 µM) in PC3 cells (Broad Connectivity Map (CMAP 2.0) [39]) was negatively correlated with the profile identified in the blood samples of patients with major depressive disorder patients (MDD) with generalized anxiety disorder (GAD) versus MDD patients without GAD (GSE98793 [40]) (**Fig. 8a**). This negative correlation of maprotiline with MDD/GAD could illustrate the fact that drugs administered to treat diseases often revert the expression of perturbed disease- associated genes to their normal levels [87, 88]. Secondly, the MDD/GAD profile was negatively correlated with the expression profile of adrenal gland cortex (**Fig. 8a**), indicating that this tissue could be critical to disease alleviation. Maprotiline-induced DGE profile was positively correlated with the profile of adrenal gland (**Fig. 8a**), indicating that maprotiline-mediated MDD/GAD alleviation may be dependent on adrenal gland, i.e. the reversal of MDD/GAD-associated expression

profile induced by maprotiline could occur in the adrenal cortex. We then asked whether the genes differentially expressed in each of these profiles converged on a common set of biological processes. Specifically, we identified the top-10 Gene Ontology (GO) biological processes enriched among the genes differentially expressed in (i) MDD/GAD versus maprotiline (in different directions), (ii) MDD/GAD versus adrenal cortex (in different directions) and (iii) maprotiline versus adrenal cortex (in the same direction). We then used the web-based tool called NaviGO [89] to group these 30 enriched biological processes into functionally cohesive networks based on semantic similarity measures of GO terms. Two such functional networks not only had top-scoring edges between the GOterms, but also contained GO terms enriched among all the three differential expression profiles (**Fig. 8b, c**). One network contained four GO terms associated with protein folding (**Fig. 8b**), and another network contained eleven GO terms representing cell cycle events (**Fig. 8c**). Interestingly, 'protein
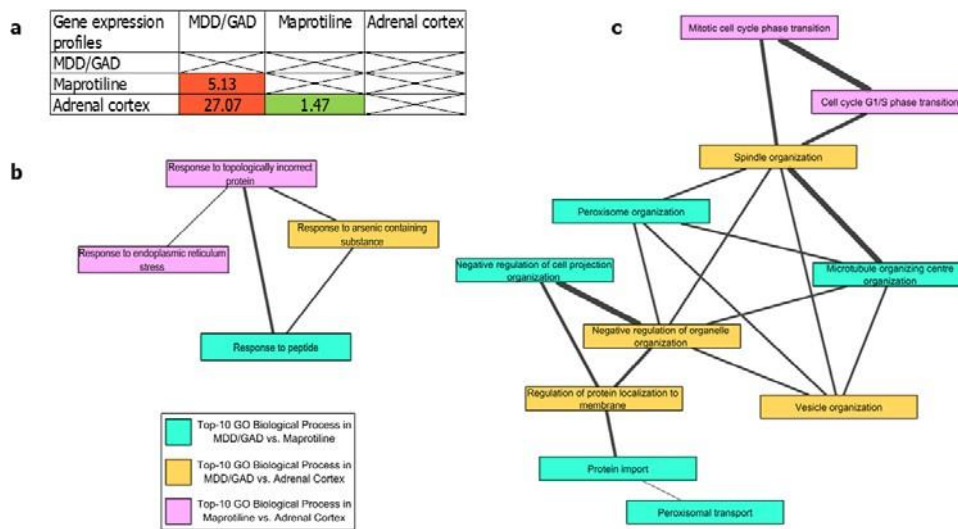
Figure 8: Relationship between MDD/GAD, maprotiline and adrenal cortex at transcriptomic and biological process levels. (a) Correlation of differential gene expression profiles associated with a comorbid condition (major depressive disorder and generalized anxiety disorder), a drug (maprotiline) anda tissue (adrenal cortex). $-\log_{10}$(p-values) indicating the overlap of the expression profiles have been shown; red and green colors indicate negative and positive correlations between the profiles respectively. Significant overlap was found among the genes that are upregulated in patients with both major depressive disorder (MDD) and generalized anxiety disorder (GAD) and downregulated on treating PC3 cells with maprotiline (p-value = 7.4E-06), among the genes that are upregulated in MDD/GAD patients and downregulated in adrenal cortex (p-value = 8.4E-28), and among the genes that are downregulated on treating PC3 cells with maprotiline and downregulated in adrenal cortex (p-value = 0.034). (b, c) The functional networks of the Gene Ontology biological processes related to (b) protein folding and (c) cell cycle events that were enriched in the three expression profiles. The GO terms associated with each of theexpression profiles have been shown using different node colors. The thickness of the edges corresponds to the Resnik semantic similarity score for GO terms (greater the thickness of the edges, greater is the similarity between the linked GO terms).

folding', 'cyclin D associated events in G1' and 'G1 phase' were independently retrieved among our top-10 Reactome pathways found to be nearest (in terms of Euclidean distance) to the DTN of anti-depressants that are not contraindicated in anxiety. Together, these results suggest that adrenal cortex

may be preferentially targeted by drugs such as maprotiline that produce beneficial effects in anxiety as well as in depression, and that their actions may converge on protein folding and cell cycle processes. Note that maprotiline has been discontinued from usage [69] and is only being cited here as a demonstrative example.

On the other hand, the networks of disease A that are contraindicated for disease B seemed to be nearest to tissues preferentially affiliated with disease A. This could indicate that these disease A-specific tissues may play a role in producing beneficial effects in disease A, while producing deleterious effects in disease B. For example, spleen was detected as a tissue highly specific to rheumatoid arthritis. The primary functions of this lymphoid organ are blood filtration, recycling of iron from old blood cells and generation of adaptive immune responses against bacterial, fungal and viral infections [90]. However, spleen has also been shown to act as a reservoir of osteoclast precursor cells, which upon resorption into bones, differentiate into osteoclasts.[91] Splenomegaly (enlargement of the spleen) has been noted in 5-10% and 52% of rheumatoid arthritis patients in separate studies (based on physical examination and imaging studies respectively) [92-94]. Rheumatoid arthritis patients are also prone to developing spontaneous splenic ruptures [95]. In our analysis, spleen was identified to be nearest (in terms of Euclidean distance) to the DTN of rheumatoid arthritis drugs that were contraindicated in osteoporosis (**Fig. 9**). This seemed to indicate that spleen mediated opposite effects in rheumatoid arthritis and osteoporosis. Anecdotal evidence seemed to support this conjecture. While splenectomy seemed to improve rheumatoid arthritis in a patient [96], it seemed to inhibit (a) attenuation of osteoporosis in a rat model [97] and (b) fracture healing in patients [98].

Table 5 summarizes the general conclusions of our study. We discovered that the DTNs of disease A drugs that are not contraindicated for a disease B may be nearest (in terms of Euclidean distance) to (a) proteins that are either uniquely found in the PPI network of disease B or shared between the PPI networks of disease A and disease B, (b) biological pathways that are associated with B or are commonly active in both the diseases, and are regulated in the same direction in both the diseases and (c) tissues showing a high enrichment of disease-B associated variants and thereby preferential affiliation with the etiology of disease B, while also being important to the pathophysiology and treatment of disease A (**Table 5**). On the other hand, disease A drugs that are contraindicated for a disease B may be nearest to (a) proteins that are either uniquely found in the PPI network of disease A or are shared between the PPI networks of disease A and disease B, (b) biological pathways that are associated with disease A or are commonly active in both the diseases, and are regulated in an

opposing manner in both the diseases and (c) tissues showing a high enrichment of disease A- associated variants and thereby preferential affiliation with the etiology of disease A, and mediating opposing effects in disease A and disease B (**Table 5**).
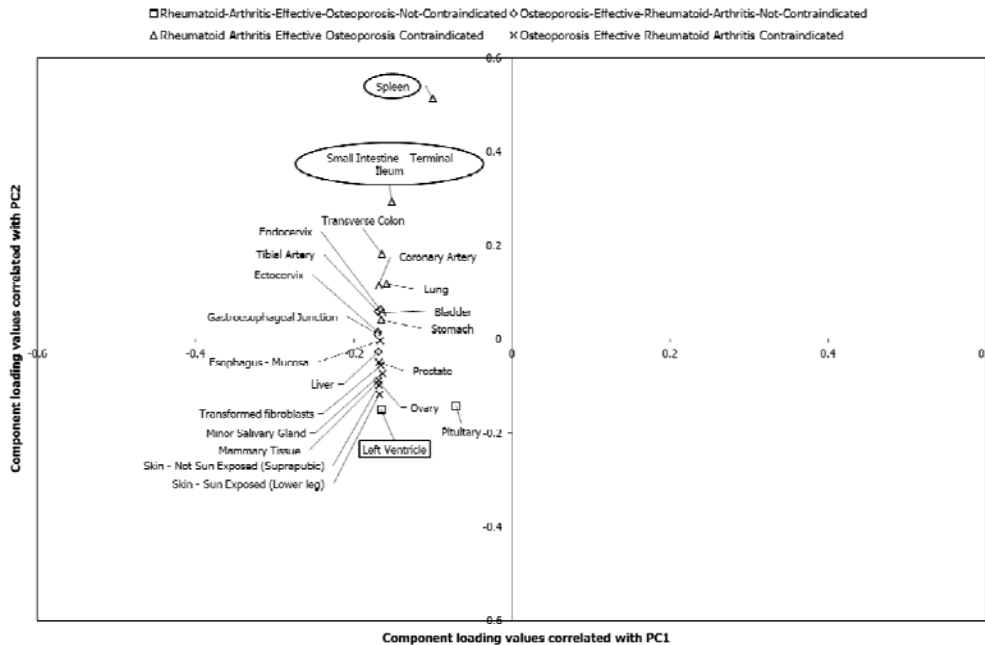
Figure 9: Tissues associated with the target networks of rheumatoid arthritis and osteoporosis drugs. Component loading values of 39 tissues associated with the drug target networks (DTNs) of rheumatoid arthritis (RA) and osteoporosis to PC1 and PC2 have been plotted along the X and Y axes respectively. PCA was performed with the p-values of enrichment of the tissues significantly associated (p-value < 0.05)with the DTNs of RA and osteoporosis. These values were transformed to $-\log_{10}P$ values, which were thenassembled into a data matrix containing tissues as rows and DTNs as columns. Unit variance scaling was applied across this matrix. Single value decomposition (SVD) with imputation was used to extract the principal components (PCs). The component loading values shown in the figure correspond to component scores of 4 DTNs along PC1 and PC2 that explain 89% and 6% of the total variance respectively. The tissues that were exclusively associated with each of the 4 DTNs among the top-ten tissues that were identified to be highly related to the DTNs, after computing the Euclidean distance between the componentloading values and the component scores, are shown as square-shaped data points for the DTN of drugs effective in RA and not contraindicated in osteoporosis, diamond-shaped data points for the DTN of drugseffective in osteoporosis and not contraindicated in RA, triangle-shaped data points for the DTN of drugs effective in RA and contraindicated in osteoporosis and cross-mark-shaped data points for the DTN of drugs effective in osteoporosis and contraindicated in RA. The tissues shown in circular and rectangular boxes were also identified to be highly specific to RA and osteoporosis respectively by TSEA-DB (due to a significant enrichment of RA/osteoporosis-associated variants). Note that spleen, which was identified to be associated with rheumatoid arthritis (disease A) drugs that are contraindicated in osteoporosis (disease B), is a tissue enriched with rheumatoid arthritis (i.e. disease A) associated variants. This corroborates our finding that disease A drugs that are contraindicated in disease B are affiliated with disease A-specific

Table 5: Disease network, pathway and tissue-level characterization of drugs that are contraindicated/not contraindicated in comorbid conditions. A ✓ has been used to indicate the

285

close affiliation of a specific category of drug target network with specific disease protein sets, disease-associated pathways and tissues.

Drug target networks	Disease PPI protein sets			Pathways			Tissues		
	Common to disease A and disease B	Unique to disease A	Unique to disease B	Common to disease A and disease B	Specific to disease A	Specific to disease B	Common to disease A and disease B	Specific to disease A	Specific to disease B
Disease A drugs <u>not</u> contraindicated in Disease B	✓		✓	✓		✓	✓		✓
Disease A drugs contraindicated in Disease B	✓	✓		✓	✓		✓	✓	

3. Discussion

Despite the increased prevalence of adverse drug reactions in comorbidities, knowledge on the mechanistic basis of drug contraindications in such conditions is limited. In our study, we attempted to characterize the biological profiles of the target networks of drugs used in specific diseases that areeither contraindicated or not contraindicated in a comorbid disease. We sought to provide an integrated interactome, pathway and tissue level view of the drug target networks.

The first key finding in our study was that the relative risk of comorbidity between diseases was proportional to their network similarity measures (**Fig. 2**). The four network similarity measures along with the relative risk were low in the case of our three negative control pairs, namely, Multiple sclerosis – Peroxisomal disorders, Schizophrenia – Rheumatoid arthritis, and Asthma – Schizophrenia. This confirmed that these were indeed non-comorbid pairs. The network similarity measures and relative risk were higher in the case of Anxiety – Depression, Asthma – Hypertension,Chronic obstructive pulmonary disorder – Heart failure, Type 2 diabetes – Obesity, Rheumatoid arthritis – Osteoporosis, and Parkinson's disease – Schizophrenia, confirming that they were comorbidities. However, these measures do not follow the same trend in the case of the comorbid pairs. The higher relative risks of Rheumatoid arthritis – Osteoporosis and Parkinson's disease – Schizophrenia (compared with the other comorbid pairs) were not accompanied by a corresponding increase in the network similarity measures. Several factors may explain these variations in our analysis. Firstly, it has been shown that relative risk overestimates the comorbid associations between rare diseases and underestimates the associations between highly prevalent diseases [43]. The number of cases in the HuDiNe database for Rheumatoid arthritis – Osteoporosis and Parkinson's disease – Schizophrenia are 24629 and 5439 respectively, which can be classified as rare occurrences when

286

compared with the other comorbid pairs. **Additional File 17: Figure S14** shows the relationship of the relative risks of the nine pairs of diseases with the individual prevalence of the diseases and the prevalence of the disease pairs as comorbidities. Secondly, the human interactome is a progressively developing network with ~85% remaining to be discovered. Therefore, the inherent incompleteness of the human PPI network, sampling biases introduced as a result of the selective discovery of PPIs, and the tendency of such incomplete networks to exhibit small overlaps [60] could have led to the underestimation of the network overlaps. Our second key finding was that druggable proteins were highly enriched among the proteins shared between the networks of two comorbid diseases (**Table 2**). Based on these results, we speculated that drug action on targets shared between the two diseases may give rise to contraindications in comorbidities. Interestingly, this hypothesis was only partially supported in our study.

The major finding in this respect was that the target network of the drugs used in the treatment of a specific disease A and contraindicated in a comorbid disease B showed preferential affiliation to proteins shared between the PPI networks of both the diseases or proteins uniquely found in the PPI network of the disease A, pathways shared by the two diseases or pathways associated with the disease A and tissues specifically associated with disease A (**Table 5**). As explained before, this was contrary to our hypothesis that these target networks would be preferentially affiliated with commonmechanisms underlying the two diseases. This hypothesis was based on the assumption that adverse events stem from drugs inducing opposite pharmacological effects in comorbid diseases by targeting effectors that are shared between the two diseases. However, our findings indicate that mechanisms underlying the pathology of disease A may contribute to contraindications in the comorbid disease B. Although further studies are required to examine the basis of this finding, it seems to indicate that the possibility of contraindications may be high when disease A drugs are highly specific to disease A in terms of the targeted PPI network, pathway and tissue. Instead, rational drug development should takeinto account the causative and correlational influences of the other comorbid conditions (disease B) that co-exist with disease A.

The target network of the drugs used in the treatment of a specific disease A and not contraindicatedin a comorbid disease B showed preferential affiliation to proteins shared between the PPI networks of both the diseases or proteins uniquely found in the PPI network of the comorbid disease B, pathways shared between the two diseases or pathways associated with the comorbid disease B and tissues specifically associated with the comorbid disease B (**Table 5**). This was contrary to our expectation that these target networks would be preferentially affiliated with biological modalities pertaining to disease A. This conjecture was based on the assumption that for a drug to be specificallyactive against a specific disease A without aggravating a comorbid disease B, it had to reverse the

phenotypes specifically associated with disease A. In this model, phenotypes of disease B were considered as 'off-targets' in line with the principles of conventional pharmacology, in which unintended effects of the drugs were attributed to interaction with pathways that may not be consequential to the pathology of disease A (i.e. pathways relevant to disease B) [13]. Our findings on the contrary indicate that the mechanisms underlying the pathology of the comorbid disease B may contribute to the therapeutic alleviation of disease A. Although further investigations may be necessary to dissect the basis for this observation, it is possible that an etiological association between the two diseases may cause their emergence or development to be interdependent. Specifically, future studies should concentrate on 3 etiological models of comorbidity [99], namely, the direct causation model, the associated risk factors model and the heterogeneity model. Disease B could be directly responsible for causing disease A in the 'disease causation model'. The comorbidity of disease A and disease B may arise from the correlation of the risk factors of disease B with the risk factors of disease A in the 'associated risk factors model'. On the other hand, comorbidity in the 'heterogeneity model' may arise not from the correlation of the risk factors associated with disease A and disease B, but from the capacity of the risk factors of disease A to cause disease B and vice versa. On applying the disease causation model to our findings, one may speculate that drugs targeting the proteins uniquely found in the disease B PPI network, and the pathways and tissues associated with disease B may alleviate disease A without aggravating disease B. The associated risk factors and heterogeneity models in this scenario would imply that the risk factors of disease B would influence the development of disease A directly, or through correlation with the risk factors of disease A. This model can be illustrated for genetic risk factors of disease B with the capacity to influence disease A. For example, the alterations in such genes would have led to pathway perturbations in specific tissues, which if counteracted by the drugs, may lead to alleviation of disease A.

Despite disease A drugs contraindicated in disease B and disease A drugs not contraindicated in disease B showing preferential affiliation with disease A and disease B respectively, it was clear, atleast in the case of the drug target and disease network analysis, that both these categories also showed affiliation with proteins shared between the two diseases (**Table 5**). This is in line with the speculation that both beneficial and adverse outcomes of drug treatment may arise from shared effectors and pathways, and that it may be difficult to delineate the separate mechanisms underlying the two outcomes [13]. Future analysis should focus on biological variables with the potential to differentially affect the functions of such shared proteins, specifically their cellular, pathway and tissue landscapes.

Our current approach has some limitations. Firstly, our study is based on 6 pairs of diseases that were selected based on literature survey. Ideally, future studies must be expanded to include all the known

pairs of comorbid disorders. Secondly, our analysis did not take the overlaps among the drug target networks into account; this would have allowed us to identify the network configurations of disease A – disease B – disease A drug not contraindicated in disease B – disease B drug not contraindicated in disease A. Secondly, although we were able to support our findings by citing evidence based on the known clinical activity of specific drugs, further investigations with the six comorbid disease pairs are essential to confirm the validity of our findings. These should focus on large-scale analysis of patient treatment data collected from observational studies and functional assays in animal models of human comorbidities.

In summary, our findings suggest that studies driven by biological modalities that influence comorbidities, such as disease PPI networks, pathways and tissue-specificity, are essential for rational drug development and minimization of adverse events. The results from our study have therapeutic applications and may directly benefit future assessments of drug contraindications in individuals with comorbidities.

4. Conclusions

We observed that the target networks of disease A drugs that were not contraindicated in disease B were mostly affiliated with the disease B network, and pathways and tissues associated with disease B. On the other hand, the target networks of disease A drugs that were contraindicated in disease B were affiliated with the disease A network, and pathways and tissues associated with disease A. This could indicate that etiological associations between the two diseases could play an active role in their therapeutic alleviation. In summary, our findings suggest that the enrichment patterns of drug target networks in pathways, tissues and the PPI networks of comorbid diseases will help identify drugs with/without contraindications in comorbidities.

3. Acknowledgments

Data availability: The condition concept names from the TWOSIDES database that were used to categorize the drugs associated with each of the comorbid disease pairs, the drug lists, the proteinstargeted by the drugs, and the genes associated with the comorbid disease pairs have been made available as **Table S1**, **Table S2**, **Table S3** and **Table S4**, respectively.

References

1.   Gadermann AM, Alonso J, Vilagut G, Zaslavsky AM, Kessler RC: **Comorbidity and diseaseburden in the National Comorbidity Survey Replication (NCS R).** *Depression and anxiety* 2012, **29:**797-806.

2.   Caughey GE, Ramsay EN, Vitry AI, Gilbert AL, Luszcz MA, Ryan P, Roughead EE: **Comorbid chronic diseases, discordant impact on mortality in older people: a 14-year longitudinal population study.** *Journal of Epidemiology & Community Health* 2010, **64:**1036-1042.

3.   Van den Akker M, Buntinx F, Metsemakers JF, Roos S, Knottnerus JA: **Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases.** *Journal of clinical epidemiology* 1998, **51:**367-375.

4.   Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L: **Prevalence of multimorbidity amongadults seen in family practice.** *The Annals of Family Medicine* 2005, **3:**223-228.

5.   Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F,Abdela J, Abdelalim A: **Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017.** *The Lancet* 2018, **392:**1859-1922.

6.   Morales DR, Lipworth BJ, Donnan PT, Jackson C, Guthrie B: **Respiratory effect of beta- blockers in people with asthma and cardiovascular disease: population-based nested case control study.** *BMC medicine* 2017, **15:**1-9.

7.   Mittmann N, Knowles SR, Koo M, Shear NH, Rachlis A, Rourke SB: **Incidence of toxic epidermal necrolysis and Stevens-Johnson Syndrome in an HIV cohort.** *American journal of clinical dermatology* 2012, **13:**49-54.

8.   Bassi PU, Osakwe AI, Ogar CK, Elagbaje C, Nwankwo BB, Balogun ST, Ntadom GN, Isah AO: **Impact of comorbidity on adverse drug reaction profile in a cohort of patients treated with artemisinin combination therapies for**

uncomplicated malaria in Nigeria. *Pharmacology research & perspectives* 2017, **5:**e00302.

9. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y: **When good drugs go bad.** *Nature* 2007, **446:**975-977.

10. Barabási A-L, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach tohuman disease.** *Nature reviews genetics* 2011, **12:**56-68.

11. Hey AJ, Tansley S, Tolle KM: *The fourth paradigm: data-intensive scientific discovery.*

Microsoft research Redmond, WA; 2009.

12. Brahmachari SK: **Introducing the medical bioinformatics in Journal of TranslationalMedicine.** Springer; 2012.

13. Chan SY, Loscalzo J: **The emerging paradigm of network medicine in the study ofhuman disease.** *Circulation research* 2012, **111:**359-374.

14. Mizutani S, Pauwels E, Stoven V, Goto S, Yamanishi Y: **Relating drug–protein interactionnetwork with drug side effects.** *Bioinformatics* 2012, **28:**i522-i528.

15. Fliri AF, Loging WT, Thadeio PF, Volkmann RA: **Analysis of drug-induced effect patternsto link structure and side effects of medicines.** *Nature chemical biology* 2005, **1:**389-397.

16. Wang X, Thijssen B, Yu H: **Target essentiality and centrality characterize drug sideeffects.** *PLoS computational biology* 2013, **9:**e1003119.

17. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P: **Drug target identification usingside-effect similarity.** *Science* 2008, **321:**263-266.

18. Brouwers L, Iskar M, Zeller G, Van Noort V, Bork P: **Network neighbors of drug targetscontribute to drug side-effect similarity.** *PloS one* 2011, **6:**e22187.

19. Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H: **Structure of protein interactionnetworks and their implications on drug design.** *PLoS Comput Biol* 2009, **5:**e1000550.

20. Cheng F, Kovács IA, Barabási A-L: **Network-based prediction of drug combinations.**

Nature communications 2019, **10:**1-11.

21. Ryaboshapkina M, Hammar M: **Tissue-specific genes as an underutilized resource in drugdiscovery.** *Scientific reports* 2019, **9:**1-12.

22. Dezső Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D, Bugrim A, Rakhmatulin E, Brennan RJ, Guryanov A: **A comprehensive functional analysis of tissue specificity of human gene expression.** *BMC biology* 2008, **6:**1-15.

23. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M: **DrugBank: a knowledgebase for drugs, drug actions and drug targets.** *Nucleic acids research* 2008, **36:**D901-D906.

24. Tatonetti NP, Patrick PY, Daneshjou R, Altman RB: **Data-driven prediction of drug effectsand interactions.** *Science translational medicine* 2012, **4:**125ra131-125ra131.

25. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM: **DGIdb: mining the druggable genome.** *Nature methods* 2013, **10:**1209-1210.

26. Keshava Prasad T, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A: **Human protein reference database— 2009 update.** *Nucleic acids research* 2008, **37:**D767-D772.

27. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a generalrepository for interaction datasets.** *Nucleic acids research* 2006, **34:**D535-D539.

28. Martin A, Ochagavia ME, Rabasa LC, Miranda J, Fernandez-de-Cossio J, Bringas R: **BisoGenet: a new tool for gene network building, visualization and analysis.** *BMC bioinformatics* 2010, **11:**1-9.

29. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI: **DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants.** *Nucleic acids research* 2016**:**gkw943.

30. Brown NJ, Frazier CR, Cauthen KR, Nozick LK: **A Unique Graph Similarity Metric for Anomaly Detection.** Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); 2019.

31. Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M: **Computing topological parameters of biological networks.** *Bioinformatics* 2008, **24:**282-284.

32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B,Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome research* 2003, **13:**2498-2504.

33. Hidalgo CA, Blumm N, Barabási A-L, Christakis NA: **A dynamic network approach for the study of human phenotypes.** *PLoS computational biology* 2009, **5:**e1000353.

34. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B: **WebGestalt 2019: gene set analysis**

toolkit with revamped UIs and APIs. *Nucleic acids research* 2019.

35.   Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B: **WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs.** *Nucleic acids research* 2019, **47:**W199-W205.

36.   Consortium G: **The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue generegulation in humans.** *Science* 2015, **348:**648-660.

37.   Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å,Kampf C, Sjöstedt E, Asplund A: **Tissue-based map of the human proteome.** *Science* 2015,**347**.

38.   Jia P, Dai Y, Hu R, Pei G, Manuel AM, Zhao Z: **TSEA-DB: a trait–tissue association mapfor human complex traits and diseases.** *Nucleic acids research* 2020, **48:**D1022-D1030.

39.   Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK: **A next generation connectivity map: L1000 platform and the first1,000,000 profiles.** *Cell* 2017, **171:**1437-1452. e1417.

40.   Leday GG, Vértes PE, Richardson S, Greene JR, Regan T, Khan S, Henderson R, Freeman TC, Pariante CM, Harrison NA: **Replicable and coupled changes in innate and adaptive immune gene expression in two case-control studies of blood microarrays in major depressive disorder.** *Biological psychiatry* 2018, **83:**70-80.

41.   Kupershmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, Shekar M, Wang H, ParkJ, Cui W, et al: **Ontology-based meta-analysis of global collections of high-throughput public data.** *PLoS One* 2010, **5**.

42.   Love MI, Anders S, Kim V, Huber W: **RNA-Seq workflow: gene-level exploratory analysisand differential expression.** *F1000Research* 2015, **4**.

43.   Metsalu T, Vilo J: **ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap.** *Nucleic acids research* 2015, **43:**W566-W570.

44.   Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L: **Uncovering disease-disease relationships through the incomplete interactome.** *Science* 2015, **347**.

45.   Vinogradov S, Gottesman II, Moises HW, Nicol S: **Negative association between schizophrenia and rheumatoid arthritis.** *Schizophrenia bulletin* 1991, **17:**669-678.

46.   Oken RJ, Schulzer M: **At issue: schizophrenia and rheumatoid arthritis: the**

negative association revisited. *Schizophrenia Bulletin* 1999, **25:**625-638.

47.   Benros ME, Pedersen MG, Rasmussen H, Eaton WW, Nordentoft M, Mortensen PB: **A nationwide study on the risk of autoimmune diseases in individuals with a personal or afamily history of schizophrenia and related psychosis.** *American Journal of Psychiatry* 2014, **171:**218-226.

48.   Boulet L: **Influence of comorbid conditions on asthma.** *European Respiratory Journal* 2009, **33:**897-906.

49.   Lamers F, van Oppen P, Comijs HC, Smit JH, Spinhoven P, van Balkom AJ, Nolen WA, Zitman FG, Beekman AT, Penninx BW: **Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA).** *The Journal of clinical psychiatry* 2011, **72:**341-348.

50.   Dogra S, Ardern CI, Baker J: **The relationship between age of asthma onset and cardiovascular disease in Canadians.** *Journal of Asthma* 2007, **44:**849-854.

51.   Christiansen SC, Schatz M, Yang S-J, Ngor E, Chen W, Zuraw BL: **Hypertension and asthma: a comorbid relationship.** *The Journal of Allergy and Clinical Immunology: In Practice* 2016, **4:**76-81.

52.   Rutten FH, Cramer M-JM, Grobbee DE, Sachs AP, Kirkels JH, Lammers J-WJ, Hoes AW: **Unrecognized heart failure in elderly patients with stable chronic obstructive pulmonary disease.** *European heart journal* 2005, **26:**1887-1894.

53.   Iversen K, Kjaergaard J, Akkan D, Kober L, Torp Pedersen C, Hassager C, Vestbo J, KjollerE, Group ELFS: **Chronic obstructive pulmonary disease in patients admitted with heartfailure.** *Journal of internal medicine* 2008, **264:**361-369.

54.   Iglay K, Hannachi H, Joseph Howie P, Xu J, Li X, Engel SS, Moore LM, Rajpathak S: **Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus.** *Current medical research and opinion* 2016, **32:**1243-1252.

55.   Pantalone KM, Hobbs TM, Chagin KM, Kong SX, Wells BJ, Kattan MW, Bouchard J, Sakurada B, Milinovich A, Weng W: **Prevalence and recognition of obesity and its associated comorbidities: cross-sectional analysis of electronic health record data from alarge US integrated health system.** *BMJ open* 2017, **7**.

56.   Llorente I, García-Castañeda N, Valero C, González-Álvaro I, Castañeda S: **Osteoporosis inRheumatoid Arthritis: Dangerous Liaisons.** *Frontiers in Medicine* 2020, **7:**802.

57.   Kuusimäki T, Al Abdulrasul H, Kurki S, Hietala J, Hartikainen S, Koponen M,

Tolppanen AM, Kaasinen V: **Increased Risk of Parkinson's Disease in Patients With SchizophreniaSpectrum Disorders.** *Movement Disorders* 2020.

58. Chang D, Keinan A: **Principal component analysis characterizes shared pathogenetics from genome-wide association studies.** *PLoS computational biology* 2014, **10:**e1003820.

59. McGuirl MR, Smith SP, Sandstede B, Ramachandran S: **Detecting shared genetic architecture among multiple phenotypes by hierarchical clustering of gene-level association statistics.** *Genetics* 2020, **215:**511-529.

60. Hakes L, Pinney JW, Robertson DL, Lovell SC: **Protein-protein interaction networks andbiology—what's the connection?** *Nature biotechnology* 2008, **26:**69-72.

61. Hopkins AL, Groom CR: **The druggable genome.** *Nature reviews Drug discovery* 2002,

1:727-730.

62. Sun J, Zhu K, Zheng WJ, Xu H: **A comparative study of disease genes and drug targets inthe human protein interactome.** In *BMC bioinformatics*. BioMed Central; 2015: 1-9.

63. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.** *Advanced drug delivery reviews* 1997, **23:**3-25.

64. Han Y, Wang C, Klinger K, Rajpal DK, Zhu C: **An integrative network-based approach for drug target indication expansion.** *PLoS one* 2021, **16:**e0253614.

65. Maruyama Y, Nishida M, Sugimoto Y, Tanabe S, Turner JH, Kozasa T, Wada T, Nagao T, Kurose H: **Gα12/13 Mediates α1-Adrenergic Receptor–Induced Cardiac Hypertrophy.** *Circulation research* 2002, **91:**961-969.

66. Lacy C: *Drug information handbook: a comprehensive resource for all clinicians and healthcare professionals.* Lexi-Comp Incorporated; 2006.

67. Ressler KJ, Nemeroff CB: **Role of serotonergic and noradrenergic systems in the pathophysiology of depression and anxiety disorders.** *Depression and anxiety* 2000, **12:**2-19.

68. Mineur YS, Obayemi A, Wigestrand MB, Fote GM, Calarco CA, Li AM, Picciotto MR: **Cholinergic signaling in the hippocampus regulates social stress resilience and anxiety- and depression-like behavior.** *Proceedings of the National Academy of Sciences* 2013, **110:**3573-3578.

69. Data E: **Orange book: Approved drug products with therapeutic equivalence evaluations.** 2017.

70. Goodwin GM: **The overlap between anxiety, depression, and obsessive-compulsive disorder.** *Dialogues in clinical neuroscience* 2015, **17:**249.

71. Pöldinger W, Sieberns S: **Depression-inducing and antidepressive effects of neuroleptics.**

Neuropsychobiology 1983, **10:**131-136.

72. Alam A, Voronovich Z, Carley JA: **A review of therapeutic uses of mirtazapine in psychiatric and medical conditions.** *The primary care companion for CNS disorders* 2013,**15**.

73. Zahodne LB, Fernandez HH: **Pathophysiology and treatment of psychosis in Parkinson's disease.** *Drugs & aging* 2008, **25:**665-682.

74. Stoner SC, Dahmen MM, Makos M, Lea JW, Carver LJ, Rasu RS: **An exploratory retrospective evaluation of ropinirole-associated psychotic symptoms in an outpatient population treated for restless legs syndrome or Parkinson's disease.** *Annals of Pharmacotherapy* 2009, **43:**1426-1432.

75. Smith SM, Vale WW: **The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress.** *Dialogues in clinical neuroscience* 2006, **8:**383.

76. Rubin RT, Phillips JJ, Sadow TF, McCracken JT: **Adrenal gland volume in major depression: increase during the depressive episode and decrease with successful treatment.** *Archives of general psychiatry* 1995, **52:**213-218.

77. Ulrich-Lai YM, Figueiredo HF, Ostrander MM, Choi DC, Engeland WC, Herman JP: **Chronic stress induces adrenal hyperplasia and hypertrophy in a subregion-specific manner.** *American journal of physiology-endocrinology and metabolism* 2006, **291:**E965- E973.

78. Qin D-d, Rizak J, Feng X-l, Yang S-c, Lü L-b, Pan L, Yin Y, Hu X-t: **Prolonged secretion of cortisol as a possible mechanism underlying stress and depressive behaviour.** *Scientific reports* 2016, **6:**1-9.

79. Gibbons JL, McHugh PR: **Plasma cortisol in depressive illness.** *Journal of Psychiatric Research* 1962.

80. Cosgriff J, Abbott R, Oakley-Browne M, Joyce PR: **Cortisol hypersecretion predicts early depressive relapse after recovery with electroconvulsive therapy.** *Biological psychiatry* 1990.

81. Nemeroff CB: **The corticotropin-releasing factor (CRF) hypothesis of depression: new findings and new directions.** *Mol Psychiatry* 1996, **1:**336-342.

82. Graeff FG: **Anxiety, panic and the hypothalamic-pituitary-adrenal axis.** *Brazilian Journalof Psychiatry* 2007, **29:**s3-s6.

83. Lenze EJ, Mantella RC, Shi P, Goate AM, Nowotny P, Butters MA, Andreescu C, ThompsonPA, Rollman BL: **Elevated cortisol in older adults with generalized anxiety disorder is reduced by treatment: a placebo-controlled evaluation of escitalopram.** *The American Journal of Geriatric Psychiatry* 2011, **19:**482-490.

84. Jazayeri S, Keshavarz SA, Tehrani-Doost M, Djalali M, Hosseini M, Amini H, Chamari M, Djazayery A: **Effects of eicosapentaenoic acid and fluoxetine on plasma cortisol, serum interleukin-1beta and interleukin-6 concentrations in patients with major depressive disorder.** *Psychiatry Research* 2010, **178:**112-115.

85. Dziurkowska E, Wesolowski M, Dziurkowski M: **Salivary cortisol in women with major depressive disorder under selective serotonin reuptake inhibitors therapy.** *Archives of women's mental health* 2013, **16:**139-147.

86. Kupershmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, Shekar M, Wang H, ParkJ, Cui W: **Ontology-based meta-analysis of global collections of high-throughput public data.** *PloS one* 2010, **5:**e13066.

87. Duran-Frigola M, Mateo L, Aloy P: **Drug repositioning beyond the low-hanging fruits.** *Current Opinion in Systems Biology* 2017, **3:**95-102.

88. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guilliams T, Latimer J, McNamee C: **Drug repurposing: progress, challenges and recommendations.** *Nature Reviews Drug Discovery* 2019, **18:**41.

89. Wei Q, Khan IK, Ding Z, Yerneni S, Kihara D: **NaviGO: interactive tool for visualizationand functional similarity and coherence analysis with gene ontology.** *Bmc Bioinformatics*2017, **18:**1-13.

90. Bronte V, Pittet MJ: **The spleen in local and systemic regulation of immunity.** *Immunity* 2013, **39:**806-818.

91. Nakamichi Y, Mizoguchi T, Arai A, Kobayashi Y, Sato M, Penninger JM, Yasuda H, Kato S,DeLuca HF, Suda T: **Spleen serves as a reservoir of osteoclast precursors through vitamin D-induced IL-34 expression in osteopetrotic op/op mice.** *Proceedings of the National Academy of Sciences* 2012, **109:**10006-10011.

92. NISHIYA K, HISAKAWA N, HOSOKAWA T, HASHIMOTO K, DOI T: **Enlarged spleendetected by abdominal ultrasonography in patients with RA.** *Annals of the rheumatic diseases* 2000, **59:**750-750.

93.   GREEN RA, FROMKE VL: **Splenectomy in Felty's syndrome.** *Annals of internal medicine* 1966, **64:**1265-1270.

94.   Isomäki H, Koivisto O, Kiviniitty K: **Splenomegaly in rheumatoid arthritis.** *Acta Rheumatologica Scandinavica* 1971, **17:**23-26.

95.   Fishman D, Isenberg DA: **Splenic involvement in rheumatic diseases.** In *Seminars in arthritis and rheumatism*. Elsevier; 1997: 141-155.

96.   Khan MA, Kushner I: **Improvement of rheumatoid arthritis following splenectomy for Felty syndrome.** *Jama* 1977, **237:**1116-1118.

97.   Sun D, Zheng X, Chen Y, Jia C, Xu S, Lin C, Zhang P, Zhang Z, Cai D, Jin D: **Enhancement of osteogenesis post splenectomy does not attenuate bone loss in ovariectomized rats.** *Journal of Orthopaedic Research* 2015, **33:**1356-1363.

98.   Xiao W-A, Wang Y, Li J-J, Yang X-X, Li Z-W, Zheng W: **Effect of immune function changes after splenectomy on fracture healing in patients with compound injury.** *Int J Clin Exp Med* 2016, **9:**1716-1723.

99.   Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M: **Defining comorbidity: implications for understanding health and health services.** *The Annals of Family Medicine* 2009, **7:**357-363.

11. General discussion

The discovery of concepts and technologies with increased capacity to explain the complexity of human diseases and therapeutics gradually led to the emergence of network biology in the early 2000s. Subsequently, network biologists focused on elucidating the organizational principles of disease-associated genes and drug targets in the human interactome and their implications on higher phenomena such as cross-disorder relationships and drug interactions (Barabási et al., 2011, Barabasi and Oltvai, 2004, Yıldırım et al., 2007, Cheng et al., 2019, Ulitsky and Shamir, 2007, Kelley and Ideker, 2005, Paci et al., 2021, Goh et al., 2007, Hidalgo et al., 2009, Costanzo et al., 2016). Although the studies conducted thus far have led to significant advancements in network biology, there are five critical issues highlighted in section 1.4 (**Fig. 7**) that have remained unaddressed by these studies and subsequent research (Sharma et al., 2015a, Sharma et al., 2018, Sun et al., 2009, Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a). These issues include the lack of an integrated conceptual framework that incorporates network formalisms to examine interactomes, insufficient exploration of interactomes leading to a scarcity of biological and clinical insights from network biology studies (Sun et al., 2009, Sharma et al., 2015a, Sharma et al., 2018, Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a), the failure to consider the biological contexts of molecular interactions (Goh et al., 2007, Hidalgo et al., 2009), inadequate addressing of the sparseness of the human interactome (Kotlyar et al., 2015, Hopf et al., 2014, Emamjomeh et al., 2014a, Garzón et al., 2016, You et al., 2013, Jia et al., 2015, Li and Ilie, 2017, Deng et al., 2003, Raja et al., 2013) and the necessity to develop methods for establishing correlational relationships between disease interactomes and DTNs (Goh et al., 2007, Cheng et al., 2019). The interactomic framework proposed in this thesis provides a comprehensive approach to overcome these limitations.

First, this framework assembles a comprehensive set of analytical methods in the form of a standard pipeline to examine disease interactomes and DTNs. The network-based frameworks proposed in previous studies focused solely on identifying the local neighbourhoods of disease-associated genes using topological parameters of the human interactome alone (Sharma et al., 2015a, Sharma et al., 2018) or in combination with real-world datasets (Sun et al., 2009), such as differential expression profiles in patients, and validating the identified disease modules by demonstrating their enrichment for disease-associated biological modalities (Sharma et al., 2015a, Sharma et al., 2018, Sun et al., 2009). These frameworks did not provide novel actionable biological information. Other research groups only employed parts of these frameworks to characterise the network modules of other diseases (Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a). As such, the frameworks, in their entirety, remained unused in the context of other disorders. Additionally, although one of these frameworks attempted to examine the overlap of the disorder of interest with other related disorders in the interactome (Sharma et al., 2015a), this characterization was limited to finding a shared sub-network and the biological processes underlying this sub-network. It did not involve a comprehensive assessment of the common sub-network, the sub-networks unique to each disease, the various biological factors influencing their common aetiology, and their etiological diversification. This thesis addressed these shortcomings by proposing a unified framework containing two separate modules that will help analyse disease interactomes and DTNs individually and in relation to one another.

Second, the pipeline helped derive critical insights into disease biology and reposition existing drugs

for new therapeutic uses. Third, the framework proposed novel approaches to extract higher-level cross-disorder and cross-DTN relationships, which led to significant biological and clinical insights. In the upcoming two sections, I discuss the wealth of biological insights into individual disorders (section 11.1) and higher-level cross-disorder relationships (section 11.2) generated using the pipeline. Following this, I discuss the repurposable drugs identified from individual disease interactomes (section 11.3) and the factors underlying drug action elucidated from examining cross-DTN relationships (section 11.4). Fourth, the integration with disease-associated multi-omics datasets allowed the identification of context-sensitive interactomes and demonstrated the validity of the interactome as a data integration model (section 11.5). Lastly, the framework circumvented the sparsity of the interactomes by adding computationally predicted PPIs (section 11.6).

11.1 Towards a better understanding of complex disease biology: the interactomic framework provided novel insights into the mechanisms underlying multiple disorders

In **Chapters 2-7**, explorative analyses using the framework helped gain new insights into a specific subtype of a congenital heart disease, two rare cancers, two viral infections, a skeletal disorder, and several neurological and neuropsychiatric disorders (**Fig. 15**). The successful implementation of the framework across these six classes of diseases demonstrates its potential generalizability to other specific disorders and disorder classes in the future.
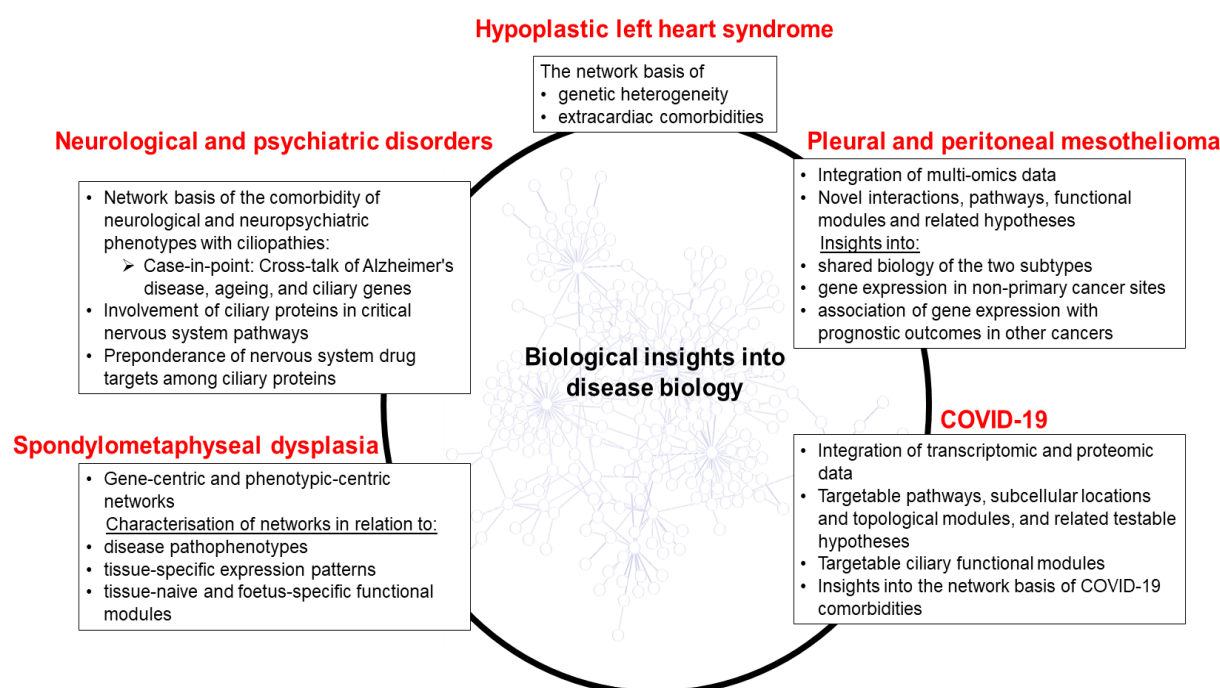
Figure 15: The biological insights gained into the mechanisms of multiple disorders using the interactomic framework. The main findings from the analysis of the following interactomes have been listed in the boxes: the interactome of hypoplastic left heart syndrome, a congenital heart disease subtype (Chapter 3), malignant pleural and peritoneal mesotheliomas (Chapter 4 and Chapter 5), the host proteins targeted by SARS-CoV-2, the causative agent for COVID-19 (Chapter 6), genes associated with the skeletal disorder Sedaghatian type spondylometaphyseal dysplasia or phenotypically/aetiologically related disorders and its causative gene *GPX4* (Chapter 7) and finally multiple neurological and psychiatric disorders (Chapter 2). The figure was created in Microsoft PowerPoint.

11.1.1 Insights into the biology of a heart disease

Hypoplastic left heart syndrome or HLHS is a complex congenital heart disease affecting 1 in 5,000 newborns (Gobergs et al., 2016). It manifests as underdevelopment of the structures on the left side of the heart, namely, atresia or critical stenosis of the mitral or aortic valves and hypoplasia of the left ventricle, ascending aorta, and aortic arch (Gobergs et al., 2016). Until 30 years ago, infants born with this condition died within the first few weeks of life, and HLHS accounts for 23% of the deaths occurring in the first week of life due to cardiac abnormalities (Gobergs et al., 2016, Šamánek et al., 1989). Surgical palliative techniques and post-operative care have significantly improved survival, with 60-70% of HLHS neonates surviving for at least five years following repair (Hamzah et al., 2020, d'Udekem et al., 2014, Alsoufi et al., 2015). However, mortality is highest in the first year of life, with 30% of infants dying or requiring a heart transplant before they are one year old (Siffel et al., 2015). The interactomic framework helped address a key issue pertaining to HLHS biology.

HLHS likely has a strong genetic basis and could have a multigenic aetiology, as indicated by its high familial aggregation with other left ventricular outflow tract obstructive defects and evidence from a large-scale mutagenesis screen in mice (McBride et al., 2009) and clinical studies (McBride et al., 2005). Exome sequencing studies (Zaidi et al., 2013) and studies involving genome sequencing and genome-wide screening by comparative genomic hybridization helped identify HLHS-associated variants (Theis et al., 2021, Reuter et al., 2020, Verma et al., 2016, Gill et al., 2009, Theis et al., 2020, Homsy et al., 2015). However, despite the recovery of genes associated with HLHS, an integrative approach to elucidate their functional consequences is still lacking. **Chapter 3** details the discovery and analysis of a cohesive HLHS interactome seeded by more than 70 HLHS-associated genes discovered in mutant mouse models (Liu et al., 2017, Li et al., 2015). It provided crucial insights into the comorbidities of HLHS, including the network proximity of diabetes, Alzheimer's disease, and liver carcinoma-associated genes to HLHS genes, suggesting a mechanistic basis for their comorbidity with HLHS (Becerra et al., 1990, Bagge et al., 2018, Komatsu et al., 2019, Kogiso and Tokushige, 2020) and the tissue-specificity of the interactome genes for sites of extracardiac anomalies (placenta, liver, and brain) (Jones et al., 2015, Weinberg and Bolande, 1970, Marino et al., 2012, Hinton et al., 2008). The HLHS interactome also shared significant overlaps with the interactomes of ciliopathy- and microcephaly-associated genes, with the shared genes enriched for involvement in intellectual disability/developmental delay and neuronal death pathways, respectively. These intersections supported two clinical observations: (i) the increased burden of ciliopathy variants in HLHS patients with developmental delay (Geddes et al., 2017) and (ii) the prevalence of neurological abnormalities among HLHS patients with microcephaly (Hangge et al., 2013). In summary, this study provided evidence for the utility of the HLHS interactome in investigating various HLHS comorbidities and the functional consequences of the genes harbouring HLHS-associated mutations. These results can directly inform and catalyse future investigations on the molecular basis of HLHS and biomedical studies seeking to improve clinical interventions in HLHS.

11.1.2 Insights into the biology of cancers

Chapter 4 and **Chapter 5** examined the pleural and peritoneal mesothelioma subtypes that affect the lining of the lungs and the abdominal cavity, respectively. Mesothelioma is a rare and aggressive cancer that originates from the mesothelial lining of several internal organs and the thoracic and abdominal cavities (Carbone et al., 2019a). Malignant pleural mesothelioma (MPM) accounts for

90% of malignant mesotheliomas, has a short median survival of approximately 1 year (Lang-Lazdunski, 2018), and is associated with asbestos exposure. After exposure, it has a long latency period and is conclusively diagnosable only after reaching the invasive phase (Wang et al., 2004). In contrast with MPM, where asbestos exposure characterises 80% of the cases, only 8% of peritoneal mesothelioma cases have a history of asbestos exposure (Robinson and Lake, 2005). Malignant peritoneal mesothelioma (MPeM) was more apparent among patients with a history of abdominal surgeries rather than asbestos exposure (Carbone et al., 2019b). MPeM has a higher median survival rate than MPM (31 months versus 14 months) (Amin et al., 2018). Two factors warrant urgent investigations into the molecular mechanisms underlying the two cancers: (i) their fatality and (ii) the limited therapeutic options, namely, pemetrexed eliciting modest clinical responses and disease stabilization (Carbone et al., 2019b).

MPM tends to cluster in families and occurs only in a small fraction of the population exposed to asbestos (Bueno et al., 2016). Additionally, MPeM appears to be associated with germline mutations more frequently than MPM (Carbone et al., 2019b). Therefore, studies suggest the involvement of a genetic component in both cancers. Most studies examine only a handful of mesothelioma-associated genes, such as *BAP1*, *CDKN2A*, and *NF2*, despite the steady generation of multi-omics datasets of both these cancers. Addressing the lack of a framework for unifying these multi-omics datasets, both **Chapter 4** and **Chapter 5** aimed to demonstrate the validity of the interactome as an integrative framework. Encouragingly, in **Chapter 4**, genetic variant, transcriptomic, and proteomic data pertaining to MPM supported 85.65% of proteins in the MPM interactome. Similarly, in **Chapter 5**, differential gene expression in pre-clinical models and human tumour specimens of MPeM supported 75.6% of proteins in the MPeM interactome. This showed that both the MPM and MPeM interactomes with disease-associated proteins and their interacting partners will help biologists, bioinformaticians, and clinicians to piece together an integrated view of how mesothelioma-associated genes from various studies are functionally linked.

An explorative analysis of the MPM interactome in **Chapter 4** helped derive valuable insights into the aetiology of MPM. Experimental validation of five computationally predicted PPIs, BAP1-PARP3, KDR-ALB, PDGFRA-ALB, CUTA-HMGB1, and CUTA-CLPS, based on their proximity to MPM-associated genes and their biological relevance helped generate testable hypotheses for future studies. The MPM interactome also showed significant enrichments for cancer-related pathways, such as NF-κB signalling, PI3/AKT signalling, VEGF signalling, and natural killer cell signalling, providing further insights into the molecular underpinnings of MPM. Most importantly, the interactome had highly significant overlaps with ten MPM-related multi-omics datasets, including MPM-associated genetic variants, genes differentially expressed or methylated in MPM or upon asbestos exposure, genes correlated with lung cancer prognosis, and exosome-derived proteins in malignant mesothelioma cell lines. Forty-eight computationally predicted interactors of MPM genes had three or more pieces of biological evidence, indicating their suitability for future studies.

The explorative analysis in **Chapter 5** shed light on the functional underpinnings of MPeM aetiology. Two findings supported the splenomegaly seen in *BAP1* knockout mice resulting from the expansion of myeloid cells in the spleen (Dey et al., 2012). Firstly, the human orthologues of mouse genes with elevated expression in the spleen and the thymus – both extramedullary haematopoietic sites that regulate lymphoid cells outside the bone marrow – showed significant enrichment in the interactome.

Secondly, the interactome contained haematopoiesis as a functional module. Chromosome segregation was the most enriched functional module in the interactome, consistent with the observation that 56% of the MPM-associated seed genes underwent chromosomal events leading to cancer. Transcriptional deregulation seemed critical to MPeM aetiology, as suggested by the related functional modules in the interactome, such as covalent chromatin modification. The interactome contained 'positive regulation of IL-6 production' as a functional module, which was relevant since MPeM patients show elevated expression of an anti-apoptotic factor called survivin (BIRC5) induced by the cytokine IL-6 (Zaffaroni et al., 2007). As previously stated, germline mutations are more prevalent among peritoneal than pleural mesothelioma patients (Carbone et al., 2019b). These mutations may predispose MPeM patients to multiple other cancers (Carbone et al., 2019b). In line with this, the interactome showed significant enrichment of genes with their expression positively correlated with specific clinical outcomes in other cancers. Lastly, the study revealed extensive overlap between the MPM and MPeM interactomes and interconnections between the genes associated with both diseases, indicating a common aetiology (discussed further in section 11.5).

11.1.3 Insights into the biology of viral infections

Chapter 6 demonstrates the use of the interactomic framework in exploring the biology of viral infections. The study involved analysing the host proteins targeted by the viruses and the neighbourhood network surrounding them. COVID-19, caused by SARS-CoV-2 (Rothan and Byrareddy, 2020), emerged as a global pandemic in 2020. Viral infections elicit a cascade of interactions among multiple genes and proteins in the host cell. This complex network can either restrict viral replication in host cells or get hijacked by the virus for its perpetuation. Several COVID-centric network biology studies (Kumar et al., 2020, Gysi et al., 2020, Zhou et al., 2020b) presented the analysis of the 'known PPI neighbourhood' of the host proteins targeted by SARS-CoV-2 (Gordon et al., 2020), i.e., an incomplete interactome consisting of only experimentally verified PPIs. Contrary to this, in **Chapter 6**, this neighbourhood is augmented with 1,941 computationally predicted PPIs. The study presents a fuller version of the host protein interactome, facilitating the discovery of previously unknown disease mechanisms and the characterisation of under-studied host proteins through functional associations of their predicted interactors.

The host protein interactome provided valuable insights into COVID-19 biology. The interactome shared large and statistically significant overlaps with SARS-CoV- and SARS-CoV-2-induced transcriptional profiles, indicating its biological validity. It also unveiled computationally predicted interactions between SARS-CoV-2-modulated host proteins and protein biomarkers with varied expression across the different stages of COVID-19 identified in independent studies. The interactome revealed genes with elevated expression in the lungs (the target tissue of the virus), the spleen (that regulates the host immune responses), and the brain and the heart (target tissues of co-morbidities among COVID-19 non-survivors) as novel interactors of the host proteins. Furthermore, it revealed topologically connected modules involved in various cellular processes. These included modules of cilium organization, nuclear transport, ribonucleoprotein complex biogenesis, endosomal transport, and epigenetic regulation of gene expression. The interactome showed enrichment for subcellular locations and cellular processes that could act as potential targets of SARS-CoV-2. For example, SARS-CoV-2-modulated host proteins shared several common interactors with ciliary proteins. The discovery of a novel interaction between the host protein NUP98 and CHMP5, a ciliary protein, and its role as a connector of two critical functional modules led to the hypothesis that this

interaction may activate an IFN-stimulated pathway with the potential to interfere with viral budding. These findings highlight the importance of investigating the function of cilia as viral entry points and modulators of viral infections and suggest potential targets for the development of therapeutics. Lastly, the interactome showed significant enrichments for genes associated with metabolic, neurological, developmental, and vascular disorders, and cancers. For example, the proteins linked to hypertension and diabetes, two common co-morbidities of COVID-19 non-survivors (Fang et al., 2020, Sidaway, 2020), directly interacted with host proteins.

11.1.4 Insights into the biology of a skeletal disorder

Chapter 7 unravelled the mechanisms underlying a rare and lethal skeletal dysplasia known as spondylometaphyseal dysplasia, Sedaghatian type or SMDS, caused by mutations in a single gene, *GPX4*. Although some initial studies described the clinical presentation of SMDS (Sedaghatian and Opitz, 1980, Opitz et al., 1987, Peeden Jr et al., 1992, Elçioglu and Hall, 1998, Koutouby et al., 2000, English et al., 2006) and characterised *GPX4* mutations (Smith et al., 2014, Aygun et al., 2012), much about the disorder remained unexplored. The study probed the neighbourhood network of *GPX4*, exploring its connections with three sets of SMDS-centric genes and a set of *GPX4*-centric genes. The SMDS-centric networks included the shortest paths between *GPX4* and other genes putatively associated with SMDS in the DisGeNET database (Piñero et al., 2016), the genes linked to other skeletal dysplasias, and the genes linked to disorders with similar phenotypes as SMDS. The gene-centric network, on the other hand, identified the shortest paths between *GPX4* and genes that impact *GPX4* expression when knocked out, knocked down, mutated, or overexpressed. Analysis of these networks revealed their enrichment for genes showing elevated expression in the brain, testis, or both, a pattern seen with *GPX4*. They also showed enrichment for genes associated with various SMDS pathophenotypes and specific functional modules active in tissue-naive and foetus-specific contexts. Lastly, the study identified seven novel interactors of *GPX4*, namely, *APBA3*, *EGR4*, *FUT5*, *GAMT*, *GTF2F1*, *MATK*, and *ZNF197*. Literature evidence suggested their biological relevance to *GPX4* and SMDS. Overall, the study expanded our understanding of SMDS, providing a network-level view of its underlying mechanisms and allowing researchers to prioritise genes and functional modules for potential therapeutic innovations.

11.1.5 Insights into the biology of neurological and neuropsychiatric disorders

Chapter 2 helped derive valuable insights into functional underpinnings of a range of neurological and neuropsychiatric disorders.

Given the frequent co-occurrence of neurological and neuropsychiatric phenotypes with ciliopathies (Guo et al., 2015, Louvi and Grove, 2011, Marley and von Zastrow, 2012, Alvarez Retuerto et al., 2008, Higginbotham et al., 2012, Torri et al., 2010, Marley and von Zastrow, 2010, Kamiya et al., 2008), there is a growing recognition of the crucial role played by the cilium in various aspects of nervous system development and function (Lee and Gleeson, 2011, Louvi and Grove, 2011, Guemez-Gamboa et al., 2014). To elucidate the underlying mechanisms of nervous system phenotypes associated with ciliopathies, in **Chapter 2** a systems-level investigation of the cilia interactome was conducted. The interactions of primary and motile ciliary proteins – experimentally verified or computationally predicted – constituted the cilia interactome. Genetic and transcriptomic datasets and functional interactions in signalling pathways ascertained the biological validity of the

interactome. The cilia interactome overlapped extensively with the interactomes of seven neuropsychiatric and neurological disorders assembled from human genes harbouring disease-associated mutations in GWA studies. These disorders included SCZ, attention-deficit hyperactivity disorder, major depressive disorder, bipolar disorder and ASDs, Alzheimer's disease and Parkinson's disease. The cilia interactome also overlapped with genes differentially expressed in six of these disorders, highlighting its potential significance in the pathogenesis of these conditions. As a case study, this chapter shows the intersection of three gene sets in a specific interactomic subspace, i.e., human ageing-associated genes, Alzheimer's disease-associated genes, and ciliary genes. Integration of this sub-network with multi-omics data revealed the potential role of ciliary sonic hedgehog signalling in hippocampal neurogenesis and memory deficits in Alzheimer's disease (Mufson et al., 2015, Smith et al., 2000, Jessberger et al., 2009, Breunig et al., 2008, Whitfield and Chakravarthy, 2009). The interactome was found to be significantly associated with cellular pathways related to neuropsychiatric processes and contained proteins targeted by approximately 100 drugs used to treat nervous system disorders. Lastly, the study proposes several novel hypotheses for examination in future studies. These relate to the involvement of ciliary PPIs in neuropsychiatric disorders, primary ciliary dyskinesia, hydrocephalus, ciliogenesis, and ciliary membrane receptor trafficking. Overall, the cilia interactome suggested that ciliary defects play a role in neuropsychiatric processes. It is a valuable resource for investigating the potential therapeutic targets for neurological and neuropsychiatric disorders.

11.2 Uncovering disease-disease relationships: a novel methodology to examine cross-disorder relationships and its relevance in disease classification and multi-scale disease modelling

In **Chapter 9**, the methodology detailed in section 1.5.2.2 to examine cross-disorder relationships helped derive critical insights into the aetiological differentiation of various anxiety disorder subtypes (**Fig. 16**), namely, generalised anxiety disorder (GAD), social anxiety disorder (SAD), obsessive-compulsive disorder (OCD), specific phobia, panic disorder (PD) and post-traumatic stress disorder (PTSD). Studies have noted a substantial overlap between functional connectivity and gene co-expression patterns within and between cortical and striatal networks (Anderson et al., 2018, Richiardi et al., 2015, Wang et al., 2015, Krienen et al., 2016, Patania et al., 2019, Mills et al., 2018). These findings gave rise to the central premise of the study, i.e., testing the regional expression patterns of the anxiety disorder interactomes may reveal brain regions governing key anxiety traits. Clustering the anxiety disorder interactomes based on the region-specific expression of their constituent genes failed to unravel any groups among the disorders. Moreover, analysis of component loadings correlated with PC1 and PC2 revealed that this approach did not capture regional specificities underlying the anxiety disorders. Therefore, contrary to our speculation, disorder groupings were not directly evident at the level of their interactomes.

However, genes co-occurring across these interactomes showed remarkable enrichment for expression in the striatum. This enrichment suggested a potential relationship between anxiety disorders and striatal gene dysfunction and presented an opportunity to refine the interactomes based on their intersection with the interactomes of two striatal compartments, striosomes and matrix. Notably, specific anxiety disorder interactomes showed significant and distinctive overlaps with the striosome and matrix interactomes, indicating that the striatal compartments could act as diverging points for the aetiological differentiation of various anxiety disorders. Anxiety disorder

groupings became apparent when the interactomes were restricted to include only genes intersecting with the striatal compartment interactomes, supporting this notion. Exploration of these sub-networks through systematic gene expression analysis revealed a branching pattern among the disorders: [GAD, OCD], [SAD, PD] and [specific phobia]. Their expression patterns in specific brain regions, including the anterior cingulate cortex, the nucleus accumbens, the amygdala, and the hippocampus, influenced these groupings, as did their involvement in the dopaminergic signalling pathway. The results indicated that the functionally distinct striatal circuits constituted by the striosome and the matrix could govern the development of anxiety disorders (**Fig. 16**). Although further investigations are necessary to characterise anxiety as an emergent property driven by specific neural circuits, our study proposes striatum and its subdivisions as candidate regions for anxiety research.

Several findings from the study also inform disorder categorisation. For example, the PTSD interactome did not show distinctive overlaps with the striosome or matrix interactomes. However, the results indicated that its mechanism could be mostly, if not entirely, different from the other anxiety disorders. Corroborating previous studies (Joshi et al., 2020), the PTSD interactome showed the highest enrichment for the hippocampus followed by the striatum, substantia nigra, and
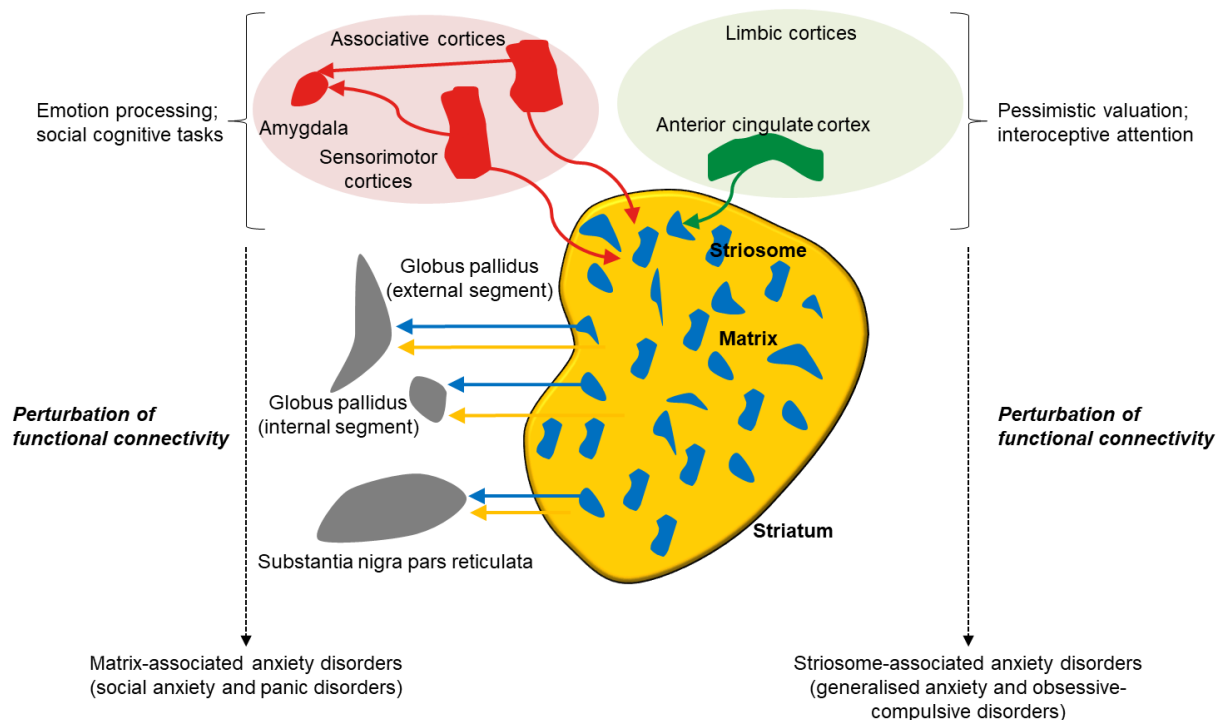
Figure 16: The aetiological distinctions of anxiety disorders associated with different striatal compartments. Genetic perturbations associated with a specific anxiety disorder may modulate the interactome in one of the two striatal compartments (striosome and matrix) or even their progenitor populations, influence their functional connectivity with other regions, and govern the route of development of critical anxiety traits. Generalised anxiety disorder- and obsessive-compulsive disorder-associated traits generated in this manner may include internally generated ruminations (involving the phenomenon of interoceptive awareness), cognitive rigidity, and pessimistic valuation, controlled by the anterior cingulate cortex, which targets the striosome compartment. Social anxiety disorder- and panic disorder-associated traits may include heightened emotional responses to neutral stimuli and excessive emotional contextualization during social information processing, controlled by the amygdala that collates information from the sensorimotor cortices, which, in turn, targets the matrix compartment. The behavioural traits characterizing striosome- and matrix-associated anxiety disorders could also arise from differential responses of these striatal compartments to reinforcement contingencies and differential involvement in resolving motivational conflicts. The figure was created in Microsoft PowerPoint.

amygdala, a pattern not seen with any other anxiety disorder. These regional specificities may justify its placement in a separate group (i.e., trauma and stress-related disorders) in the Diagnostic and Statistical Manual of Mental Disorders, unlike the case of OCD, which has been placed in a distinct group (i.e., OCD and related disorders) despite its close aetiological relationship with other anxiety disorders, as suggested by our results.

The methodology of **Chapter 9** can be adapted to uncover distinct subgroups and converging themes within other sets of disorders. It will facilitate the transformation of large amounts of disease-associated genetic data into cross-disorder network relationships. These relationships could help refine disease classification systems and offer a perspective grounded in the genetic structures of diseases in the 'lumping' versus 'splitting' debate in disorder categorization (Thaxton et al., 2022). Furthermore, the value of multi-scale disease modelling is becoming increasingly clear (Totah, 2016, Anderson et al., 2018, Richiardi et al., 2015, Wang et al., 2015, Krienen et al., 2016, Patania et al., 2019, Mills et al., 2018). In this context, the analytical approach proposed in the thesis proves helpful by condensing the multigenic complexities of diseases into single data points using multivariate techniques (as detailed in section 1.5.2.2). This approach simplifies the exploration of the various levels of biological organization that contribute to disease aetiology, as seen in **Chapter 9** with regional gene expression patterns, signalling pathways, local neuronal circuitries, and functional connectivity patterns in the brain.

11.3 Discovering new disease indications for existing drugs: the interactome provided candidate repurposable drugs for multiple disorders

Three approaches were employed for drug repurposing (**Fig. 17**). The first approach, as exemplified in **Chapter 8** (**Fig. 17a**; detailed earlier in **Fig. 12**) for the treatment of SCZ, involved conducting comparative transcriptomic analysis followed by a series of network analyses and correlation with clinical trial data. The advantage of combining the comparative transcriptomic analysis approach with the interactomic framework is that it allows identification of sub-networks that demonstrate
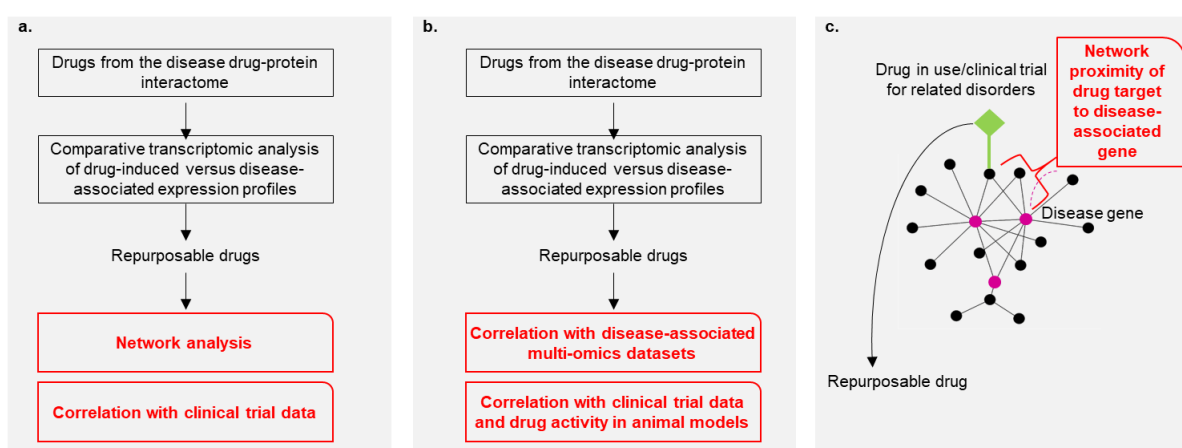
Figure 17: The three approaches used for drug repurposing. a. Comparative transcriptomic analysis of drugs identified from the disease-associated drug-protein interactome is followed first by network analysis of the targets of repurposable drugs and the genes with opposite expression in drug- versus disease-associated expression profiles, and then, compared with clinical trial results. **b.** Comparative transcriptomic analysis of drugs is followed by correlation with diverse disease-associated datasets, and utilization of information on drug activity from animal models and clinical trials. **c.** The targets of drugs that are being tested in clinical trials or are already in use for related disorders are examined for their proximity to disease-associated genes. The figure was created in Microsoft PowerPoint. The network diagram was created in Cytoscape.

interconnections between the drug targets and disease-associated proteins in the human interactome. These sub-networks could be used to investigate the mechanistic basis for the negative correlation observed between the drugs and the disease at the transcriptomic level. For example, in **Chapter 8**, the network analysis showed that the protein targets of acetazolamide were embedded in a network containing genes associated with various neuropsychiatric disorders, indicating that it could be potentially used in SCZ. Additionally, signalling pathways dysfunctional in SCZ were enriched in the network of genes that show a negative correlation in acetazolamide-induced and SCZ-associated expression profiles. In **Chapter 7**, an analysis of specific functional modules containing *GPX4* suggested that the drugs (identified from the repurposing analysis) that exploit the functional associations of *GPX4* with ferroptosis, mediated via the gene *MAPK13*, and with bone development, mediated via the gene *ZNF197*, could be examined for clinical utility in the context of SMDS.

The above approach has to be adjusted based on the varying availability of disease-associated datasets. The thesis proposes two alternate approaches for drug repurposing. The first of these approaches, as described in **Chapters 4-7**, is applicable for cases wherein disease-associated multi-omics datasets are available as necessary (**Fig. 17b**). This approach involved identifying repurposable drugs through comparative transcriptomic analysis and assessing their validity through correlation with disease-associated multi-omics datasets and drug activity in clinical settings and animal models (**Table 1**). In several of these Chapters, the latter set of correlations has helped ascertain the validity of the repurposing approach. For example, in **Chapter 5**, more than 70% of the repurposable drugs for MPeM identified using comparative transcriptome analysis are effective against peritoneal mesothelioma, pleural mesothelioma, peritoneal metastasis or primary peritoneal cancer in clinical trials, animal models or cell lines. The second approach involves examining the network proximity of drug targets to identify repurposable drugs when disease-associated datasets are unavailable (**Fig. 17c**). In **Chapter 7**, the unavailability of SMDS-associated datasets prompted the use of the datasets of two related, albeit potentially aetiologically distinct, skeletal dysplasias for comparative transcriptomic analysis, namely, chondroplasia and immune-osseous dysplasia. In this scenario, a supplemental method involved identifying the drugs – in use or clinical trials against different forms of skeletal dysplasias (Marzin and Cormier-Daire, 2020) – with targets closely connected to *GPX4* (the disease-associated gene). This analysis helped show that the targets of the drug resveratrol – earlier identified using comparative analysis with the expression profile of immuno-osseous dysplasia – interacted with *GPX4* via intermediate interactors.

Table 1: The details of the drug repurposing analysis involving comparative transcriptomic analysis and additional validation criteria.

Chapter	Disorder	Drug dataset	Disease dataset	Validation criteria of repurposable drugs
Chapter 4	Malignant pleural mesothelioma (MPM) interactome	Differential transcriptomes of drugs targeting the MPM interactome from Connectivity Map	Differential expression profiles in lung cancer	Affected genes of high differential expression in pleural mesothelioma tumours or cell lines (GSE51024(45) and GSE2549(46)), or underwent prior clinical testing in lung cancer
Chapter 5	Malignant peritoneal mesothelioma	Differential transcriptomes of drugs targeting the	Differential expression profiles of peritoneal mesothelioma	Effectiveness against peritoneal/pleural mesothelioma and/or

	(MPeM) interactome	MPeM interactome from Connectivity Map	(Fridlender et al., 2012, Kim et al., 2006, Blackshear et al., 2014)	peritoneal metastasis/primary peritoneal cancer in clinical trials, animal models or cell lines.
Chapter 6	SARS-CoV-2-modulated host protein interactome	Differential transcriptomes of drugs targeting the host protein interactome from Connectivity Map	Differential expression profiles of four SARS datasets and one COVID-19 dataset (Blanco-Melo et al., 2020a, Reghunathan et al., 2005)	Effectiveness in COVID-19 clinical trials; broad-spectrum antiviral properties; proven activity against SARS-CoV-2 or SARS-CoV/MERS-CoV in cell-based assays.
Chapter 7	SMDS interactome (constructed using *GPX4* and other genes with putative disease associations)	Differential transcriptomes of drugs targeting the SMDS interactome from Connectivity Map	Differential expression profiles of chondroplasia and immune-osseous dysplasia (Cameron et al., 2011, Baradaran-Heravi et al., 2012)	Examination of the drug targets in their PPI network and functional modules
Chapter 8	Schizophrenia	Differential transcriptomes of drugs targeting the SCZ interactome from Connectivity Map	Differential tissue or blood sample transcriptomes of SCZ patients	Examination of the biological properties and the PPI networks of (i) the targets of repurposable drugs and (ii) genes with opposite expression in drug- versus disease-associated expression profiles; biological properties of other drugs with the same targets; correlation with NIH Clinical Trials data.

Although the comparative transcriptome approach has resulted in some valuable results in the past, it has several limitations. The CMAP drug-associated gene expression profiles analysed in **Chapters 4-8** were induced in cancer cell lines (Lamb et al., 2006) and not in cell lines relevant to mesothelioma, COVID-19, SMDS, or SCZ. For increased biological validity, analyses should involve gene expression profiles induced by drugs in neuronal cell lines such as SH-SY5Y, in patient-derived stem cells, or animal models for SCZ, human pleural/peritoneal cancer cell lines or animal models for mesothelioma or in human bronchial epithelial cells or human lung cancer cell lines for COVID-19. However, previous studies show that drug data in cancer cell lines is valuable for repurposing drugs for non-cancer diseases. Specific examples include repurposing topiramate, an anti-epileptic drug, for inflammatory bowel disease (Dudley et al., 2011), SCZ drugs (Zhao and So, 2018), and bipolar disorder drugs (Kidnapillai et al., 2018).

Altogether, the thesis proposed multiple approaches for drug repurposing based on the interactomic framework and demonstrated their efficacy for five disorders. The selection of the two repurposable drugs identified in **Chapter 8** for clinical trials (Nimgaonkar, 2019, Nimgaonkar, 2022), namely, acetazolamide and cromoglycate, supports the validity of the approaches. Further investigations of the repurposable drugs identified in **Chapters 4-8** for mesothelioma, COVID-19, SMDS, and SCZ (**Fig. 18**) could help advance clinical translation for these diseases.

Understanding the precise mechanisms of the action of repurposable drugs through experimental investigations is crucial to optimise their therapeutic potential. Additionally, utilizing appropriate model systems, such as animal or in vitro models, can provide valuable insights into drug efficacy
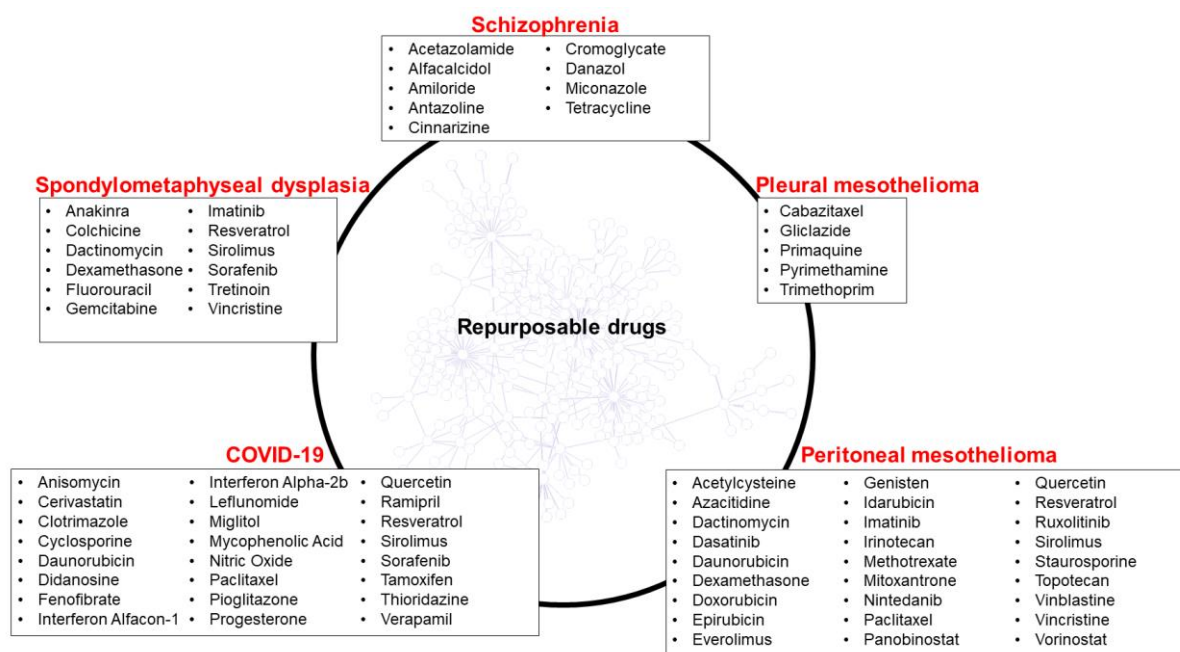
Figure 18: The repurposable drugs identified using the interactomic framework. The drugs identified as repurposable candidates for the following five disorders have been listed in the boxes: schizophrenia (Chapter 8), malignant pleural mesothelioma (Chapter 4), malignant peritoneal mesothelioma (Chapter 5), COVID-19 (Chapter 6) and spondylometaphyseal dysplasia (Chapter 7). The figure was created in Microsoft PowerPoint.

and safety. It is also imperative to consider potential side effects and evaluate the overall safety profile of repurposable drugs. Although the thesis does not include any pre-clinical study on the drugs identified in the chapters described above, **Appendix section 14.2** can be used as a template (detailed in section 1.5.1.6) to examine the activity of the repurposable drugs in neurological and psychiatric disorders.

Appendix section 14.2 explores the potential of *Convolvulus pluricaulis* (CP), a crude drug extract, in treating cognitive impairments characteristic of Alzheimer's disease. It examines the acetylcholinesterase (AChE) inhibitory activity of CP in a zebrafish model of cognitive impairment induced by scopolamine. CP is a perennial herb with anti-amnesiac and anxiolytic properties and contains alkaloids, anthocyanins, coumarins, flavonoids, phytosterols, and triterpenoids (Nahata et al., 2009, Bihaqi et al., 2011, Malik et al., 2016, Bihaqi et al., 2009). Biochemical, bioinformatics, and behavioural tests reveal that CP inhibits AChE in a manner similar to the positive control isoxazole (AChE inhibitor). CP improves avoidance response retention in adult zebrafish compared to isoxazole. Isoxazole directly binds Ser203 of the catalytic triad on the human AChE, which catalyses the breakdown of acetylcholine. The active components of CP – scopoletin and kaempferol – were found to bind to another amino acid of the catalytic triad, the anionic subsite of the catalytic centre, i.e., His447, and the peripheral anionic site. CP combined with scopolamine enhances AChE inhibition and depletes ACh levels, similar to isoxazole combined with scopolamine. However, both combinations have adverse effects on the peripheral cholinergic system. In summary, **Appendix section 14.2** suggests investigating CP for alleviating cognitive deficits in Alzheimer's disease.

11.4 Uncovering the disease-associated factors underlying drug interactions: a novel methodology to examine cross-drug target network relationships

In Chapter 10, the methodology detailed in section 1.5.2.4 helped identify the potential mechanisms of adverse drug reactions (ADRs) in comorbid diseases, i.e., pairs of comorbid diseases ('disease A' and 'disease B'), and their FDA-approved drugs. The study examined three pairs of non-comorbid diseases (negative controls) and six pairs of comorbid diseases. The four categories of drugs compiled for each disease pair based on their clinical activity included disease A drugs that are (i) contraindicated and (ii) not contraindicated in disease B, and disease B drugs that are (iii) contraindicated and (iv) not contraindicated in disease A. The study – which examined the four corresponding DTNs for each disorder pair – produced three key findings (**Table 2**).

Table 2: Disease network, pathway and tissue-level characterization of drugs that are contraindicated/not contraindicated in comorbid conditions. A tick mark (✓) has been used to indicate the close affiliation of a specific category of drug target network with specific disease protein sets, disease-associated pathways, and tissues. Disease PPI protein sets, pathways and tissues that are common to disease A and disease B, unique to disease A and unique to disease B have been marked in tables a, b, and c, respectively.

a.

Drug target networks	Disease PPI protein sets common to **A** and **B**	Pathways common to **A** and **B**	Tissues common to **A** and **B**
Disease **A** drugs **_not_** contraindicated in Disease **B**	✓	✓	✓
Disease **A** drugs contraindicated in Disease **B**	✓	✓	✓

b.

Drug target networks	Disease PPI protein sets unique to **A**	Pathways unique to **A**	Tissues unique to **A**
Disease **A** drugs **_not_** contraindicated in Disease **B**			
Disease **A** drugs contraindicated in Disease **B**	✓	✓	✓

c.

Drug target networks	Disease PPI protein sets unique to **B**	Pathways unique to **B**	Tissues unique to **B**
Disease **A** drugs **_not_** contraindicated in Disease **B**	✓	✓	✓
Disease **A** drugs contraindicated in Disease **B**			

First, disease B-associated PPI networks, pathways, and tissues showed enrichment for the DTNs of disease A drugs not contraindicated in disease B. This finding was contrary to the expected preferential affiliation of these DTNs with biological modalities of disease A, assuming that for a drug to be specifically active against disease A without aggravating a comorbid disease B, it had to reverse the phenotypes associated with disease A. In this model, phenotypes of disease B were considered as 'off-targets' in line with the principles of conventional pharmacology, in which unintended effects of the drugs were attributed to interaction with pathways inconsequential to disease A pathology (i.e., the signalling pathways relevant to disease B) (Chan and Loscalzo, 2012). Our findings to the contrary indicate that the mechanisms underlying the pathology of the comorbid disease B may contribute to the therapeutic alleviation of disease A. Two disorders can emerge and develop interdependently based on etiological associations. Future studies should concentrate on such etiological models of comorbidity (Valderas et al., 2009). For example, the risk factors of disease B could either influence disease A development directly or through correlation with the risk factors of disease A, according to the 'heterogeneity' and 'associated risk factors' models,

respectively (Valderas et al., 2009). This model could explain why our study connected disease B-associated PPI networks, pathways, and tissues with disease A drugs not contraindicated in disease B. For example, the alterations in disease B-associated genes may perturb pathways in specific tissues, whose counteraction by disease A drugs could help alleviate the symptoms of disease A.

Second, disease A-associated PPI networks, pathways, and tissues showed enrichment for the DTNs of disease A drugs contraindicated in disease B. Although further studies are required to examine the basis of this finding, it suggests that contraindications may arise when drugs used in disease A are highly specific to the biological modalities associated with disease A, i.e., disease A-associated PPI network, pathway, and tissue. Therefore, rational drug development should involve causative and correlational influences of comorbid conditions (i.e., disease B).

Third, supporting previous findings, both categories of drugs used to treat primary conditions (whether contraindicated or not in a comorbid condition) showed enrichment for proteins shared between the two disease networks, and it could be challenging to delineate the separate mechanisms underlying the two outcomes. Historically, drug design relied on findings from studies that described genetic and pharmacological modulation of specific targets and pathways, which elicited measurable changes in pathophenotypes (Chan and Loscalzo, 2012). This framework suggested that side effects arise from unintended manipulation of 'off-targets' in signalling pathways unrelated to the disease. However, both beneficial and adverse outcomes of drug treatment in complex disorders (and their distinct pathophenotypes) may arise from shared effectors and pathways, albeit active in varied combinations in specific cells and tissues (Chan and Loscalzo, 2012). Future analysis should focus on biological variables that differentially affect the functions of such shared proteins, e.g., their cellular, pathway, and tissue landscapes.

In summary, the findings suggest that the pathway membership and the tissue-specificity of the DTNs and their overlap with disease PPI networks will influence contraindications in comorbidities. Examining these biological modalities is critical for rational drug development and minimizing adverse events. The results from the study have therapeutic applications and could directly benefit future assessments of drug contraindications in individuals with comorbidities. The methodology of the study can be applied to examine distinct comorbid pairs or to all known pairs of comorbid and non-comorbid disorders.

The generalisability of the methodologies that helped derive the results detailed in section 11.3 and section 11.4 – namely, the repurposable drugs and the relationship of drug action with disease-associated biological factors – holds promise for their integration into the clinical developmental pipeline, as shown in **Fig. 6**. Information on drug safety and drug efficacy for the majority of repurposable drugs will be available from studies involving preclinical models and early-stage trials. Therefore, the findings from drug repurposing efforts could directly inform phase I trials focused on evaluating drug safety and phase II trials examining drug efficacy. The integration of multi-omics datasets with DTNs, as seen in **Chapter 10**, plays a pivotal role in unravelling the mechanisms underlying the action of drugs. Such comprehensive analyses will help characterise their therapeutic and non-therapeutic effects during phase I trials. They will also help understand the factors regulating drug efficacy in phase II trials.

11.5 The interactome as an integration platform: revealing the biological themes underlying heterogeneous, sparse and cross-species disease-associated gene sets

The primary benefit of the interactome framework over the GWA framework lies in its ability to integrate disease-associated data from diverse sources. This thesis highlights this integration power in three distinct ways.

First, the network framework helped localise the heterogeneous genetic structures of several diseases to specific interactomic subspaces and reveal the broad biological themes underlying them. The HLHS study in **Chapter 3**, the mesothelioma studies in **Chapter 4** and **Chapter 5**, and the COVID-19 study in **Chapter 6** contain striking examples. To seed the HLHS interactome, we used 74 genes harbouring HLHS-associated mutations identified from eight independent HLHS mutant mouse lines (Liu et al., 2017, Li et al., 2015). Notably, none of the mouse lines shared any HLHS-associated variations. Our study defined a coherent network of HLHS genes consisting of 248 known interactors and 377 novel interactors of 72 HLHS genes. Genes constitutively expressed in all the tissues or 'housekeeping' genes comprised more than 60% of the network, including 51% of the HLHS-associated seed genes. This preponderance of housekeeping genes in the HLHS interactome suggested that the genetic heterogeneity in the HLHS mouse lines stemmed from the limited transmission of mutations in these genes regulating essential cellular functions, whose perturbation may result in lethality or reduction in reproductive fitness (Zaidi and Brueckner, 2017).

In **Chapter 4** and **Chapter 5**, diverse gene sets derived from genetic, transcriptomic, and pharmacological datasets helped construct the interactomes of two mesothelioma subtypes, i.e., MPM and MPeM. The subtypes differ in their association with germline mutations, history of asbestos exposure, and post-operative complications (Robinson and Lake, 2005) and have different median survival rates (Amin et al., 2018). However, in **Chapter 5**, we showed that 989 genes co-occurred commonly in MPM and MPeM interactomes, which is an approximately 3-fold higher enrichment of high statistical significance than expected. Further, this sub-network contained an intricately interconnected network of PPIs containing 38 MPM- and MPeM-associated seed genes. These results helped define a shared interactomic subspace underlying the two distinct mesothelioma subtypes showing genetic and environmental heterogeneity.

In the COVID-19 study in **Chapter 6**, 332 host proteins interacting with SARS-CoV-2 viral proteins were used to construct the SARS-CoV-2-modulated host protein interactome (Gordon et al., 2020). This protein set had limited congruence with the genes differentially expressed upon SARS-CoV-2 infection in the A549 cell line (Blanco-Melo et al., 2020b). However, on the development of the host protein interactome, it became clear that despite an overlap of only two genes between the datasets, several differentially expressed genes were direct interactors of several host proteins, and 808 intermediate interactors connected 108 differentially expressed genes with 285 host proteins. Furthermore, analysis of the shared interactomic space of the two datasets revealed the enrichment of several immune-related pathways. Similarly, SARS-CoV (Pfefferle et al., 2011) and SARS-CoV-2 shared only four host targets. However, interactome analysis showed that several host proteins targeted by SARS-CoV-2 directly interacted with several host proteins targeted by SARS-CoV. Additionally, intermediate interactors connected 29 SARS-CoV-modulated host proteins with 249 SARS-CoV-2-modulated host proteins. Lastly, this shared interactomic space showed enrichment for

proteins localised to the mitochondria and involved in regulating mitochondrion organization and mitophagy, suggesting that these viruses could commonly target the mitochondria (Khan et al., 2015). It is important to note that integrative frameworks like the interactome are particularly essential during global health crises such as COVID-19. In these scenarios, generating clinically actionable results is possible only by examining shared biology with other viral agents by rapidly integrating emerging and existing multi-omics disease data.

Second, in addition to integrating heterogeneous data, the interactome model can reveal the broader genomic context of a disease, even when only a limited number of disease-associated genes are known, as seen with *GPX4* in **Chapter 7**. This ability becomes particularly valuable for Mendelian disorders, as demonstrated in this chapter. Specifically, exploring the interdependencies between the causative gene and other genes with potential disease associations within the interactome will help comprehend the phenotypic complexity of the disorder.

Lastly, the starting point for interactome construction could be any accessible cross-species genetic data mapped to orthologous human data, e.g., the ciliary proteins derived from Chlamydomonas proteome data (Pazour et al., 2005) in **Chapter 2**, the genes differentially expressed in the two striatal compartments of various species (rat, mouse, ferret, cat, monkey, and human) (Crittenden and Graybiel, 2011) in **Chapter 9**, and the HLHS-associated genes recovered from a large-scale mouse mutagenesis screen (Liu et al., 2017, Li et al., 2015) in **Chapter 3**.

Independent validation of the disease interactomes using multi-omics human disease data is crucial to ascertain their biological relevance, given the heterogeneous, and (as highlighted above) often cross-species, origins of the seed disease-associated genes. In **Chapters 2-6**, the interactomes and related multi-omics datasets intersected with high statistical significance (**Table 3**), helping to retrieve context-sensitive sub-networks from the disease interactomes. These intersections confirmed the biological validity of the interactomes for two reasons. Firstly, the discovery of genes previously associated with the disease in other multi-omics studies within the interactomic landscape was unbiased. Secondly, these genes were closely connected to the curated set of genes that seeded the disease interactome. These close connections indicated the presence of interdependent functional relationships in the form of signalling pathways and cellular processes, often previously described in experimental investigations of the disease.

Table 3: The multi-omics datasets used for interactome validation split into 4 data types, namely, genetic, transcriptomic, methylation and proteomic data.

Chapter	Interactome	Genetic dataset	Transcriptomic dataset	Methylation dataset	Proteomic dataset
Chapter 2	Ciliary protein interactome	Regulators of the ciliary sonic hedgehog pathway identified in a CRISPR genetic screen(Breslow et al., 2018)	Genes differentially expressed in bronchial biopsies of primary ciliary dyskinesia patients (Geremek et al., 2014), gene expression datasets of patients with MDD (Lanz et al., 2015), SCZ (Maycox et al., 2009), BD (Harris et	n/a	Proteome isolated from the primary cilia of mouse kidney cells

			al., 2008), ASD (Kong et al., 2012), Alzheimer's disease (Miller et al., 2013), Parkinson's disease, and non-syndromic intellectual disability		
Chapter 3	Hypoplastic left heart syndrome interactome	n/a	Genes differentially expressed in cardiomyocytes differentiated from the induced pluripotent stem cells of HLHS patients (Yang et al., 2017, Theis et al., 2020) and in the HLHS-affected right ventricle of patients (Ricci et al., 2012)	n/a	n/a
Chapter 4	Malignant pleural mesothelioma interactome	n/a	Differential transcriptomic datasets of tumours (and control adjacent tissues), either partitioned into subtypes or not (Crispi et al., 2009, Suraokar et al., 2014) and in comparison with etiologically different thoracic tumours (De Rienzo et al., 2013)	Differential methylation profiles in tumours versus normal adjacent tissues (Christensen et al., 2009)	Proteomes in cancer cell lines and tissues (Greening et al., 2016, Großerueschkamp et al., 2017)
Chapter 5	Malignant peritoneal mesothelioma interactome	n/a	Differential gene expression profiles of pre-clinical models and human cancer cell lines and tumour specimens (Fridlender et al., 2012, Kim et al., 2006, Blackshear et al., 2014, Shukla et al., 2009, Dragon et al., 2015)	n/a	n/a
Chapter 6	SARS-CoV-2-modulated host protein interactome	n/a	Gene expression profiles induced by SARS-CoV-2 in human cell lines (Blanco-Melo et al., 2020b) and COVID-19 patients (Carapito et al., 2021, Arunachalam et al., 2020)	n/a	Protein biomarkers isolated from COVID-19 patients (Messner et al., 2020)

11.6 Circumventing the sparseness of the interactome: the need to augment the existing network with computationally predicted interactions

In **Chapters 2-7**, PPIs predicted using the HiPPIP computational model augmented the disease interactomes. The addition of these new PPIs helped link disease-associated genes across various

high-throughput studies in novel ways and gain novel insights into disease biology as described in section 11.1 and the functional associations of under-characterised proteins such as *GPX4* (**Chapter 7**). **Chapter 6** contains a specific example of how the addition of predicted PPIs helped gain additional biological insights. With the inclusion of novel PPIs, 323 proteins in the SARS-CoV-2-modulated host protein interactome could be organised into 21 topological modules, whereas without novel PPIs, 199 proteins could be organised into 18 modules. The inclusion of novel PPIs yielded a higher scaled modularity score compared to excluding them. Importantly, the incorporation of novel PPIs revealed additional topological modules enriched for specific biological functions, such as cilium organisation, which were not identified using known PPIs alone. Overall, the integration of novel PPIs significantly improved our understanding of COVID-related biological processes by unveiling hidden modules.

The disease interactomes in each of these chapters have been made available as downloadable files in various formats (Cytoscape and Excel files) to facilitate their investigation in conjunction with transcriptomic/proteomic data. Seventeen of the PPIs predicted by this model have been tested experimentally and shown to be true PPIs, namely, 8 PPIs validated by Co-IP (Zhu et al., 2014, Ganapathiraju et al., 2016), 5 PPIs validated by in vitro pull-down and mass spectrometry (**Chapter 4**) and 4 PPIs validated by co-localization (Ganapathiraju et al., 2016). Although large-scale experimental confirmation is necessary to ascertain the reliability of the predicted novel PPIs in each of the disease interactomes in **Chapters 2-7**, studies indicate that HiPPIP outperforms other state-of-the-art algorithms in PPI prediction (Dunham and Ganapathiraju, 2022).The novel PPIs predicted by HiPPIP show only limited overlap with recently released interactome maps (Luck et al., 2020b, Huttlin et al., 2020) (**Chapter 6**). This could be due to the distinct nature of the various high-throughput biotechnological methods employed in interactome mapping, with each capturing different portions of the interactome due to their complementary strengths (Luck et al., 2020b). Therefore, there is a clear need to develop more comprehensive approaches that encompass both experimental and computational methods to effectively map disease interactomes. Furthermore, **Chapters 2-7** explored the intersections between different sets of disease-associated genes and proteins identified using various mapping techniques. While the overlap between these sets was marginal, extensive interconnections were discovered through direct and intermediate interactors, which were either experimentally determined or computationally predicted. This finding underscores the value of combining different techniques and utilising computational methods to construct an integrated view of the disease interactome.

11.7 Thesis overview and conclusion

The advancements in disease gene discovery leading up to the post-genomic era and the progress in drug discovery leading up to the systems therapeutics era, naturally led to the emergence of network biology concepts (Barabási and Albert, 1999, Albert and Barabási, 2002, Jeong et al., 2001, Barabási et al., 2011, Barabasi and Oltvai, 2004). These concepts offered a better understanding of the complex mechanisms underlying disease development and manifestation. Despite significant advancements in network biology, the formalisms failed to get incorporated into conventional disease and drug investigations due to the lack of an integrated conceptual framework to examine interactomes (Sun et al., 2009, Sharma et al., 2015a, Sharma et al., 2018, Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a). Previous studies also failed to adequately explore the complex web of molecular interactions underlying diseases, resulting in a lack of biologically insightful and clinically actionable results (Sun et al., 2009, Sharma et al., 2015a, Sharma et al., 2018, Sakai et al., 2011, Lim et al., 2006, Ganapathiraju et al., 2016, Malavia et al., 2017a). The biological contexts of the molecular interactions, e.g., the tissue-level and cellular contexts, were not considered in the analyses, limiting the biological relevance of several network biology studies (Goh et al., 2007, Hidalgo et al., 2009). Additionally, due to the challenges involved in experimental validation of PPIs, the human interactome remains incomplete, requiring alternative computational methods to fill the gaps (Kotlyar et al., 2015, Hopf et al., 2014, Emamjomeh et al., 2014a, Garzón et al., 2016, You et al., 2013, Jia et al., 2015, Li and Ilie, 2017, Deng et al., 2003, Raja et al., 2013). Furthermore, the higher-level aspects of disease causation and drug action, encoded within the genetic structures of diseases, such as cross-disorder relationships and the biological factors influencing drug interactions in multiple disorders (Goh et al., 2007, Cheng et al., 2019), have received relatively less attention due to a lack of easily adaptable conceptual advances.

My thesis addressed these five limitations by introducing a conceptual framework to investigate disease interactomes and DTNs, individually and in relation to each other. This framework provided valuable insights into the biological mechanisms underlying specific disorders and the relationships between multiple diseases, including heart diseases, cancers, viral infections, skeletal disorders, and neurological and psychiatric disorders. Its success in these six disease classes showcases its potential applicability to other diseases in the future. The methodology outlined in the thesis to extract disease-disease relationships can be adapted to uncover subgroups and common themes among different disorders and refine the categorisation of diseases based on their genetic structures. It involved reframing the multigenic complexities of diseases into multivariate data analytical problems, facilitating the study of various biological levels and their contributions to disease aetiology. Furthermore, the framework yielded clinically actionable results, such as repurposable drugs for a psychiatric disorder, two cancers, COVID-19, and a skeletal disorder, and insights into how the aetiological associations between six pairs of comorbidities affect drug action in these diseases. The methodologies that helped extract these results will provide valuable insights into drug safety and efficacy, highlighting their potential for integration into the clinical development pipeline. Lastly, the framework allowed the extraction of context-specific sub-networks from the interactomes by intersecting them with disease-associated multi-omics datasets. In this process, the framework additionally demonstrated its ability to integrate diverse, sparse, and cross-species disease-associated genetic data. It also incorporated computationally predicted PPIs into the existing network of experimentally verified interactions to overcome the limitations of interactome sparseness.

In summary, the thesis proposes a comprehensive interactome-based framework to uncover hidden patterns within the polygenic architecture of diseases. This framework involves piecing together an integrated view of disease interactomes and DTNs at various levels of biological organisation. The effectiveness of the framework has been ascertained in multiple disorders. Therefore, it can be adapted in future studies to model disease biology and drug action using emerging multi-omics disease data and advance our understanding of many diseases and their treatment.

12 Future work

The rapidly emerging disease-associated multi-omics data will require interpretation to understand their functional implications on diseases using integrative frameworks, such as the human interactome. This thesis addresses the central problem of the lack of an integrated conceptual framework to examine such multi-omics data. The transformation of the proposed pipeline into a fully automated computational pipeline will help resolve the historical outpacing of functional interpretation of data by novel data generation (Woodsmith and Stelzl, 2017) to some extent. This transformation will enable the generation of biological insights into disease mechanisms and the identification of repurposable drugs from this data.

The strategies described in the thesis will help investigate the polygenic landscape of Mendelian disorders with complex phenotypes, which are not accounted for by current theories focusing on their monogenic underpinnings, in future studies. This examination will contribute to a more nuanced understanding of the regulatory gene neighbourhoods of single causative genes. Although the studies in this thesis described the extraction of biologically active sub-networks from disease interactomes for further examination, future studies should adopt a better approach, which involves stratifying the PPIs based on the level of disease-associated evidence associated with them, the individual genes coding for the interacting proteins, or the proteins themselves. Furthermore, the interactomes should be able to model dynamic events that affect disease development, such as PPI perturbations and gene-environment interactions. Additionally, integrating large-scale spatial and temporal transcriptomic data into the interactome will help identify hidden regional and developmental modules that influence disease development. Future works will expand the methodology for comparative interactome analysis to more groups of disorders and focus on revealing their relationships, potentially prompting re-evaluations of current disease classifications. The findings on the effects of aetiologically related conditions on the clinical activity of drugs can inform the drug repurposing methodology. This integrated perspective will help identify candidate drugs with minimal risk of adverse events and a reduced tendency to worsen comorbid conditions in future studies.

Interactome analysis is a valuable tool for gaining biological insights into neuropsychiatric disorders by bypassing the various levels in the linear paradigm of biological discovery (Totah, 2016). Hence, highlighting the need to integrate the proposed pipeline into routine examinations of the genetic structures of psychiatric disorders will be another focal point for future works. Finally, the interactomic framework can contribute to our preparedness for viral diseases with the potential to expand into global pandemics, such as COVID-19. Analysing pan-viral interactomes (as shown in (Ghavasieh et al., 2021)) – incorporating both experimentally verified and computationally predicted PPIs of virus-modulated host proteins – will help us understand the proximity of virus families and the shared and unique themes underlying their host invasion mechanisms. It will also help us pre-emptively discover new drugs or reposition existing drugs for viral diseases.

13. References

ALBERT, R. & BARABÁSI, A.-L. 2002. Statistical mechanics of complex networks. Reviews of modern physics, 74, 47.

ALSOUFI, B., MORI, M., GILLESPIE, S., SCHLOSSER, B., SLESNICK, T., KOGON, B., KIM, D., SACHDEVA, R. & KANTER, K. 2015. Impact of patient characteristics and anatomy on results of Norwood operation for hypoplastic left heart syndrome. The Annals of thoracic surgery, 100, 591-598.

ALTSHULER, D., DALY, M. J. & LANDER, E. S. 2008. Genetic mapping in human disease. science, 322, 881-888.

ALVAREZ RETUERTO, A. I., CANTOR, R. M., GLEESON, J. G., USTASZEWSKA, A., SCHACKWITZ, W. S., PENNACCHIO, L. A. & GESCHWIND, D. H. 2008. Association of common variants in the Joubert syndrome gene (AHI1) with autism. Hum Mol Genet, 17, 3887-96.

AMIN, W., LINKOV, F., LANDSITTEL, D. P., SILVERSTEIN, J. C., BASHARA, W., GAUDIOSO, C., FELDMAN, M. D., PASS, H. I., MELAMED, J. & FRIEDBERG, J. S. 2018. Factors influencing malignant mesothelioma survival: a retrospective review of the National Mesothelioma Virtual Bank cohort. F1000Research, 7.

ANDERSON, K. M., KRIENEN, F. M., CHOI, E. Y., REINEN, J. M., YEO, B. T. & HOLMES, A. J. 2018. Gene expression links functional networks across cortex and striatum. Nature communications, 9, 1-14.

ARUNACHALAM, P. S., WIMMERS, F., MOK, C. K. P., PERERA, R. A., SCOTT, M., HAGAN, T., SIGAL, N., FENG, Y., BRISTOW, L. & TSANG, O. T.-Y. 2020. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science, 369, 1210-1220.

ATHAUDA, D., MACLAGAN, K., SKENE, S. S., BAJWA-JOSEPH, M., LETCHFORD, D., CHOWDHURY, K., HIBBERT, S., BUDNIK, N., ZAMPEDRI, L. & DICKSON, J. 2017. Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. The Lancet, 390, 1664-1675.

AYGUN, C., CELIK, F. C., NURAL, M. S., AZAK, E., KUCUKODUK, Ş., OGUR, G. & INCESU, L. 2012. Simplified gyral pattern with cerebellar hypoplasia in Sedaghatian type spondylometaphyseal dysplasia: a clinical report and review of the literature. American Journal of Medical Genetics Part A, 158, 1400-1405.

BAGGE, C. N., HENDERSON, V. W., LAURSEN, H. B., ADELBORG, K., OLSEN, M. & MADSEN, N. L. 2018. Risk of dementia in adults with congenital heart disease: population-based cohort study. Circulation, 137, 1912-1920.

BALLARD, C. 2002. Advances in the treatment of Alzheimer's disease: benefits of dual cholinesterase inhibition. European neurology, 47, 64-70.

BARABÁSI, A.-L. & ALBERT, R. 1999. Emergence of scaling in random networks. science, 286, 509-512.

BARABÁSI, A.-L., GULBAHCE, N. & LOSCALZO, J. 2011. Network medicine: a network-based approach to human disease. Nature reviews genetics, 12, 56-68.

BARABASI, A.-L. & OLTVAI, Z. N. 2004. Network biology: understanding the cell's functional organization. Nature reviews genetics, 5, 101-113.

BARADARAN-HERAVI, A., CHO, K. S., TOLHUIS, B., SANYAL, M., MOROZOVA, O., MORIMOTO, M., ELIZONDO, L. I., BRIDGEWATER, D., LUBIENIECKA, J. & BEIRNES, K. 2012. Penetrance of biallelic SMARCAL1 mutations is associated with environmental and genetic

disturbances of gene expression. Human molecular genetics, 21, 2572-2587.

BARRETT, T., WILHITE, S. E., LEDOUX, P., EVANGELISTA, C., KIM, I. F., TOMASHEVSKY, M., MARSHALL, K. A., PHILLIPPY, K. H., SHERMAN, P. M. & HOLKO, M. 2012. NCBI GEO: archive for functional genomics data sets—update. Nucleic acids research, 41, D991-D995.

BASPINAR, A., CUKUROGLU, E., NUSSINOV, R., KESKIN, O. & GURSOY, A. 2014. PRISM: a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes. Nucleic acids research, 42, W285-W289.

BEADLE, G. W. & TATUM, E. L. 1941. Genetic control of biochemical reactions in Neurospora. Proceedings of the National Academy of Sciences, 27, 499-506.

BECERRA, J. E., KHOURY, M. J., CORDERO, J. F. & ERICKSON, J. D. 1990. Diabetes mellitus during pregnancy and the risks for specific birth defects: a population-based case-control study. Pediatrics, 85, 1-9.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57, 289-300.

BIHAQI, S. W., SHARMA, M., SINGH, A. P. & TIWARI, M. 2009. Neuroprotective role of Convolvulus pluricaulis on aluminium induced neurotoxicity in rat brain. Journal of ethnopharmacology, 124, 409-415.

BIHAQI, S. W., SINGH, A. P. & TIWARI, M. 2011. In vivo investigation of the neuroprotective property of Convolvulus pluricaulis in scopolamine-induced cognitive impairments in Wistar rats. Indian journal of pharmacology, 43, 520.

BLACKSHEAR, P. E., PANDIRI, A. R., TON, T.-V. T., CLAYTON, N. P., SHOCKLEY, K. R., PEDDADA, S. D., GERRISH, K. E., SILLS, R. C. & HOENERHOFF, M. J. 2014. Spontaneous mesotheliomas in F344/N rats are characterized by dysregulation of cellular growth and immune function pathways. Toxicologic pathology, 42, 863-876.

BLANCO-MELO, D., NILSSON-PAYANT, B. E., LIU, W.-C., MØLLER, R., PANIS, M., SACHS, D., ALBRECHT, R. A. & TENOEVER, B. R. 2020a. SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. bioRxiv, 2020.03.24.004655.

BLANCO-MELO, D., NILSSON-PAYANT, B. E., LIU, W.-C., UHL, S., HOAGLAND, D., MØLLER, R., JORDAN, T. X., OISHI, K., PANIS, M. & SACHS, D. 2020b. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell.

BLASCHE, S. & KOEGL, M. 2013. Analysis of Protein–Protein Interactions Using LUMIER Assays. Virus-Host Interactions: Methods and Protocols, 17-27.

BOTSTEIN, D. & RISCH, N. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature genetics, 33, 228-237.

BOTSTEIN, D., WHITE, R. L., SKOLNICK, M. & DAVIS, R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics, 32, 314.

BRANON, T., HAN, S. & TING, A. 2017. Beyond Immunoprecipitation: Exploring New Interaction Spaces with Proximity Biotinylation. ACS Publications.

BRAVO, À., PIÑERO, J., QUERALT, N., RAUTSCHKA, M. & FURLONG, L. I. BeFree: a text mining system to extract relations between genes, diseases and drugs for translational research. SMBM 2014, 79.

BREEN, G., LI, Q., ROTH, B. L., O'DONNELL, P., DIDRIKSEN, M., DOLMETSCH, R., O'REILLY,

P. F., GASPAR, H. A., MANJI, H. & HUEBEL, C. 2016. Translating genome-wide association findings into new therapeutics for psychiatry. Nature neuroscience, 19, 1392-1396.

BRESLOW, D. K., HOOGENDOORN, S., KOPP, A. R., MORGENS, D. W., VU, B. K., KENNEDY, M. C., HAN, K., LI, A., HESS, G. T. & BASSIK, M. C. 2018. A CRISPR-based screen for Hedgehog signaling provides insights into ciliary function and ciliopathies. Nature genetics, 50, 460.

BREUNIG, J. J., SARKISIAN, M. R., ARELLANO, J. I., MOROZOV, Y. M., AYOUB, A. E., SOJITRA, S., WANG, B., FLAVELL, R. A., RAKIC, P. & TOWN, T. 2008. Primary cilia regulate hippocampal neurogenesis by mediating sonic hedgehog signaling. Proceedings of the National Academy of Sciences, 105, 13127-13132.

BROUWERS, L., ISKAR, M., ZELLER, G., VAN NOORT, V. & BORK, P. 2011. Network neighbors of drug targets contribute to drug side-effect similarity. PloS one, 6, e22187.

BUENO, R., STAWISKI, E. W., GOLDSTEIN, L. D., DURINCK, S., DE RIENZO, A., MODRUSAN, Z., GNAD, F., NGUYEN, T. T., JAISWAL, B. S. & CHIRIEAC, L. R. 2016. Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. Nature genetics, 48, 407-420.

BUNIELLO, A., MACARTHUR, J. A. L., CEREZO, M., HARRIS, L. W., HAYHURST, J., MALANGONE, C., MCMAHON, A., MORALES, J., MOUNTJOY, E. & SOLLIS, E. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic acids research, 47, D1005-D1012.

CACHEIRO, P., HAENDEL, M. A., SMEDLEY, D., CONSORTIUM, I. M. P. & INITIATIVE, M. 2019. New models for human disease from the International Mouse Phenotyping Consortium. Mammalian Genome, 30, 143-150.

CALDERWOOD, M. A., VENKATESAN, K., XING, L., CHASE, M. R., VAZQUEZ, A., HOLTHAUS, A. M., EWENCE, A. E., LI, N., HIROZANE-KISHIKAWA, T. & HILL, D. E. 2007. Epstein–Barr virus and virus human protein interaction maps. Proceedings of the National Academy of Sciences, 104, 7606-7611.

CAMERON, T. L., BELL, K. M., TATARCZUCH, L., MACKIE, E. J., RAJPAR, M. H., MCDERMOTT, B. T., BOOT-HANDFORD, R. P. & BATEMAN, J. F. 2011. Transcriptional profiling of chondrodysplasia growth plate cartilage reveals adaptive ER-stress networks that allow survival but disrupt hypertrophy. PloS one, 6, e24600.

CAMPILLOS, M., KUHN, M., GAVIN, A.-C., JENSEN, L. J. & BORK, P. 2008. Drug target identification using side-effect similarity. Science, 321, 263-266.

CAO, C. & MOULT, J. 2014. GWAS and drug targets. BMC genomics, 15, 1-14.

CARAPITO, R., LI, R., HELMS, J., CARAPITO, C., GUJJA, S., ROLLI, V., GUIMARAES, R., MALAGON-LOPEZ, J., SPINNHIRNY, P. & LEDERLE, A. 2021. Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. Science Translational Medicine, eabj7521.

CARBONE, M., ADUSUMILLI, P. S., ALEXANDER, H. R., JR., BAAS, P., BARDELLI, F., BONONI, A., BUENO, R., FELLEY-BOSCO, E., GALATEAU-SALLE, F., JABLONS, D., MANSFIELD, A. S., MINAAI, M., DE PERROT, M., PESAVENTO, P., RUSCH, V., SEVERSON, D. T., TAIOLI, E., TSAO, A., WOODARD, G., YANG, H., ZAUDERER, M. G. & PASS, H. I. 2019a. Mesothelioma: Scientific clues for prevention, diagnosis, and therapy. CA Cancer J Clin, 69, 402-429.

CARBONE, M., ADUSUMILLI, P. S., ALEXANDER JR, H. R., BAAS, P., BARDELLI, F., BONONI,

A., BUENO, R., FELLEY-BOSCO, E., GALATEAU-SALLE, F. & JABLONS, D. 2019b. Mesothelioma: scientific clues for prevention, diagnosis, and therapy. CA: a cancer journal for clinicians, 69, 402-429.

CEDRES, S., MONTERO, M., MARTINEZ, P., MARTINEZ, A., RODRIGUEZ-FREIXINOS, V., TORREJON, D., GABALDON, A., SALCEDO, M., Y CAJAL, S. R. & FELIP, E. 2012. Exploratory analysis of activation of PTEN–PI3K pathway and downstream proteins in malignant pleural mesothelioma (MPM). Lung Cancer, 77, 192-198.

CERAMI, E., DEMIR, E., SCHULTZ, N., TAYLOR, B. S. & SANDER, C. 2010. Automated network analysis identifies core pathways in glioblastoma. PloS one, 5.

CHAN, S. Y. & LOSCALZO, J. 2012. The emerging paradigm of network medicine in the study of human disease. Circulation research, 111, 359-374.

CHANG, D. & KEINAN, A. 2014. Principal component analysis characterizes shared pathogenetics from genome-wide association studies. PLoS computational biology, 10, e1003820.

CHATTOPADHYAY, A. & GANAPATHIRAJU, M. K. 2017. Demonstration Study: A Protocol to Combine Online Tools and Databases for Identifying Potentially Repurposable Drugs. Data, 2, 15.

CHAUSSABEL, D., ALLMAN, W., MEJIAS, A., CHUNG, W., BENNETT, L., RAMILO, O., PASCUAL, V., PALUCKA, A. K. & BANCHEREAU, J. 2005. Analysis of significance patterns identifies ubiquitous and disease-specific gene-expression signatures in patient peripheral blood leukocytes. Annals of the New York Academy of Sciences, 1062, 146-154.

CHAUTARD, E., FATOUX-ARDORE, M., BALLUT, L., THIERRY-MIEG, N. & RICARD-BLUM, S. 2010. MatrixDB, the extracellular matrix interaction database. Nucleic acids research, 39, D235-D240.

CHEN, F., SHI, Q., PEI, F., VOGT, A., PORRITT, R. A., GARCIA JR, G., GOMEZ, A. C., CHENG, M. H., SCHURDAK, M. E. & LIU, B. 2021. A systems-level study reveals host-targeted repurposable drugs against SARS-CoV-2 infection. Molecular Systems Biology, 17, e10239.

CHENG, F., KOVÁCS, I. A. & BARABÁSI, A.-L. 2019. Network-based prediction of drug combinations. Nature communications, 10, 1-11.

CHIRAC, P., MAILLET, D., LEPRÊTRE, F., ISAAC, S., GLEHEN, O., FIGEAC, M., VILLENEUVE, L., PÉRON, J., GIBSON, F. & GALATEAU-SALLÉ, F. 2016. Genomic copy number alterations in 33 malignant peritoneal mesothelioma analyzed by comparative genomic hybridization array. Human pathology, 55, 72-82.

CHIU, I. M., MORIMOTO, E. T., GOODARZI, H., LIAO, J. T., O'KEEFFE, S., PHATNANI, H. P., MURATET, M., CARROLL, M. C., LEVY, S. & TAVAZOIE, S. 2013. A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. Cell reports, 4, 385-401.

CHOI, H., LARSEN, B., LIN, Z.-Y., BREITKREUTZ, A., MELLACHERUVU, D., FERMIN, D., QIN, Z. S., TYERS, M., GINGRAS, A.-C. & NESVIZHSKII, A. I. 2011. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nature methods, 8, 70-73.

CHOI, S. J., MUKAI, J., KVAJO, M., XU, B., DIAMANTOPOULOU, A., PITYCHOUTIS, P. M., GOU, B., GOGOS, J. A. & ZHANG, H. 2017. A Schizophrenia-related deletion leads to KCNQ2-dependent abnormal dopaminergic modulation of prefrontal cortical interneuron activity. Cerebral Cortex, 28, 2175-2191.

CHRISTENSEN, B. C., MARSIT, C. J., HOUSEMAN, E. A., GODLESKI, J. J., LONGACKER, J.

L., ZHENG, S., YEH, R.-F., WRENSCH, M. R., WIEMELS, J. L. & KARAGAS, M. R. 2009. Differentiation of lung adenocarcinoma, pleural mesothelioma, and nonmalignant pulmonary tissues using DNA methylation profiles. Cancer research, 69, 6315-6321.

CHU, Q., RATHORE, A., DIEDRICH, J. K., DONALDSON, C. J., YATES III, J. R. & SAGHATELIAN, A. 2017. Identification of Microprotein–Protein Interactions via APEX Tagging. Biochemistry.

COELHO, E. D., ARRAIS, J. P., MATOS, S., PEREIRA, C., ROSA, N., CORREIA, M. J., BARROS, M. & OLIVEIRA, J. L. 2014. Computational prediction of the human-microbial oral interactome. BMC systems biology, 8, 24.

BREUER, K., FOROUSHANI, A. K., LAIRD, M. R., CHEN, C., SRIBNAIA, A., LO, R., WINSOR, G. L., HANCOCK, R. E., BRINKMAN, F. S. & LYNN, D. J. 2013. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic acids research, 41, D1228-D1233.

CONSORTIUM, C.-D. G. O. T. P. G. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. The Lancet, 381, 1371-1379.

CONSORTIUM, G. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348, 648-660.

CONSORTIUM, G. O. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic acids research, 32, D258-D261.

COSTANZO, M., VANDERSLUIS, B., KOCH, E. N., BARYSHNIKOVA, A., PONS, C., TAN, G., WANG, W., USAJ, M., HANCHARD, J. & LEE, S. D. 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science, 353, aaf1420.

CRISPI, S., CALOGERO, R. A., SANTINI, M., MELLONE, P., VINCENZI, B., CITRO, G., VICIDOMINI, G., FASANO, S., MECCARIELLO, R. & COBELLIS, G. 2009. Global gene expression profiling of human pleural mesotheliomas: identification of matrix metalloproteinase 14 (MMP-14) as potential tumour target. PLoS One, 4, e7016.

CRITTENDEN, J. R. & GRAYBIEL, A. M. 2011. Basal Ganglia disorders associated with imbalances in the striatal striosome and matrix compartments. Frontiers in neuroanatomy, 5, 59.

CROFT, D., MUNDO, A. F., HAW, R., MILACIC, M., WEISER, J., WU, G., CAUDY, M., GARAPATI, P., GILLESPIE, M. & KAMDAR, M. R. 2014. The Reactome pathway knowledgebase. Nucleic acids research, 42, D472-D477.

D'UDEKEM, Y., IYENGAR, A. J., GALATI, J. C., FORSDICK, V., WEINTRAUB, R. G., WHEATON, G. R., BULLOCK, A., JUSTO, R. N., GRIGG, L. E. & SHOLLER, G. F. 2014. Redefining expectations of long-term survival after the Fontan procedure: twenty-five years of follow-up from the entire population of Australia and New Zealand. Circulation, 130, S32-S38.

DE RIENZO, A., RICHARDS, W. G., YEAP, B. Y., COLEMAN, M. H., SUGARBAKER, P. E., CHIRIEAC, L. R., WANG, Y. E., QUACKENBUSH, J., JENSEN, R. V. & BUENO, R. 2013. Sequential binary gene ratio tests define a novel molecular diagnostic strategy for malignant pleural mesothelioma. Clinical Cancer Research, 19, 2493-2502.

DEMONTIS, D., WALTERS, R. K., MARTIN, J., MATTHEISEN, M., ALS, T. D., AGERBO, E., BALDURSSON, G., BELLIVEAU, R., BYBJERG-GRAUHOLM, J. & BÆKVAD-HANSEN, M. 2019. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature genetics, 51, 63.

DENG, M., ZHANG, K., MEHTA, S., CHEN, T. & SUN, F. 2003. Prediction of protein function using protein–protein interaction data. Journal of Computational Biology, 10, 947-960.

DEY, A., SESHASAYEE, D., NOUBADE, R., FRENCH, D. M., LIU, J., CHAURUSHIYA, M. S., KIRKPATRICK, D. S., PHAM, V. C., LILL, J. R. & BAKALARSKI, C. E. 2012. Loss of the tumor suppressor BAP1 causes myeloid transformation. Science, 337, 1541-1546.

DI SANTO, S. G., PRINELLI, F., ADORNI, F., CALTAGIRONE, C. & MUSICCO, M. 2013. A meta-analysis of the efficacy of donepezil, rivastigmine, galantamine, and memantine in relation to severity of Alzheimer's disease. Journal of Alzheimer's Disease, 35, 349-361.

DIETZ, H. C., CUTTING, C. R., PYERITZ, R. E., MASLEN, C. L., SAKAI, L. Y., CORSON, G. M., PUFFENBERGER, E. G., HAMOSH, A., NANTHAKUMAR, E. J. & CURRISTIN, S. M. 1991. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature, 352, 337-339.

DRAGON, J., THOMPSON, J., MACPHERSON, M. & SHUKLA, A. 2015. Differential susceptibility of human pleural and peritoneal mesothelial cells to asbestos exposure. Journal of cellular biochemistry, 116, 1540-1552.

DUDLEY, J. T., SIROTA, M., SHENOY, M., PAI, R. K., ROEDDER, S., CHIANG, A. P., MORGAN, A. A., SARWAL, M. M., PASRICHA, P. J. & BUTTE, A. J. 2011. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Science translational medicine, 3, 96ra76-96ra76.

DUFFY, Á., VERBANCK, M., DOBBYN, A., WON, H.-H., REIN, J. L., FORREST, I. S., NADKARNI, G., ROCHELEAU, G. & DO, R. 2020. Tissue-specific genetic features inform prediction of drug side effects in clinical trials. Science Advances, 6, eabb6242.

DUNHAM, B. & GANAPATHIRAJU, M. K. 2022. Benchmark Evaluation of Protein–Protein Interaction Prediction Algorithms. Molecules, 27, 41.

DUNHAM, W. H., MULLIN, M. & GINGRAS, A. C. 2012. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. Proteomics, 12, 1576-1590.

DURAN-FRIGOLA, M., MATEO, L. & ALOY, P. 2017. Drug repositioning beyond the low-hanging fruits. Current Opinion in Systems Biology, 3, 95-102.

EGBE, A., LEE, S., HO, D., UPPU, S. & SRIVASTAVA, S. 2014. Prevalence of congenital anomalies in newborns with congenital heart disease diagnosis. Annals of pediatric cardiology, 7, 86.

ELÇIOGLU, N. & HALL, C. M. 1998. Spondylometaphyseal dysplasia–sedaghatian type. American journal of medical genetics, 76, 410-414.

EMAMJOMEH, A., GOLIAEI, B., TORKAMANI, A., EBRAHIMPOUR, R., MOHAMMADI, N. & PARSIAN, A. 2014a. Protein-protein interaction prediction by combined analysis of genomic and conservation information. Genes & genetic systems, 89, 259-272.

EMAMJOMEH, A., GOLIAEI, B., ZAHIRI, J. & EBRAHIMPOUR, R. 2014b. Predicting protein–protein interactions between human and hepatitis C virus via an ensemble learning method. Molecular Biosystems, 10, 3147-3154.

ENGLISH, S. J., GAYATRI, N., ARTHUR, R. & CROW, Y. J. 2006. Sedaghatian spondylometaphyseal dysplasia with pachygyria and absence of the corpus callosum. American Journal of Medical Genetics Part A, 140, 1854-1858.

ERGÜN, A., LAWRENCE, C. A., KOHANSKI, M. A., BRENNAN, T. A. & COLLINS, J. J. 2007. A network biology approach to prostate cancer. Molecular systems biology, 3, 82.

EVANS, J. M., DONNELLY, L. A., EMSLIE-SMITH, A. M., ALESSI, D. R. & MORRIS, A. D. 2005. Metformin and reduced risk of cancer in diabetic patients. Bmj, 330, 1304-1305.

FANG, L., KARAKIULAKIS, G. & ROTH, M. 2020. Are patients with hypertension and diabetes

mellitus at increased risk for COVID-19 infection? Lancet Respir Med.

FARRELL, M., WERGE, T., SKLAR, P., OWEN, M. J., OPHOFF, R., O'DONOVAN, M. C., CORVIN, A., CICHON, S. & SULLIVAN, P. F. 2015. Evaluating historical candidate genes for schizophrenia. Molecular psychiatry, 20, 555-562.

FIELDS, S. & SONG, O.-K. 1989. A novel genetic system to detect protein–protein interactions. Nature, 340, 245-246.

FLIRI, A. F., LOGING, W. T., THADEIO, P. F. & VOLKMANN, R. A. 2005. Analysis of drug-induced effect patterns to link structure and side effects of medicines. Nature chemical biology, 1, 389-397.

FOSTER, J. M., RADHAKRISHNA, U., GOVINDARAJAN, V., CARREAU, J. H., GATALICA, Z., SHARMA, P., NATH, S. K. & LOGGIE, B. W. 2010. Clinical implications of novel activating EGFR mutations in malignant peritoneal mesothelioma. World journal of surgical oncology, 8, 88.

FRIDLENDER, Z. G., SUN, J., MISHALIAN, I., SINGHAL, S., CHENG, G., KAPOOR, V., HORNG, W., FRIDLENDER, G., BAYUH, R. & WORTHEN, G. S. 2012. Transcriptomic analysis comparing tumor-associated neutrophils with granulocytic myeloid-derived suppressor cells and normal neutrophils. PloS one, 7, e31524.

FUKUTO, T. R. 1990. Mechanism of action of organophosphorus and carbamate insecticides. Environmental health perspectives, 87, 245.

GALLAGHER, M. D. & CHEN-PLOTKIN, A. S. 2018. The post-GWAS era: from association to function. The American Journal of Human Genetics, 102, 717-730.

GANAPATHIRAJU, M. K., THAHIR, M., HANDEN, A., SARKAR, S. N., SWEET, R. A., NIMGAONKAR, V. L., LOSCHER, C. E., BAUER, E. M. & CHAPARALA, S. 2016. Schizophrenia interactome with 504 novel protein–protein interactions. npj Schizophrenia, 2, 16012.

GARZÓN, J. I., DENG, L., MURRAY, D., SHAPIRA, S., PETREY, D. & HONIG, B. 2016. A computational interactome and functional annotation for the human proteome. Elife, 5, e18715.

GEDDES, G. C., STAMM, K., MITCHELL, M., MUSSATTO, K. A. & TOMITA-MITCHELL, A. 2017. Ciliopathy variant burden and developmental delay in children with hypoplastic left heart syndrome. Genetics in Medicine, 19, 711-714.

GEREMEK, M., ZIĘTKIEWICZ, E., BRUINENBERG, M., FRANKE, L., POGORZELSKI, A., WIJMENGA, C. & WITT, M. 2014. Ciliary genes are down-regulated in bronchial tissue of primary ciliary dyskinesia patients. PLoS One, 9, e88216.

GEYMAN, J. P. 1983. The Oslerian tradition and changing medical education: A reappraisal. Western Journal of Medicine, 138, 884.

GHAVASIEH, A., BONTORIN, S., ARTIME, O., VERSTRAETE, N. & DE DOMENICO, M. 2021. Multiscale statistical physics of the pan-viral interactome unravels the systemic nature of SARS-CoV-2 infections. Communications Physics, 4, 83.

GIACOMINI, K. M., KRAUSS, R. M., RODEN, D. M., EICHELBAUM, M., HAYDEN, M. R. & NAKAMURA, Y. 2007. When good drugs go bad. Nature, 446, 975-977.

GILL, H. K., PARSONS, S. R., SPALLUTO, C., DAVIES, A. F., KNORZ, V. J., BURLINSON, C. E., NG, B. L., CARTER, N. P., OGILVIE, C. M. & WILSON, D. I. 2009. Separation of the PROX1 gene from upstream conserved elements in a complex inversion/translocation patient with hypoplastic left heart. European journal of human genetics, 17, 1423-1431.

GOBERGS, R., SALPUTRA, E. & LUBAUA, I. 2016. Hypoplastic left heart syndrome: a review. Acta medica Lituanica, 23, 86.

GOH, K. I., CUSICK, M. E., BARABÁSI, A. L. & VIDAL, M. 2007. Drug—target network. Nat

Biotechnol, 25, 1119-26.

GOH, K.-I., CUSICK, M. E., VALLE, D., CHILDS, B., VIDAL, M. & BARABÁSI, A.-L. 2007. The human disease network. Proceedings of the National Academy of Sciences, 104, 8685-8690.

GORDON, D. E., JANG, G. M., BOUHADDOU, M., XU, J., OBERNIER, K., WHITE, K. M., O'MEARA, M. J., REZELJ, V. V., GUO, J. Z. & SWANEY, D. L. 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature, 1-13.

GREENING, D. W., JI, H., CHEN, M., ROBINSON, B. W., DICK, I. M., CREANEY, J. & SIMPSON, R. J. 2016. Secreted primary human malignant mesothelioma exosome signature reflects oncogenic cargo. Scientific reports, 6, 32643.

GREIG, N. H., UTSUKI, T., YU, Q.-S., ZHU, X., HOLLOWAY, H. W., PERRY, T., LEE, B., INGRAM, D. K. & LAHIRI, D. K. 2001. A new therapeutic target in Alzheimer's disease treatment: attention to butyrylcholinesterase. Current medical research and opinion, 17, 159-165.

GRIFFITH, M., GRIFFITH, O. L., COFFMAN, A. C., WEIBLE, J. V., MCMICHAEL, J. F., SPIES, N. C., KOVAL, J., DAS, I., CALLAWAY, M. B. & ELDRED, J. M. 2013. DGIdb: mining the druggable genome. Nature methods, 10, 1209-1210.

GROßERUESCHKAMP, F., BRACHT, T., DIEHL, H. C., KUEPPER, C., AHRENS, M., KALLENBACH-THIELTGES, A., MOSIG, A., EISENACHER, M., MARCUS, K. & BEHRENS, T. 2017. Spatial and molecular resolution of diffuse malignant mesothelioma heterogeneity by integrating label-free FTIR imaging, laser capture microdissection and proteomics. Scientific reports, 7, 1-12.

GROVE, J., RIPKE, S., ALS, T. D., MATTHEISEN, M., WALTERS, R. K., WON, H., PALLESEN, J., AGERBO, E., ANDREASSEN, O. A. & ANNEY, R. 2019. Identification of common genetic risk variants for autism spectrum disorder. Nature genetics, 51, 431.

GUEMEZ-GAMBOA, A., COUFAL, N. G. & GLEESON, J. G. 2014. Primary cilia in the developing and mature brain. Neuron, 82, 511-21.

GUO, J., HIGGINBOTHAM, H., LI, J., NICHOLS, J., HIRT, J., GHUKASYAN, V. & ANTON, E. S. 2015. Developmental disruptions underlying brain abnormalities in ciliopathies. Nat Commun, 6, 7857.

GUPTA, G. D., COYAUD, É., GONÇALVES, J., MOJARAD, B. A., LIU, Y., WU, Q., GHEIRATMAND, L., COMARTIN, D., TKACH, J. M. & CHEUNG, S. W. 2015. A dynamic protein interaction landscape of the human centrosome-cilium interface. Cell, 163, 1484-1499.

GUSELLA, J. F., WEXLER, N. S., CONNEALLY, P. M., NAYLOR, S. L., ANDERSON, M. A., TANZI, R. E., WATKINS, P. C., OTTINA, K., WALLACE, M. R. & SAKAGUCHI, A. Y. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. Nature, 306, 234-238.

GYSI, D. M., VALLE, Í. D., ZITNIK, M., AMELI, A., GAN, X., VAROL, O., SANCHEZ, H., BARON, R. M., GHIASSIAN, D. & LOSCALZO, J. 2020. Network medicine framework for identifying drug repurposing opportunities for covid-19. arXiv preprint arXiv:2004.07229.

HAMOSH, A., SCOTT, A. F., AMBERGER, J. S., BOCCHINI, C. A. & MCKUSICK, V. A. 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic acids research, 33, D514-D517.

HAMP, T. & ROST, B. 2015. Evolutionary profiles improve protein–protein interaction prediction from sequence. Bioinformatics, 31, 1945-1950.

HAMZAH, M., OTHMAN, H. F., BALOGLU, O. & ALY, H. 2020. Outcomes of hypoplastic left heart syndrome: analysis of National Inpatient Sample Database 1998–2004 versus 2005–2014. European journal of pediatrics, 179, 309-316.

HAN, Y.-C., SONG, J.-M., WANG, L., SHU, C.-C., GUO, J. & CHEN, L.-L. 2016. Prediction and characterization of protein-protein interaction network in Bacillus licheniformis WX-02. Scientific reports, 6, 19486.

HANDEN, A. & GANAPATHIRAJU, M. K. 2015. LENS: web-based lens for enrichment and network studies of human proteins. BMC medical genomics, 8, 1-11.

HANGGE, P. T., CNOTA, J. F., WOO, J. G., HINTON, A. C., DIVANOVIC, A. A., MANNING, P. B., ITTENBACH, R. F. & HINTON, R. B. 2013. Microcephaly is associated with early adverse neurologic outcomes in hypoplastic left heart syndrome. Pediatric research, 74, 61-67.

HARRIS, L. W., WAYLAND, M., LAN, M., RYAN, M., GIGER, T., LOCKSTONE, H., WUETHRICH, I., MIMMACK, M., WANG, L. & KOTTER, M. 2008. The cerebral microvasculature in schizophrenia: a laser capture microdissection study. PLoS One, 3, e3964.

HASE, T., TANAKA, H., SUZUKI, Y., NAKAGAWA, S. & KITANO, H. 2009. Structure of protein interaction networks and their implications on drug design. PLoS Comput Biol, 5, e1000550.

HATTINGH, C. J., IPSER, J., TROMP, S., SYAL, S., LOCHNER, C., BROOKS, S. J. B. & STEIN, D. J. 2013. Functional magnetic resonance imaging during emotion recognition in social anxiety disorder: an activation likelihood meta-analysis. Frontiers in human neuroscience, 6, 347.

HAWRYLYCZ, M. J., LEIN, E. S., GUILLOZET-BONGAARTS, A. L., SHEN, E. H., NG, L., MILLER, J. A., VAN DE LAGEMAAT, L. N., SMITH, K. A., EBBERT, A. & RILEY, Z. L. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature, 489, 391.

HAYWARD, C. & BROCK, D. J. 1997. Fibrillin-1 mutations in Marfan syndrome and other type-1 fibrillinopathies. Human mutation, 10, 415.

HIDALGO, C. A., BLUMM, N., BARABÁSI, A.-L. & CHRISTAKIS, N. A. 2009. A dynamic network approach for the study of human phenotypes. PLoS computational biology, 5, e1000353.

HIGGINBOTHAM, H., EOM, T. Y., MARIANI, L. E., BACHLEDA, A., HIRT, J., GUKASSYAN, V., CUSACK, C. L., LAI, C., CASPARY, T. & ANTON, E. S. 2012. Arl13b in primary cilia regulates the migration and placement of interneurons in the developing cerebral cortex. Dev Cell, 23, 925-38.

HINTON, R. B., ANDELFINGER, G., SEKAR, P., HINTON, A. C., GENDRON, R. L., MICHELFELDER, E. C., ROBITAILLE, Y. & BENSON, D. W. 2008. Prenatal head growth and white matter injury in hypoplastic left heart syndrome. Pediatric research, 64, 364-369.

HOMSY, J., ZAIDI, S., SHEN, Y., WARE, J. S., SAMOCHA, K. E., KARCZEWSKI, K. J., DEPALMA, S. R., MCKEAN, D., WAKIMOTO, H. & GORHAM, J. 2015. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science, 350, 1262-1266.

HOPF, T. A., SCHÄRFE, C. P., RODRIGUES, J. P., GREEN, A. G., KOHLBACHER, O., SANDER, C., BONVIN, A. M. & MARKS, D. S. 2014. Sequence co-evolution gives 3D contacts and structures of protein complexes. Elife, 3, e03430.

HULOVATYY, Y., SOLAVA, R. W. & MILENKOVIĆ, T. 2014. Revealing missing parts of the interactome via link prediction. PloS one, 9, e90073.

HUNG, Y. P., DONG, F., TORRE, M., CRUM, C. P., BUENO, R. & CHIRIEAC, L. R. 2020. Molecular characterization of diffuse malignant peritoneal mesothelioma. Modern Pathology, 1-11.

HUNG, Y. P., DONG, F., WATKINS, J. C., NARDI, V., BUENO, R., DAL CIN, P., GODLESKI, J. J., CRUM, C. P. & CHIRIEAC, L. R. 2018. Identification of ALK rearrangements in malignant peritoneal mesothelioma. JAMA oncology, 4, 235-238.

HUNTER, T. & SEFTON, B. M. 1980. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. Proceedings of the National Academy of Sciences, 77, 1311-1315.

HUTTLIN, E. L., BRUCKNER, R. J., NAVARRETE-PEREA, J., CANNON, J. R., BALTIER, K., GEBREAB, F., GYGI, M. P., THORNOCK, A., ZARRAGA, G. & TAM, S. 2020. Dual Proteome-scale Networks Reveal Cell-specific Remodeling of the Human Interactome. bioRxiv.

IRION, U. & NÜSSLEIN-VOLHARD, C. 2022. Developmental genetics with model organisms. Proceedings of the National Academy of Sciences, 119, e2122148119.

JEONG, H., MASON, S. P., BARABÁSI, A.-L. & OLTVAI, Z. N. 2001. Lethality and centrality in protein networks. Nature, 411, 41-42.

JESSBERGER, S., CLARK, R. E., BROADBENT, N. J., CLEMENSON, G. D., CONSIGLIO, A., LIE, D. C., SQUIRE, L. R. & GAGE, F. H. 2009. Dentate gyrus-specific knockdown of adult neurogenesis impairs spatial and object recognition memory in adult rats. Learning & memory, 16, 147-154.

JIA, P., DAI, Y., HU, R., PEI, G., MANUEL, A. M. & ZHAO, Z. 2020. TSEA-DB: a trait–tissue association map for human complex traits and diseases. Nucleic acids research, 48, D1022-D1030.

JIA, J., LIU, Z., XIAO, X., LIU, B. & CHOU, K.-C. 2015. iPPI-Esml: an ensemble classifier for identifying the interactions of proteins by incorporating their physicochemical properties and wavelet transforms into PseAAC. Journal of theoretical biology, 377, 47-56.

JIN, S. C., HOMSY, J., ZAIDI, S., LU, Q., MORTON, S., DEPALMA, S. R., ZENG, X., QI, H., CHANG, W. & SIERANT, M. C. 2017. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nature genetics, 49, 1593.

JONES, H. N., OLBRYCH, S. K., SMITH, K. L., CNOTA, J. F., HABLI, M., RAMOS-GONZALES, O., OWENS, K. J., HINTON, A. C., POLZIN, W. J. & MUGLIA, L. J. 2015. Hypoplastic left heart syndrome is associated with structural and vascular placental abnormalities and leptin dysregulation. Placenta, 36, 1078-1086.

JOSEPH, N. M., CHEN, Y.-Y., NASR, A., YEH, I., TALEVICH, E., ONODERA, C., BASTIAN, B. C., RABBAN, J. T., GARG, K. & ZALOUDEK, C. 2017. Genomic profiling of malignant peritoneal mesothelioma reveals recurrent alterations in epigenetic regulatory genes BAP1, SETD2, and DDX3X. Modern Pathology, 30, 246-254.

JOSHI, S. A., DUVAL, E. R., KUBAT, B. & LIBERZON, I. 2020. A review of hippocampal activation in post-traumatic stress disorder. Psychophysiology, 57, e13357.

KAMIYA, A., TAN, P. L., KUBO, K., ENGELHARD, C., ISHIZUKA, K., KUBO, A., TSUKITA, S., PULVER, A. E., NAKAJIMA, K., CASCELLA, N. G., KATSANIS, N. & SAWA, A. 2008. Recruitment of PCM1 to the centrosome by the cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses. Arch Gen Psychiatry, 65, 996-1006.

KEILHAUER, E. C., HEIN, M. Y. & MANN, M. 2015. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). Molecular & Cellular Proteomics, 14, 120-135.

KELLEY, R. & IDEKER, T. 2005. Systematic interpretation of genetic interactions using protein networks. Nature biotechnology, 23, 561.

KESHAVA PRASAD, T., GOEL, R., KANDASAMY, K., KEERTHIKUMAR, S., KUMAR, S., MATHIVANAN, S., TELIKICHERLA, D., RAJU, R., SHAFREEN, B. & VENUGOPAL, A. 2008. Human protein reference database—2009 update. Nucleic acids research, 37, D767-D772.

METSALU, T. & VILO, J. 2015. ClustVis: a web tool for visualizing clustering of multivariate data

using Principal Component Analysis and heatmap. Nucleic acids research, 43, W566-W570.

MUNGALL, C. J., MCMURRY, J. A., KÖHLER, S., BALHOFF, J. P., BORROMEO, C., BRUSH, M., CARBON, S., CONLIN, T., DUNN, N. & ENGELSTAD, M. 2017. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic acids research, 45, D712-D722.

KHAN, M., SYED, G. H., KIM, S.-J. & SIDDIQUI, A. 2015. Mitochondrial dynamics and viral infections: a close nexus. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1853, 2822-2833.

KIDNAPILLAI, S., BORTOLASCI, C. C., UDAWELA, M., PANIZZUTTI, B., SPOLDING, B., CONNOR, T., SANIGORSKI, A., DEAN, O. M., CROWLEY, T. & JAMAIN, S. 2018. The use of a gene expression signature and connectivity map to repurpose drugs for bipolar disorder. The World Journal of Biological Psychiatry, 1-9.

KIM, Y., TON, T.-V., DEANGELO, A. B., MORGAN, K., DEVEREUX, T. R., ANNA, C., COLLINS, J. B., PAULES, R. S., CROSBY, L. M. & SILLS, R. C. 2006. Major carcinogenic pathways identified by gene expression analysis of peritoneal mesotheliomas following chemical treatment in F344 rats. Toxicology and applied pharmacology, 214, 144-151.

KINNINGS, S. L., LIU, N., BUCHMEIER, N., TONGE, P. J., XIE, L. & BOURNE, P. E. 2009. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. PLoS computational biology, 5, e1000423.

KLEIN, R. J., ZEISS, C., CHEW, E. Y., TSAI, J.-Y., SACKLER, R. S., HAYNES, C., HENNING, A. K., SANGIOVANNI, J. P., MANE, S. M. & MAYNE, S. T. 2005. Complement factor H polymorphism in age-related macular degeneration. Science, 308, 385-389.

KOGISO, T. & TOKUSHIGE, K. 2020. Fontan-associated liver disease and hepatocellular carcinoma in adults. Scientific Reports, 10, 1-14.

KÖHLER, S., BAUER, S., HORN, D. & ROBINSON, P. N. 2008. Walking the interactome for prioritization of candidate disease genes. The American Journal of Human Genetics, 82, 949-958.

KOMATSU, H., INUI, A., KISHIKI, K., KAWAI, H., YOSHIO, S., OSAWA, Y., KANTO, T. & FUJISAWA, T. 2019. Liver disease secondary to congenital heart disease in children. Expert review of gastroenterology & hepatology, 13, 651-666.

KONG, S. W., COLLINS, C. D., SHIMIZU-MOTOHASHI, Y., HOLM, I. A., CAMPBELL, M. G., LEE, I.-H., BREWSTER, S. J., HANSON, E., HARRIS, H. K. & LOWE, K. R. 2012. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One, 7, e49475.

KOTLYAR, M., PASTRELLO, C., MALIK, Z. & JURISICA, I. 2019. IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species. Nucleic acids research, 47, D581-D589.

KOTLYAR, M., PASTRELLO, C., PIVETTA, F., SARDO, A. L., CUMBAA, C., LI, H., NARANIAN, T., NIU, Y., DING, Z. & VAFAEE, F. 2015. In silico prediction of physical protein interactions and characterization of interactome orphans. Nature methods, 12, 79-84.

KOUTOUBY, A., HABIBULLAH, J. & MOINUDDIN, F. A. 2000. Spondylometaphyseal dysplasia: Sedaghatian type. American journal of medical genetics, 90, 199-202.

KRÄMER, A., GREEN, J., POLLARD JR, J. & TUGENDREICH, S. 2014. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics, 30, 523-530.

KRIENEN, F. M., YEO, B. T., GE, T., BUCKNER, R. L. & SHERWOOD, C. C. 2016. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network

architecture in the human brain. Proceedings of the National Academy of Sciences, 113, E469-E478.

KRISHNAN, A., ZHANG, R., YAO, V., THEESFELD, C. L., WONG, A. K., TADYCH, A., VOLFOVSKY, N., PACKER, A., LASH, A. & TROYANSKAYA, O. G. 2016. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nature neuroscience, 19, 1454-1462.

KRUGLYAK, L. 1997. The use of a genetic map of biallelic markers in linkage studies. Nature genetics, 17, 21-24.

KRUGLYAK, L., DALY, M. J., REEVE-DALY, M. P. & LANDER, E. S. 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. American journal of human genetics, 58, 1347.

KUMAR, A., MAJUMDER, M. M., MARTÍ, J. M. L., PARSONS, A., MATTILA, P., PORKKA, K., KALLIONIEMI, O. & HECKMAN, C. A. 2014. The Use of RNA Sequencing to Identify Disease-Specific Gene Expression Signatures and Critical Regulatory Networks Across Hematologic Malignancies. Am Soc Hematology.

KUMAR, N., MISHRA, B., MEHMOOD, A., ATHAR, M. & MUKHTAR, M. S. 2020. Integrative network biology framework elucidates molecular mechanisms of sars-cov-2 pathogenesis.

KUPERSHMIDT, I., SU, Q. J., GREWAL, A., SUNDARESH, S., HALPERIN, I., FLYNN, J., SHEKAR, M., WANG, H., PARK, J., CUI, W., WALL, G. D., WISOTZKEY, R., ALAG, S., AKHTARI, S. & RONAGHI, M. 2010. Ontology-based meta-analysis of global collections of high-throughput public data. PLoS One, 5.

LADDACH, A., NG, J. C.-F., CHUNG, S. S. & FRATERNALI, F. 2018. Genetic variants and protein–protein interactions: a multidimensional network-centric view. Current opinion in structural biology, 50, 82-90.

LAITINEN, O., HYTÖNEN, V., NORDLUND, H. & KULOMAA, M. 2006. Genetically engineered avidins and streptavidins. Cellular and Molecular Life Sciences CMLS, 63, 2992-3017.

LAMB, J., CRAWFORD, E. D., PECK, D., MODELL, J. W., BLAT, I. C., WROBEL, M. J., LERNER, J., BRUNET, J.-P., SUBRAMANIAN, A. & ROSS, K. N. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. science, 313, 1929-1935.

LAMBERT, J.-P., PICAUD, S., FUJISAWA, T., HOU, H., SAVITSKY, P., UUSKÜLA-REIMAND, L., GUPTA, G. D., ABDOUNI, H., LIN, Z.-Y. & TUCHOLSKA, M. 2019. Interactome rewiring following pharmacological targeting of BET bromodomains. Molecular cell, 73, 621-638. e17.

LANG-LAZDUNSKI, L. 2018. Malignant pleural mesothelioma: some progress, but still a long way from cure. Journal of thoracic disease, 10, 1172.

LANZ, T. A., JOSHI, J. J., REINHART, V., JOHNSON, K., GRANTHAM II, L. E. & VOLFSON, D. 2015. STEP levels are unchanged in pre-frontal cortex and associative striatum in post-mortem human brain samples from subjects with schizophrenia, bipolar disorder and major depressive disorder. PloS one, 10, e0121744.

LE SAGE, V., CINTI, A. & MOULAND, A. J. 2016. Proximity-Dependent Biotinylation for Identification of Interacting Proteins. Current protocols in cell biology, 17.19. 1-17.19. 12.

LEE, J. E. & GLEESON, J. G. 2011. Cilia in the nervous system: linking cilia function and neurodevelopmental disorders. Curr Opin Neurol, 24, 98-105.

LEE, P. H., ANTTILA, V., WON, H., FENG, Y.-C. A., ROSENTHAL, J., ZHU, Z., TUCKER-DROB, E. M., NIVARD, M. G., GROTZINGER, A. D. & POSTHUMA, D. 2019. Genomic relationships,

novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell, 179, 1469-1482. e11.

LEVEY, D. F., GELERNTER, J., POLIMANTI, R., ZHOU, H., CHENG, Z., ASLAN, M., QUADEN, R., CONCATO, J., RADHAKRISHNAN, K. & BRYOIS, J. 2020. Reproducible genetic risk loci for anxiety: results from~ 200,000 participants in the Million Veteran Program. American Journal of Psychiatry, 177, 223-232.

LI, G. & DE CLERCQ, E. 2020. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nature Publishing Group.

LI, J., SHI, M., MA, Z., ZHAO, S., EUSKIRCHEN, G., ZISKIN, J., URBAN, A., HALLMAYER, J. & SNYDER, M. 2014. Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. Molecular systems biology, 10, 774.

LI, Y. & ILIE, L. 2017. SPRINT: Ultrafast protein-protein interaction prediction of the entire human interactome. arXiv preprint arXiv:1705.06848.

LI, Y., KLENA, N. T., GABRIEL, G. C., LIU, X., KIM, A. J., LEMKE, K., CHEN, Y., CHATTERJEE, B., DEVINE, W. & DAMERLA, R. R. 2015. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. Nature, 521, 520-524.

LIAO, Y., WANG, J., JAEHNIG, E. J., SHI, Z. & ZHANG, B. 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic acids research.

LIBERZON, A., SUBRAMANIAN, A., PINCHBACK, R., THORVALDSDÓTTIR, H., TAMAYO, P. & MESIROV, J. P. 2011. Molecular signatures database (MSigDB) 3.0. Bioinformatics, 27, 1739-1740.

LICATA, L., BRIGANTI, L., PELUSO, D., PERFETTO, L., IANNUCCELLI, M., GALEOTA, E., SACCO, F., PALMA, A., NARDOZZA, A. P. & SANTONICO, E. 2012. MINT, the molecular interaction database: 2012 update. Nucleic acids research, 40, D857-D861.

LIM, J., HAO, T., SHAW, C., PATEL, A. J., SZABÓ, G., RUAL, J.-F., FISK, C. J., LI, N., SMOLYAR, A. & HILL, D. E. 2006. A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell, 125, 801-814.

LIU, E. M., LUNA, A., DONG, G. & SANDER, C. 2020. netboxr: Automated discovery of biological process modules by network analysis in R. Plos one, 15, e0234669.

LIU, X., YAGI, H., SAEED, S., BAIS, A. S., GABRIEL, G. C., CHEN, Z., PETERSON, K. A., LI, Y., SCHWARTZ, M. C. & REYNOLDS, W. T. 2017. The complex genetics of hypoplastic left heart syndrome. Nature genetics, 49, 1152.

LOOS, R. J. 2020. 15 years of genome-wide association studies and no signs of slowing down. Nature Communications, 11, 5900.

LOSCALZO, J. 2023. Molecular interaction networks and drug development: Novel approach to drug target identification and drug repositioning. The FASEB Journal, 37, e22660.

LOSCALZO, J. & BARABASI, A. L. 2011. Systems biology and the future of medicine. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 3, 619-627.

LOUVI, A. & GROVE, E. A. 2011. Cilia in the CNS: the quiet organelle claims center stage. Neuron, 69, 1046-60.

LUCK, K., KIM, D.-K., LAMBOURNE, L., SPIROHN, K., BEGG, B. E., BIAN, W., BRIGNALL, R., CAFARELLI, T., CAMPOS-LABORIE, F. J. & CHARLOTEAUX, B. 2020a. A reference map of the human binary protein interactome. Nature, 580, 402-408.

LUCK, K., KIM, D.-K., LAMBOURNE, L., SPIROHN, K., BEGG, B. E., BIAN, W., BRIGNALL, R., CAFARELLI, T., CAMPOS-LABORIE, F. J. & CHARLOTEAUX, B. 2020b. A reference map of the

human binary protein interactome. Nature, 1-7.

LUO, X., YOU, Z., ZHOU, M., LI, S., LEUNG, H., XIA, Y. & ZHU, Q. 2015. A highly efficient approach to protein interactome mapping based on collaborative filtering framework. Scientific reports, 5, 7702.

MACARTHUR, J., BOWLER, E., CEREZO, M., GIL, L., HALL, P., HASTINGS, E., JUNKINS, H., MCMAHON, A., MILANO, A. & MORALES, J. 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research, 45, D896-D901.

MACDONALD, M. E., AMBROSE, C. M., DUYAO, M. P., MYERS, R. H., LIN, C., SRINIDHI, L., BARNES, G., TAYLOR, S. A., JAMES, M. & GROOT, N. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell, 72, 971-983.

MAERTZDORF, J., WEINER, J., MOLLENKOPF, H.-J., NETWORK, T., BAUER, T., PRASSE, A., MÜLLER-QUERNHEIM, J., KAUFMANN, S. H., BELLINGER, O. & DIEL, R. 2012. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proceedings of the National Academy of Sciences, 109, 7853-7858.

MAGENIS, R. E., MASLEN, C. L., SMITH, L., ALLEN, L. & SAKAI, L. Y. 1991. Localization of the fibrillin (FBN) gene to chromosome 15, band q21. 1. Genomics, 11, 346-351.

MAHER, B. 2008. Personal genomes: The case of the missing heritability. Nature, 456, 18-21.

MALAVIA, T., CHAPARALA, S., WOOD, J., CHOWDARI, K., PRASAD, K., MCCLAIN, L., JEGGA, A., GANAPATHIRAJU, M. K. & NIMGAONKAR, V. 2017a. Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epidemiological, genomic and protein interaction data npj Schizophrenia, in Press.

MALAVIA, T. A., CHAPARALA, S., WOOD, J., CHOWDARI, K., PRASAD, K. M., MCCLAIN, L., JEGGA, A. G., GANAPATHIRAJU, M. K. & NIMGAONKAR, V. L. 2017b. Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epidemiological, genomic, and protein interaction data. npj Schizophrenia, 3, 11.

MALIK, J., KARAN, M. & VASISHT, K. 2016. Attenuating effect of bioactive coumarins from Convolvulus pluricaulis on scopolamine-induced amnesia in mice. Natural product research, 30, 578-582.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R. & CHAKRAVARTI, A. 2009. Finding the missing heritability of complex diseases. Nature, 461, 747-753.

MARINO, B. S., LIPKIN, P. H., NEWBURGER, J. W., PEACOCK, G., GERDES, M., GAYNOR, J. W., MUSSATTO, K. A., UZARK, K., GOLDBERG, C. S. & JOHNSON JR, W. H. 2012. Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association. Circulation, 126, 1143-1172.

MARLEY, A. & VON ZASTROW, M. 2010. DISC1 regulates primary cilia that display specific dopamine receptors. PLoS One, 5, e10902.

MARLEY, A. & VON ZASTROW, M. 2012. A simple cell-based assay reveals that diverse neuropsychiatric risk genes converge on primary cilia. PLoS One, 7, e46647.

MARTIN, A., OCHAGAVIA, M. E., RABASA, L. C., MIRANDA, J., FERNANDEZ-DE-COSSIO, J. & BRINGAS, R. 2010. BisoGenet: a new tool for gene network building, visualization and analysis. BMC bioinformatics, 11, 1-9.

MARZIN, P. & CORMIER-DAIRE, V. 2020. New perspectives on the treatment of skeletal

dysplasia. Therapeutic advances in endocrinology and metabolism, 11, 2042018820904016.

MASTERS, C. L., BATEMAN, R., BLENNOW, K., ROWE, C. C., SPERLING, R. A. & CUMMINGS, J. L. 2015. Alzheimer's disease. Nature Reviews Disease Primers, 1, 15056.

MAYCOX, P. R., KELLY, F., TAYLOR, A., BATES, S., REID, J., LOGENDRA, R., BARNES, M. R., LARMINIE, C., JONES, N. & LENNON, M. 2009. Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. Molecular psychiatry, 14, 1083.

MCBRIDE, K. L., PIGNATELLI, R., LEWIN, M., HO, T., FERNBACH, S., MENESSES, A., LAM, W., LEAL, S. M., KAPLAN, N. & SCHLIEKELMAN, P. 2005. Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: segregation, multiplex relative risk, and heritability. American journal of medical genetics Part A, 134, 180-186.

MCBRIDE, K. L., ZENDER, G. A., FITZGERALD-BUTT, S. M., KOEHLER, D., MENESSES-DIAZ, A., FERNBACH, S., LEE, K., TOWBIN, J. A., LEAL, S. & BELMONT, J. W. 2009. Linkage analysis of left ventricular outflow tract malformations (aortic valve stenosis, coarctation of the aorta, and hypoplastic left heart syndrome). European journal of human genetics, 17, 811-819.

MCGUIRL, M. R., SMITH, S. P., SANDSTEDE, B. & RAMACHANDRAN, S. 2020. Detecting shared genetic architecture among multiple phenotypes by hierarchical clustering of gene-level association statistics. Genetics, 215, 511-529.

MESSNER, C. B., DEMICHEV, V., WENDISCH, D., MICHALICK, L., WHITE, M., FREIWALD, A., TEXTORIS-TAUBE, K., VERNARDIS, S. I., EGGER, A.-S. & KREIDL, M. 2020. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. Cell systems, 11, 11-24. e4.

MILLER, J. A., WOLTJER, R. L., GOODENBOUR, J. M., HORVATH, S. & GESCHWIND, D. H. 2013. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. Genome medicine, 5, 48.

MILLS, B. D., GRAYSON, D. S., SHUNMUGAVEL, A., MIRANDA-DOMINGUEZ, O., FECZKO, E., EARL, E., NEVE, K. A. & FAIR, D. A. 2018. Correlated gene expression and anatomical communication support synchronized brain activity in the mouse functional connectome. Journal of Neuroscience, 38, 5774-5787.

MITRA, K., CARVUNIS, A.-R., RAMESH, S. K. & IDEKER, T. 2013. Integrative approaches for finding modular structure in biological networks. Nature Reviews Genetics, 14, 719-732.

MIZUTANI, S., PAUWELS, E., STOVEN, V., GOTO, S. & YAMANISHI, Y. 2012. Relating drug–protein interaction network with drug side effects. Bioinformatics, 28, i522-i528.

MORISSET, S., TRAIFFORT, E. & SCHWARTZ, J.-C. 1996. Inhibition of histamine versus acetylcholine metabolism as a mechanism of tacrine activity. European journal of pharmacology, 315, R1-R2.

MOSS, D. E., PEREZ, R. G. & KOBAYASHI, H. 2017. Cholinesterase inhibitor therapy in Alzheimer's disease: The limits and tolerability of irreversible CNS-selective acetylcholinesterase inhibition in primates. Journal of Alzheimer's Disease, 55, 1285-1294.

MUFSON, E. J., MAHADY, L., WATERS, D., COUNTS, S. E., PEREZ, S. E., DEKOSKY, S., GINSBERG, S. D., IKONOMOVIC, M. D., SCHEFF, S. & BINDER, L. 2015. Hippocampal plasticity during the progression of Alzheimer's disease. Neuroscience, 309, 51-67.

MULLINS, N., FORSTNER, A. J., O'CONNELL, K. S., COOMBES, B., COLEMAN, J. R., QIAO, Z., ALS, T. D., BIGDELI, T. B., BØRTE, S. & BRYOIS, J. 2021. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology.

Nature Genetics, 53, 817-829.

NAHATA, A., PATIL, U. & DIXIT, V. 2009. Anxiolytic activity of Evolvulus alsinoides and Convulvulus pluricaulis in rodents. Pharmaceutical biology, 47, 444-451.

NG, S. B., TURNER, E. H., ROBERTSON, P. D., FLYGARE, S. D., BIGHAM, A. W., LEE, C., SHAFFER, T., WONG, M., BHATTACHARJEE, A. & EICHLER, E. E. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature, 461, 272-276.

NIMGAONKAR, V. 2019. Cromoglicate Adjunctive Therapy for Outpatients With Schizophrenia (CATOS). ClinicalTrials.gov.

NIMGAONKAR, V. 2022. Acetazolamide for Treatment Resistant Schizophrenia (APTS). ClinicalTrials.gov.

OKADA, Y. 2014. From the era of genome analysis to the era of genomic drug discovery: a pioneering example of rheumatoid arthritis. Clinical genetics, 86, 432-440.

OLSEN, U. D. J. G. I. H. T. B. E. P. P. R. P. W. S. S. T. D. N. C. J.-F., 9, R. G. S. C. S. Y. F. A. H. M. Y. T. T. A. I. T. K. C. W. H. T. Y., GENOSCOPE, 10, C. U.-W. J. H. R. S. W. A. F. B. P. B. T. P. E. R. C. W. P., DEPARTMENT OF GENOME ANALYSIS, I. O. M. B. R. A. P. M. N. G. T. S. R. A., 11, G. S. C. S. D. R. D.-S. L. R. M. W. K. L. H. M. D. J. & 15, B. G. I. H. G. C. Y. H. Y. J. W. J. H. G. G. J. 2001. Initial sequencing and analysis of the human genome. nature, 409, 860-921.

OPITZ, J. M., SPRANGER, J. W., STÖSS, H. R., PESCH, H. J., AZADEH, B. & REYNOLDS, J. F. 1987. Sedaghatian congenital lethal metaphyseal chondrodysplasia—observations in a second Iranian family and histopathological studies. American journal of medical genetics, 26, 583-590.

ORKIN, S. H. 1986. Reverse genetics and human disease. Cell, 47, 845-850.

OSLER, W. & MCCRAE, T. 1892. The principles and practice of medicine, New York, D. Appleton & Co, 278.

PACI, P., FISCON, G., CONTE, F., WANG, R.-S., FARINA, L. & LOSCALZO, J. 2021. Gene co-expression in the interactome: moving from correlation toward causation via an integrated approach to disease module discovery. NPJ systems biology and applications, 7, 3.

PARDIÑAS, A. F., HOLMANS, P., POCKLINGTON, A. J., ESCOTT-PRICE, V., RIPKE, S., CARRERA, N., LEGGE, S. E., BISHOP, S., CAMERON, D. & HAMSHERE, M. L. 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nature genetics, 50, 381.

PATANIA, A., SELVAGGI, P., VERONESE, M., DIPASQUALE, O., EXPERT, P. & PETRI, G. 2019. Topological gene expression networks recapitulate brain anatomy and function. Network Neuroscience, 3, 744-762.

PAZOUR, G. J., AGRIN, N., LESZYK, J. & WITMAN, G. B. 2005. Proteomic analysis of a eukaryotic cilium. Journal of Cell Biology, 170, 103-113.

PEARN, J. 2011. Differentiating diseases: the centrum of differential diagnosis in the evolution of Oslerian medicine. Fetal and Pediatric Pathology, 30, 1-15.

PEEDEN JR, J., RIMOIN, D., LACHMAN, R., DYER, M., GERARD, D. & GRUBER, H. 1992. Spondylometaphyseal dysplasia, Sedaghatian type. American journal of medical genetics, 44, 651-656.

PFEFFERLE, S., SCHÖPF, J., KÖGL, M., FRIEDEL, C. C., MÜLLER, M. A., CARBAJO-LOZOYA, J., STELLBERGER, T., VON DALL'ARMI, E., HERZOG, P. & KALLIES, S. 2011. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. PLoS pathogens, 7.

PILLAI, K., POURGHOLAMI, M. H., CHUA, T. C. & MORRIS, D. L. 2013. MUC1 has prognostic

significance in malignant peritoneal mesothelioma. The International journal of biological markers, 28, 303-312.

PIÑERO, J., BRAVO, À., QUERALT-ROSINACH, N., GUTIÉRREZ-SACRISTÁN, A., DEU-PONS, J., CENTENO, E., GARCÍA-GARCÍA, J., SANZ, F. & FURLONG, L. I. 2016. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic acids research, gkw943.

PIÑERO, J., QUERALT-ROSINACH, N., BRAVO, À., DEU-PONS, J., BAUER-MEHREN, A., BARON, M., SANZ, F. & FURLONG, L. I. 2015. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database, 2015.

PLANAS-IGLESIAS, J., MARIN-LOPEZ, M. A., BONET, J., GARCIA-GARCIA, J. & OLIVA, B. 2013. iLoops: a protein–protein interaction prediction server based on structural features. Bioinformatics, 29, 2360-2362.

POIRIER, J. 2002. Evidence that the clinical effects of cholinesterase inhibitors are related to potency and targeting of action. International journal of clinical practice. Supplement, 6-19.

PORRAS, P., BARRERA, E., BRIDGE, A., DEL-TORO, N., CESARENI, G., DUESBURY, M., HERMJAKOB, H., IANNUCCELLI, M., JURISICA, I. & KOTLYAR, M. 2020. Towards a unified open access dataset of molecular interactions. Nature communications, 11, 6144.

PUJANA, M. A., HAN, J.-D. J., STARITA, L. M., STEVENS, K. N., TEWARI, M., AHN, J. S., RENNERT, G., MORENO, V., KIRCHHOFF, T. & GOLD, B. 2007. Network modeling links breast cancer susceptibility and centrosome dysfunction. Nature genetics, 39, 1338-1349.

PUSHPAKOM, S., IORIO, F., EYERS, P. A., ESCOTT, K. J., HOPPER, S., WELLS, A., DOIG, A., GUILLIAMS, T., LATIMER, J. & MCNAMEE, C. 2019. Drug repurposing: progress, challenges and recommendations. Nature Reviews Drug Discovery, 18, 41.

RAJA, K., SUBRAMANI, S. & NATARAJAN, J. 2013. PPInterFinder—a mining tool for extracting causal relations on human proteins from literature. Database, 2013.

REGENMORTEL, M. H. V. 2004. Reductionism and complexity in molecular biology: Scientists now have the tools to unravel biological complexity and overcome the limitations of reductionism. EMBO reports, 5, 1016-1020.

REGHUNATHAN, R., JAYAPAL, M., HSU, L.-Y., CHNG, H.-H., TAI, D., LEUNG, B. P. & MELENDEZ, A. J. 2005. Expression profile of immune response genes in patients with severe acute respiratory syndrome. BMC immunology, 6, 2.

REUTER, M. S., CHATURVEDI, R. R., LISTON, E., MANSHAEI, R., AUL, R. B., BOWDIN, S., COHN, I., CURTIS, M., DHIR, P. & HAYEEMS, R. Z. 2020. The Cardiac Genome Clinic: implementing genome sequencing in pediatric heart disease. Genetics in Medicine, 22, 1015-1024.

RICCI, M., XU, Y., HAMMOND, H. L., WILLOUGHBY, D. A., NATHANSON, L., RODRIGUEZ, M. M., VATTA, M., LIPSHULTZ, S. E. & LINCOLN, J. 2012. Myocardial alternative RNA splicing and gene expression profiling in early stage hypoplastic left heart syndrome. PloS one, 7, e29784.

RICHIARDI, J., ALTMANN, A., MILAZZO, A.-C., CHANG, C., CHAKRAVARTY, M. M., BANASCHEWSKI, T., BARKER, G. J., BOKDE, A. L., BROMBERG, U. & BÜCHEL, C. 2015. Correlated gene expression supports synchronous activity in brain networks. Science, 348, 1241-1244.

RIGAUT, G., SHEVCHENKO, A., RUTZ, B., WILM, M., MANN, M. & SÉRAPHIN, B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nature biotechnology, 17, 1030-1032.

RIPKE, S., NEALE, B. M., CORVIN, A., WALTERS, J. T., FARH, K.-H., HOLMANS, P. A., LEE, P., BULIK-SULLIVAN, B., COLLIER, D. A. & HUANG, H. 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature, 511, 421.

ROBINSON, B. W. & LAKE, R. A. 2005. Advances in malignant mesothelioma. New England Journal of Medicine, 353, 1591-1603.

ROLLAND, T., TAŞAN, M., CHARLOTEAUX, B., PEVZNER, S. J., ZHONG, Q., SAHNI, N., YI, S., LEMMENS, I., FONTANILLO, C. & MOSCA, R. 2014. A proteome-scale map of the human interactome network. Cell, 159, 1212-1226.

ROTH, B. L., SHEFFLER, D. J. & KROEZE, W. K. 2004. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nature reviews Drug discovery, 3, 353-359.

ROTHAN, H. A. & BYRAREDDY, S. N. 2020. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun, 102433.

ROUILLARD, A. D., GUNDERSEN, G. W., FERNANDEZ, N. F., WANG, Z., MONTEIRO, C. D., MCDERMOTT, M. G. & MA'AYAN, A. 2016. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database, 2016.

ROUX, K. J., KIM, D. I., RAIDA, M. & BURKE, B. 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. Journal of cell biology, 196, 801-810.

RUAL, J.-F., VENKATESAN, K., HAO, T., HIROZANE-KISHIKAWA, T., DRICOT, A., LI, N., BERRIZ, G. F., GIBBONS, F. D., DREZE, M. & AYIVI-GUEDEHOUSSOU, N. 2005. Towards a proteome-scale map of the human protein–protein interaction network. Nature, 437, 1173-1178.

SAHNI, N., YI, S., TAIPALE, M., BASS, J. I. F., COULOMBE-HUNTINGTON, J., YANG, F., PENG, J., WEILE, J., KARRAS, G. I. & WANG, Y. 2015. Widespread macromolecular interaction perturbations in human genetic disorders. Cell, 161, 647-660.

SAKAI, Y., SHAW, C. A., DAWSON, B. C., DUGAS, D. V., AL-MOHTASEB, Z., HILL, D. E. & ZOGHBI, H. Y. 2011. Protein interactome reveals converging molecular pathways among autism disorders. Science translational medicine, 3, 86ra49-86ra49.

ŠAMÁNEK, M., SLAVÍK, Z., ZBOŘILOVÁ, B., HROBOŇOVÁ, V., VOŘÍŠKOVÁ, M. & ŠKOVRÁNEK, J. 1989. Prevalence, treatment, and outcome of heart disease in live-born children: a prospective analysis of 91,823 live-born children. Pediatric cardiology, 10, 205-211.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences, 74, 5463-5467.

SANO, T., VAJDA, S. & CANTOR, C. R. 1998. Genetic engineering of streptavidin, a versatile affinity tag. Journal of Chromatography B: Biomedical Sciences and Applications, 715, 85-91.

SCHADT, E. E. 2009. Molecular networks as sensors and drivers of common human diseases. Nature, 461, 218-223.

SCHORK, N. J. 2015. Personalized medicine: time for one-person trials. Nature, 520, 609-611.

SEDAGHATIAN, M. & OPITZ, J. M. 1980. Congenital lethal metaphyseal chondrodysplasia: a newly recognized complex autosomal recessive disorder. American journal of medical genetics, 6, 269-274.

SHAPIRA, S. D., GAT-VIKS, I., SHUM, B. O., DRICOT, A., DE GRACE, M. M., WU, L., GUPTA, P. B., HAO, T., SILVER, S. J. & ROOT, D. E. 2009. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell, 139, 1255-1267.

SHARMA, A., KITSAK, M., CHO, M. H., AMELI, A., ZHOU, X., JIANG, Z., CRAPO, J. D., BEATY,

T. H., MENCHE, J. & BAKKE, P. S. 2018. Integration of molecular interactome and targeted interaction analysis to identify a COPD disease network module. Scientific reports, 8, 14439.

SHARMA, A., MENCHE, J., HUANG, C. C., ORT, T., ZHOU, X., KITSAK, M., SAHNI, N., THIBAULT, D., VOUNG, L. & GUO, F. 2015a. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. Human molecular genetics, 24, 3005-3020.

SHARMA, K., SCHMITT, S., BERGNER, C. G., TYANOVA, S., KANNAIYAN, N., MANRIQUE-HOYOS, N., KONGI, K., CANTUTI, L., HANISCH, U.-K. & PHILIPS, M.-A. 2015b. Cell type–and brain region–resolved mouse brain proteome. Nature neuroscience, 18, 1819.

SHUKLA, A., MACPHERSON, M. B., HILLEGASS, J., RAMOS-NINO, M. E., ALEXEEVA, V., VACEK, P. M., BOND, J. P., PASS, H. I., STEELE, C. & MOSSMAN, B. T. 2009. Alterations in gene expression in human mesothelial cells correlate with mineral pathogenicity. American journal of respiratory cell and molecular biology, 41, 114-123.

SIDAWAY, P. 2020. COVID-19 and cancer: what we know so far. Nature Reviews Clinical Oncology, 1-1.

SIFFEL, C., RIEHLE-COLARUSSO, T., OSTER, M. E. & CORREA, A. 2015. Survival of children with hypoplastic left heart syndrome. Pediatrics, 136, e864-e870.

SIROTA, M., DUDLEY, J. T., KIM, J., CHIANG, A. P., MORGAN, A. A., SWEET-CORDERO, A., SAGE, J. & BUTTE, A. J. 2011. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Science translational medicine, 3, 96ra77-96ra77.

SMITH, A. C., MEARS, A. J., BUNKER, R., AHMED, A., MACKENZIE, M., SCHWARTZENTRUBER, J. A., BEAULIEU, C. L., FERRETTI, E., MAJEWSKI, J. & BULMAN, D. E. 2014. Mutations in the enzyme glutathione peroxidase 4 cause Sedaghatian-type spondylometaphyseal dysplasia. Journal of medical genetics, 51, 470-474.

SMITH, C. L. & EPPIG, J. T. 2012. The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. Mammalian genome, 23, 653-668.

SMITH, T. D., ADAMS, M. M., GALLAGHER, M., MORRISON, J. H. & RAPP, P. R. 2000. Circuit-specific alterations in hippocampal synaptophysin immunoreactivity predict spatial learning impairment in aged rats. Journal of Neuroscience, 20, 6587-6593.

SMOOT, M. E., ONO, K., RUSCHEINSKI, J., WANG, P.-L. & IDEKER, T. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics, 27, 431-432.

SPIRIN, V. & MIRNY, L. A. 2003. Protein complexes and functional modules in molecular networks. Proceedings of the national Academy of sciences, 100, 12123-12128.

SRIVATSAN, S. R., MCFALINE-FIGUEROA, J. L., RAMANI, V., SAUNDERS, L., CAO, J., PACKER, J., PLINER, H. A., JACKSON, D. L., DAZA, R. M. & CHRISTIANSEN, L. 2020. Massively multiplex chemical transcriptomics at single-cell resolution. Science, 367, 45-51.

STAHL, E. A., BREEN, G., FORSTNER, A. J., MCQUILLIN, A., RIPKE, S., TRUBETSKOY, V., MATTHEISEN, M., WANG, Y., COLEMAN, J. R. & GASPAR, H. A. 2019. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nature genetics, 51, 793.

STARK, C., BREITKREUTZ, B.-J., REGULY, T., BOUCHER, L., BREITKREUTZ, A. & TYERS, M. 2006. BioGRID: a general repository for interaction datasets. Nucleic acids research, 34, D535-D539.

STARRUß, J., DE BACK, W., BRUSCH, L. & DEUTSCH, A. 2014. Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology. Bioinformatics, 30, 1331-1332.

STAYTON, P. S., FREITAG, S., KLUMB, L. A., CHILKOTI, A., CHU, V., PENZOTTI, J. E., TO, R., HYRE, D., LE TRONG, I. & LYBRAND, T. P. 1999. Streptavidin–biotin binding energetics. Biomolecular engineering, 16, 39-44.

STREBHARDT, K. & ULLRICH, A. 2008. Paul Ehrlich's magic bullet concept: 100 years of progress. Nature Reviews Cancer, 8, 473-480.

SUBRAMANIAN, A., NARAYAN, R., CORSELLO, S. M., PECK, D. D., NATOLI, T. E., LU, X., GOULD, J., DAVIS, J. F., TUBELLI, A. A. & ASIEDU, J. K. 2017. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell, 171, 1437-1452. e17.

SUN, Z., LUO, J., ZHOU, Y., LUO, J., LIU, K. & LI, W. 2009. Exploring phenotype-associated modules in an oral cavity tumor using an integrated framework. Bioinformatics, 25, 795-800.

SURAOKAR, M., NUNEZ, M., DIAO, L., CHOW, C., KIM, D., BEHRENS, C., LIN, H., LEE, S., RASO, G. & MORAN, C. 2014. Expression profiling stratifies mesothelioma tumors and signifies deregulation of spindle checkpoint pathway and microtubule network with therapeutic implications. Annals of Oncology, 25, 1184-1192.

SZKLARCZYK, D., GABLE, A. L., NASTOU, K. C., LYON, D., KIRSCH, R., PYYSALO, S., DONCHEVA, N. T., LEGEAY, M., FANG, T. & BORK, P. 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research, 49, D605-D612.

TANIGUCHI, T., PALMIERI, M. & WEISSMANN, C. 1978. Qβ DNA-containing hybrid plasmids giving rise to Qβ phage formation in the bacterial host. Nature, 274, 223-228.

TATONETTI, N. P., PATRICK, P. Y., DANESHJOU, R. & ALTMAN, R. B. 2012. Data-driven prediction of drug effects and interactions. Science translational medicine, 4, 125ra31-125ra31.

THAXTON, C., GOLDSTEIN, J., DISTEFANO, M., WALLACE, K., WITMER, P. D., HAENDEL, M. A., HAMOSH, A., REHM, H. L. & BERG, J. S. 2022. Lumping versus splitting: How to approach defining a disease to enable accurate genomic curation. Cell genomics, 2, 100131.

THEIS, J. L., HU, J. J., SUNDSBAK, R. S., EVANS, J. M., BAMLET, W. R., QURESHI, M. Y., O'LEARY, P. W. & OLSON, T. M. 2021. Genetic Association between Hypoplastic Left Heart Syndrome and Cardiomyopathies. Circulation: Genomic and Precision Medicine, 14, e003126.

THEIS, J. L., VOGLER, G., MISSINATO, M. A., LI, X., NIELSEN, T., ZENG, X.-X. I., MARTINEZ-FERNANDEZ, A., WALLS, S. M., KERVADEC, A. & KEZOS, J. N. 2020. Patient-specific genomics and cross-species functional analysis implicate LRP2 in hypoplastic left heart syndrome. Elife, 9, e59554.

THOMAS, D. W., BURNS, J., AUDETTE, J., CARROLL, A., DOW-HYGELUND, C. & HAY, M. 2016. Clinical development success rates 2006–2015. San Diego: Biomedtracker/Washington, DC: BIO/Bend: Amplion.

TORRI, F., AKELAI, A., LUPOLI, S., SIRONI, M., AMANN-ZALCENSTEIN, D., FUMAGALLI, M., DAL FIUME, C., BEN-ASHER, E., KANYAS, K., CAGLIANI, R., COZZI, P., TROMBETTI, G., STRIK LIEVERS, L., SALVI, E., ORRO, A., BECKMANN, J. S., LANCET, D., KOHN, Y., MILANESI, L., EBSTEIN, R. B., LERER, B. & MACCIARDI, F. 2010. Fine mapping of AHI1 as a schizophrenia susceptibility gene: from association to evolutionary evidence. FASEB J, 24, 3066-82.

TOTAH, N. 2016. Complexity and heterogeneity in psychiatric disorders: opportunities for computational psychiatry. Conference Proceedings.

TREPTE, P., BUNTRU, A., KLOCKMEIER, K., WILLMORE, L., ARUMUGHAN, A., SECKER, C., ZENKNER, M., BRUSENDORF, L., RAU, K. & REDEL, A. 2015. DULIP: a dual luminescence-

based co-immunoprecipitation assay for interactome mapping in mammalian cells. Journal of molecular biology, 427, 3375-3388.

UGURLUER, G., CHANG, K., GAMEZ, M. E., ARNETT, A. L., JAYAKRISHNAN, R., MILLER, R. C. & SIO, T. T. 2016. Genome-based mutational analysis by next generation sequencing in patients with malignant pleural and peritoneal mesothelioma. Anticancer research, 36, 2331-2338.

UHLÉN, M., FAGERBERG, L., HALLSTRÖM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, Å., KAMPF, C., SJÖSTEDT, E. & ASPLUND, A. 2015. Tissue-based map of the human proteome. Science, 347.

ULITSKY, I. & SHAMIR, R. 2007. Pathway redundancy and protein essentiality revealed in the Saccharomyces cerevisiae interaction networks. Molecular systems biology, 3, 104.

UNIPROT CONSORTIUM 2019. UniProt: a worldwide hub of protein knowledge. Nucleic acids research, 47, D506-D515.

VALDERAS, J. M., STARFIELD, B., SIBBALD, B., SALISBURY, C. & ROLAND, M. 2009. Defining comorbidity: implications for understanding health and health services. The Annals of Family Medicine, 7, 357-363.

VARGHESE, S., CHEN, Z., BARTLETT, D. L., PINGPANK, J. F., LIBUTTI, S. K., STEINBERG, S. M., WUNDERLICH, J. & ALEXANDER JR, H. R. 2011. Activation of the phosphoinositide-3-kinase and mammalian target of rapamycin signaling pathways are associated with shortened survival in patients with malignant peritoneal mesothelioma. Cancer, 117, 361-371.

VERMA, S. K., DESHMUKH, V., NUTTER, C. A., JAWORSKI, E., JIN, W., WADHWA, L., ABATA, J., RICCI, M., LINCOLN, J. & MARTIN, J. F. 2016. Rbfox2 function in RNA metabolism is impaired in hypoplastic left heart syndrome patient hearts. Scientific reports, 6, 1-10.

VIDAL, M., CUSICK, M. E. & BARABÁSI, A.-L. 2011. Interactome networks and human disease. Cell, 144, 986-998.

VON MERING, C., KRAUSE, R., SNEL, B., CORNELL, M., OLIVER, S. G., FIELDS, S. & BORK, P. 2002. Comparative assessment of large-scale data sets of protein–protein interactions. Nature, 417, 399-403.

WANG, G.-Z., BELGARD, T. G., MAO, D., CHEN, L., BERTO, S., PREUSS, T. M., LU, H., GESCHWIND, D. H. & KONOPKA, G. 2015. Correspondence between resting-state activity and brain gene expression. Neuron, 88, 659-666.

WANG, X., THIJSSEN, B. & YU, H. 2013. Target essentiality and centrality characterize drug side effects. PLoS computational biology, 9, e1003119.

WANG, Z. & ZHANG, J. 2007. In search of the biological significance of modular structures in protein networks. PLoS computational biology, 3.

WANG, Z. J., REDDY, G. P., GOTWAY, M. B., HIGGINS, C. B., JABLONS, D. M., RAMASWAMY, M., HAWKINS, R. A. & WEBB, W. R. 2004. Malignant pleural mesothelioma: evaluation with CT, MR imaging, and PET. Radiographics, 24, 105-119.

WEINBERG, A. G. & BOLANDE, R. P. 1970. The Liver in Congenital Heart Disease: Effects of Infantile Coarctation of the Aorta and the Hypoplastic Left Heart Syndrome in Infancy. American Journal of Diseases of Children, 119, 390-394.

WELLCOME TRUST CASE CONTROL CONSORTIUM 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447, 661-678.

WHALEN, P. J., JOHNSTONE, T., SOMERVILLE, L. H., NITSCHKE, J. B., POLIS, S., ALEXANDER, A. L., DAVIDSON, R. J. & KALIN, N. H. 2008. A functional magnetic resonance

imaging predictor of treatment response to venlafaxine in generalized anxiety disorder. Biological psychiatry, 63, 858-863.

WHITFIELD, J. F. & CHAKRAVARTHY, B. R. 2009. The neuronal primary cilium: driver of neurogenesis and memory formation in the hippocampal dentate gyrus? Cellular signalling, 21, 1351-1355.

WILLSEY, A. J., MORRIS, M. T., WANG, S., WILLSEY, H. R., SUN, N., TEERIKORPI, N., BAUM, T. B., CAGNEY, G., BENDER, K. J. & DESAI, T. A. 2018. The psychiatric cell map initiative: a convergent systems biological approach to illuminating key molecular pathways in neuropsychiatric disorders. Cell, 174, 505-520.

WISHART, D. S., KNOX, C., GUO, A. C., CHENG, D., SHRIVASTAVA, S., TZUR, D., GAUTAM, B. & HASSANALI, M. 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic acids research, 36, D901-D906.

WOODSMITH, J. & STELZL, U. 2017. Understanding disease variants through the lens of protein interactions. Cell systems, 5, 544-546.

WRAY, N. R., RIPKE, S., MATTHEISEN, M., TRZASKOWSKI, M., BYRNE, E. M., ABDELLAOUI, A., ADAMS, M. J., AGERBO, E., AIR, T. M. & ANDLAUER, T. M. 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature genetics, 50, 668.

WU, G., DAWSON, E., DUONG, A., HAW, R. & STEIN, L. 2014. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis. F1000Research, 3.

XENARIOS, I., SALWINSKI, L., DUAN, X. J., HIGNEY, P., KIM, S.-M. & EISENBERG, D. 2002. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research, 30, 303-305.

XU, S., XIAO, Q., COSMANESCU, F., SERGEEVA, A. P., YOO, J., LIN, Y., KATSAMBA, P. S., AHLSEN, G., KAUFMAN, J. & LINAVAL, N. T. 2018a. Interactions between the Ig-Superfamily proteins DIP-α and Dpr6/10 regulate assembly of neural circuits. Neuron, 100, 1369-1384. e6.

YANG, C., XU, Y., YU, M., LEE, D., ALHARTI, S., HELLEN, N., AHMAD SHAIK, N., BANAGANAPALLI, B., SHEIKH ALI MOHAMOUD, H. & ELANGO, R. 2017. Induced pluripotent stem cell modelling of HLHS underlines the contribution of dysfunctional NOTCH signalling to impaired cardiogenesis. Human molecular genetics, 26, 3031-3045.

YıLDıRıM, M. A., GOH, K.-I., CUSICK, M. E., BARABÁSI, A.-L. & VIDAL, M. 2007. Drug—target network. Nature biotechnology, 25, 1119-1126.

YOU, Z.-H., LEI, Y.-K., ZHU, L., XIA, J. & WANG, B. 2013. Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. BMC bioinformatics, 14, S10.

ZAFFARONI, N., COSTA, A., PENNATI, M., DE MARCO, C., AFFINI, E., MADEO, M., ERDAS, R., CABRAS, A., KUSAMURA, S. & BARATTI, D. 2007. Survivin is highly expressed and promotes cell survival in malignant peritoneal mesothelioma. Analytical Cellular Pathology, 29, 453-466.

ZAHIRI, J., MOHAMMAD-NOORI, M., EBRAHIMPOUR, R., SAADAT, S., BOZORGMEHR, J. H., GOLDBERG, T. & MASOUDI-NEJAD, A. 2014. LocFuse: human protein–protein interaction prediction via classifier fusion using protein localization information. Genomics, 104, 496-503.

ZAIDI, S. & BRUECKNER, M. 2017. Genetics and genomics of congenital heart disease. Circulation research, 120, 923-940.

ZAIDI, S., CHOI, M., WAKIMOTO, H., MA, L., JIANG, J., OVERTON, J. D., ROMANO-

ADESMAN, A., BJORNSON, R. D., BREITBART, R. E. & BROWN, K. K. 2013. De novo mutations in histone-modifying genes in congenital heart disease. Nature, 498, 220-223.

ZHAO, K. & SO, H.-C. 2018. Drug repositioning for schizophrenia and depression/anxiety disorders: A machine learning approach leveraging expression data. IEEE journal of biomedical and health informatics.

ZHENG, X., XING, X.-H. & ZHANG, C. 2017. Targeted mutagenesis: A sniper-like diversity generator in microbial engineering. Synthetic and systems biotechnology, 2, 75-86.

ZHOU, Y., HOU, Y., SHEN, J., HUANG, Y., MARTIN, W. & CHENG, F. 2020a. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discovery, 6, 14.

ZHOU, Y., HOU, Y., SHEN, J., KALLIANPUR, A., ZEIN, J., CULVER, D. A., FARHA, S., COMHAIR, S., FIOCCHI, C. & GACK, M. U. 2020b. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. ChemRxiv.

ZHU, J., ZHANG, Y., GHOSH, A., CUEVAS, R. A., FORERO, A., DHAR, J., IBSEN, M. S., SCHMID-BURGK, J. L., SCHMIDT, T. & GANAPATHIRAJU, M. K. 2014. Antiviral activity of human OASL protein is mediated by enhancing signaling of the RIG-I RNA sensor. Immunity, 40, 936-948.

ZUK, O., HECHTER, E., SUNYAEV, S. R. & LANDER, E. S. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences, 109, 1193-1198.

14. Appendix

14.1 Workflow for interactome-based framework to translate disease genetic data into biological and clinical insights

14.1.1 Compilation of disease-associated genes and construction of disease interactomes
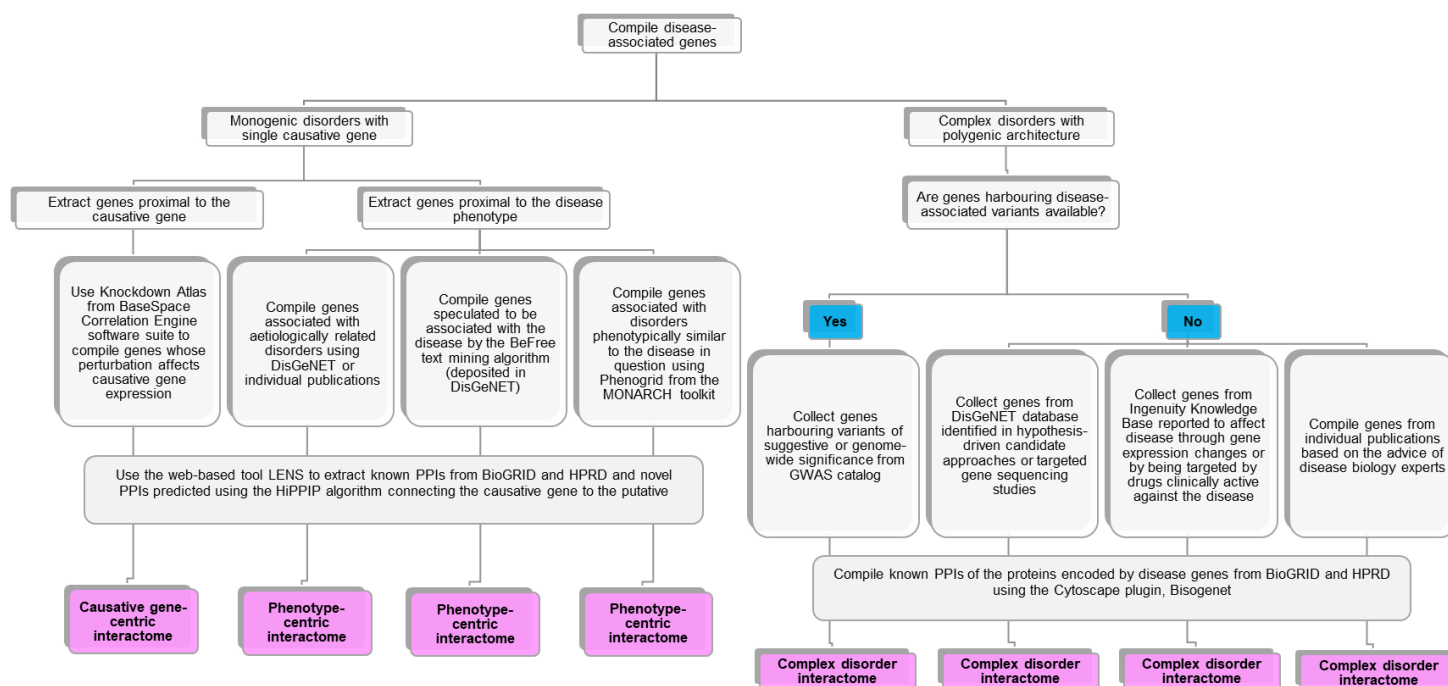
Appendix – Figure 1: Workflow for disease gene compilation and disease interactome construction. The workflow has been divided into two sections, one for single gene disorders and another for complex disorders with polygenic aetiology.

An overview of the methods in this section has been presented in **Appendix – Fig. 1**.

In the case of monogenic disorders exhibiting complex phenotypes unexplained by a single known causative gene, the initial step involves compiling two sets of disease-associated genes.

1. The first set comprises genes proximal to the known causative gene. To accomplish this, input the gene symbol into the Knockdown Atlas feature of the Illumina BaseSpace Correlation Engine software suite (http://www.illumina.com/basespacecorrelationengine) (Kupershmidt et al., 2010). This tool systematically compiles genes that, when subjected to knockout, knockdown, mutation, or overexpression, result in the upregulation or downregulation of the causative gene expression. For technical guidance on using the Correlation Engine, refer to https://sapac.illumina.com/products/by-type/informatics-products/basespace-correlation-engine.html. Note that BaseSpace Correlation Engine is a licensed software.

2. The second gene set includes genes proximal to the disease phenotype. This set comprises three subsets.

a. The first subset encompasses genes putatively associated with the disease, identified by the BeFree text mining algorithm (Bravo et al., 2015) retrievable from the DisGeNET database (https://www.disgenet.org/home/) (Piñero et al., 2015). DisGeNET compiles information on disease-associated human genetic variations, genetic associations in model organisms, disease-related differential gene expression, post-translational modifications, and gene-dependent responses to therapeutics. Input the disease name into the search field and download "Evidences for GDAs" as an XLSX file. Sort the results in the "original DB" field to obtain genes identified by the BeFree algorithm. BeFree utilises a kernel-based approach based on morphosyntactic and dependency information to identify potential gene-disease associations from MEDLINE abstracts (Bravo et al., 2015).

b. To compile the second subset, gather disorders aetiologically related to the disease in question based on expert advice. Utilise the DisGeNET database to compile genes associated with these related disorders, employing a gene-disease association (GDA) score of ≥ 0.7 from the "Evidences for GDAs" file. The GDA score considers factors like the number of supporting publications, types and number of database sources, and validations in model organisms. A stringent GDA threshold of ≥ 0.7 is typically used to select disease-associated genes from the DisGeNET database.

c. The third subset can be compiled using the Phenogrid feature from the MONARCH toolkit (https://monarchinitiative.org/) (Cacheiro et al., 2019), identifying disorders sharing >70% phenotypic similarity with the disease in question. Please refer to (Mungall et al., 2017), to learn how to use Phenogrid. The phenogrid algorithm identifies the common phenotypes between two given diseases. It then assesses the information content in each of these phenotypes (gene and disease associations) to assign a specific strength to the similarity observed between the diseases.

In total, one will be able to compile four gene sets: one causative gene-centric set and three phenotype-centric sets.

3. Subsequently, use the web-based tool LENS (http://severus.dbmi.pitt.edu/LENS/) to extract four separate networks containing the shortest path lengths in the human interactome connecting the causative gene to each of the four gene sets. Export these four networks to Cytoscape (https://cytoscape.org/) (Smoot et al., 2011) as SIF or XLSX files. For guidance on using LENS, refer to (Handen and Ganapathiraju, 2015). Note that LENS contains experimentally validated PPIs from BioGRID (https://thebiogrid.org/) (Stark et al., 2006) and HPRD (http://www.hprd.org/) (Keshava Prasad et al., 2009) databases, as well as predicted PPIs using the HiPPIP algorithm (Ganapathiraju et al., 2016).

4. For complex disorders with polygenic architectures, compile information on genes harbouring disease-associated variants from the GWAS catalogue (https://www.ebi.ac.uk/gwas/) (Buniello et al., 2019). Download the GWAS dataset (All associations v1.0.2 - with added ontology annotations, GWAS Catalog study accession numbers and genotyping technology) as a TSV file. Select the specific complex disorder in the field Disease/Trait and compile both Reported Genes and Mapped Genes with variants at levels suggestive of genome-wide significance (p-value < 1E-05) or those reaching genome-wide significance (p-value < 1E-08).

5. If GWAS information is unavailable, one of three alternative approaches can be employed.
   a. Firstly, use the DisGeNET database to compile genes with GDA ≥ 0.7 identified through candidate gene or targeted sequencing approaches ("Evidences for VDAs" file).
   b. Secondly, employ Ingenuity Pathway Analysis (IPA) software (https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/) to search for the disease name and retrieve genes causally related to the disease (Krämer et al., 2014). IPA retrieves genes from the Ingenuity Knowledge Base, which includes approximately 5 million experimentally curated findings from biomedical literature or other databases. Note that IPA is a licensed software.
   c. Thirdly, based on advice from experts in disease biology, compile disease-associated genes from individual publications.

6. The resulting gene set, compiled through one of the above four approaches, serves as the starting point to seed disease interactomes. Assemble interactomes by extracting PPIs of proteins encoded by disease-associated genes from the PPI repositories BioGRID and HPRD using the Cytoscape plugin, BisoGenet (Martin et al., 2010) (https://apps.cytoscape.org/apps/bisogenet). Ensure the network building options used in BisoGenet are set to organism - Homo sapiens, biorelation type - protein-protein interaction, data sources - BioGRID and HPRD, and method - input nodes and their neighbours up to a distance of 1.

14.1.2 Disease interactome validation using multi-omics patient datasets

An overview of the methods in this section has been presented in **Appendix – Fig. 2**.

The subsequent phase involves validating the constructed disease interactomes. To achieve this, differential gene expression and methylation datasets from patients can be utilised.

1. Employ the Curated Studies feature of the BaseSpace Correlation Engine to select relevant transcriptomic and methylation datasets from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2012). This feature allows one to examine pre-processed publicly available gene expression datasets from GEO.

2. Once the pertinent study is chosen, users can incorporate differentially expressed genes from the selected study into the Meta Analysis feature of the Correlation Engine.

3. Subsequently, upload a text file containing the list of interactome genes and check the intersection of this list with the differentially expressed/methylated genes.
   a. For transcriptomic studies, genes with a fold change >2 and <1/2 are considered significantly overexpressed and underexpressed, respectively, at a p-value < 0.05.
   b. For methylation studies, genes with methylation values >1 and <1 are regarded as hypermethylated and hypomethylated, respectively, at a p-value < 0.05.
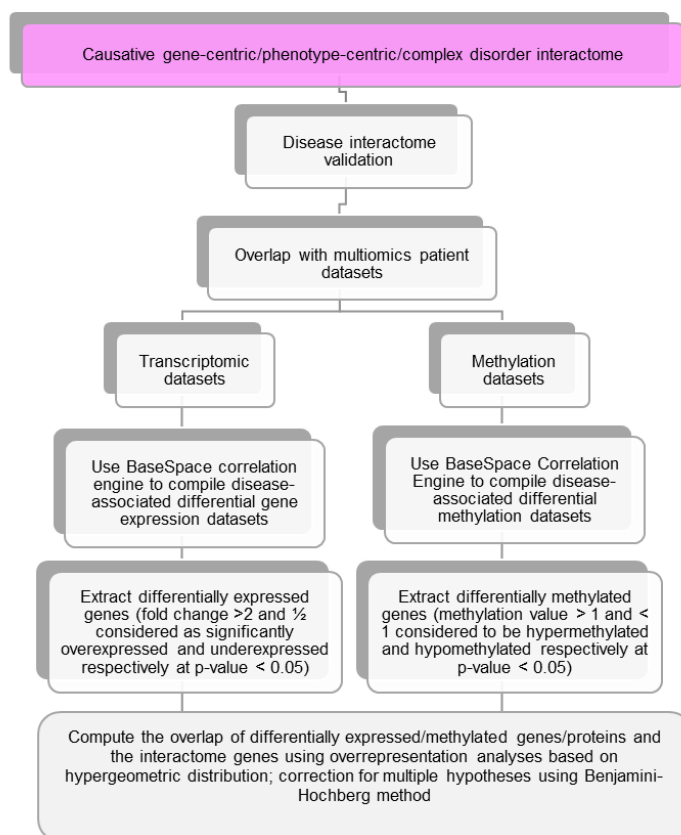
Appendix – Figure 2: Workflow for disease interactome validation. The workflow outlines analyses with transcriptomic and methylation datasets. Note that genetic and proteomic datasets can also be used for validation, as detailed below.

4. If available, one can assess the overlap with proteomic and external genetic datasets (i.e., those that are not utilised during the interactome construction), obtained from independent publications. Compile a list of differentially expressed proteins, derived from computations described in the publications, and genes reported to harbour genetic variants along with their corresponding evidences (i.e., proteomic/genetic evidence). Organise this information as an XLSX file and import it into the Cytoscape file containing the network as a Table. Subsequently, mark these evidences as Node Attributes. Utilise the Filter function in Cytoscape to identify the number of genes in the interactome with proteomic/genetic evidence.

5. Upon retrieving the number of intersecting genes, a hypergeometric p-value calculator (e.g., https://systems.crump.ucla.edu/hypergeometric/index.php) can be employed to compute the statistical significance of the overlap of differentially expressed/methylated genes with interactome genes. Input the four parameters (N, M, s, k) into the calculator as described below. In the hypergeometric test, the p-value is derived from the probability of k successes in s draws (without replacement) from a finite population of size N containing exactly M objects with an interesting feature:

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{s-k}}{\binom{N}{s}}$$

N = Total number of genes assayed in the experiment (retrieve this number from the GPL file attached to the selected study in GEO; the GEO platform file describes the list of genes assessed

346

in the experiment)

M = Number of differentially expressed/methylated genes

s = Number of genes in the interactome

k = Number of common genes between K and s (differentially expressed/methylated genes in the interactome)

6. Alternatively, the cumulative distribution function (CDF) of the hypergeometric distribution can also be calculated using Microsoft Excel:

   IF(k>=((s*M)/N),1-HYPGEOM.DIST(k-1,s,M,N,TRUE),HYPGEOM.DIST(k,s,M,N,TRUE)).

   a. The formula assesses whether the observed value, k, exceeds or equals the expected threshold value for enrichment calculated as (s*M)/N. If k surpasses this threshold, the cumulative probability is calculated using HYPGEOM.DIST(k-1, s, M, N, TRUE). If k is below the threshold, the cumulative probability is calculated using HYPGEOM.DIST(k, s, M, N, TRUE).

   b. The expected threshold for enrichment represents the expected number of successes in a sample of size s drawn from a population of N items, containing M successes.

   c. The direction of enrichment is calculated as IF(k=expected,"match",IF(k<expected,"de-enriched","enriched")). This formula categorises the relationship between the observed value k and the expected value, i.e. (s*M)/N. It indicates whether k matches the expected value, is lower than expected (de-enriched), or is higher than expected (enriched).

   d. The fold change of enrichment (or enrichment ratio) is calculated as IF(k<expected,expected/k,k/expected). The formula quantifies the magnitude of change between the observed value k and the expected value. It expresses how many times greater or smaller k is compared to the expected value. If k is less than the expected values, the fold change is calculated as expected/k. If k is greater than the expected value, the fold change is calculated as k/expected.

14.1.3 Disease interactome characterisation

An overview of the methods in this section has been presented in **Appendix – Fig. 3**.

For functional and phenotypic enrichment analyses, utilise WebGestalt (https://www.webgestalt.org/) (Liao et al., 2019). WebGestalt is a comprehensive web-based tool designed for gene set enrichment analysis and functional annotation of large gene lists. Offering a wide range of analysis options, including Gene Ontology (GO) analysis, pathway analysis (e.g., KEGG and Reactome), and disease enrichment (e.g., OMIM and DisGeNET) analysis, WebGestalt assists researchers in uncovering biological insights from high-throughput omics data through its user-friendly interface. Additionally, it provides interactive visualisations to aid in the interpretation of enrichment results.

1. Prepare the list of interactome genes, ensuring they are in a compatible format such as official gene symbols or Entrez Gene IDs.
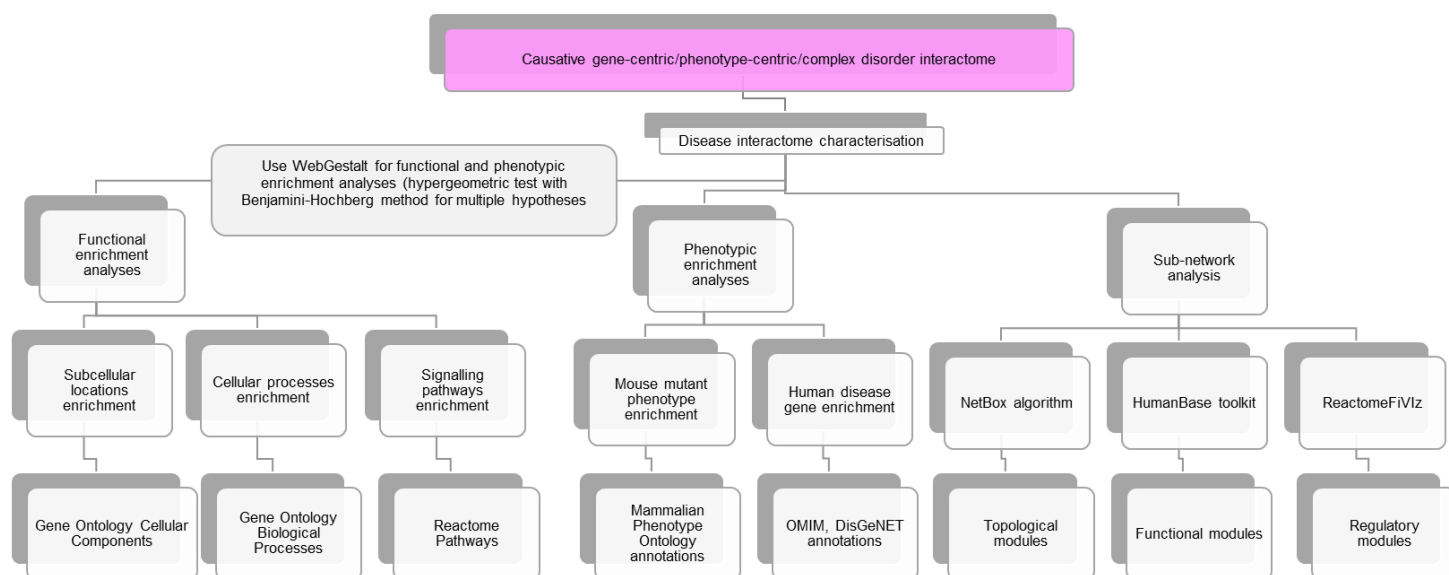
Appendix – Figure 3: Workflow for disease interactome characterisation. The workflow has been divided into three sections, namely, functional enrichment analyses, phenotypic enrichment analyses and sub-network analyses. Note that the former two analyses are performed using WebGestalt, whereas the latter is performed using three separate toolkits.

2. Open the WebGestalt website in your web browser. To learn how to use WebGestalt, refer: https://www.webgestalt.org/WebGestalt_2019_Manual.pdf

3. Choose the type of analysis to perform, such as GO analysis, pathway analysis, or disease enrichment analysis using the Functional Database feature.

4. Upload your gene list to WebGestalt by either copying and pasting into the provided text box or uploading a file.

5. Choose the reference database for enrichment analysis, considering options like the entire genome, custom gene sets, or pre-defined gene sets based on specific databases.

6. Adjust parameters including statistical method for multiple testing correction, significance threshold, and minimum number of genes required for significance. Consider the appropriate settings based on the size of your input gene list and the desired stringency of the analysis.

7. Initiate the analysis by clicking Submit. WebGestalt will compute enrichment based on your data and parameters.

8. Review the enriched categories (GO terms, pathways, etc.) along with p-values and enrichment scores. Focus on significantly enriched categories, typically those with adjusted p-values below your chosen threshold.

9. Explore interactive visualisations provided by WebGestalt. These include: (i) bar charts ranked based on the enrichment scores of the enriched biological annotations and coloured based on

levels of FDR-corrected p-values, (ii) scatter plots that help explore the enriched annotations based on both the enrichment scores and FDR-corrected p-values associated with them, and (iii) enrichment maps that help explore relationships between enriched categories and visualise sub-networks, e.g., for GO enrichments.

10. Download results table containing enriched categories and statistics in TXT format for further analyses.

For sub-network analyses of the interactome, one can use NetBox, HumanBase and ReactomeFiViz for the detection of topological, functional and regulatory modules, respectively.

NetBox (Liu et al., 2020) is a computational algorithm designed for identifying topological modules within PPI networks; refer to https://bioc.ism.ac.jp/packages/3.16/bioc/vignettes/netboxr/inst/doc/netboxrTutorial.html for technical assistance on using NetBox. It employs a community detection approach to partition the network into cohesive modules based on network topology, thereby revealing functional units or complexes within the interactome. These modules represent groups of proteins that exhibit dense connectivity and are likely to participate in similar biological processes or pathways. By uncovering these modules, NetBox facilitates the exploration of protein interactions and their role in cellular functions, disease mechanisms, and drug discovery.

1. To begin, obtain the PPI network data representing the human interactome. This data can be obtained from publicly available databases such as BioGRID and HPRD (as well as STRING or IMEx).

2. Next, clean and preprocess the PPI network data to eliminate low-confidence interactions, duplicate entries, and non-human proteins if present.

3. Format the pre-processed PPI network data into a suitable file format compatible with NetBox (e.g., edge list, adjacency matrix).

4. Execute the NetBox algorithm using the prepared input data. Configure parameters such as the resolution parameter (gamma) to control the granularity of module detection. Specify the desired number of modules or allow NetBox to automatically determine the optimal number based on network topology.

5. Retrieve the identified topological modules generated by the NetBox algorithm. Each module represents a cohesive group of proteins with dense intra-module connectivity and potentially shared biological functions.

6. Analyse the properties of the identified modules, such as their size, density, and internal connectivity.

7. Conduct functional enrichment analysis on the proteins within each module to elucidate the biological processes, molecular functions, and cellular components overrepresented within the

modules. This can be achieved using tools for functional enrichment analysis such as WebGestalt.

8. Visualise the topological modules using network visualisation tools such as Cytoscape to gain insights into their structure and connectivity patterns.

9. Validate the functional implications of the modules through literature review, comparison with experimental results, and integration with omics data.

In the HumanBase toolkit (https://hb.flatironinstitute.org/) (Krishnan et al., 2016), community detection is utilised to identify cohesive gene clusters by combining an input gene list with a relevant tissue selection. These clusters comprise genes that share local network neighbourhoods, thus creating functional modules. Functional modules are delineated using tissue-specific networks, which utilise extensive data collections to forecast gene interactions and link genes with specific GO biological processes. Specifically, HumanBase employs shared k-nearest-neighbours and the Louvain community-finding algorithm to cluster genes sharing similar network neighbourhoods and Gene Ontology biological processes into functional modules. The p-values of the terms enriched in the modules are calculated using Fisher's exact test and the Benjamini-Hochberg method. The modules discovered by this approach hold the potential to encompass higher-order, tissue-specific functionalities.

1. To commence, choose the analysis type specific to functional modules and input the list of interactome genes. Select particular tissue-specific co-expression networks (e.g., nervous system) or the global tissue-naïve co-expression network for mapping the interactome genes.

2. Retrieve the identified functional modules generated by the co-expression network analysis. These modules represent groups of genes with correlated expression patterns and are likely to participate in similar biological processes or pathways. Adjust tissue-specific networks or parameters such as minimum module size as necessary.

3. Analyse the properties of the identified functional modules, including module size, density, and functional coherence. Explore the biological significance of the modules by examining the enrichment of Gene Ontology terms. These results, including the module membership details of the interactome genes, can be exported in TSV format.

4. The detected modules can be mapped to the Cytoscape file of the interactome, and the network topology, node attributes, and connectivity patterns can be explored to gain insights into module structure.

ReactomeFIViz (https://apps.cytoscape.org/apps/reactomefiplugin) (Wu et al., 2014) is a Cytoscape plugin designed to explore and visualise functional interaction networks derived from the Reactome database (https://reactome.org/) (Croft et al., 2014). Integrating molecular interaction data with pathway information from Reactome, ReactomeFIViz enables users to identify regulatory modules and functional interactions within biological pathways. By capitalising on the extensive biological knowledge encoded in Reactome, ReactomeFIViz empowers researchers to uncover complex

regulatory relationships and gain insights into the functional organisation of biological systems.

1. Begin by launching Cytoscape, an open-source platform for visualising and analysing biological networks. Install the ReactomeFIViz plugin within Cytoscape by navigating to the Cytoscape App Store, searching for ReactomeFIViz, and installing the plugin. The plugin provides access to the extensive collection of pathway information and molecular interactions in Reactome.

2. Choose the Gene Set/Mutation Analysis function of ReactomeFIViz, input the list of interactome genes, and under FI network construction parameters, select Fetch FI Annotations and Use Linker Genes, if necessary.

3. Visualise the retrieved functional interaction networks using built-in visualisation tools of Cytoscape. Explore network topology, node attributes, and functional annotations to gain insights into regulatory relationships within the pathway.

4. Analyse the regulatory modules to uncover key regulatory relationships and functional associations within the pathway. Interpret the results in the context of existing biological knowledge and experimental data. Perform additional analyses or validations on the identified regulatory modules, such as functional enrichment analysis using WebGestalt.

14.1.4 Identification and validation of repurposable drugs

An overview of the methods in this section has been presented in **Appendix – Fig. 4**.

1. First, identify drugs targeting at least one protein in the disease interactome by accessing DrugBank (https://go.drugbank.com/) (Wishart et al., 2008) via WebGestalt (i.e., perform an enrichment analysis of the interactome with Drug Bank as the Functional Database). Utilise the Redundancy Reduction feature in WebGestalt to prioritise drugs exhibiting highly significant overlaps with the interactome. This feature employs an affinity propagation algorithm, clustering sets of genes in the interactome targeted by specific drugs using the Jaccard index as the similarity metric. It identifies a representative for each cluster, ensuring that only individual drugs targeting a statistically significant number of proteins in the interactome are prioritised for further analysis.

2. For comparative transcriptomic analysis of drug-induced (from Connectivity Map https://www.broadinstitute.org/connectivity-map-cmap (Subramanian et al., 2017)) versus disease-associated expression profiles (from GEO), leverage BaseSpace Correlation Engine. This tool allows users to study the impact of diseases and/or drugs on thousands of pre-processed publicly available gene expression datasets. Refer to the protocol described in (Chattopadhyay and Ganapathiraju, 2017) to obtain the list of drugs whose expression profiles are negatively correlated with disease-associated expression profiles. The Correlation Engine
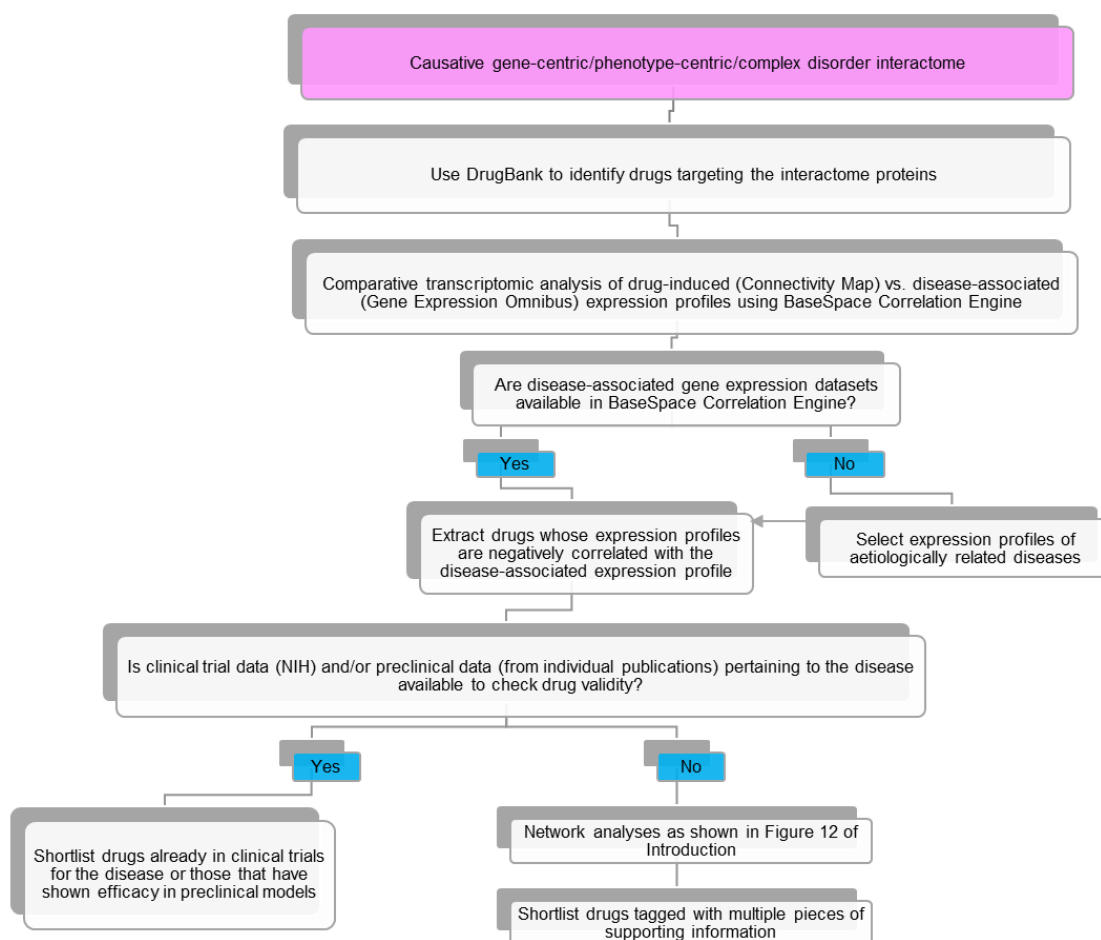
Appendix – Figure 4: Workflow for repurposable drug identification. The most important component of this workflow is comparative transcriptomic analysis of drug and disease profiles, which will be followed by validation analyses.

generates a correlation score based on the strength of the overlap between the drug and disease datasets. Statistical criteria, including correction for multiple hypothesis testing, are applied, and correlated datasets are ranked by statistical significance. A numerical score of 100 is assigned to the most significant result, and other scores are normalised with respect to this top-ranked result.

3. In cases where specific disease differential expression profiles are unavailable, use expression profiles of aetiologically related disorders.

4. Identify the intersection of drugs identified in steps 1 and 2 to pinpoint drugs targeting proteins in the interactome while also exhibiting negative correlation with the disease.

5. Subsequently, assess the validity of the shortlisted drugs through two approaches. If clinical trial data is available for the disease in the NIH Clinical Trials database (https://clinicaltrials.gov/), check whether a clinical trial is underway or has been successfully completed by filling the Intervention/Treatment field with the name of the shortlisted drug and the Condition/Disease field with the name of the disease. Additionally, check whether literature evidence supports the efficacy of the shortlisted drugs in preclinical animal models or cell lines.

6. In case such clinical data is unavailable, conduct a series of gene set and network analyses to evaluate the validity of the shortlisted drugs. Compile the targets of the shortlisted drugs from

DrugBank and subject them to functional and phenotypic enrichment analyses as outlined in section 14.1.3. Similarly, perform these analyses with the targets and their first-order interactors (i.e., drug target networks), and also with the specific genes displaying anti-correlated expression in drug versus disease expression datasets and their network.

14.1.5 Comparative disease interactome analysis

An overview of the methods in this section has been presented in **Appendix – Fig. 5**.

To compile disease-associated genes and construct the disease interactome, follow the steps outlined in section 14.1.1. The steps for comparative interactome analysis have been elucidated in relation to the biological modality of regional gene expression in the human brain. Specifically, the procedure has been detailed with respect to addressing the question of how one can derive subgroupings among specific psychiatric disorders based on the regional expression patterns of their interactomes in the brain.

1. Begin by preparing four transcriptomic datasets in the form of gene matrix transposed (GMT) files.
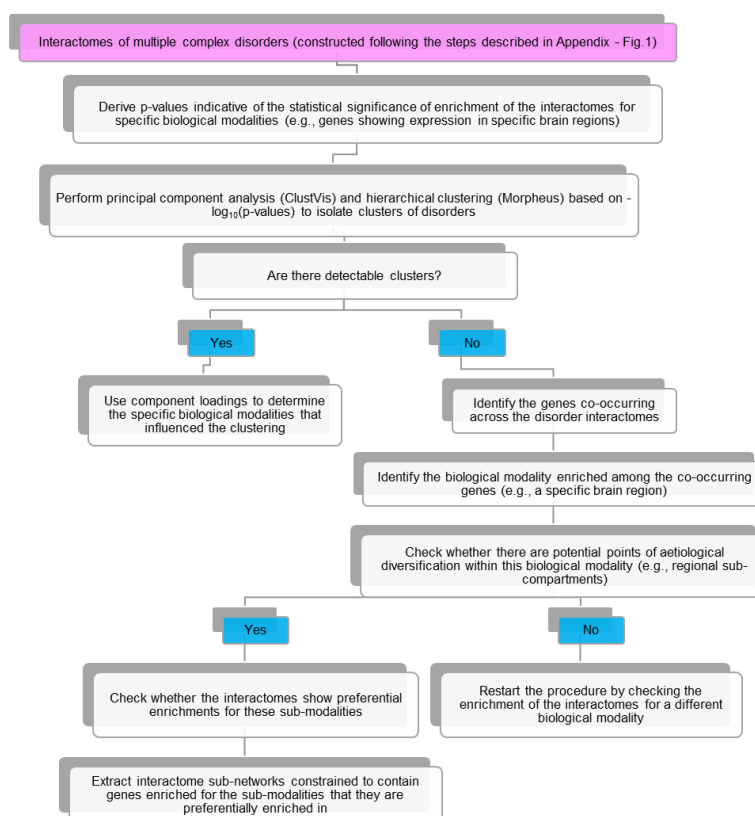
Appendix – Figure 5: Workflow for comparative disease interactome analysis. The workflow includes contingencies for situations where a particular biological modality fails to uncover disorder subgroupings.

    a. Download the BrainSpan Atlas dataset (https://www.brainspan.org/static/download.html; file: RNA-Seq Gencode v10 summarised to genes) (Hawrylycz et al., 2012) and identify genes with logRPKM > 2 in 26 brain regions that are not housekeeping genes (RPKM = Reads Per Kilobase per Million mapped reads). This entails identifying 9,638 genes (https://www.proteinatlas.org/humanproteome/tissue/housekeeping) detected in all tissues

with TPM ≥ 1, as specified by the Human Protein Atlas (TPM = Transcripts Per Million) (https://www.proteinatlas.org/) (Uhlén et al., 2015). The GMT file will include data from various brain regions: amygdaloid complex (3,231 genes), anterior (rostral) cingulate (medial prefrontal) cortex (5,362 genes), caudal ganglionic eminence (3,677 genes), cerebellar cortex (3,542 genes), cerebellum (3,601 genes), dorsal thalamus (3,410 genes), dorsolateral prefrontal cortex (3,473 genes), hippocampus (hippocampal formation) (3,628 genes), inferolateral temporal cortex – area TEv, area 20 (7,000 genes), lateral ganglionic eminence (3,698 genes), medial ganglionic eminence (3,678 genes), mediodorsal nucleus of thalamus (3,442 genes), occipital neocortex (3,790 genes), orbital frontal cortex (3,506 genes), parietal neocortex (3,724 genes), posterior (caudal) superior temporal cortex – area 22c (3,443 genes), posteroventral (inferior) parietal cortex (3,369 genes), primary auditory cortex (core) (3,437 genes), primary motor cortex – area M1, area 4 (3,431 genes), primary motor-sensory cortex (3,628 genes), primary somatosensory cortex – area S1, areas 3,1,2 (3,402 genes), primary visual cortex-striate cortex – area V1/17 (3,436 genes), striatum (3,557 genes), temporal neocortex (3,611 genes), upper (rostral) rhombic lip (3,528 genes) and ventrolateral prefrontal cortex (3,467 genes).

b. Download RNA-seq data from the brains of adult donors available in GTEx (https://gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression; file: GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct.gz) (Consortium, 2015). Include genes with high or medium expression (TPM ≥ 9) in 13 brain regions in the GMT file, excluding housekeeping genes. The brain regions include: amygdala (1,953 genes), anterior cingulate cortex-BA24 (2,269 genes), caudate nucleus (2,229 genes), cerebellar hemisphere (3,978 genes), cerebellum (3,968 genes), cortex (2,706 genes), frontal cortex-BA9 (2,872 genes), hippocampus (1,949 genes), hypothalamus (2,374 genes), nucleus accumbens (2,464 genes), putamen (1,892 genes), spinal cord-cervical c-1 (2,408 genes) and substantia nigra (1,949 genes).

c. Create a third GMT file using genes showing high expression in each of 516 prenatal brain structures relative to other structures (Hawrylycz et al., 2012). Collect these expression profiles from 4 human prenatal samples spanning 4 time points from the Harmonizome database (download from: https://maayanlab.cloud/Harmonizome/dataset/Allen+Brain+Atlas+Prenatal+Human+Brain+Tissue+Gene+Expression+Profiles) (Rouillard et al., 2016).

d. Lastly, compile a GMT file containing the genes highly expressed in each of the 232 structures available in the Allen Brain Atlas microarray dataset (Hawrylycz et al., 2012) relative to others, as described by (Rouillard et al., 2016), in the form of a GMT file (https://human.brain-map.org/static/download; files to be downloaded: H0351.2001, H0351.2002, H0351.1009, H0351.1012, H0351.1015, H0351.1016). Normalised microarray data will be available for 29,130 genes from the brains of six healthy donors (IDs 9861, 10021, 12876, 14380, 15496 and 15697), each dissected into 363-946 samples – yielding a total of 3702 samples – from 13 areas in the brain, namely, cerebellum, cerebral nuclei, diencephalon, frontal lobe, insular cortex, limbic lobe, medulla oblongata, midbrain, occipital lobe, parietal lobe, pons, temporal lobe and white matter. Note that the (i) probes matching multiple genes should be excluded, (ii) if the same gene has been detected by multiple probes, the expression levels should be averaged across the probes, and (iii) 232 brain regions (414 when the samples from the left

and right hemispheres are treated separately) should be grouped into 13 brain areas based on the structured vocabulary used in Allen Brain Atlas to classify them.

2. Perform an overrepresentation analysis to determine the enrichment of different disease interactomes for brain regions in each of the four transcriptomic datasets.
   a. Upload each of the four GMT files using the Functional Database feature in WebGestalt, upload each of the interactome gene lists under the Gene List section, select the reference set as genome and adjust the Advanced Parameters as per requirements (e.g. adjust the maximum number of genes for a category depending on the gene list sizes in the GMT files).
   b. Choose the multiple test adjustment procedure as Benjamini-Hochberg and retrieve the statistical significances as FDR-corrected p-values for every brain region represented in the transcriptomic dataset (which will depend on the varying number of sampled brain regions across the four datasets).

3. Compile negative log-transformed p-values of regional enrichments into a data matrix with brain regions as rows and the disorders as columns.

4. Use the ClustVis web-based tool for principal component analysis, inputting the data matrix for analysis. To learn how to use ClustVis, refer to (Metsalu and Vilo, 2015).
   a. Pre-process the data matrix by excluding rows and columns with more than 70% missing values, and centre the $-\log_{10}P$ values using unit variance scaling.
   b. Apply the singular value decomposition (SVD) method with imputation to extract principal components, predicting and filling missing values iteratively.
   c. Transform original variables ($-\log_{10}P$) into uncorrelated variables (principal components), ranking them by the percentage of total variance explained.
   d. Plot component scores (possibly PC1 and PC2 explaining majority of the observed expression variance) of each tissue on a 2D plane, with PC1 on the X-axis and PC2 on the Y-axis.
   e. Conduct an initial assessment of potential clusters based on the plotted component scores. For independent validation of the detected clusters, the data matrix can be hierarchically clustered using Morpheus (https://software.broadinstitute.org/morpheus/) (Starruß et al., 2014). Pairwise distances in the data matrix can be calculated using Pearson correlation and closely linked clusters can be identified using the average linkage method.
   f. Extract factor/component loadings corresponding to brain regions contributing to selected principal components.
   g. Plot the loading of each brain region on the X-axis (PC1) and Y-axis (PC2) to analyse their influence on grouping patterns.
   h. Analyse results, interpret patterns, and draw conclusions regarding the relationships between brain regions and disorders based on the contribution to principal components.

5. If no cross-disorder relationships are detected, use the Advanced Merge Function in the Tools section of Cytoscape (specifically, the intersection function) to extract the common sub-network shared across the disorder interactomes. Check for the enrichment of shared genes for specific brain regions using Webgestalt following the same procedure detailed above.

6. If significant leads are found, check whether potential points for aetiological diversification can be identified (e.g., histochemically delineable region sub-compartments).

7. Compile marker genes for region sub-compartments if available from individual publications, then construct their respective interactomes using the procedures described in section 14.1.1.

8. Check whether the sub-compartment interactomes show preferential intersections with the disorder interactomes based on hypergeometric tests, as described in section 14.1.2.

9. If they do, use these intersecting subnetworks for principal component analysis and to detect cross-disorder relationships. Repeat the steps described above.

10. If this fails, examine the enrichment of disorder interactomes for other biological modalities (e.g., signalling pathways) and repeat the above procedure.

14.1.6 Comparative drug target network analysis

An overview of the methods in this section has been presented in **Appendix – Fig. 6**.

The steps for conducting comparative drug target network analysis will be explained in the context of drug contraindications in comorbidities. Specifically, the process has been outlined to address how different biological factors influence the target networks of drugs used for a primary condition, whether it is contraindicated or not, in a comorbid condition.
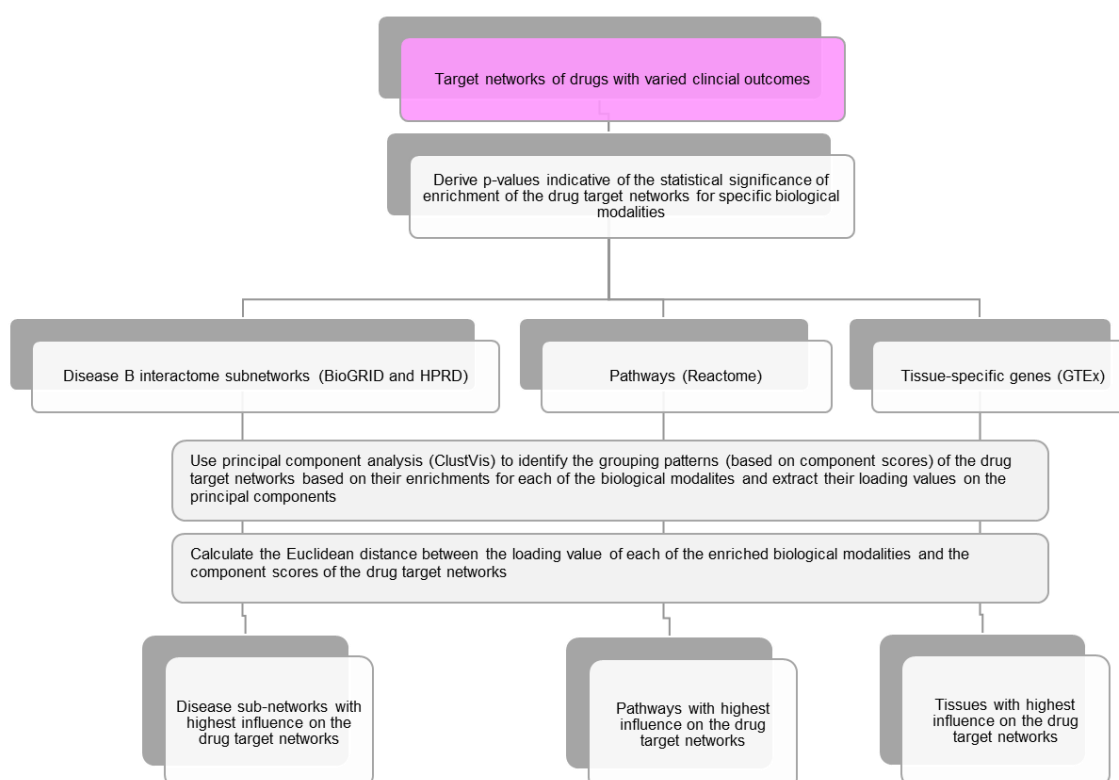
Appendix – Figure 6: Workflow for comparative drug target network analysis. This workflow will help identify the biological modalities underlying target networks of drugs with different outcomes in clinically related diseases.

1. Compile genes associated with comorbid disease pairs (test disease sets) and non-comorbid disease pairs (negative control disease pairs) using the procedures detailed in section 14.1.1.

2. Assemble PPI networks for disease-associated genes using HPRD and BioGRID databases as described in section 14.1.1. Utilise the Cytoscape plugin, Bisogenet, with specified options including organism (Homo sapiens), biorelation type (protein-protein interaction), data sources (BioGRID and HPRD), and method (input nodes and their neighbours up to a distance of 1).

3. Utilise the Drug Bank database to gather lists of drugs indicated for each disease in both the comorbid and non-comorbid pairs.

4. Employ the TWOSIDES database (https://www.tatonettilab.org/resources/nsides/) (Tatonetti et al., 2012) to categorise drugs based on their effects on disease pairs. Group drugs into four categories: (a) drugs for disease A that are contraindicated and (b) not contraindicated for disease B, as well as (c) drugs for disease B that are contraindicated and (d) not contraindicated for disease A. Identify drugs associated with specific adverse effects using condition concept names in the TWOSIDES database. For instance, to identify anxiolytic drugs that may cause depression, select the condition concept names depression, major depression, depressive symptom, depression suicidal, depression postoperative, postpartum depression, depressive delusion, and agitated depression. The list of anxiolytic drugs (compiled from the Drug Bank database) can be compared with the list of drugs associated with these condition concept names. The matching drugs can be compiled into groups 'a' and 'c', for example, "drugs effective in anxiety and contraindicated in depression". Groups 'b' and 'd' drugs can be compiled in a similar manner.

5. Compile proteins targeted by drugs from the Drug Bank database by querying the DGIdb web portal (https://old.dgidb.org/) (Griffith et al., 2013). Assemble drug target networks (DTNs) containing the drug targets and their first-order interactors. Compile PPIs of drug targets from HPRD and BioGRID using Bisogenet with the same options as for disease network construction.

6. To characterise the four DTNs associated with each disease pair, examine three types of data: (i) disease PPI networks, (ii) biological pathways, and (iii) tissue gene expression.
   a. For each disease pair, calculate the overlaps of the four DTNs with proteins that are (a) common to both disease A and disease B networks, (b) unique to disease A network, and (c) unique to disease B network using the Advanced Merge Functions in the Tools section of Cytoscape. Then, add these protein subsets into a GMT file and use it to compute their enrichment for the four DTNs associated with respective disease pairs using WebGestalt, and retrieve FDR-corrected p-values (p-values corrected for multiple hypotheses using the Benjamini-Hochberg method).
   b. Use WebGestalt to compute the distribution of genes involved in specific signalling pathways (functional database: Reactome) in the DTNs. Retrieve FDR-corrected p-values for the enrichment of the pathways in the four DTNs associated with each disease pair.
   c. Use the GTEx database to compute the enrichment of DTNs for genes expressed in specific tissues and retrieve FDR-corrected p-values. Create GMT files for hypergeometric tests in WebGestalt. GTEx contains RNA-sequencing data from 53 postnatal human tissues. Genes

showing high or medium expression (TPM ≥ 9) in specific tissues can be included in the analysis, provided they are not housekeeping genes, i.e., those detected in all tissues with TPM ≥ 1, as described in the Human Protein Atlas.

7. Use PCA to capture relationships of the four DTNs for each disease pair based on enrichment in disease sub-networks, biological pathways, and tissues.
    a. Perform PCA using ClustVis (https://biit.cs.ut.ee/clustvis/) (Metsalu and Vilo, 2015) with a data matrix containing DTNs (columns) versus the specific disease protein sets, pathways, or tissues (rows). For example, for the data modality 'disease protein set', the rows would be 'common to both networks', 'unique to disease A network', and 'unique to disease B network', while for the data modality 'tissue', the members would be 'amygdala', 'aorta', 'lungs', etc. Each cell will contain $-\log_{10}$ transformed p-values, indicating the enrichment of a disease protein set/pathway/tissue in a specific DTN.
    b. Pre-process the data matrix for missing values and apply log transformation and unit variance scaling, as described in section 14.1.5.

8. All generated PCs can be considered for further analysis, and the PC scores of the DTNs can be used to identify their grouping patterns.

9. Extract the component loading values denoting the weights of each biological modality on the PCs. Component loadings depict the correlation of the original variables ($-\log_{10}P$ values) in our data matrix with each of the extracted PCs. Their magnitudes can be used to assess the influence of different biological modalities on the 4 DTNs separated along the PCs.

10. Calculate the Euclidean distance between the PC scores of each of the DTNs and the corresponding component loadings of the biological modalities (see code deposited in Github: https://github.com/KalyaniBindu/Interactome-Frameworks/blob/main/euclidean-distance.txt) This will yield a list of specific disease protein sets/pathways/tissues closely related to each of the 4 DTNs associated with each disease pair.

11. To interpret the relevance of specific disease protein sets and pathways to specific target networks of drugs with particular clinical activity, seek expert advice. For interpretation of tissues closely related to specific DTNs with different clinical outcomes, identify tissues enriched with disease-specific single nucleotide polymorphisms using TSEA-DB (https://bioinfo.uth.edu/TSEADB/) (Jia et al., 2020), a reference database for information on disease-associated tissues from GTEx and GWAS catalogue. Retrieve the top-3 tissues showing significant enrichment for disease-associated variants for each disease from TSEA-DB and compare these results with tissues obtained from the Euclidean distance analysis.

Github page: https://github.com/KalyaniBindu/Interactome-Frameworks

14.2 Novel insights on acetylcholinesterase inhibition by Convolvulus pluricaulis, scopolamine and their combination in zebrafish

The experimental chapter is based on the following peer-reviewed publication:
Karunakaran, Kalyani B., Anand Thiyagaraj, and Kirankumar Santhakumar. Novel insights on acetylcholinesterase inhibition by Convolvulus pluricaulis, scopolamine and their combination in zebrafish. Natural Products and Bioprospecting 12, no. 1 (2022): 1-15.

Summary of this chapter

In this chapter, I demonstrate how preclinical studies in animal models can be conducted to test the efficacy of a candidate drug and the mechanisms underlying its pharmacological actions, using a crude drug for Alzheimer's disease as an example. I investigated the anti-acetylcholinesterase (AChE) effects of *Convolvulus pluricaulis* (CP) on the biochemical and behavioural deficits induced by scopolamine in zebrafish, using isoxazole as a positive control for AChE inhibition. I used Ellman's assay, Karnovsky staining, and hydroxylamine methods to quantify AChE activity and acetylcholine levels in 168 hours post-fertilization larvae. I performed molecular docking of human AChE with the active components of CP (scopoletin and kaempferol) to elucidate their inhibitory mechanisms. I performed the passive avoidance response test to examine avoidance response acquisition in adult zebrafish. I used the wrMTrck software to examine the locomotor patterns of the larvae. CP-treated larvae showed similar AChE inhibition patterns as those treated with isoxazole. CP also improved the retention of avoidance response in adult zebrafish compared with isoxazole and elicited specific locomotor responses in the larvae. CP's active components bound to different residues in the catalytic site and the peripheral anionic site of the human AChE. Combining CP with scopolamine significantly increased AChE inhibition and depleted ACh levels in larvae, similar to isoxazole. Overall, the study proposed the examination of CP for its ability to modulate cognitive deficits in Alzheimer's disease.

Contribution to this chapter (85%)

- Designed the study and developed the methodology of the project, including zebrafish rearing, embryo collection, plant extract preparation, standardisation of lethal concentration, Ellman's assay, Karnovsky staining, hydroxylamine assay, passive avoidance response test, assessment of larval locomotor activity and protein-small molecule rigid body docking
- Performed all the experimental and bioinformatic analyses, and derived the conclusions
- Conceptualised and wrote the manuscript and prepared all the figures, tables and supplementary files

Natural Products and
Bioprospecting

ORIGINAL ARTICLE

Open Access

Novel insights on acetylcholinesterase inhibition by *Convolvulus pluricaulis*, scopolamine and their combination in zebrafish

Kalyani Bindu Karunakaran[1], Anand Thiyagaraj[2] and Kirankumar Santhakumar[2,3*]

Abstract

Acetylcholinesterase (AChE) inhibitors increase the retention of acetylcholine (ACh) in synapses. Although they alleviate cognitive deficits in Alzheimer's disease, their limited benefits warrant investigations of plant extracts with similar properties. We studied the anti-AChE activity of *Convolvulus pluricaulis* (CP) in a zebrafish model of cognitive impairment induced by scopolamine (SCOP). CP is a perennial herb with anti-amnesiac and anxiolytic properties. It contains alkaloid, anthocyanin, coumarin, flavonoid, phytosterol and triterpenoid components. Isoxazole (ISOX) was used as a positive control for AChE inhibition. CP-treated 168 hpf larvae showed a similar pattern of AChE inhibition (in the myelencephalon and somites) as that of ISOX-treated larvae. CP was superior to ISOX as evidenced by the retention of avoidance response behavior in adult zebrafish. Molecular docking studies indicated that ISOX binds Ser203 of the catalytic triad on the human AChE. The active components of CP—scopoletin and kaempferol—were bound by His447 of the catalytic triad, the anionic subsite of the catalytic center, and the peripheral anionic site. This suggested the ability of CP to mediate both competitive and non-competitive modes of inhibition. Surprisingly, SCOP showed AChE inhibition in larvae, possibly mediated via the choline-binding sites. CP + SCOP induced a concentration-dependent increase in AChE inhibition and ACh depletion. Abnormal motor responses were observed with ISOX, CP, ISOX + SCOP, and CP + SCOP, indicative of undesirable effects on the peripheral cholinergic system. Our study proposes the examination of CP, SCOP, and CP + SCOP as potential AChE inhibitors for their ability to modulate cognitive deficits.

Keywords: Alzheimer's disease, Acetylcholinesterase, Zebrafish, *Convolvulus pluricaulis*, Scopolamine, Isoxazole

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by memory loss, behavioral changes, and impaired cognition and language [1]. Around two-thirds of dementia cases have been attributed to AD, and it has an estimated prevalence of 10–30% in the population aged 65 years and more [1]. AD has a long preclinical phase of around 20 years andthe average survival time for a person diagnosed with

the disease is 8–10 years [2]. The two most distinctive hallmarks of AD are the accumulation of amyloid-beta plaques in the brain and the aggregation of tau proteins into neurofibrillary tangles within neurons [1]. Accumulation of amyloid-beta plaques leads to widespread non-specific degeneration of neurons, which in turn, affects various neurotransmitter systems including the cholinergic, monoaminergic, and glutamatergic systems [1]. Cognitive deficits seen in AD such as impaired memory and learning are often attributed to depleted levels of the neurotransmitter acetylcholine (ACh) in the basal forebrain cholinergic system, which is involved in arousal, memory coding, and storage and retrieval of working memory [3]. Therefore, maintaining substantial levels of

*Correspondence: kirankus@srmist.edu.in
[3] Zebrafish Genetics Laboratory, Department of Genetic Engineering, SRM Institute of Science and Technology, Kattankulathur 603 203, India
Full list of author information is available at the end of the article

ACh in the neuronal synapses of the cholinergic system may be integral to rescuing these cognitive deficits [4].

Depleted ACh levels could result from the dramatic loss of cholinergic neurons in the basal forebrain, reduced cholinergic innervation to the hippocampus and neocortex, reduced levels of the enzyme choline acetyltransferase (ChAT), and increased activity of the enzyme acetylcholinesterase (AChE) [2, 5–8]. AChE and ChAT modulate the levels of ACh in the central and peripheral cholinergic systems [2]. AChE catalyzes the breakdown of ACh into acetate and choline, while ChAT synthesizes ACh from choline and acetyl-CoA. ACh catabolism mediated by AChE serves two purposes. Firstly, it allows ACh to be continually replenished through the reuptake of choline into the synaptic knob. Secondly, it prevents neuronal hyperexcitability arising from enhanced ACh levels at the neuronal synapse [9]. However, excessive AChE activity may cause cataclysmic degradation of ACh leading to cognitive effects in AD patients such as memory impairment. Drugs such as donepezil, rivastigmine, and galantamine, which reversibly inhibit AChE by forming hydrolyzable carbamylated compounds with it, are widely used for symptomatic alleviation of AD [10–15]. Unfortunately, the cognitive benefit that they confer lasts only for ~ 2 years. Their actions on the peripheral cholinergic system produce side effects such as gastrointestinal disturbances, convulsions, nausea, vomiting, bradycardia, and muscle weakness, further limiting their efficacy [16]. Despite these disadvantages, a small section of the people treated with these drugs experience cognitive improvement [17], and a vast majority of people experience a delay of cognitive decline by 6–9 months [1].

Cholinesterase inhibitors increase the synaptic residence time of ACh and enhance postsynaptic cholinergic signaling [2, 18]. The exact mechanism by which this enhanced signaling translates into improved cognitive and behavioral effects remains undiscovered. Characterization of this mechanism may help us discover drugs that modulate the dementia component of AD more effectively. Detailed studies that dissect the nature of AChE inhibition and describe the influence of ACh receptor binding, ChAT activity, and enzyme localization on AChE activity are required for this characterization. Monitoring behavioral and locomotor responses to drug compounds in an appropriate model system, and studying the docking of active components of these drug compounds on AChE will provide insights into the nature of AChE inhibition.

The structure of human AChE has been characterized using chemical and kinetic studies. The active site of the enzyme contains two subsites. The breakdown of ACh into acetate and choline is catalyzed within the esteratic subsite [19]. This subsite contains the catalytic triad

of three amino acids, namely, serine (Ser203), histidine (His447), and glutamate (Glu334) [20, 21]. The anionic subsite is a choline-binding pocket and interacts with the charged quaternary amine of ACh, cationic substrates, and inhibitors [19]. Apart from these, AChE also contains a distinct 'peripheral' anionic site at the active site entrance. This site has been implicated in substrate inhibition and allosteric regulation of ACh hydrolysis at the esteratic subsite [22].

Zebrafish (*Danio rerio*) is primarily used as a genetic model system for studying developmental and disease processes. They have biochemical and behavioral responses comparable to mammalian systems, making them suitable for drug testing. They exhibit comparable brain macro-organization, cellular morphology, neuromediator systems, and sensitivity to several classes of neurotropic drugs [23]. Localization of cholinoceptive (i.e. AChE-immunoreactive) neurons, cholinergic (i.e. ChAT-immunoreactive) neurons and AChE activity are well characterized in zebrafish [24]. AChE is the solitary cholinesterase in zebrafish [25]. AChE expression is initially found in 4 hours post-fertilization (hpf) embryos, and increases by 210-folds in 144 hpf larvae [26]. The 16 hpf embryos exhibit abrupt movements, reflecting spontaneous ACh release at the developing synaptic junctions. AChE is expressed in somites and several bilateral clusters in the presumptive brain at this stage. 21 hpf larvae become sensitive to touch stimulus and exhibit uncoordinated movements less frequently [26]. 27 hpf larvae show coordinated escape movements induced by tactile stimuli [27]. AChE localizes as large clusters in the epiphysis, forebrain, midbrain, hindbrain, and the seven rhombomeres at this stage. The 168 hpf (free-swimming) larvae exhibit the fully mature pattern of AChE activity [28].

Convolvulus pluricaulis (CP) is a perennial herb that has been previously studied for its anti-amnesiac and anxiolytic properties in rodents [29–32]. Aqueous CP extract has shown significant AChE inhibition in the cortex and hippocampus of male Wistar rats with scopolamine-induced cognitive impairment [30]. Two active components of CP, namely, scopoletin and scopolin, have significantly reduced scopolamine-induced amnesia in a dose-dependent manner in mice [31]. CP, in combination with rivastigmine, has inhibited aluminium-induced elevation of AChE activity in rats [32].

In this study, we studied the inhibitory mechanism of CP in zebrafish with scopolamine-induced cognitive impairment using biochemical assays, behavioral tests, and bioinformatics methods. CP showed higher avoidance response retention than isoxazole (positive control for AChE inhibition) in adult zebrafish. It exhibited inhibitory activity in the same regions as that of isoxazole in

168 hpf larvae and a concentration-dependent increase in this activity when used in combination with scopolamine. Two constituents of CP (scopoletin and kaempferol) were bound by active as well as allosteric sites of human AChE. Overall, our study proposes further investigations of CP as a modulator of cognitive brain function.

2 Results and discussion

2.1 Optimization of treatment concentrations

The binomial response (death/no death) of 24 hpf embryos to *Convolvulus pluricaulis* (CP) was recorded over 48 h, and the concentration-probits curve was plotted to determine lethal concentration (LC_{50}) [33]. For

CP, LC_{50} was determined to be 0.4708 ± 0.089 mg/mL (Fig. 1). Scopolamine (SCOP), an anti-cholinergic ligand that prevents the binding of ACh to its receptor [34], was used to induce cognitive impairment in zebrafish; this is a well-established pharmacological model of cognitive impairment. SCOP has been shown to impair both retention of learned response and acquisition of passive avoidance response in zebrafish; these cognitive deficits were rescued by the AChE inhibitor physostigmine [35]. Isoxazole (ISOX) was used as the positive control for AChE inhibition [36]. Several ISOX derivatives have exhibited inhibitory activity in vitro against AChE isolated from electric eel, rat brain, and human serum [37]. Molecular docking studies with AChE extracted from electric eel [38] and their ability to rescue scopolamine-induced amnesia in mice [39] further ascertained the utility of these compounds as AChE inhibitors. Concentrations of SCOP and ISOX were determined for 25 larvae (i.e.

Fig. 1 Lethal concentration (LC_{50}) of *Convolvulus pluricaulis*. The graph depicts the relationship between probit mortality and the concentration of *Convolvulus pluricaulis* (CP) (depicted as solid circles). The binomial response (death/no death) of 24 hpf embryos exposed to the plant extract was recorded over 48 h. The solid line represents the third-order polynomial equation that modeled the responses, i.e. $y = 8.54x^3 + 1.61x^2 + 0.55x + 5.31$, where y is 5.00 probits. LC_{50} of 0.4708 ± 0.089 mg/ml was obtained by calculating the inverse log value of x, and standard error as $(\log LC_{84} - \log LC_{16})/\sqrt{2N}$

40 mg of tissue), based on an estimated 200 µM of SCOP and 31.2 mM ISOX for fishes weighing ~1.2 g [35, 40,41]. The least toxic concentrations of the compounds were chosen for the treatment of 25 larvae, namely 6.68 µM for SCOP, 1.04 mM of ISOX, and 0.38 mg/mL for CP. The larvae were treated with the AChE inhibitor one hour before SCOP treatment.

2.2 The activity of *Convolvulus pluricaulis* in zebrafish larvae and adults

The effect of CP on AChE activity and ACh levels in 168 hpf zebrafish larvae treated with SCOP was studied using Ellman's assay and the hydroxylamine method (Fig. 2). We also employed the Karnovsky staining method for qualitative analysis of AChE activity in 168 hpf larvae [42]. The ACh level in untreated control

larvae was found to be 87.86 ± 1.61 µM. Karnovsky staining revealed a mature pattern of AChE activity in these larvae (Fig. 3A). On treatment with the AChE inhibitor, ISOX, the level only slightly increased to 88.33 ± 3.12 µM, and the inhibitory activity was found to be $13.48\% \pm 1.92$ (p-value $= 1.97E-02$). Clearance of the Karnovsky stain was interpreted as AChE inhibition. Visual inspection revealed that ISOX showed AChE inhibition in the myelencephalon and the somites containing sensory interneurons and motor neurons (Fig. 3C). With CP, the inhibitory level was $9.76\% \pm 2.94$ and the ACh level was at 79.83 ± 13.44 µM, which was not significantly different from the levels in untreated larvae. However, CP also showed stain clearance in the myelencephalon and the somites (Fig. 3D). Overall, the pattern of clearance shown by ISOX and CP in these regions seemed to be similar (Fig. 3C, D), indicating that CP may exert an inhibitory effect comparable to that of the positive control.

Using the wrMTrck software optimized for zebrafish, we inspected the locomotor patterns of the larvae treated with the different test compounds (Fig. 4) [43, 44]. Visual inspection revealed that the movement patterns of larvae treated with individual compounds such as ISOX, SCOP, and CP were distinguishable from the paths of untreated control larvae (Fig. 4A–C, E). The overlapping paths of the treated larvae compared to the non-overlapping paths of control larvae indicated inadequate sensory reception and motor control in the treated larvae. We had found the inhibitory activity of CP to be localized to the mye-

lencephalon and the somites containing small sensory interneurons and large motor neurons (Fig. 3). Myelencephalon regulates the anti-predatory escape response in zebrafish larvae, the stimuli for which are conveyed by the hair cells of the lateral line system innervated by large motor neurons [45]. This C-start startle escape response regulated by Mauthner cells in the zebrafish hindbrain is modulated by a form of non-associative memory called
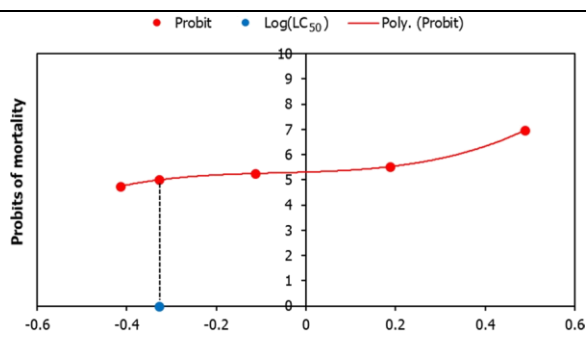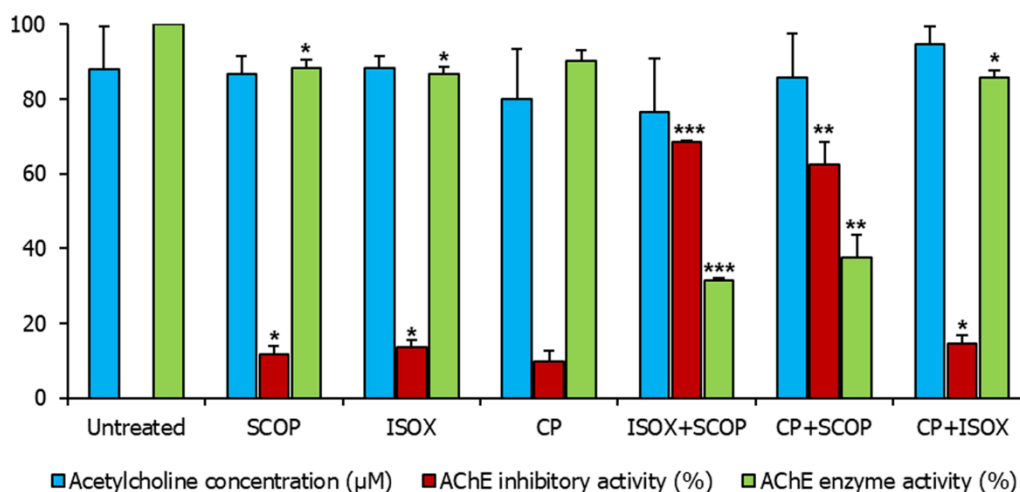
Fig. 2 Effect of test compound(s) on acetylcholinesterase activity and acetylcholine levels. Acetylcholine (ACh) levels (blue bars), acetylcholinesterase (AChE) inhibitory activity (red bars), and acetylcholinesterase enzyme activity (green bars) in 168 hpf larvae unexposed to any test compound (i.e. 'untreated'), treated with scopolamine (SCOP), isoxazole (ISOX), *Convolvulus pluricaulis* (CP) and their combinations (ISOX + SCOP, CP + SCOP, and CP + ISOX) have been shown. Both the ACh level (μM) and AChE activity (%) have been expressed as mean ± standard deviation. Three independent experiments were conducted to determine ACh level and AChE activity, and in each of these experiments, three sets of 15-25 embryos were treated with the least toxic concentrations of SCOP (6.68 μM), ISOX (1.04 mM), and CP (0.38 mg/ml) and their combinations. The larvae were treated with the AChE inhibitor (i.e. ISOX or CP) one hour before treatment with SCOP. *, **, and *** indicate p-value < 0.5, < 0.01 and < 0.001 of two-tailed unpaired t-test for untreated versus test groups respectively
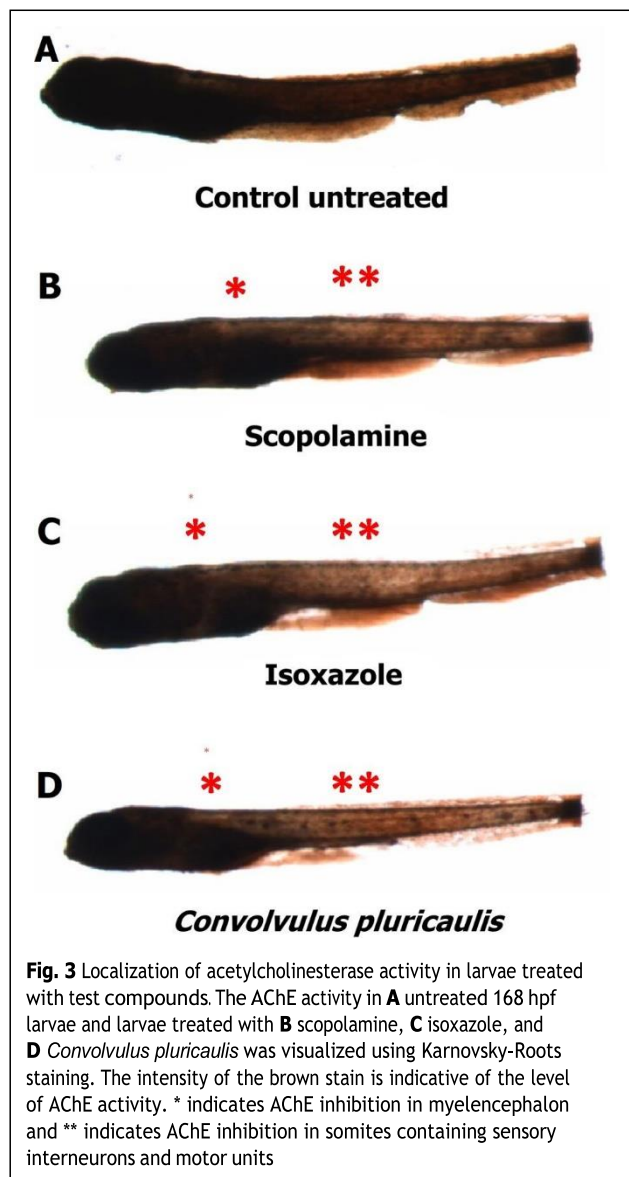
habituation [46]. As expected, we found abnormal body bends and deregulated coordination of motor responses in CP-treated larvae (Fig. 4E). This also shows that CP can enter and act on the lateral line system of zebrafish larvae, producing abnormal locomotor responses, and thereby limiting its value as a therapeutic candidate for cognitive impairment in AD. However, further studies are required to examine these motor responses in detail.

A passive avoidance response test was used to test the acquisition of avoidance response in adult zebrafish. As expected, SCOP-treated adults failed to acquire the response (p-value = 8.2E−03) (Fig. 5A). ISOX-treated fishes showed greater acquisition of response com-

pared with untreated fishes during the training session (p-value = 3.55E−02) but did not retain the memory in the test session (Fig. 5B). Although CP-treated fishes showed lower acquisition of response than con- trol during the training session, they showed higher retention of memory than isoxazole in the test session (p-value < 1E−04), indicating that CP may be superior to the positive control in memory retention (Fig. 5B).

We studied the binding of the two active components of CP—scopoletin, and kaempferol—on human AChE using molecular docking (SwissDock [47], see Sect. 4). The crystal structure of human AChE was extracted from PDB ID: 4PQE (Fig. 6 and Table 1). In the absence of its substrate ACh, AChE was bound by the positive control ISOX at its catalytic site—specifically at Ser203

[20, 21]—with a binding energy of − 14.03. When AChE was bound by ACh at its primary binding site Hsd405 in the choline-binding pocket, ISOX was found at Ser203 with a lower binding energy of − 14.11. Scopoletin (an active component of CP) was found to bind to Glu313 (via tropane, the central structure of scopoletin), the sitebound by SCOP (Fig. 6B, C and Table 1). Scopoletin also establishes contact with His447 of the catalytic triad [20,21] and Glu202 in the peripheral anionic site (Table 1). Kaempferol was found to bind to Glu202 and Tyr72 in the peripheral anionic site (Fig. 6B, D and Table 1).

The binding mode of scopoletin on AChE—specifically, its ability to occupy the ACh-binding anionic sub- site (in the catalytic center)—suggests that it may act as a competitive inhibitor of ACh. This observation is sup- ported by previous studies that demonstrated its AChE inhibitory activity in vitro [31, 48] and ability to increase extracellular ACh concentration in rat brains to a level comparable to that of galantamine [48], a compound often used as a positive control for AChE inhibition. However, scopoletin may enhance ACh levels via mechanisms other than AChE inhibition, such as agonistic activity on nicotinic ACh receptors, which increases ACh release from synaptosomes [49]. Although our results indicate that competitive inhibition of AChE by scopoletin may increase ACh levels, further investigations may be necessary to elucidate the exact mechanisms. Besides, the disease-modifying effects of scopoletin possibly

Fig. 3 Localization of acetylcholinesterase activity in larvae treated with test compounds. The AChE activity in **A** untreated 168 hpf larvae and larvae treated with **B** scopolamine, **C** isoxazole, and **D** *Convolvulus pluricaulis* was visualized using Karnovsky-Roots staining. The intensity of the brown stain is indicative of the level of AChE activity. * indicates AChE inhibition in myelencephalon and ** indicates AChE inhibition in somites containing sensory interneurons and motor units

Fig. 4 Swimming patterns of larvae and adults in response to test compounds. The locomotion tracking plots (or 'traces') extracted from wrMTrck for **A** untreated, **B** scopolamine (SCOP)-treated, **C** isoxazole (ISOX)-treated, **D** ISOX + SCOP-treated, **E** *Convolvulus pluricaulis* (CP)-treated, and **F** CP + SCOP-treated larvae. The 168 hpf larvae were treated with the least toxic concentrations of SCOP (6.68 μM), ISOX (1.04 mM), and CP (0.38 mg/mL) and their combinations. Note that each larva is represented by a single trace. Three independent experiments were conducted with each of the treatment conditions to determine the locomotion patterns of the larvae. Three minutes long videos (2 min for acclimatization and 1 min for test response) of 3 larvae (per Petri dish placed on a light source) were shot in a dark chamber
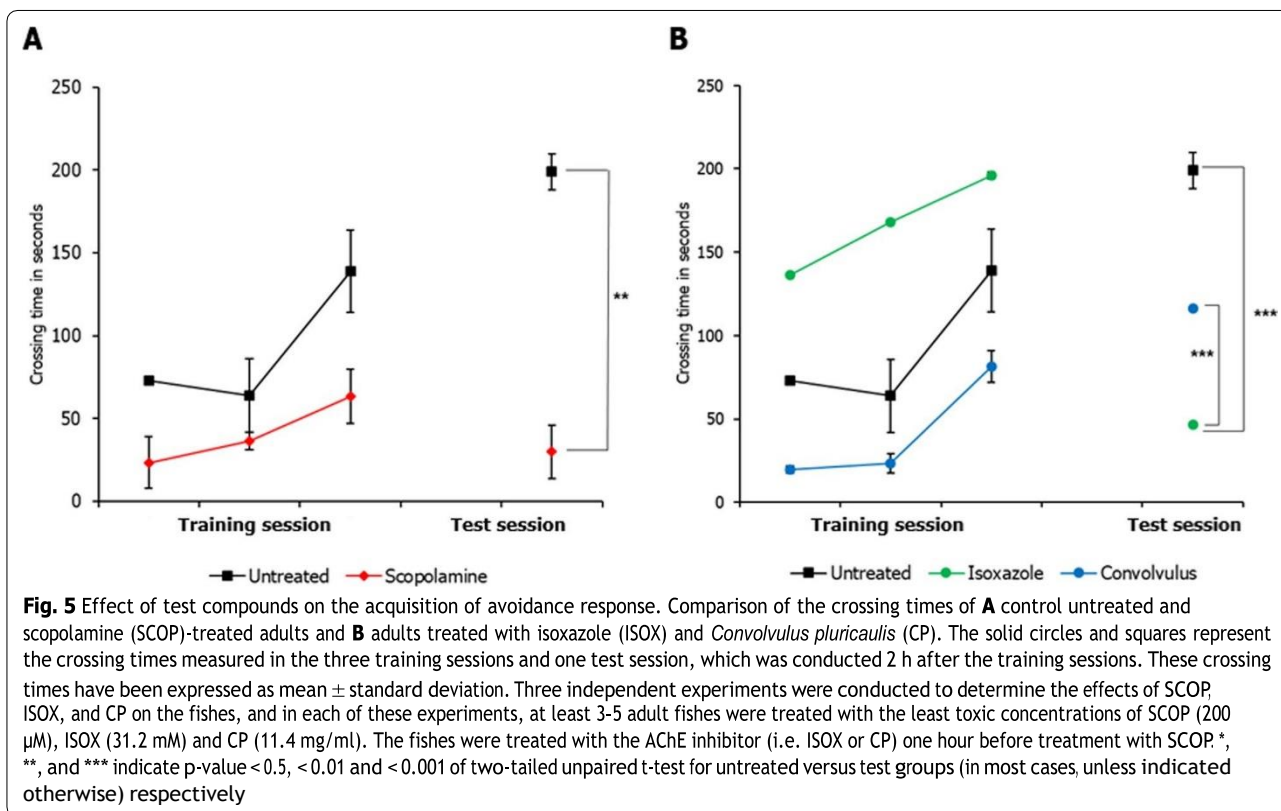
mediated by its additional interactions with the peripheral anionic and esteratic sites of the enzyme also need to be examined further. Unlike scopoletin, kaempferol (a flavonoid) may allosterically modulate the conformation of the catalytic triad or block ACh entry into the active site of the enzyme by binding to the peripheral anionic site as a non-competitive inhibitor, a feature exhibited by flavonoids [50]. Kaempferol has been shown to strongly inhibit AChE in a previous study [51]. Overall, in contrast with the positive control ISOX that only acts on the catalytic triad, the active components of CP may bind to the choline-binding pocket and the peripheral anionic site and mediate both competitive and non-competitive modes of inhibition.
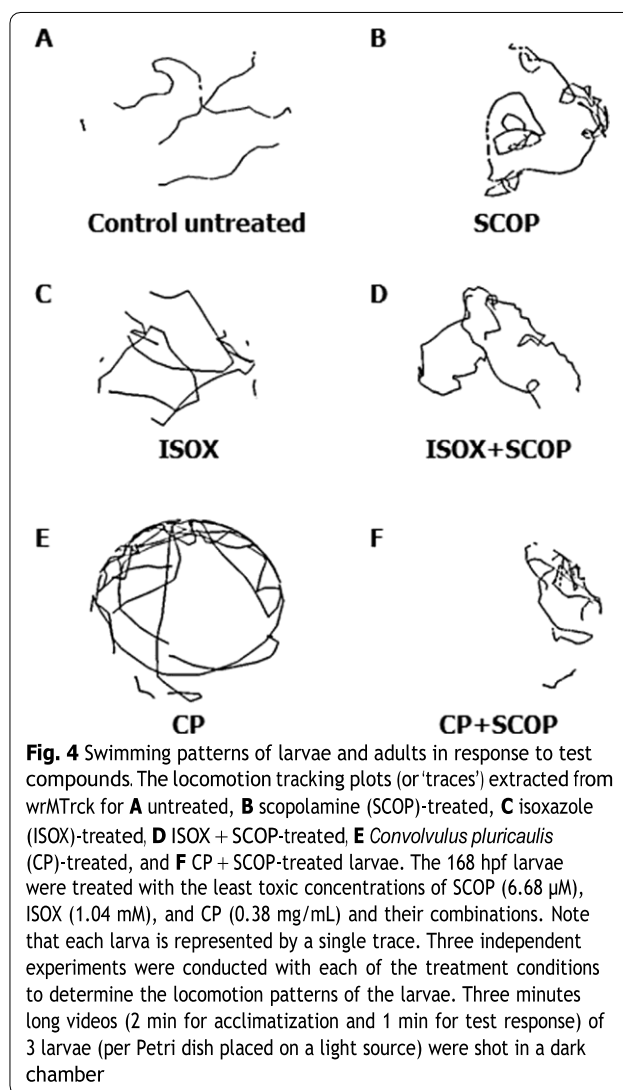
It must be noted that our docking analysis was limited to only two phytoconstituents of CP (scopoletin and kaempferol) that have shown anti-AChE activity in separate studies. A more comprehensive analysis including alkaloid, anthocyanin, triterpenoid and phytosterol com-ponents of CP [52] should be conducted to fully characterize the AChE inhibitory activity of CP.

2.3 The inhibitory activity of scopolamine in zebrafish larvae

An inhibitory activity of 11.68% ± 2.28 was noted in SCOP-treated larvae compared to untreated larvae (p-value = 3.61E−02); the ACh level was 86.54 ± 4.90 μM (Fig. 2). This was supported by qualitative analysis for AChE activity as well (Fig. 3B). To the best of our knowl-edge, our study is the first to report the AChE inhibitory

Fig. 5 Effect of test compounds on the acquisition of avoidance response. Comparison of the crossing times of **A** control untreated and scopolamine (SCOP)-treated adults and **B** adults treated with isoxazole (ISOX) and *Convolvulus pluricaulis* (CP). The solid circles and squares represent the crossing times measured in the three training sessions and one test session, which was conducted 2 h after the training sessions. These crossing times have been expressed as mean ± standard deviation. Three independent experiments were conducted to determine the effects of SCOP, ISOX, and CP on the fishes, and in each of these experiments, at least 3-5 adult fishes were treated with the least toxic concentrations of SCOP (200 µM), ISOX (31.2 mM) and CP (11.4 mg/ml). The fishes were treated with the AChE inhibitor (i.e. ISOX or CP) one hour before treatment with SCOP. *, **, and *** indicate p-value < 0.5, < 0.01 and < 0.001 of two-tailed unpaired t-test for untreated versus test groups (in most cases, unless indicated otherwise) respectively
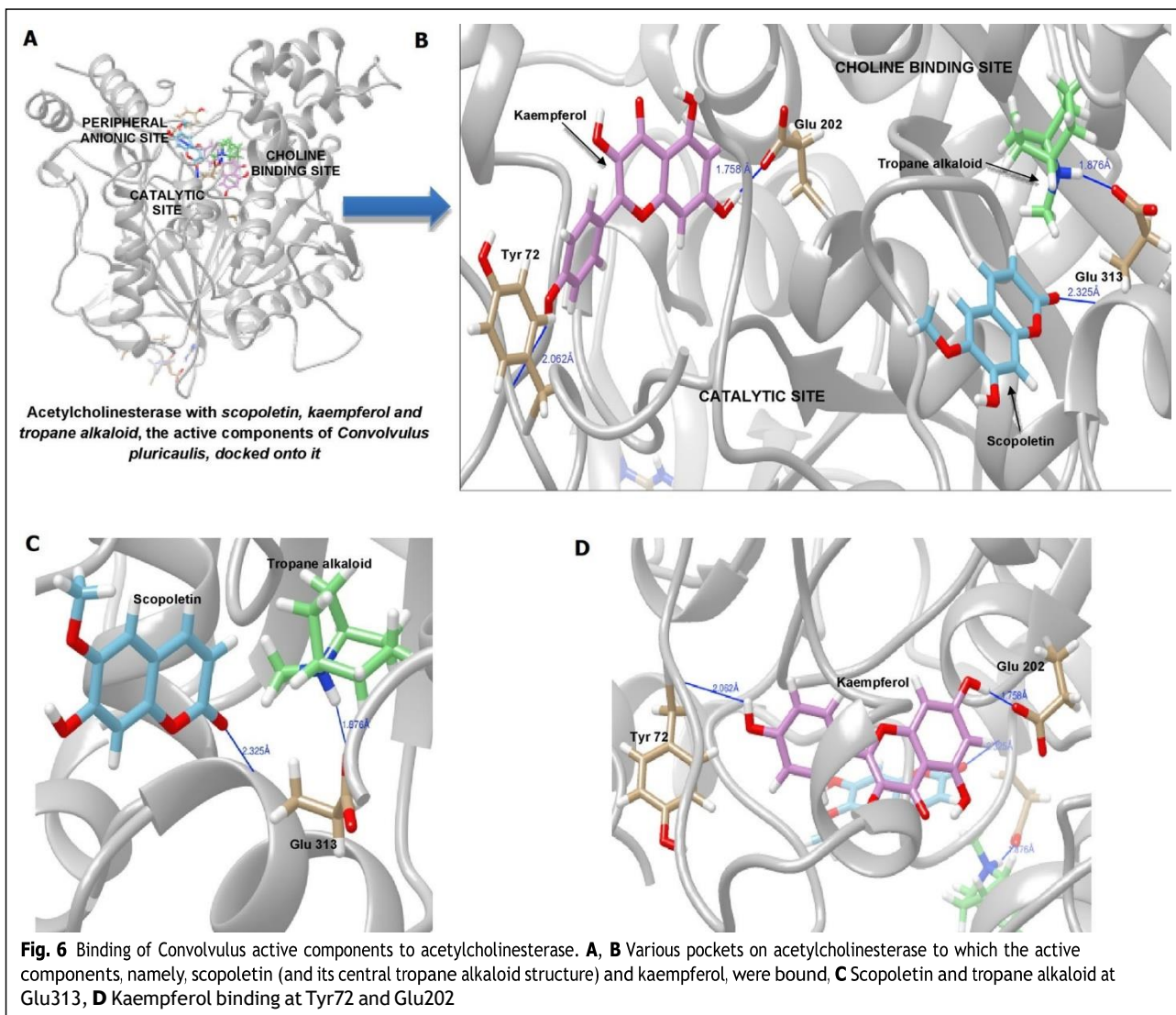
activity of SCOP in zebrafish. However, several studies in the past have demonstrated the AChE inhibitory activity of SCOP or its analogs in other animal models. Cholinergic ligands such as atropine, hyoscyamine, and gallamine have been known to show inhibitory activity on AChE [53, 54]. For example, atropine—which is structurally similar to SCOP—inhibits AChE in the presence of low concentrations of acetylthiocholine iodide ($K_i = 4 \times 10^{-3}$) [53]. It was shown in guinea pigs that SCOP administered at a concentration of 1.94 µg/h for 6 days inhibited red blood cell AChE by 18.7% ± 3.7 and plasma cholinesterase by 44.1% ± 3.1 [54]. SCOP showed competitive AChE inhibition at a concentration of 0.25×10^{-2} M and mixed type AChE inhibition at a concentration of 0.5×10^{-2} M in synaptosomal fractions isolated from rat brain [55].

Using molecular docking, we found that, in the absence of ACh, AChE was bound by SCOP at Gly234 and Thr238 in the peripheral anionic site with a binding energy of 223.762, the former residue being the one that mediates the inhibitory effect of galantamine, an AChE inhibitor [56]. In the presence of a single molecule of ACh at Hsd405, SCOP established hydrogen bonds with Glu313, the same site with which atropine interacts at a binding energy of − 4.475. Atropine is structurally identical to SCOP, except for a highly reactive epoxide group that the latter bears. These epoxides are highly reactive

(compared to simple ethers) because of their ring strain. Nucleophilic attack of the electrophilic C in the C–O bond causes the ring to open. The energy of a simple epoxide, ethylene oxide, changes from − 8.95 to 181.04, after protonation [57]. In the presence of ACh molecules at the primary (Hsd405) and secondary (Phe295) sites (Fig. 7C, D), the binding energy of SCOP and AChE is 219.043 (Fig. 7B, E). SCOP may be primed to be bound by AChE after protonation of the epoxide group. This hypothesis should be validated through a dynamic simulation of the binding of SCOP to AChE. Nevertheless, our data indicate that the AChE inhibitory activity of SCOP may be mediated via its binding to the two ACh-binding sites, the anionic subsite of the catalytic center and the peripheral anionic site.

2.4 The combinatorial effect of *Convolvulus pluricaulis* and scopolamine

A combination of ISOX and SCOP (ISOX + SCOP) increased the inhibitory activity of AChE to 68.45% ± 0.5 (p-value < 1E−04) compared to either of their inhibitory activities in isolation (Fig. 2). Similarly, a combination of CP and SCOP (CP + SCOP) increased the inhibitory activity to 62.5% ± 6.065 (p-value = 9.3E−03) compared to their inhibitory activities in isolation (Fig. 2). To investigate the combinatorial effect of ISOX + SCOP and CP + SCOP

Fig. 6 Binding of Convolvulus active components to acetylcholinesterase. **A**, **B** Various pockets on acetylcholinesterase to which the active components, namely, scopoletin (and its central tropane alkaloid structure) and kaempferol, were bound, **C** Scopoletin and tropane alkaloid at Glu313, **D** Kaempferol binding at Tyr72 and Glu202

further, we increased the concentration of these compounds in tandem (Fig. 8). The increase in inhibitory activity with increasing concentrations was significant inCP + SCOP (p-value < 1E−02), but not in ISOX + SCOP(Fig. 8). Hence, CP, and not ISOX, seemed to significantly enhance AChE inhibition in a synergistic and concentration-dependent manner with SCOP. Contrary to an expected increase in ACh levels in this scenario, a significant drop in ACh level (p-value < 1E−03) was noted with CP + SCOP (Fig. 8). The drop in ACh levels in the different concentrations of ISOX + SCOP and CP + SCOP was significant compared to untreated control (Fig. 8). SCOP, which is synergistically inhibiting AChE together with CP,may be rendered unavailable for binding with the ACh receptor. In this scenario, the ACh pool enhanced as a result of the inhibitory activity mediated by CP + SCOP

may be rapidly depleted through its binding with the ACh receptor. SCOP has been previously shown to decrease cerebral ACh levels by 31% at a concentration of 0.63 mg/kg [58]. It has also been noted that SCOP is more potent than atropine in reducing ACh levels [58]. Further experi- ments are necessary to test our speculations.

We studied the locomotor patterns of larvae treated with ISOX + SCOP and CP + SCOP. Larvae treated withISOX + SCOP covered increased distances in comparisonwith ISOX-treated larvae (p-value = 8.25E−02) (Fig. 9). Larvae treated with SCOP had executed a whirling motion in an uncoordinated direction with high speed over a short distance (Fig. 4B). This locomotor type, known as corkscrew swimming, is commonly observedas part of a seizure phenotype [59]. This would have arisen from the inhibitory activity of SCOP in sensory

366

Table 1 Details on critical interactions of acetylcholinesterase with isoxazole, scopolamine and active components of *Convolvulus pluricaulis*, namely, scopoletin and kaempferol, and the central structure of scopoletin (tropane alkaloid)

Target	Ligand	Hydrogen bond forming residues	Binding energy
AChE	Acetylcholine	Hsd 405	− 13.5878
AChE + acetylcholine	Acetylcholine	Phe 295	− 12.2174
AChE	Isoxazole	Ser 203	− 14.0329
AChE + scopolamine	Isoxazole	Phe 295	− 12.7298
AChE + scopolamine	Isoxazole	Hsd 405	− 12.9197
AChE + acetylcholine	Isoxazole	Ser 203	− 14.1166
AChE + acetylcholine	Isoxazole	Phe 295	− 12.7298
AChE + acetylcholine + scopolamine	Isoxazole	Ser 203	− 13.9634
AChE	Scopolamine	Gly 234, Thr 238	223.762
AChE + isoxazole	Scopolamine	Lys 53	226.227
AChE + acetylcholine	Scopolamine	Glu 313	223.469
AChE + acetylcholine + acetylcholine	Scopolamine	Glu 313	219.043
AChE + acetylcholine + isoxazole	Scopolamine	Glu 313	222.987
AChE + acetylcholine + acetylcholine	Atropine	Glu 313	− 4.475
AChE	Galantamine	Glu 313	0.951452
AChE	Galantamine	Lys 53, Glu 185	6.7258
AChE + acetylcholine + acetylcholine + scopolamine	Isoxazole	Ser 203	− 14.1095
AChE	Scopoletin	Glu 313	23.1182
AChE	Scopoletin	His 447 (HSP)	17.3092
AChE	Scopoletin	Glu 202	17.3475
AChE	Tropane alkaloids	Glu 313	− 16.981
AChE	Kaempferol	Glu 202, Tyr 72	10.6597

interneurons and motor units, rendering cholinergic transmission uncoordinated at the cholinergic synapse. In ISOX + SCOP and CP + SCOP, this whirling motion characteristic of SCOP added up to the locomotor repertoire of the treated larvae (Fig. 4D, F). Larvae treated with ISOX + SCOP (Fig. 4D) executed a zigzag motion often produced in response to alarming stimuli (chemical cuein our case) [59]. However, ISOX + SCOP-treated adults seemed to exhibit boldness, a behavior often accompanied by an increased approach towards novel objects [59]. They spent more time in the central portion of the tank than the peripheral areas compared to untreatedand ISOX-treated adults (Additional file 1: Fig. S1). This could indicate that distinct behavioral repertoires characterize the different life stages of the zebrafish. Never-

theless, ISOX + SCOP and CP + SCOP disrupt motor response patterns characteristic of specific stages of life.

We studied the binding of ISOX and SCOP on AChE using molecular docking. When AChE was bound by SCOP at Glu313 in the choline-binding pocket, ISOX was bound by the enzyme at the primary and secondaryACh binding sites at a higher energy of − 12.91 to − 12.72.However, in the presence of ACh at Hsd405, SCOP was bound by the enzyme at Glu313 and ISOX at Ser203 with a lower binding energy of − 13.96. An even lower

binding energy of − 14.1 was observed with AChE binding two molecules of ACh at Hsd405 and Phe295 (Fig. 7C,D and Table 1), SCOP at Glu313 and ISOX at Ser203 (Fig. 7B, F and Table 1), indicating that this configurationof ISOX + SCOP at the various subsites in the catalytic center may be responsible for its synergistic activity on the enzyme. Similar studies should also be conducted withCP + SCOP to elucidate their mechanism of inhibition.

3 Conclusions

This study was undertaken to gain new mechanistic understanding into the modes of AChE inhibition of CP, known to have anti-AChE activity. CP-treated 168 hpf larvae showed a similar pattern of AChE inhibition (in the myelencephalon and somites) as that of the larvae treated with the AChE inhibitor ISOX, which was used as a positive control. Additionally, CP improved the retention of avoidance response in adult zebrafish compared with ISOX. ISOX was found to directly bind Ser203 of the cata-lytic triad on the human AChE. The active components of CP—scopoletin and kaempferol—were found to bind to His447 of the catalytic triad, the anionic subsite of the catalytic center, and the peripheral anionic site. Unexpectedly,SCOP, which was used in our study to induce cognitive
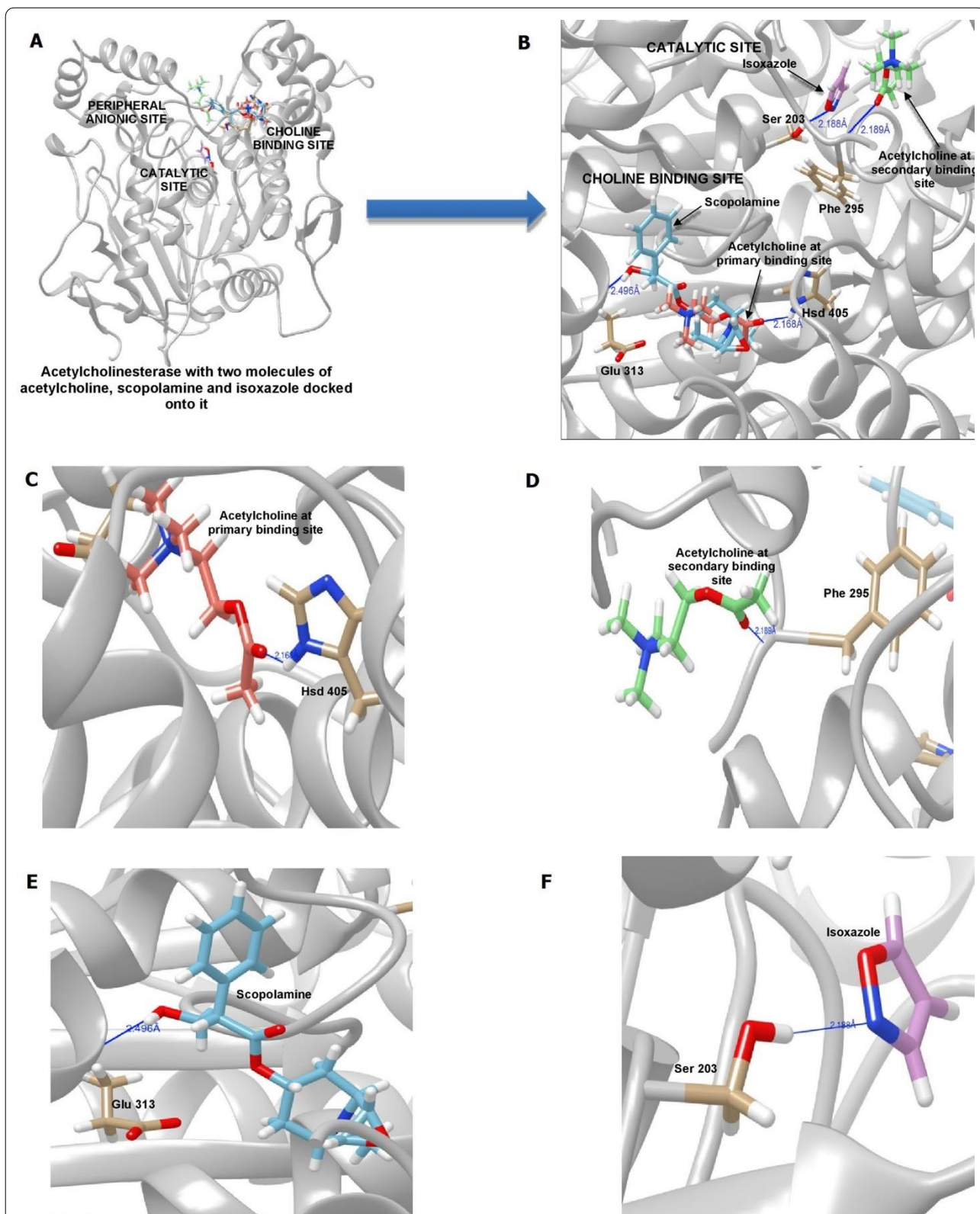
367

Fig. 7 Binding of isoxazole and scopolamine to acetylcholinesterase. **A**, **B** Various pockets on acetylcholinesterase to which its substrate acetylcholine, isoxazole and scopolamine were bound, **C** Acetylcholine at Hsd405, the primary binding site, in the choline-binding pocket, **D** Acetylcholine at Phe295, the secondary binding site, near the catalytic site, **E** Scopolamine at Glu313 in the choline-binding pocket, **F** Isoxazole binding Ser203, a catalytic triad residue
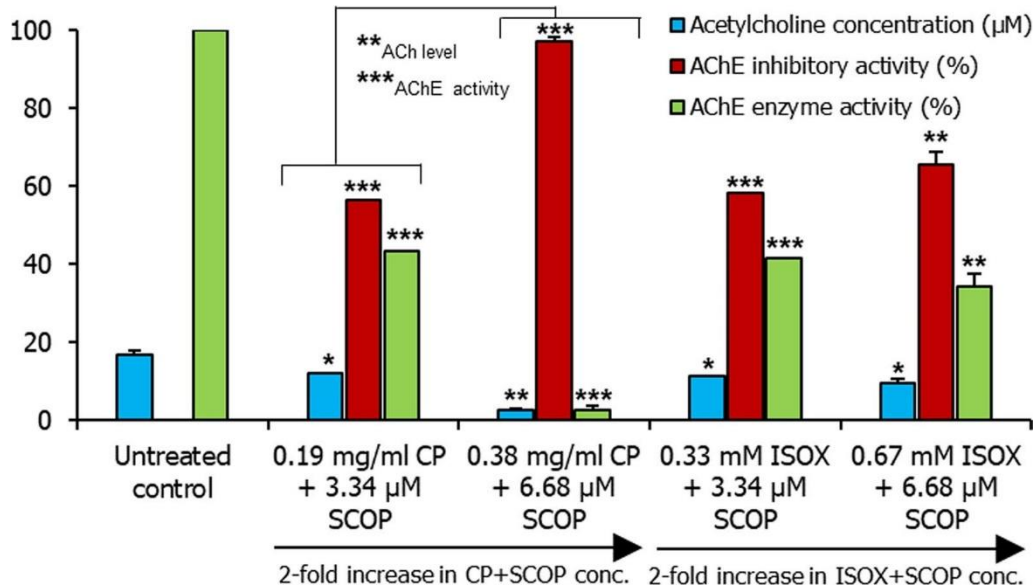
Fig. 8 Concentration-dependent effect of CP + SCOP and ISOX + SCOP on acetylcholinesterase activity and acetylcholine levels. Acetylcholine (ACh) levels (blue bars), acetylcholinesterase (AChE) inhibitory activity (red bars), and acetylcholinesterase enzyme activity (green bars) in 168 hpf larvae unexposed to any test compound (i.e. 'untreated') and treated with the combinations of scopolamine (SCOP) with *Convolvulus pluricaulis* (CP) and isoxazole (ISOX), i.e. CP + SCOP and ISOX + SCOP, in varying concentrations have been shown. Two concentrations were tested for each of these combinations. For CP + SCOP, these were 0.19 mg/ml of CP + 3.34 µM of SCOP and 0.38 mg/mL of CP + 6.68 µM of SCOP. For ISOX + SCOP, these were 0.52 mM of ISOX + 3.34 µM of SCOP and 1.04 mM of ISOX + 6.68 µM of SCOP. Both the ACh level (µM) and AChE activity (%) have been expressed as mean ± standard deviation. Three independent experiments were conducted to determine ACh level and AChE activity, and in each of these experiments, three sets of 15-25 embryos were treated with the said concentrations. The larvae were treated with the AChE inhibitor (i.e. ISOX or CP) 1 h before treatment with SCOP. *, **, and *** indicate p-value < 0.5, < 0.01 and < 0.001 of two-tailed unpaired t-test for untreated versus test groups respectively

impairment in zebrafish, showed AChE inhibition in 168 hpf larvae, possibly mediated via the anionic subsiteof the catalytic center and the peripheral anionic site, as

Fig. 9 Distance covered by the larvae treated by the test compounds. Histograms of the distances covered by untreated, scopolamine (SCOP)-treated, isoxazole (ISOX)-treated, *Convolvulus pluricaulis* (CP)-treated, ISOX +SCOP-treated, and CP +SCOP-treated larvae. Three independent experiments were conducted with each of the treatment conditions. *, **, and *** indicate p-value < 0.5, < 0.01 and < 0.001 of two-tailed unpaired t-test for untreated versus test groups (unless indicated otherwise) respectively

indicated by docking studies with human AChE. Interestingly, CP + SCOP significantly increased AChE inhibition and depleted ACh levels compared with untreated
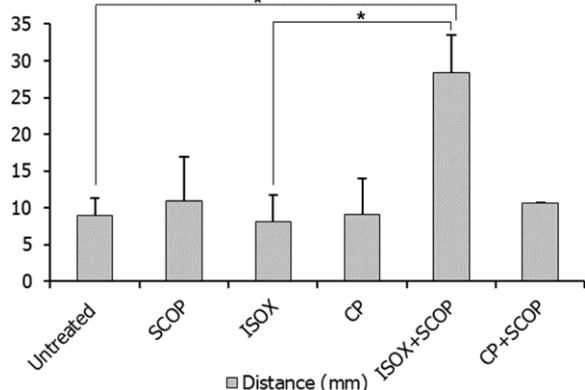
larvae, a pattern that was also observed albeit in a statistically non-significant way in ISOX + SCOP. Abnormalmotor responses were observed individually with ISOX and CP, and in their combinations with SCOP, indicative of undesirable effects on the peripheral cholinergic system. Our study proposes the examination of CP, SCOP, andCP + SCOP as potential AChE inhibitors for their ability tomodulate cognitive deficits in Alzheimer's disease.

4 Materials and methods

4.1 Collection of zebrafish embryos

Adult wild-type zebra fish were maintained in tanks as per standard conditions [60]. Spawning was set up every5 to 6 days in large troughs, usually 2 to 3 h after feeding.

The natural mating ratio of zebrafish is 1 female:2 males. Females and males were housed for spawning either in this ratio or in equal numbers. Fertilization occurs in theearly hours of the morning. Eggs were collected at the4 hours post-fertilization (hpf) stage and transferred to large Petri plates containing the E3 medium (1–5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$,

10–5% Methylene Blue). The developmental stage of the embryos was observed under a microscope [60]. Embryos showing asynchronies in development stages and abnormal development (detectable after 10 hpf) were segregated from the rest. Dead embryos were separated every 4 h. Swimming larvae at 168 hpf were used for this study.
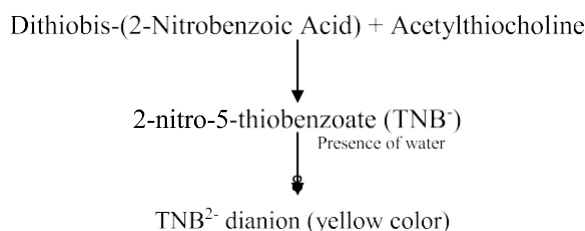
4.2 Preparation of plant extract

An aqueous solution of the root of *Convolvulus pluricaulis* (CP) was prepared by agitating 2 g of macerated root in 100 ml of distilled water at 150 rpm at room temperature for 24 h. The solution was clarified by filtration, frozen overnight at – 20 °C, and lyophilized for 18 h to obtain 18.5 mg of the extract.

4.3 Determination of lethal concentration, 50% (LC$_{50}$) of the plant extract

24 hpf embryos were treated with the plant extract in 96 well plates according to Zebrafish Embryo Toxicity Test (ZFET) Protocol Standards [61]. ZFET allows assessment of the phenotypes manifested by fish embryos on treatment with chemicals. The lethal dose for a plant preparation is the particular concentration at which half of the embryos treated with the preparation are alive. Binomial response (death/no death) of 24 hpf embryos to CP was recorded over 48 h and concentration-probits mortality curve was plotted to determine LC$_{50}$ [33].

4.4 Ellman's assay

Ellman's assay is a spectrophotometric method that quantifies AChE activity in terms of μ moles of acetylthiocholine iodide (ATCh-I) hydrolyzed per minute per mg of the plant preparation [62]. A key player in this assay is the DTNB reagent also known as Ellman's reagent. The mechanism by which it quantifies the amount of substrate hydrolyzed is depicted below:

Dithiobis-(2-Nitrobenzoic Acid) + Acetylthiocholine

$\downarrow$

2-nitro-5-thiobenzoate (TNB$^-$)
Presence of water

$\downarrow$

TNB^{2-} dianion (yellow color)

The reagents for Ellman's assay included 0.1 M phosphate buffer saline (PBS) at pH 7.0, 0.075 M acetylthiocholine iodide (ATCh) as substrate, and 0.01 M DTNB (dithiobis-(2-nitrobenzoic acid)), prepared by dissolving 39.6 mg in 10 ml phosphate buffer (0.1 M) at pH 7.0 and adding 15 mg of sodium bicarbonate. The digestion buffer was prepared by adding 20 mM Tris/HCl at pH 7.0,
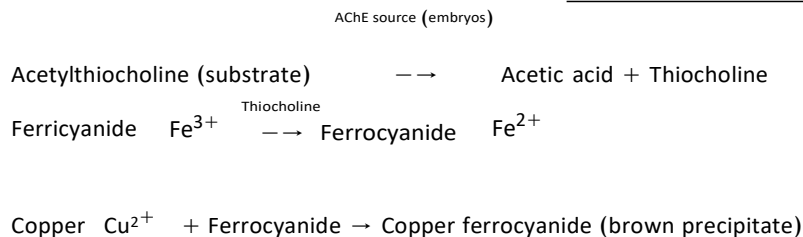
5 mM EDTA and 1% Triton X-100. Embryos were euthanized using a mixture of 1 mL of clove oil and 9 mL of absolute ethanol. 1 mL of this solution was then dissolved in 50 mL of tap water. This solution was then transferred to a Petri dish with 20 embryos to be euthanized. 15–25 embryos were suspended in 150–250 μL digestion buffer (20 mM Tris/HCl pH 7.0, 5 mM EDTA, 1% Triton X-100) and homogenized by pipetting the suspension in and out. The homogenate was centrifuged at 1500 rpm for 15 min. The supernatant diluted in 0.1 M phosphate buffer at pH 7.0 was used as the enzyme (AChE) source. A blank reaction mixture was prepared with the phosphate buffer, substrate, plant extract solutions at specific concentrations, and DTNB reagent. Test reaction mixtures were prepared with the phosphate buffer, substrate, enzyme source, plant extract solutions at specific concentrations, and DTNB reagent. Absorbance was measured at 412 nm.

4.5 Hydroxylamine method for acetylcholine estimation

Hydroxylamine reacts in a strongly alkaline medium with the substrate (ACh) forming acetohydroxamic acid and methanol. Acidification of this mixture with HCl and the addition of Fe^{3+} ions result in a red-brown complex, ACh-acetohydroxamic product, which can be detected through colorimetry [63]. The solutions used for this assay included 2 M aqueous hydroxylamine hydrochloride, 3.5 M aqueous potassium hydroxide, conc. HCl/H$_2$O (in 1:2 ratio), 0.37 M Fe^{3+} (as ferric nitrate or ferric chloride) in aqueous 0.1 M HCl and a standard aqueous solution of 4 mM. Embryos were euthanized and washed. 15–25 embryos were suspended in 150–250 μL digestion buffer and homogenized by pipetting the suspension in and out. The homogenate was centrifuged at 1500 rpm for 15 min. The supernatant diluted in 0.1 M phosphate buffer at pH 7.0 served as the source for the enzyme (AChE) and the substrate (ACh). The blank reaction mixture was prepared with the phosphate buffer and the enzyme and substrate source. Test reaction mixtures were pre- pared with the phosphate buffer, the enzyme and substrate source, and the plant extract solution at specific concentrations. The reaction mixture was vigorously mixed with aqueous hydroxylamine hydrochloride and aqueous potassium hydroxide in the ratio of 1:1. The rapid change in pH stops hydrolysis. The resulting mixture was then mixed for 2 min to allow conversion of ACh to acetohydroxamic acid. The pH was then changed to 1.2 by adding conc. HCl/H$_2$O and aqueous ferric nitrate or ferric chloride. Absorbance was measured at 540 nm. The concentration of ACh was calculated using its molar absorption coefficient (ε(Ach, 540 nm, 25 °C) = 785 M^{-1} cm^{-1}).

4.6 Karnovsky's direct coloring thiocholine method for cholinesterases

Karnovsky's staining method was used for the visualization of localized AChE activity [42]. The mechanism by which this staining method generates a color reaction in response to AChE activity is depicted below:

$$\text{Acetylthiocholine (substrate)} \xrightarrow{\text{AChE source (embryos)}} \text{Acetic acid + Thiocholine}$$

$$\text{Ferricyanide} \quad Fe^{3+} \xrightarrow{\text{Thiocholine}} \text{Ferrocyanide} \quad Fe^{2+}$$

$$\text{Copper} \quad Cu^{2+} + \text{Ferrocyanide} \rightarrow \text{Copper ferrocyanide (brown precipitate)}$$

The working solution of the stain normally contains 60 mM sodium acetate, 5 mM sodium citrate, 4.7 mM copper (III) sulfate, 0.5 mM potassium ferricyanide, 1.7 mM acetylthiocholine iodide, and distilled water. However, since this composition failed to stain the embryos even after repeated trials and the use of freshly prepared solutions, we had to optimize the staining solution. Copper ions inhibit acetylcholinesterase and cause neurotoxicity in zebrafish [64]. Based on this, we hypothesized that decreasing the concentration of copper sulfate in the staining solution may remove or alleviate any inhibitory effect that it may have on the enzyme. We stained the embryos with staining solutions containing 4.7 mM (the original concentration in the Karnovsky method), 4.5 mM, and 4.3 mM of copper sulfate. No staining was observed in embryos treated with the solution contain-ing 4.7 mM copper sulfate. Faint staining was observed with 4.5 mM copper sulfate. The expected intensity and pattern of staining were observed with 4.3 mM of cop-per sulfate. 6.8 g of sodium acetate, 12.905 g of sodium citrate, and 12.4845 g of cupric sulfate, each dissolved in 100 mL distilled water, 8.231 g of potassium ferricyanide in 50 ml distilled water, and 0.72295 g of acetylthiocholine iodide in 5 mL distilled water were used to prepare 0.5 M stock solutions for Karnovsky's staining. 1 mL of paraformaldehyde (PFA) was used to fix 10 embryos in a vial. Before staining the embryos, PFA was removed from the vials. The embryos were washed with $1\times$ PBS solution thrice for 5 min each. The embryos were then incu-bated in the stain. After 3 h, the stained embryos were washed thrice with $1\times$ PBS-Tween (prepared by dissolving 140 µl of Tween 20 in 14 mL $1\times$ PBS) and inspected under the microscope.

4.7 Assessment of larval locomotor activity

The settings of wrMTrck, a freely available ImageJ plugin originally developed for examining multiple behavioral parameters in the nematode *Caenorhabditis elegans*, have been optimized in our lab for zebrafish larvae [43].

This software was used to study the locomotor patterns of treated larvae and calculate their average speed, distance, and body bend. Videos of 3 larvae swimming in a Petri dish were shot in a dark chamber by placing them on a light source. The videos were 3 min long; 2 min was allowed for acclimatization and 1 min for test response.

4.8 Passive avoidance response test

Adult zebrafish treated with a test compound was trans-ferred to an experimental chamber. This chamber con-sisted of a dark compartment and a lit compartment separated by a movable door. The fish was placed in the dark chamber and allowed to acclimatize for 3 min. After this, the door was opened. A stone was dropped in front of the fish 3 s after it crosses the door. The stone served as the shock stimulus in this scenario. Crossing time estimated as the time taken by the fish to cross the door from the moment the door was opened was recorded after every such 'trial'. Three such trials made up 'one training session' for the acquisition of the avoidance response. Two hours after a training session, a 'test session' consisting of a single trial was conducted to assess the extent of learned avoidance response retention. The fish was exposed to the test compound in the experimental chamber during both sessions [35].

4.9 Protein-small molecule rigid body docking

The crystallographic structure of human AChE was collected from the PDB database (PDB ID: 4PQE) [65]. 4PQE has been widely used as a model for human AChE in molecular docking studies [66–69]. Docking on human AChE (PDB ID: 4PQE) was performed using SwissDock [47]. Before docking, this structure was prepared—by repairing incomplete residues, deleting water molecules, adding hydrogen atoms, and assigning partial charges—using the 'Dock Prep' function in UCSF Chimera [70]. The 3D structures of the ligands were collected from the ZINC database [71], namely, acetylcholine (ZINC3079336), isoxazole (ZINC1420779), scopolamine (ZINC100196329), atropine (ZINC100009278), galantamine (ZINC491073), scopoletin/buxuletin (ZINC57733), and kaempferol (ZINC3869768). SwissDock generates binding poses of the ligands in the vicinity of target cavities and computes the summation of various types of energy. The docked poses of the ligands were visualized and

curated using the 'View Dock' function in Chimera. Docked poses with minimal binding energy were chosen for further examination. Note that our AChE model (4PQE) neither had modified residues nor any ligands associated with it. We selected such a structure in order to perform an exploratory analysis with our ligands of interest, and detect different binding poses in all possible binding pockets within AChE. Despite employing the 'blind docking' approach, we were able to replicate key binding interactions, for example, (i) the interaction of ACh with Phe 295, which determines substrate specificity of the acyl pocket [72], and (ii) the binding of atropine and scopolamine—which are structurally identical— to the same residue (Glu 313). All the inter- actions reported in Figs. 6B–D, 7B–F and Table 1 fall within the range of hydrogen bonding limit (< 2.5 Å) [73].

Abbreviations
ACh: Acetylcholine; AChE: Acetylcholinesterase; AD: Alzheimer's disease; ATCh-I: Acetylthiocholine iodide; ChAT: Choline acetyltransferase; CP: *Convolvulus pluricaulis*; DTNB: 5,5'-Dithiobis-(2-nitrobenzoic acid); hpf: Hours post-fertilization; ISOX: Isoxazole; PBS: Phosphate buffer saline; PFA: Paraformaldehyde; SCOP: Scopolamine; ZFET: Zebrafish Embryo Toxicity Test.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s13659-022-00332-5.

Additional file 1: Figure S1. Time spent by adult fishes in each of the three areas of the tank. Adult zebrafish treated with test compound(s) were transferred to an experimental chamber divided equally into three areas and the extent of lateralization in swimming was measured as time spent in each of the three areas (shown as histograms). Areas 1 and 3 are peripheral, whereas area 2 is central. The time spent in the three areas was determined by analyzing 3 min long videos. 2 min was allowed for acclimatization and 1 min for test response. The fish was either untreated or treated with scopolamine (SCOP), isoxazole (ISOX), and ISOX $+$ SCOP. The fish was exposed to 200 μM of SCOP and/or 31.2 mM of ISOX.

Author details
[1]Supercomputer Education and Research Centre, Indian Institute of Science, Bengaluru, India. [2]Department of Genetic Engineering, SRM Institute of Science and Technology, Kattankulathur 603 203, India. [3]Zebrafish Genetics Laboratory, Department of Genetic Engineering, SRM Institute of Science and Technology, Kattankulathur 603 203, India.

References
1. Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. Nat Rev Dis Primers. 2015;1:15056.
2. Ferreira-Vieira HT, Guimaraes MI, Silva RF, Ribeiro MF. Alzheimer's disease: targeting the cholinergic system. Curr Neuropharmacol. 2016;14(1):101–15.
3. Ballinger EC, Ananth M, Talmage DA, Role LW. Basal forebrain choliner-gic circuits and signaling in cognition and cognitive decline. Neuron. 2016;91(6):1199–218.
4. McGleenon B, Dynan K, Passmore A. Acetylcholinesterase inhibitors in Alzheimer's disease. Br J Clin Pharmacol. 1999;48(4):471.
5. Paul S, Jeon WK, Bizon JL, Han J-S. Interaction of basal forebrain cholin-ergic neurons with the glucocorticoid system in stress regulation and cognitive impairment. Front Aging Neurosci. 2015;7:43.
6. García-Ayllón M-S, Riba-Llena I, Serra-Basante C, Alom J, Boopathy R, Sáez-Valero J. Altered levels of acetylcholinesterase in Alzheimer plasma. PLoS ONE. 2010;5(1):e8701.
7. Perry R, Wilson I, Bober M, Atack J, Blessed G, Tomlinson B, et al. Plasma and erythrocyte acetylcholinesterase in senile dementia of Alzheimer type. Lancet. 1982;319(8264):174–5.
8. Atack J, Perry E, Perry R, Wilson I, Bober M, Blessed G, et al. Blood acetyl- and butyrylcholinesterases in senile dementia of Alzheimer type. J NeurolSci. 1985;70(1):1–12.
9. Soreq H, Seidman S. Acetylcholinesterase—new roles for an old actor. Nat Rev Neurosci. 2001;2(4):294–302.
10. Moss DE, Perez RG, Kobayashi H. Cholinesterase inhibitor therapy in Alzheimer's disease: the limits and tolerability of irreversible CNS-selective acetylcholinesterase inhibition in primates. J Alzheimers Dis. 2017;55(3):1285-94.
11. Greig NH, Utsuki T, Yu Q-S, Zhu X, Holloway HW, Perry T, et al. A new therapeutic target in Alzheimer's disease treatment: attention to butyryl-cholinesterase. Curr Med Res Opin. 2001;17(3):159–65.
12. Ballard C. Advances in the treatment of Alzheimer's disease: benefits of dual cholinesterase inhibition. Eur Neurol. 2002;47(1):64–70.
13. Poirier J. Evidence that the clinical effects of cholinesterase inhibitors are related to potency and targeting of action. Int J Clin Pract Suppl. 2002;127:6–19.
14. Fukuto TR. Mechanism of action of organophosphorus and carbamate insecticides. Environ Health Perspect. 1990;87:245.
15. Morisset S, Traiffort E, Schwartz J-C. Inhibition of histamine versus acetyl-choline metabolism as a mechanism of tacrine activity. Eur J Pharmacol. 1996;315(1):R1–2.
16. Colovic MB, Krstic DZ, Lazarevic-Pasti TD, Bondzic AM, Vasic VM. Acetyl-cholinesterase inhibitors: pharmacology and toxicology. Curr Neurophar- macol. 2013;11(3):315–35.
17. Di Santo SG, Prinelli F, Adorni F, Caltagirone C, Musicco M. A meta-analysis of the efficacy of donepezil, rivastigmine, galantamine, and

memantine in relation to severity of Alzheimer's disease. J Alzheimers Dis. 2013;35(2):349-61.

18. Singh R, Sadiq NM. Cholinesterase inhibitors. StatPearls: StatPearls Publishing; 2019.

19. Dvir H, Silman I, Harel M, Rosenberry TL, Sussman JL. Acetylcholinesterase:from 3D structure to function. Chem Biol Interact. 2010;187(1–3):10–22.

20. Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, et al. Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. Science. 1991;253(5022):872–9.

21. Ordentlich A, Barak D, Kronman C, Ariel N, Segall Y, Velan B, et al. Func- tional characteristics of the oxyanion hole in human acetylcholinesterase.J Biol Chem. 1998;273(31):19509–17.

22. Bourne Y, Taylor P, Radić Z, Marchot P. Structural insights into ligand interactions at the acetylcholinesterase peripheral anionic site. EMBO J. 2003;22(1):1–12.

23. Kalueff AV, Stewart AM, Gerlai R. Zebrafish as an emerging model for studying complex brain disorders. Trends Pharmacol Sci. 2014;35(2):63–75.

24. Clemente D, Porteros Á, Weruaga E, Alonso JR, Arenzana FJ, Aijón J, et al. Cholinergic elements in the zebrafish central nervous system: histochemical and immunohistochemical analysis. J Comp Neurol.2004;474(1):75–107.

25. Bertrand C, Chatonnet A, Takke C, Yan Y, Postlethwait J, Toutant J-P, et al. Zebrafish acetylcholinesterase is encoded by a single gene localized on linkage group 7 gene structure and polymorphism; molecular forms and expression pattern during development. J Biol Chem. 2001;276(1):464–74.

26. Parveen M, Kumar S. Recent trends in the acetylcholinesterase system. Amsterdam: IOS Press; 2005.

27. Jin Y, Liu Z, Peng T, Fu Z. The toxicity of chlorpyrifos on the early life stageof zebrafish: a survey on the endpoints at development, locomotor behavior, oxidative stress and immunotoxicity. Fish Shellfish Immunol. 2015;43(2):405–14.

28. Holmberg A, Schwerte T, Pelster B, Holmgren S. Ontogeny of the gut motility control system in zebrafish Danio rerio embryos and larvae. J ExpBiol. 2004;207(23):4085–94.

29. Nahata A, Patil U, Dixit V. Anxiolytic activity of *Evolvulus alsinoides* and *Convulvulus pluricalis* in rodents. Pharm Biol. 2009;47(5):444-51.

30. Bihaqi SW, Singh AP,Tiwari M. In vivo investigation of the neuroprotective property of *Convolvulus pluricaulis* in scopolamine-induced cognitive impairments in Wistar rats. Indian J Pharmacol. 2011;43(5):520.

31. Malik J, Karan M, Vasisht K. Attenuating effect of bioactive coumarins from *Convolvulus pluricaulis* on scopolamine-induced amnesia in mice.Nat Prod Res. 2016;30(5):578–82.

32. Bihaqi SW, Sharma M, Singh AP,Tiwari M. Neuroprotective role of *Convolvulus pluricaulis* on aluminium induced neurotoxicity in rat brain. JEthnopharmacol. 2009;124(3):409–15.

33. Finney DJ, Tattersfield F. Probit analysis. Cambridge: Cambridge UniversityPress; 1952.

34. Kopelman M, Corn T. Cholinergic 'blockade'as a model for choliner-gic depletion: a comparison of the memory deficits with those of Alzheimer-type dementia and the alcoholic Korsakoff syndrome. Brain. 1988;111(5):1079-110.

35. Kim Y-H, Lee Y, Kim D, Jung MW, Lee C-J. Scopolamine-induced learn- ing impairment reversed by physostigmine in zebrafish. Neurosci Res. 2010;67(2):156–61.

36. . !!! INVALID CITATION !!! .

37. Rangappa KS. New cholinesterase inhibitors: synthesis and structure–activity relationship studies of 1, 2-benzisoxazole series and novel imidazolyl-d2-isoxazolines. J Phys Org Chem. 2005;18(8):773–8.

38. Gutiérrez M, Matus MF, Poblete T, Amigo J, Vallejos G, Astudillo L. Isoxa- zoles: synthesis, evaluation and bioinformatic design as acetylcholinester-ase inhibitors. J Pharm Pharmacol. 2013;65(12):1796–804.

39. Anand P, Singh B. Synthesis and evaluation of novel 4-[(3H, 3aH, 6aH)-3-phenyl)-4, 6-dioxo-2-phenyldihydro-2H-pyrrolo [3, 4-d] isoxazol-5 (3H, 6H, 6aH)-yl] benzoic acid derivatives as potent acetylcholinesterase inhib-itors and anti-amnestic agents. Bioorg Med Chem. 2012;20(1):521-30.

40. Hamilton TJ, Morrill A, Lucas K, Gallup J, Harris M, Healey M, et al. Estab-lishing zebrafish as a model to study the anxiolytic effects of scopola- mine. Sci Rep. 2017;7(1):1–9.

41. Cho H, Lee C-J, Choi J, Hwang J, Lee Y. Anxiolytic effects of an acetylcho-linesterase inhibitor, physostigmine, in the adult zebrafish. Anim Cells Syst. 2012;16(3):198–206.

42. Karnovsky MJ, Roots L. A" direct-coloring" thiocholine method for cho-linesterases. J Histochem Cytochem. 1964;12(3):219–21.

43. Selvaraj V, Santhakumar K. Analyzing locomotor activity in Zebrafishlarvae using wrMTrck. Zebrafish. 2017;14(3):287–91.

44. Nussbaum-Krammer CI, Neto MF, Brielmann RM, Pedersen JS, Morimoto RI. Investigating the spreading and toxicity of prion-like proteins using the metazoan model organism *C. elegans*. J Vis Exp. 2015;95:e52321.

45. Roberts AC, Bill BR, Glanzman DL. Learning and memory in zebrafish larvae. Front Neural Circuits. 2013;7:126.

46. Gerlai R. Associative learning in zebrafish (*Danio rerio*). Methods Cell Biol. 2011;101:249–70.

47. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small mol- ecule docking web service based on EADock DSS. Nucleic Acids Res. 2011;39(suppl_2):W270–7.

48. Rollinger JM, Hornick A, Langer T, Stuppner H, Prast H. Acetylcholinest- erase inhibitory activity of scopolin and scopoletin discovered by virtualscreening of natural products. J Med Chem. 2004;47(25):6248–54.

49. Hornick A, Lieb A, Vo N, Rollinger J, Stuppner H, Prast H. The coumarin scopoletin potentiates acetylcholine release from synaptosomes, ampli- fies hippocampal long-term potentiation and ameliorates anticholiner-gic-and age-impaired memory. Neuroscience. 2011;197:280–92.

50. Roseiro LB, Rauter AP, Serralheiro MLM. Polyphenols as acetylcholinester- ase inhibitors: structural specificity and impact on human disease. Nutr Aging. 2012;1(2):99–111.

51. Balkis A, Tran K, Lee YZ, Balkis KN, Ng K. Screening flavonoids for inhibitionof acetylcholinesterase identified baicalein as the most potent inhibitor. J Agric Sci. 2015;7(9):26.

52. Balkrishna A, Thakur P, Varshney A. Phytochemical profile, pharmacologi-cal attributes and medicinal properties of convolvulus prostrates—a cognitive enhancer herb for the management of neurodegenerative etiologies. Front Pharmacol. 2020;11:171.

53. Kato G, Tan E, Yung J. Acetylcholinesterase. Kinetic studies on the mecha- nism of atropine inhibition. J Biol Chem. 1972;247(10):3186–90.

54. Wetherell J, Hall T, Passingham S. Physostigmine and hyoscine improves protection against the lethal and incapacitating effects of nerve agent poisoning in the guinea-pig. Neurotoxicology. 2002;23(3):341–9.

55. Venkov L, Iancheva N. Effects of scopolamine and pilocarpine on the activity of acetylcholinesterase in rat brain synaptosomal fractions. ProgBrain Res. 1979;49:495.

56. Woodruff-Pak DS, Vogel RW, Wenk GL. Galantamine: effect on nicotinic receptor binding, acetylcholinesterase inhibition, and learning. Proc Natl Acad Sci. 2001;98(4):2089–94.

57. Ketchum JS, Sidell FR, Crowell EB, Aghajanian GK, Hayes AH. Atropine, scopolamine, and ditran: comparative pharmacology and antagonists inman. Psychopharmacologia. 1973;28(2):121–45.

58. Giarman N, Pepeu G. The influence of centrally acting cholinolyticdrugs on brain acetylcholine levels. Br J Pharmacol Chemother. 1964;23(1):123–30.

59. Kalueff AV, Gebhardt M, Stewart AM, Cachat JM, Brimmer M, Chawla JS, et al. Towards a comprehensive catalog of zebrafish behavior 1.0 and beyond. Zebrafish. 2013;10(1):70–86.

60. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. Stages of embryonic development of the zebrafish. Dev Dyn. 1995;203(3):253–310.

61. Busquet F, Strecker R, Rawlings JM, Belanger SE, Braunbeck T, Carr GJ, et al.OECD validation study to assess intra-and inter-laboratory reproducibility of the zebrafish embryo toxicity test for acute aquatic toxicity testing. Regul Toxicol Pharmacol. 2014;69(3):496–511.

62. Ellman GL, Courtney KD, Andres V Jr, Featherstone RM. A new and rapid colorimetric determination of acetylcholinesterase activity. Biochem Pharmacol. 1961;7(2):88–95.

63. Hestrin S. The reaction of acetylcholine and other carboxylic acid deriva-tives with hydroxylamine, and its analytical application. J Biol Chem. 1949;180(1):249–61.

64. Tilton FA, Bammler TK, Gallagher EP. Swimming impairment and acetyl-cholinesterase inhibition in zebrafish exposed to copper or chlorpyrifos separately, or as mixtures. Comp Biochem Physiol C: Toxicol Pharmacol. 2011;153(1):9–16.

65. Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, PersikovaI, et al. RCSB Protein Data Bank: enabling biomedical research and drug discovery. Protein Sci. 2020;29(1):52–65.

66. Čadež T, Kolić D, Šinko G, Kovarik Z. Assessment of four organophospho- rus pesticides as inhibitors of human acetylcholinesterase and butyryl- cholinesterase. Sci Rep. 2021;11(1):1–11.

67. Ghosh S, Jana K, Ganguly B. Revealing the mechanistic pathway of cho- linergic inhibition of Alzheimer's disease by donepezil: a metadynamics simulation study. Phys Chem Chem Phys. 2019;21(25):13578–89.

68. Junaid M, Islam N, Hossain MK, Ullah MO, Halim MA. Metal based donepezil analogues designed to inhibit human acetylcholinesterase forAlzheimer's disease. PLoS ONE. 2019;14(2):e0211935.

69. Bosak A, Opsenica DM, Šinko G, Zlatar M, Kovarik Z. Structural aspects of 4-aminoquinolines as reversible inhibitors of human acetylcholinesteraseand butyrylcholinesterase. Chem Biol Interact. 2019;308:101–9.

70. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC,et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–12.

71. Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005;45(1):177–82.

72. Ordentlich A, Barak D, Kronman C, Flashner Y, Leitner M, Segall Y, et al. Dissection of the human acetylcholinesterase active center determi- nants of substrate specificity. Identification of residues constituting theanionic site, the hydrophobic site, and the acyl pocket. J Biol Chem. 1993;268(23):17083–95.

73. Dannenberg J. An introduction to hydrogen bonding by George A. Jeffrey (University of Pittsburgh). New York: Oxford University Press. 1997.ix+303. pp. $60.00. ISBN 0-19-509549-9. ACS Publications; 1998.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in pub-lished maps and institutional affiliat

Appendix 14.3 – Table 1: Protein-protein interaction (PPI) databases. The table shows the list of 23 databases that contain PPI information, along with data on 26 other related biological modalities. A coloured cell indicates that a particular biological modality is documented in the specific PPI database. <u>Abbreviations:</u> GO: Gene Ontology; PTM: post-translational modification; MeSH: medical subject headings.

PPI databases	Binding dynamics	Binding orientation	Linked diseases	Domain data	Drug targets	Enzymatic modification	PPI experiment type	Gene co-expression	Genomic neighbourhood	GO	Interaction detection method	Interaction sites	Interaction type
DIP											■		■
HPRD							■						
STRING								■	■				
Reactome													■
I2D				■				■		■			
MINT	■			■		■							
IntAct											■		
MIPS							■					■	
CORUM										■			
BioGRID							■						
CCSB		■											
InWeb													
MiMI													
HiPPIE				■			■			■			
iRefWeb											■		
HitPredict				■						■			
IMID			■										
HAPPI			■		■								■
DroID				■				■		■			
DRYGIN										■			
APID I				■						■			

InnateDB												▓	▓
MGI													▓

PPI databases	Interologous data	Cellular localization	MeSH	Network features	Pathways	Phenotype	Phylogeny	Protein architecture	Protein complexes	PTMs	Sequence information	Structural annotations	Tissue expression
DIP													
HPRD								▓		▓			▓
STRING	▓						▓						
Reactome													
I2D													
MINT		▓											
IntAct													
MIPS													
CORUM									▓				
BioGRID										▓			
CCSB													
InWeb	▓						▓						
MiMI			▓	▓									
HiPPIE				▓									
iRefWeb				▓									
HitPredict											▓	▓	
IMID				▓	▓								
HAPPI													
DroID						▓							
DRYGIN				▓					▓				
APID				▓									
InnateDB				▓								▓	
MGI				▓									

Interactome-based framework to translate disease genetic data into biological and clinical insights

University of Reading

School of Chemistry, Food and Pharmacy
Department of Pharmacology