School of Pharmacy

Systems Biology approaches for neurodegeneration and macroautophagy

Nikoleta Vavouraki

A thesis submitted for the degree of Doctor of Philosophy

July 2022

School of Pharmacy

Systems Biology approaches for neurodegeneration and macroautophagy

Nikoleta Vavouraki

A thesis submitted for the degree of Doctor of Philosophy

July 2022

62,868 words

Supervisors:

Prof Patrick Lewis, School of Pharmacy, University of Reading

Dr Claudia Manzoni, School of Pharmacy, University College London

Prof Marcus Tindall, Department of Mathematics, University of Reading

Declaration of original authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Signed: Nikoleta Vavouraki

Date: 4 July 2022

“Remember to look up at the stars and not down at your feet. Try to make sense of what you see and wonder about what makes the universe exist. Be curious. And however difficult life may seem, there is always something you can do and succeed at.

It matters that you don't just give up.”

Stephen Hawking

Abstract

Neurodegenerative diseases have been identified and studied for decades but disease-modifying drugs are still unavailable for their majority. Their genetic and clinical complexity renders the identification of the precise molecular disease mechanism challenging. Holistic approaches that allow the analysis of diseases in a systems level, studying multiple genes and their protein products simultaneously, could aid in the endeavour to find treatment for neurodegenerative diseases, such as the Hereditary Spastic Paraplegias (HSPs) and Parkinson's disease (PD).

Firstly, a protein-protein interaction network (PPIN) analysis was performed centred on proteins derived from genes that lead to HSPs, revealing that their majority share at least one interactor. This suggests that they participate in common biological processes and pathways. Enrichment analysis highlighted membrane trafficking and vesicle mediated pathways as important for the HSPs. Furthermore, the clinical complexity of the disease led to the investigation of potential mechanistic differences of the disease depending on the mode of inheritance, type of HSP, and clinical features. The analysis of the latter also utilised basic machine learning tools (principal component analysis and hierarchical clustering) and suggested the existence of 2 subgroups of HSPs with divergent disease mechanisms.

To investigate how a fundamental cellular process can contribute to disease, macroautophagy was studied, as it is associated with multiple neurodegenerative diseases. This connection was investigated initially by creating 5 PPINs (macroautophagy, PD, Alzheimer's disease, Amyotrophic lateral sclerosis, and Frontotemporal dementia), and examining their overlap. As the intersection between all studied neurodegenerative diseases and macroautophagy was extensive, I focused on the relationship between macroautophagy and PD. This required the creation of a mathematical model of the initial stages of macroautophagy, in which differential protein amounts were used to simulate a healthy person *versus* a person with PD. Interestingly, this distinction in amounts was sufficient to simulate differential kinetics of macroautophagy.

Acknowledgements

The research of this Thesis would not have been possible without the help of many people. Firstly, I would like to thank my supervisors Prof Patrick Lewis, Dr Claudia Manzoni, and Prof Marcus Tindall, who gave me the opportunity to do my PhD in this project and supported me in great extend the last four years both with technical aspects of my project and with other challenges. I would also like to thank all my colleagues in our extended research group that included the group of Dr Eva Kevei for their valuable feedback throughout the lab meetings and coffee breaks.

A special thank you to the funding bodies that supported the work in this Thesis: University of Reading, University College London, UKRI-EPSCR, UKRI-MRC, and Dolby Foundation.

In addition, I would like to thank colleagues that have contributed to the work presented in this Thesis. Drs James Tomkins and Claudia Manzoni developed PINOT and supporting material, including scripts and dictionaries used for enrichment analysis. Prof Henry Houlden, Dr Eleanna Kara, and Prof John Hardy offered their valuable feedback for the clinical aspect of the analysis of the Hereditary Spastic Paraplegias and genetic analyses of neurodegenerative diseases. Prof Sharon Tooze offered her valuable insight into the construction of the list of key autophagy entities.

I would also like to thank the people of Hopkins, both the pre- and post-COVID group of 2018-2020 and 2021-2022, respectively. You made the office a fun place to be. Also thank you to the "permanent residents" of the building that kept me company on the days that we had to work long hours. Dr Francesco Tamagnini, thank you for the countless interesting and thought-provoking conversations, and your mentoring.

Finally, thank you to my family and friends that supported me in this journey and in life in general. A big thank you to all my people (you know who you are). Σας ευχαριστώ πολύ!

Publications

1. Zhao, Y., **Vavouraki, N.**, Lovering, R. C., Escott-Price, V., Harvey, K., Lewis, P. A., and Manzoni C. (2022). Tissue-specific LRRK2 Protein-Protein Interactomes in the Physiological Scenario. (manuscript in preparation)

2. **Vavouraki, N.,** Tomkins, J. E., Kara, E., Houlden, H., Hardy, J., Tindall, M. J., Lewis, P. A. and Manzoni, C. (2021). Integrating protein networks and machine learning for disease stratification in the Hereditary Spastic Paraplegias. iScience 24, 102484. 10.1016/j.isci.2021.102484.

3. Cogo, S., Tomkins, J. E., **Vavouraki, N.**, Forcellato, F., Franchin, C., Tessari, I., Arrigoni, G., Cendron, L., Manzoni, C., Civiero, L., et al. (2020). PKA-mediated phosphorylation of SPG11/spatacsin regulates binding with a subset of 14-3-3 proteins. bioRxiv 2020.09.09.289009. 10.1101/2020.09.09.289009.

4. Tomkins, J. E., Ferrari, R., **Vavouraki, N.**, Hardy, J., Lovering, R. C., Lewis, P. A., Mcguffin, L. J. and Manzoni, C. (2020). PINOT: an intuitive resource for integrating protein-protein interactions. Cell Commun Signal 18, 92. 10.1186/s12964-020-00554-5.

Table of Contents

List of Figures

Chapter 1: Introduction

Chapter 2: PPIN analysis of HSPs

Chapter 3: Further analysis of HSPs based on clinical features

List of Tables

Chapter 1: Introduction

Chapter 2: PPIN analysis of HSPs

Chapter 3: Further analysis of HSPs based on clinical features

Chapter 4: Network analysis of macroautophagy and its interplay with neurodegenerative diseases

Chapter 5: A process diagram of macroautophagy

Chapter 6: A mathematical model of macroautophagy

Abbreviations

AD.....Alzheimer's disease

AIM.....ATG8-interacting motif

ALS.....Amyotrophic lateral sclerosis

AMP.....Adenosine monophosphate

AMPK.....Adenosine monophosphate-activated protein kinase

ARN.....Autophagy regulatory network

ATG.....Autophagy-related genes

BioGrid.....Biological general repository for interaction datasets

cAMP.....Cyclic-adenosine monophosphate

DIP.....Database of interacting proteins

DNA.....Deoxyribonucleic acid

ELISA.....Enzyme-linked immunosorbent assay

EPOD.....Early onset, peripheral neuropathy, optical atrophy, and dementia/mental retardation

ER.....Endoplasmic reticulum

Expasy.....Expert protein analysis system

FTD..... Frontotemporal dementia

GAP.....GTPase activating protein

GO.....Gene ontology

GO-BP.....Gene ontology terms for biological process

GO-CC.....Gene ontology terms for cellular component

HC.....Hierarchical clustering

HD.....Huntington's disease

HIPPIE.....Human integrated protein-protein interaction reference

HSP.....Hereditary spastic paraplegia

HUPO-PSI.....Human proteome organization proteomics standards initiative

HUPO-PSI-MI XML.....Human proteome organization proteomics standards initiative-molecular interaction extensible markup language

IIH.....Inter-interactome hub

IMEx.....International molecular exchange

IP3R .....Inositol-trisphosphate receptor

iPSCs .....Induced pluripotent stem cells

JNK .....c-Jun N-terminal kinase

LIR.....LCE-interacting region

LRS.....LC3 recognition sequence

LSA.....Local sensitivity analysis

MA.....Macroautophagy

MAM.....Mitochondria associated endoplasmic reticulum membrane

MIMIx.....Molecular interaction experiment

MINT.....Molecular interaction database

MIQL.....Molecular interaction query language

MIST .....Molecular interaction search tool

MITAB.....Molecular interaction tabular

MPTP.....1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine

ND.....Neurodegenerative disease

ODE.....Ordinary differential equation

PCA.....Principal component analysis

PCR.....Polymerase chain reaction

PD.....Parkinson's disease

PE.....Phosphatidylethanolamine

PI3P .....Phosphatidyl inositol 3-phosphate

PINOT.....Protein interaction network online tool

PPI.....Protein-protein interaction

PPIN.....Protein-protein interaction network

PSICQUIC.....Proteomics standards initiative common query interface

REM.....Rapid eye movement

RNA.....Ribonucleic acid

STRING.....Search tool for the retrieval of interacting genes/proteins

TNF.....Tumour necrosis factor

TS.....Thinning of corpus callosum and seizures

UniProt.....Universal protein

1st Chapter:

Introduction

1. Introduction

1.1 Systems Biology

A definition of a "system" is that it is a regularly interacting or interdependent group of items forming a unified whole [1], or that it is "a network of mutually dependent and thus interconnected components comprising a unified whole" [2]. Of all the systems in the world, biological ones comprised of the greatest complexity. This can be noticed by the contrast of achievements of the respective fields of physics and biology. The understanding of the laws of physics has grown exponentially in the last decades, allowing humans to walk on the moon and only 52 years later to fly a helicopter (Ingenuity) on Mars, which is more than 300 million kms away [3]. However, many aspects of the biological mechanisms governing our own bodies are yet to be understood and several diseases are still incurable.

The response of a biological system to a stimulus is more difficult to predict than that of a physical system due to its higher level of complexity. Lowering of the environmental temperature of a gas in a container, will lead to the highly predictable change of its volume (Ideal gas law [4]), whereas the response to the same stimulus by a cell or an organism can vary significantly, as it might lead to no observable change or conversely to its death. One of the reasons for the lack of linearity in the response of biological systems is the presence of functional loops, which add complexity, allowing their behaviour to vary largely, depending on the details of the system [5]. An elegant example demonstrating this issue is described by Voit [5], in which he hypothesizes a linear system activated by an input. The input leads to the formation of the first component of the system, X, which in collaboration with E forms Y. Then Y forms Z, and Z produces an output (Fig 1-1A). Increasing or decreasing the input will lead to a simple increase or decrease of the quantities of the components of the system. However, when a loop is added starting from Z, linking it to E through TF and G (Fig 1-1B), then the outcome depends on the model's parameters. If the effect of Z on TF is weak then the system behaves in similar way to previously described. If it is strong, then this may lead to oscillations in the concentrations of the system entities, which dampen over time. This result was calculated using modelling and would have been more time-consuming to compute based on intuition. Thus, systems biology approaches, such as modelling, can provide valuable insight, when the complexity of the system renders intuition insufficient for reliably predicting the outcome.

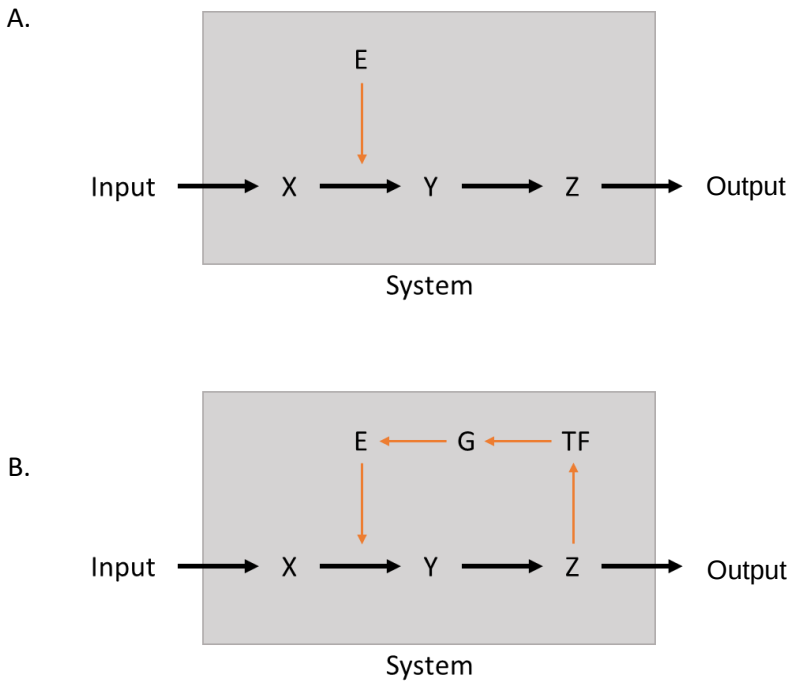


Figure 1-1. Linear and non-linear biological systems.

*(**A**) The Input activates X, which together with E is produces Y. Then, Y produces Z, which creates an output. (**B**) An additional positive feedback-loop is present compared to the model of A. The loop is formed between from Z to E, increasing the systems' complexity and diversifying the potential outcomes, based on the parameters of the system. A linear system (**A**) has predictable outcomes, even based simply on intuition, in contrast to non-linear systems (**B**).*

Systems biology is founded on the premise that biology can be more deeply understood from an integrated-systems perspective. Expertise from multiple branches of sciences, such as chemistry, physics, mathematics, and computer sciences, can facilitate in the increased understanding of the function of groups of genes, proteins, lipids, and other components that dynamically affect the functions of cells and organisms.

In the beginning of the 20th century, the need for a systems biology approach was highlighted, by Ludwig von Bertalanffy, who argued that organisms have a systemic nature for three decades [6, 7], and Mihajho Mesarovic, who first used the term "Systems Biology" [8]. Also, François Jacob stated in 1974 that "Every object that biology studies is a system of systems" [2]. However, the collaboration of biology with other disciplines like mathematics and computer science continued to be sparse and mainly limited to data management for a long time.

Over the last few decades, the appreciation of systems biology approaches has grown significantly. The study of the human body and disease has shifted from attempting to understand how our

systems and organs function and dysfunction, to studying the complicated relationships between the over 20,000 genes that are expressed differently in various cell types. It has also become evident that in addition to the relationships between genes, other factors also influence their behaviour, such as transcription factors, metabolites, RNAs, and epigenetic changes of the DNA.

Systems biology underlines that the features of a system are not a simple product of the addition of the properties of its components. Instead, there are multiple regulatory feedback loops among the components. The final properties of the system are, therefore, the complex result of the combination of all these modulatory effects. As a result, the system can have emergent behaviours, which are uniquely possessed by the whole system and not by the individual components. Examples that highlight this point are the emergent properties of the central nervous system. Understanding and predicting the features of action potentials in a neuron are insufficient for deciphering the enigma of memory storage and acquisition, and other functions of the brain. Even though obtaining information about the mechanistic details of neurons is essential, adopting in parallel a more holistic approach might be required in order to understand such a complex system.

1.1.1 Omics

Omics in biology is a term used to refer to the production or study of large (and comprehensive) data sets in a high-throughput fashion. More specific terms with this suffix include genomics and proteinomics, which correspond to global analyses of genes and proteins, respectively in a given cell, tissue or sample.

The appreciation of the value of omics approaches has risen in the last decades, which is evident by their incorporation in an increasing amount of research. Interestingly, the PubMed results for “genome” start in 1943 [9] and their exponential growth begins in the 1990s, while “omic” follows a similar trend, with its first result in 2000 [10] and the start of its exponential growth in early 2020s.

A key factor that prompted the rise of omics was the Human Genome Project, which aimed to sequence and map most of the human genome. An international team of researchers was assembled for this massive endeavour, which started at the end of 1990 and was completed in early 2003. This project managed to sequence 99% of gene-containing human DNA with an 99.99% accuracy, to sequence the complete genome of other organisms, such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster*, and to create

whole-genome drafts for additional organisms, like mouse and rat [11, 12]. These results were combined with a private investigation led by J. Craig Venter to create the human reference sequence [13, 14]. Combining the human reference sequence to the genome-wide map of common variability produced by the International HapMap project and the 1,000 Genomes project [15, 16] interestingly demonstrated the great genetic similarity between any two humans (99.5%). At the end of the Human Genome Project the gap of knowledge between genotype and phenotype was more apparent, which highlighted the importance of further research [17].

Analysing the genome can reveal important information about a person's health. Inside a human cell, there are approximately 20,000 pairs of genes [18]. Among people a gene can be present in different forms, some with positive, and/or negative effect on health [19, 20]. Studying the genome of groups of people can reveal associations between changes in the genome sequence and the risk of developing a certain disease, and can highlight the importance of a specific pathway in the disease mechanism. The functionality of a gene is usually expressed by its product, such as a protein. Therefore, complementally functional analysis of gene products can be insightful for this endeavour [21].

The rise of genomics was followed by the rise of proteinomics, the study of group of proteins, instead of their study in isolation or in combination with just a few proteins, as mainly done previously. Proteinomics can be considered more complicated than genomics. DNA has a code of 4 nucleotides and is almost identical in different cells and developmental stages in an organism. In contrast, proteins have a code of 20 amino acids, with varying sequences, lengths, and 3D structures, leading to an over 50 times increase in the number of possible protein configurations compared to the number of genes [22, 23]. This increased expansion of complexity is mainly due to alternative splicing of exons and post translational modifications [24, 25].

Proteinomics cover a wide range of approaches, including proteomics, structural proteomics and interaction proteomics [26]. Proteomics focuses on the identification and quantification of proteins, mainly through Mass Spectrometry [27]. This approach is valuable in describing differentiations of proteins among compartments, cells and tissues, and their associations with diseases for the development of diagnostic markers [28-30]. In addition, spatiotemporal proteomics can be performed allowing the study of changes, for instance, during cell signalling [28, 31, 32]. Structural proteomics focuses on the 3D structure and protein modifications. Visualising the structure of proteins (or parts of proteins) can aid in inferring their function(s). In addition, structural changes due to disease-causing mutations can hint the disease mechanism. Combining data from health and disease states can aid the development of drugs that can structurally and

functionally modify the proteins of interest [33]. Lastly, interaction proteomics focuses on the associations of proteins with other components of biological systems, such as DNA, RNA and lipids. Interactions of proteins with DNA are necessary for multiple biological processes, including the transcription of genes, during which transcription factors interact with promoters. Proteins also interact with RNAs, such as ribosomal RNAs that are responsible for proteinosynthesis. Lipids can be attached to proteins to modify their properties. For instance, phosphatidylethanolamine can be attached to LC3I to produce LC3II, which is key for the progression of macroautophagy [34]. Furthermore, proteins interact with other proteins to exert their functions. Examples of such functions include protein modifications, like phosphorylation, ubiquitylation, and truncation/proteolysis. The protein-protein interaction proteomic analyses will be discussed in more detail in Section 1.2.

The main challenge of proteinomics is technical. In comparison to genomic technologies, those of proteinomics are less developed and scalable hindering the progress of the field [28]. However, recent developments have allowed the identification and qualification of approximately 7,000 proteins from mammalian systems, requiring the reasonable measuring time of 48h and a few thousand cells [35-37]. Based on the notion that human cells express around 10,000 proteins [28], this advancement of proteinomics appears to allow comprehensive expression analysis of human proteomes. Proteinomic analyses are advantageous as they inform about the product of gene expression, which is more directly associated with the biological process compared to the mRNA, or gene. In addition, proteinomics can inform about the presence and concentration of a protein in a specific location (e.g., organelle) in contrast to genomics and transcriptomics [28] .

1.1.2 Big Data

Improvements in the technologies and methodologies used in biology have resulted in the ability to produce large amounts of data through a single experiment. Such high-throughput studies have allowed the collection of genomic data, such as genome-wide expression, and proteinomic data, such as the identification and quantification of proteins and other biomarkers from a single sample. The advances on this field are notable. For instance, even a decade ago, it was unthinkable that a detailed 3D map of an animal and human cell could be obtained, let alone from one experiment, but this has now become a reality facilitating single cell multiomic analyses [38, 39].

Multiple efforts on gathering genomic and proteomic information have been conducted. For instance, the 100,000 Genomes Project, a patient-centric effort, sequenced the DNA of 100,000

people with either rare diseases or cancer by the end of 2018 [40]. Such large amounts of genetic data allow the conducting of powerful genetic studies that can identify genetic alterations, which are associated with diseases. In January 2022 there were over 325,000 unique associations of single nucleotide polymorphism and traits from over 5,500 studies in the Genome Wide Association Studies (GWAS) catalogue, providing valuable insight into the mechanisms of diseases. This has resulted in over 500 genetic associations with Parkinson's disease from 67 studies, and in over 1,200 associations with Alzheimer's disease from 123 studies so far [41]. Other studies are focusing on collecting multiomic data, like the Precision Medicine Initiative that aims to collect a multiomic data set from over 1 million people to study a wide range of diseases [42, 43].

Proteinomics have a number of inherited challenges which need to be tackled. The heterogeneity of the identity and quantity of proteins, and their alternative splicing products and post translational modifications increases the complexity of their analysis. In addition, there is a lack of methods to sequence proteins on the omics scale, in contrast to genes [26].

Nonetheless, collective efforts through multiple groups and organisations of researchers have provided a number of guidelines and repositories that accelerate the progress of proteinomics. For instance, a common format and unique vocabulary for proteomic data was created and is dynamically being updated but the Human Proteome Organization Proteomics Standards Initiative (HUPO-PSI) [44, 45]. A consortium, ProteomeXchange, was formed in 2006 to tackle the difficulty to interpret and compare data from various repositories, based on differences in the data submission process and the level of detail they provide. Since then, it actively encourages re-using of published data and adopting open data policies [46, 47]. A protein's structure is valuable in multiple aspects. It can be indicative of a protein's function and central for the designing of new drugs. Therefore, the need of a repository of protein structure led to the formation of Protein Data Bank in 1977, which published its 10,000th structure in 1999. It is an open access digital data resource that includes more than 185,000 structures (as of 15/01/2022) of proteins, nucleic acids, and complex assemblies, and provides tools for research and education [48, 49]. Another central endeavour has been conducted to merge amino acid sequence databases to reduce inconsistencies and the amount of different identifiers, such as by the work of Universal Protein (i.e., UniProt) [50].

Another aspect of proteinomics that benefited from the organisation of the research community is the protein-protein interaction (PPI) analysis. This subject will be analysed in more detail in Section 1.2, but briefly, there is a high number of databases that holds PPI information. On one hand, this provides a collection of protein-protein interactions per specific area of interest, allowing the user to have easy access to a range of data around a certain subject. However, the inconsistency of the

level of detail and the amount of information that each database holds, renders necessary the investment of additional time for both the people or algorithms that curate the research papers for different databases, and for the end-user, who is required to collect, collate, and filter the interaction data from multiple sources. Namely, some of the groups of researchers that have facilitated in solving the aforementioned and additional issues are the Human Proteome Organization Proteomics Standards Initiative [51] and the International Molecular Exchange consortium [52].

This era of big data has accelerated our understanding of biological systems but also demands the development of sophisticated tools for data analysis. Machine learning and bioinformatic tools have been developed to mine and/or analyse large quantities of data fast and consistently, and detect trends in data that would have been undetected using simple analytical methods. The large amount of available databases and tools has led to the creation of resource portals such as Expasy (Expert Protein Analysis System), which is allowing access to over 160 databases and software tools for various omic data [53], and Bioconductor through which over 2,100 software packages are available (on 4th May 2022) [54].

Big data can be analysed using multiple approaches, with one of them being the production of networks. For instance, protein-protein interaction networks (PPINs) are valuable tools for summarizing the complexity of biological systems. On some occasions, they can be further translated into a computational model, which can then be mathematically analysed. The graphical representation of the system, produced by the PPIN or the model, can provide new insights. Therefore, studying protein networks and creating quantitative models that incorporate biological pathways and their connections are effective means to investigate, for example, how the mutations or therapeutic interventions might modulate the overall system [55].

1.1.3 Future trends of Systems Biology

Even though valuable insight has been generated with the use of genomic and proteomic approaches in biology, the reality in live human cells and human bodies is even more complicated. For instance, proteins can regulate genes, but also metabolites can regulate proteins. The interdependence of these cellular components is not captured by studying each omic separately, leading to the rise of multiomic approaches [56]. Integrating data from different omics to create a more informative global picture of a biological process/disease [57] has been applied in multiple projects, such as tumour classification and prediction of aggressiveness [58], and the study of

COVID-19 [59] and of neurodegenerative diseases, such as Parkinson's disease [60]. In addition, there are multiple recent reviews that have highlighted the need of multiomic approaches in a variety of topics, such as the study of endometriosis ([61], tuberculosis [62], atherosclerosis [63], and Parkinson's disease [64].

The rise of multiomic analysis is creating the need of developing pipelines for data integration and allowing access to multiple data sets. Custom pipelines might be tailored to the specific needs of each research group but if there are not available and valuable to the broader scientific community, they have limited capacity to speed up this endeavour. Publicly available multiomic tools are being developed. For instance, through GeneAnalytics the user can collect multiple omics data for their genes of interest [65], however, there is a subscription fee.

Omic and multi-omic data analyses are powerful approaches that can be applied to improve our understanding of biological systems, but require careful selection of appropriate data. The objectives, assumptions, and details of the pipeline behind the collection of each data set need to be known and can largely affect the compatibility of each pair of data sets. For instance, if the focus of the study is the understanding of the disease, then the disease mechanism needs to be taken into consideration to ensure -for example- that the time point(s) to which the data sets correspond could reflect the primary events linked to disease aetiology [66].

The merging of data sets itself can also present its challenges, as it is not a simple process of layering information. Across databases and data sets, there can be different terminologies, quality filters, and algorithms impeding data harmonization. Therefore, pipelines integrating a variety of different types of data are a necessity for this endeavour [26].

Aside from its limitations, omic data have strong potential, which can be more fully accessed with open access data policies, and collaborations of people from different areas of biological research and expertise from other scientific backgrounds. Such systems biology approaches can accelerate the booming of knowledge around health and disease and thus the discovery of treatment and prevention methods to ameliorate the quality of life of people with conditions/diseases.

Together with multi-omics, other applications of systems biology can contribute to the progress of scientific understanding. For example, systems biology plays an increasingly important role in multiple stages of drug development, such as the choice of drug targets, the optimal dosing strategy, and the extrapolation of data from animals to humans [5, 67]. In addition, since creating reliable and powerful biological systems through well designed alterations is another goal of systems biology, it could lead to progress towards personalised medical treatments [5]. This

endeavour commences with modelling natural systems, in which considerable progress has already been made [68-72].

Significant effort has been placed on understanding the function of neurons and brain using systems biology approaches. One of the first steps in this direction was the publication of the work of Hodgkin and Huxley in modelling neurons, almost 70 years ago [73]. Since then, a variety of models of neuronal function and networks has been published. Some models are at the level of a single neuron or a pair of neurons and include more details. For instance, models have focused on the dynamics of neurotransmitters in a neuron, and on the signal transduction from one neuron to another through the synapse [74-78]. Other models address larger systems in a less detailed manner, such as the function and interaction of different brain areas [79]. However, there are exceptions in this choice between smaller or larger systems with more or less details, respectively. A noteworthy project that models a large system in a detailed manner is the Blue Brain project. It aims to "build biologically detailed digital reconstructions and simulations of the mouse brain", based on supercomputers [80]. Multiple milestones have already been achieved including the creation of a full cell atlas of all neurons and glia in the mouse brain, published in 2018 [81], and of the architecture of the system among neurons, glia and blood vessels published in 2021 [82].

The potential of mathematical modelling is increasing with the further development of computers and programming languages. Since the 1950s, computational power and speed have doubled every 1.5 years. In parallel, multiple higher level programming languages, like MATLAB and SBLM, increase the accessibility of coding/programming to a wider audience, such as scientists who lack a computer science background. [5]

1.1.4 Examples of Systems biology approaches in this Thesis

There are different types of models that can be created in systems biology. Some are static and represent all the information and relationships of a model irrespective of time and space, while others are dynamic, in which concentrations and events can occur in both time and space. A paradigm of a static model is a protein-protein interaction network, in which all interactions between the components are incorporated, while an example of a dynamic model is a mathematical model formulated to describe macroautophagy related processes in time. Both approaches will be adopted and explored in Chapters 2-4, and 5-6, respectively.

1.2 Protein-protein interaction networks

1.2.1 Protein-protein interactions

Protein-protein interactions (PPIs) represent a key mechanism whereby proteins exert their functions. They can indicate the functional connection among genes and proteins, leading to the formation of hypotheses regarding the function of newly discovered proteins, new functions of proteins, and of the overall common processes and functions in which a group of various proteins are involved (holistic view/approach) [83].

Protein-protein interactions are useful to conceptualize the function of the proteins of interest. The "guilt by association" principle, in this context, supports the idea that interacting proteins are more likely to belong to the same pathway, and thus to have similar functions [84]. A paradigm that demonstrates the power of applying this principle is the accuracy of prediction of characteristics of novel proteins of *S. cerevisiae*. The development of an algorithm to create a multiple layer network was centred on the proteins of interest and their interacting proteins [85]. In this way, it was possible to predict the cellular role and localization of the proteins of interest with high accuracy, as well as their biochemical function. Therefore, by studying PPINs we can assert the function of a new protein, as well as illuminating additional functions of a previously studied protein.

Furthermore, understanding the dynamic connections between different proteins is particularly valuable when trying to intervene and manipulate the identified mechanism of a disease. For example, if it is known that a specific protein-protein interaction (along with its supported function) is lost in a disease, then it can be targeted to prevent or at least decelerate disease progression.

Moreover, PPIs can be used to analyse communal disease pathways when dealing with complex disorders with unidentified mechanisms, where many different genes are responsible for the manifestation of the same clinical phenotype. The analysis of the genes' interactomes (with particular interest in their overlaps and commonalities) might help in detecting shared functions responsible for univocal disease mechanisms. So, systems biology and more specifically protein-protein interaction networks provide a valuable tool to study complex processes and their relationship with diseases, such as the association of macroautophagy and neurodegeneration that is studied in Chapters 4-6 of this Thesis.

The interactome of a protein provides us with useful knowledge, however the conclusions from such an in silico analysis should be further scrutinized and always requires direct functional

validation. This is because, for instance, some interactions might not actually occur in the human body, due to differences in the tissue, space, time, and level of expression of each protein [86].

1.2.2 PPI databases

The first database of PPIs, the Database of Interacting Proteins (DIP) [87], was created in 1999. Since then, many more have been developed, due to the usefulness of information extracted from PPIs. The great variation among PPI databases highlights the importance of understanding their different types before selecting the database that fits with the type of desired data and the goal of each study. The main categories of PPI databases are primary and secondary. Briefly, primary databases store PPI data exclusively from peer-reviewed publications, while secondary databases collect and analyse data from primary ones [88].

1.2.2.1 Primary databases

Most primary databases include information regarding the identity of the interacting proteins, the detection method, and some experimental details, however they vary in many aspects of the process of collecting the data (i.e., curation). A characteristic of primary databases is whether they contain all the interactions discovered in a paper (i.e., archival database) (e.g., IntAct [89]), or only the interactions that are related to a specific subject (i.e., thematic database), such as immunology (e.g., InnateDB [90]). The depth of the curation can also vary, with the main resulting categories being: rapid-level curations, Minimum Information required to report a Molecular Interaction Experiment (MIMIx)-level curations [91], and International Molecular Exchange (IMEx)-level [52, 92] curations, with the latter having the most detailed information, as all the available details are recorded.

Another characteristic of a primary database is the data extraction process. Some databases use data that are manually curated from peer-reviewed papers that support the interactions, such as IntAct, Molecular INTeraction database (MINT) [89] and Biological General Repository for Interaction Datasets (BioGrid) [93]. Manual curation refers to the process of manual acquisition of data concerning interactions from trained experts [88]. These databases are reliable for their accuracy [88] but they require large efforts to be maintained, and the curation process is slow and expensive. In addition, manual curation is prone to inconsistencies, as not all databases have identical curating guidelines [94]. In order to resolve this issue, the IMEx consortium [52] (which will be discussed in more detail in Section 1.2.2.3) was formed with the aim of setting guidelines to

the manual curation process, in order to gain consistency among databases and to reduce the instances where one publication was curated more than once. Nonetheless, manually curated databases are usually more trusted by the scientific community due to the expertise and valuable intuition of the curators.

Other primary databases collect information on PPIs using automated methods or use algorithms allowing predictions. Automated methods use in silico models that scan the literature, i.e., text mining (e.g., PESCADOR [95] and iHOP [96]). On the other hand, databases that predict interactions between proteins, can use a variety of information -such as their localization, expression, and interactions of orthologous proteins- to evaluate the possibility of each interaction. Predictive databases can combine information of manually curated primary databases with predicted results, like Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [97], in order to increase their coverage.

1.2.2.2 Secondary databases

A PPI database is secondary when the sources of information are primary databases. Some examples are MIST (Molecular Interaction Search Tool) [98] and HIPPIE (Human Integrated Protein-Protein Interaction rEference) [99]. The range and quantity of information provided from this type of databases is usually higher than those of primary databases, as they tend to combine data from multiple primary databases, however the confidence in the results is more limited.

Some problems arise from the initial step of selecting the original data that will be processed. For example, the selection criteria of the primary databases or other sources used are often not fully transparent to the user. In addition, users might be unable to select which primary databases they want to include. These issues hinder the flexibility that is required to select data that correlate with the goal of each study. Furthermore, data from primary databases are usually stored in secondary databases. Thus, the users might be unable to incorporate the latest research findings into their analysis, limiting the rate of progress in the field. Moreover, in order to present the data in a homogeneous way, each secondary database decides the format of the final merge. This process leads to data loss, when data from primary sources is not in a format enabling the merging. Finally, the rules of merging might be unclear, removing the ability of the users to identify how the data were collected, and thus if the data are suitable for the goal of their project.

1.2.2.3 IMEx consortium and MITAB

There are more than 100 PPI-related databases. As each database uses a different way to collect and annotate PPIs, the user faces great difficulty in processing each data set and combining them for a greater coverage. This highlighted the need for curating methods to be optimized and standardized, which required a partnership among public interaction data providers.

The first step was to agree on the same format of curations. These guidelines were put into place in the early 2000s when protein interaction data providers developed a common file format, the Human Proteome Organization Proteomics Standards Initiative-Molecular Interaction eXtensible Markup Language (HUPO-PSI-MI XML) format, which was later supplemented by Molecular Interaction Tabular (MITAB), a simplified tabular format [100, 101]. Additionally, controlled vocabularies for the description of interactions were agreed to be used to increase accuracy and consistency of PPI entries in databases. This standardisation of data representation in proteinomics aimed to enable the validation, exchange and comparison of data across databases [101].

The next step was to limit redundancies by adopting a single curation strategy. In late 2005, some molecular interaction databases (e.g., IntAct, MINT, DIP) decided to co-ordinate their strategies, creating the International Molecular Exchange (IMEx) consortium [52, 92]. Some databases are full members, and others are observer members, meaning that they collaborate with the full members but have not yet fully committed to annotating interactions in accordance with the IMEx guidelines. Presently (as of 13th June 2022), the full members are DIP, IntAct, MINT, I2D, MatrixDB, UniProt group, InnateDB, Swiss-Prot group SIB, and UCL-BHF UCL London, EMBL-EBI and the observers are BioGrid and PrimesDB. All members work towards reducing the instances where more than one group of people curate the same publication, and also towards increasing the total number of publications curated and the quality and consistency of the curation. Initially, each journal was allocated to a single database. Although this is realised, sometimes databases curate additional journals. For this purpose, IMEx Central was implemented to allow databases to request publications in journals that are not allocated to other databases. Consequently, each publication has a unique IMEx accession number and is curated only once. Furthermore, it is interesting that IMEx allows the request of the input of an interaction by users, e.g., if the interaction is well-known in their field but absent from databases [52, 92].

1.2.2.4 PSICQUIC

PSI Common Query Interface (PSICQUIC) is a standard interface that was developed to ease and provide flexibility in the use of multiple primary databases and was established in 2011. PSICQUIC

allows direct computational access to multiple molecular interaction data resources that comply with the PSI-MI directives and potentially the IMEx manuals. It allows the users to access the primary databases of interest without having to replicate the query in each single database. The query must follow the molecular interaction query language (MIQL) and can be more complicated than the protein name, e.g., complex construct [102].

PSICQUIC also offers open-source client libraries and a code to facilitate the access to its registry and services via programming. Moreover, its collaboration with the IMEx consortium is evident, as it provides the option to search the query in the IMEx databases, allowing for uniqueness of the resulting interactions [102]. However, users must be cautious, as the results might be a mixed set of interactions, varying from experimentally proven direct interactions to predicted interactions.

1.2.3 Creating and analysing PPINs

1.2.3.1 Overview of PPINs

The most common way to visualise protein interactions is via networks. Networks are graphs in which proteins are represented as nodes and their interactions as edges [88]. Edges connect only two nodes and they can be directed, undirected or bidirectional. While the direction in the edges may be important for other types of networks, such as transcriptional regulatory networks, edges are undirected in protein-protein interaction networks (PPINs), as no flow of information is represented in these graphs. Edges can have scores, named edge weights, with highest weights usually representing higher confidence for the true nature of the interaction. A case where such an annotation could be beneficial is the existence of different types of data in a network. Direct experimental evidence suggesting the existence of a protein-protein interaction can have a higher edge weight, than of an interaction inferred from interactions of homologous proteins in a different species.

There are two types of network-based approaches: top-down or bottom-up. In the former, the creation of a larger unbiased network is followed by the mapping of genes of interest in the network. In contrast, the network in bottom-up approaches is build based on the genes of interest, which creates a more focused but also potentially more biased view of the disease [103].

The complexity of the network can vary depending on the aim of the analysis. If the interest is in the direct connections within a specific group of proteins, then the network consists of just the proteins of interest and their relationships and is called layer 0. In the 1st layer, direct interactors of

the group of interest are also included. Finally, in the 2nd layer of a network the direct interactors of the direct interactors of the proteins of interest are also included. Therefore, in the 2nd layer networks two proteins of interest can be linked directly, or through one, two or three interactors.

In PPINs it is in some cases essential to identify important nodes. There are multiple ways to estimate the significance of a node in a network, which are named centralities. The selection of a centrality by the researcher is dependent on the aim of each study.

Degree centrality is based on the degree of a node, which is the number of nodes to which a node connects. This is indicative of the level of connectivity within the network and the speed of the transfer of any signal from one side of the network to another in the case of directed graphs. Plotting the degree of a PPIN's nodes to show the degree distribution, can highlight if there are any nodes that interact with a lot of nodes of the graph and thus could play a vital role in the system that is being studied.

It is important to note here that in biological networks the degree distribution is usually different than in a randomly generated network. Biological networks tend to follow a power law distribution instead of a binomial distribution (Fig 1-2) because a few nodes connect with a lot more nodes acting as hubs. Such networks that follow a power law distribution are termed scale-free networks and they usually have a shortest distance between any two nodes (named shortest path). The shortest path is an indicator of how quickly information can be pass from one element of the system to the other, highlighting that biological networks are usually efficient in that aspect.

In other networks, the connectivity of a node to a group of nodes with a specific characteristic can be of higher significance compared to the node degree. For instance, if the network is constructed based on disease genes it can be more intuitive to find if a node interacts with a lot of disease proteins, then with other proteins not associated with the disease of interest. This is the basis for the concept of inter-interactome hub degree, which demonstrates the degree of each node to a specific subgroup of components of the graph [104].

Another type of measurement for the importance of a node is closeness centrality. This measurement is estimated based on the shortest path between two nodes (termed geodesic distance). A node with high closeness centrality is a few steps (i.e., nodes) away from all the other nodes in the graph, not just those with which it directly interacts.

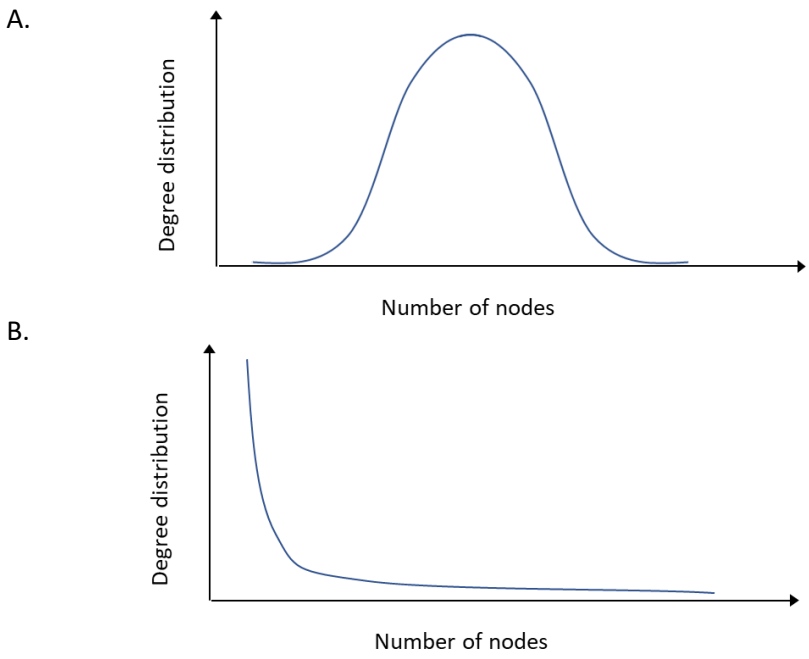


Figure 1-2. Degree distribution in random and biological networks

*(**A**) The degree distribution in random networks usually follows a binomial distribution. (**B**) In biological systems the degree distribution within protein-protein interaction networks usually follows a power law distribution.*

Additional centralities include, betweenness centrality and eigenvector centrality. Betweenness centrality defines a node as important based on the number of times a node is part of the shortest path between all other pairs of nodes in the graph. Eigenvector centrality is a measure of the centralities of a node's neighbours, and it equates to their sum. A node could have a high eigenvector centrality because it connects to multiple nodes with a medium/low level of centralities, or it connects fewer nodes with high centralities. Nodes with high eigenvector centrality have a higher influence over the whole network, so it can be useful for them to be identified [105].

A variety of software packages and plugins has been developed to offer user-friendly visualization strategies and analysis tools, providing an insight into the biology behind the interactions. Through those, several attributes can be encoded within the network by altering different visualisation parameters, such as the confidence of the interaction that can change the width of each edge. This flexibility allows visualisation of much information in each network and complex data analysis. Some examples are Cytoscape [106], NAViGaTOR [107], and packages from Bioconductor [54] and R [108]. Cytoscape is amongst the most widely-used, due to being open-source, having modular design and offering flexibility and extensibility. It allows the loading of multiple PSI-MI and txt files,

which it visualises into a network, while offering the ability to annotate and analyse the network further through a vast number of available plugins [88, 106].

1.2.3.2 Graph theory

Graph theory relates to the properties of graphs, such as the connectivity of its components. It theorises abstract ideas and methods, which can be utilised for the visualisation and analysis of networks.

Graph theory was developed as a branch of mathematics devoted in the study of social networks but it is widely applied in the analysis of PPIs [109]. Its history originates from Leonard Euler in the 18th century, who was attempting to solve a puzzle regarding a river in his city, Königsberg, Prussia. The river was going through the city and could be crossed through seven bridges. The question he posed to himself was whether he could cross each bridge just once. His background in mathematics allowed him to solve the puzzle and additionally to develop the theoretical framework to calculate when a unique route exists, which became the foundation of graph theory.

Biological networks that can be studied with graph theory are PPINs, metabolomic networks, genetic interactions, gene/transcriptional regulatory networks, and cell signalling networks. In this Thesis, the focus will be on PPINs, whose information is located mainly on their topology and the connectivity of the nodes. It is important to note that, at variance with other types of networks, PPINs are incomplete, due to the large number of as yet undetected and uncharacterised PPIs. In fact, they are based on the knowledge within published papers, which is biased towards genes/proteins that are more widely studied (i.e., ascertainment bias). Consequently, if a gene/protein is not represented in a PPIN it may either be because it is genuinely not relevant for that network, or because the linking interaction has not been discovered yet. Consequently, not all graph theory principles/definitions apply to PPIN analysis.

1.2.3.3 Annotations and analysis of PPINs

It is useful for the user to have the option to include annotations regarding the edges and nodes within the network of interest. The annotations of edges can show information relating to for instance their detection method, associated parameters, and confidence scores. These can be sourced from each used PPI entry and can be helpful for either highlighting the important interactions or contrary for filtering out the less significant ones. The annotations of nodes can inform, for instance the protein or gene identifier, expression levels, subcellular localization, and

molecular function. There is a multiplicity of databases providing such information, but the most broadly used is the database of the Gene Ontology (GO) project [88], which holds information regarding functions of gene products.

1.2.3.3.1 Gene Ontology project

The GO project was initiated in 1998 as a collaboration between FlyBase, Saccharomyces Genome Database and Mouse Genome Informatics project [110]. Since then, it has grown to include more databases. One of its central aims is to provide a consistent description of gene products. The vocabularies used, called ontologies, are very restricted to specific GO terms, in order to increase the accuracy and consistency of curation. The ontologies are hierarchically organized and cover three biological domains: (i) Cellular Component (i.e., subcellular localization), (ii) Biological Process (i.e., biological goal of a protein's function), and (iii) Molecular Function (i.e., protein activity) [110]. Using data from the GO project in PPIN analyses enables the functional interpretation of the produced network and the informative grouping/clustering of proteins [88].

1.2.3.3.2 Functional annotations

As previously discussed, PPIs are useful as they can provide clues regarding the cellular function of a protein, based on the "guilt-by-association" principle [84]. Therefore, it is essential to understand as much as possible about the function of the proteins of the network in order to gain mechanistic insight into the system being studied. Different databases might classify proteins in different functional categories but the Biological Process domain of the GO project is one of the most widely used for this purpose [88]. Such an annotation of functions to genes/proteins is called functional annotation.

Functional annotation is more informative for the specific set of proteins of interest when it is based on enrichment, which is a statistical assessment. A GO term is enriched when it is more frequent in the sample set in comparison with the reference set, which is usually the annotated human genome [111]. The enrichment based on functional annotations (i.e., GO terms for Biological Processes) is called functional enrichment but it can also be based on other information, such as the involvement of a protein in a pathway, in which case it is called pathway enrichment. Functional enrichment suggests functional specificity and -taken together with the rest of the enriched terms- can provide valuable indications regarding the mechanism of the system that is being studied.

Numerous tools are available for this type of analysis. Some examples are g:Profiler [112], Gene Ontology (which works through PANTHER [113]) and WebGestalt [114]. Their main differences lie in the specificities of the algorithms and statistical analyses [111]. For higher confidence in the

results from such an analysis, it is recommended to use multiple tools and their most recent portal [115].

1.2.4 Future trends

Network analysis can be utilised to face the recent challenge of handling large amounts of data through its combination with machine learning techniques [116]. There are multiple explanations as to what machine learning is but briefly it refers to a computational approach that recognises patterns with increased accuracy [117]. Such methods can facilitate multiple aspects of biological research. Examples include the identification of DNA patterns and binding sites of transcriptional factors in large data sets, the analysis of networks to identify groups of components based on topological connections or similarities in features [118], and the dimensionality reduction after enrichment analysis [103].

1.2.5 Limitations of protein-protein interaction network analysis

The use of PPIs for the analysis of a biological process has some intrinsic drawbacks. Firstly, the human PPI data is still incomplete for most -if not all- proteins. Additionally, the number of detected interactors of each protein is biased, as some proteins are studied more intensively than others (i.e., ascertainment bias). So, these proteins might falsely appear more important for a specific process or to be involved in more processes than others. Furthermore, most of the interactions have been detected in a variety of conditions, cell lines and tissues, which, when not specified in the PPI databases (which unfortunately is the norm), it fails to inform if a set of interactions occurs in the same tissue or cell type, or in the tissue or cell-type of interest, e.g., neurons. Moreover, a number of interactions are claimed to be detected, whereas the selected method detects only co-localization. Therefore, the results of PPI analyses should be interpreted with caution.

An increasing amount of information about PPIs is being published in peer-reviewed literature and collected in various primary databases. However, the unsystematic collection of PPIs in such databases has created a considerable problem when the user aims to combine all the available data concerning a protein of interest. In this regard, the IMEx consortium [52] has aided considerably with a set of guidelines about what information and in what format should be stored in PPI databases. Unfortunately, not all databases abide by the IMEx instructions, hindering the

efforts to merge PPI data from different sources. Secondly, the information in the databases concerning a single interaction might not be complete and is often duplicated across databases, so it is difficult to merge information from different databases in a unique manner. The aforementioned issues create a technical challenge on the use of PPIs, since strategies have to be implemented for converting and merging data sourced from different primary databases. Additionally, a strong quality control must be applied for filtering all the annotations curated in such databases.

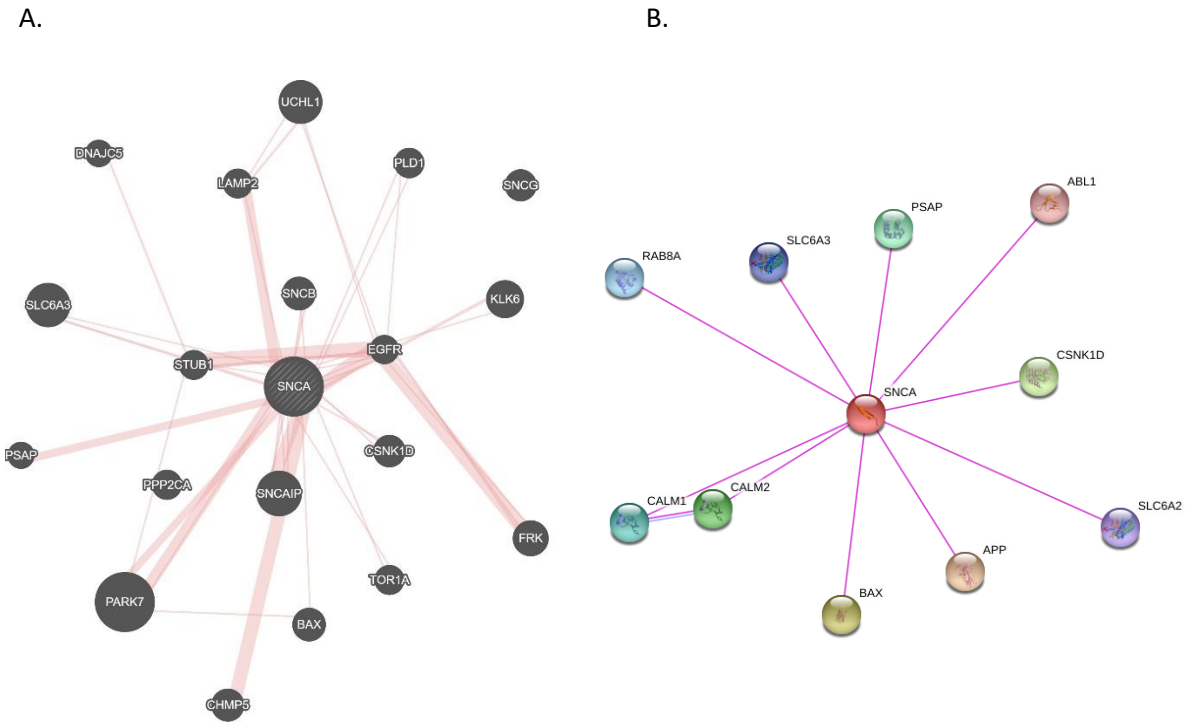


Figure 1-3. Differences in the results of different PPI tools

*The protein-protein interaction networks of the product of SNCA as produced by two tools: GeneMania v3.6.0 (**A**) and STRING v10.5 (**B**).*

There are freely available online tools for extracting PPIs from multiple databases. However, the users must carefully select which tool to use. PPIs from databases that use text mining and PPI prediction are prone to type II errors (i.e., false positive results), so the user must be cautious when interpreting results including such data. Secondly, the criteria for the quality control of the interactions, as well as the type of the data is regularly not clearly stated, hurdling the selection of the appropriate data for each study, e.g., curated data *versus* predicted data. Some of the most widespread used tools, like STRING [119] and GeneMania [120], together with providing a list of unique interactions also score the confidence of each interaction based on a list of criteria.

However, the criteria of confidence are often set and can not be modified by the user. In addition, the list of interactors of the same protein can vary in a great extent among them, in relation to the number and identity of interactors. An example is shown in Fig 1-3, where the interactome of *SNCA* is produced. The differences are noticeable even when the same type of data is selected (curated PPIs), as there were 17 and 10 proteins in the *SNCA* interactome produced by GeneMania [120] and STRING [119], respectively. Interestingly, only 23% of the total SNCA interactors showed an overlap between the two tools. These differences are incremented further when additional proteins are chosen as seeds for the networks.

1.3 Mathematical modelling

Models in biology vary substantially. A cell culture can be a model of cell growth. A *Drosophila melanogaster* strain with a specific mutation can be a model of a human disease. A phylogenetic tree can be a model of the evolution of a group of species. Even though these models are useful, they can be complemented by theoretical models, which can be used to study more conditions and experiments in a faster and cheaper manner, while not being limited by experimental setups.

A mathematical model applied to a biological system is defined an "artificial construct in the language of mathematics that represents a process or phenomenon in biology" [5]. Such models can vary in their features. For instance, they can be small and simple or large and complicated, and mathematically elegant or detailed [5]. Irrespective of their types, mathematical models aim to produce knowledge about a system, predict and extrapolate data, and help researchers formulate new hypotheses and guide them to conduct future experiments.

In systems biology, mathematical modelling is an approach that is frequently used to study complex biological systems. It aims at quantifying and predicting dynamic biological processes in order to decipher the regulatory concepts and to unravel the complexity of the system [121]. Their development is based on a combination of the current understanding of a specific system and of the intuitive understanding of biological systems. It is also based on mathematical theories and techniques and requires the creation of an abstract mathematical representation of the system of interest. Then, the model is solved, either analytically or computationally. By transforming experimental knowledge into a mathematical model, hypotheses can be tested in silico against previously acquired data and the general understanding of the system from the scientific community [122]. These in silico experimentations can lead to predictions, which are testable in the wet lab.

1.3.1 Main components of models

1.3.1.1 Variables and Parameters

Mathematical models comprise a number of different elements. An integral part of a model are the variables, which can represent single entities or groups of entities (e.g., proteins and complexes of proteins). Variables that depend on time are called dependent variables, whilst time is an independent variable. Variables can be interconnected through processes and interactions. Another essential part of the model is its set of parameters, which are constant values (i.e., independent of other variables in the model), such as the rates of a reaction.

1.3.1.2 Ordinary Differential Equations and Law of Mass Action

Ordinary Differential Equations, or else ODEs, are mathematical equations used to describe and predict the variation of variables in time (more details can be found in A1 of Appendix A).

The Law of Mass Action [123, 124] states that the rate of a reaction is proportional to the product of the concentrations of the substrates, and the ODEs for a simple reaction in the form of:

$$X \xrightarrow{k1} Y$$

with k_1, the rate constant that expresses the speed with which X is being converted to Y,

are the following:

$$\frac{dY}{dt} = k_1 X, and\ \frac{dX}{dt} = -k_1 X$$

These equations can be solved analytically to give:

$$X(t) = X_0 e^{-k_1 t}, \text{ and } Y(t) = (1 - e^{k_1 t})\, X_0,$$

with e representing the exponential function.

In biology not all enzymatic reactions are that simple to only have one substrate and one product. More information on such cases can be found in A2 Section of Appendix A.

In this Thesis the Law of Mass Action will be used to formulate a system of ODEs to model a set of reactions.

1.3.2 The mathematical modelling process

The first main step of mathematical modelling is to identify the scientific question the model will attempt to answer, which will help in making any assumptions. According to the goal of the model, a list of the main components of the system and an abstract representation of it can be drawn. In this step, the availability of data regarding the process and entities of the model needs to be assessed. Then, based on both the aim of the model and the quantity and quality of available data, the type of model is selected and drafted. Mathematics is then applied to formulate a mathematical model which can then be solved either analytically or computationally. Dependent upon models and how will they compare with known knowledge, further revisions of the model may be required. The final version of the model can be analysed to test hypotheses and can lead to the formation of new hypotheses, eliciting new understanding and providing guidance for future experiments [5]. In this section the steps in formulating, informing a model with data (parameterising), solving, testing and revising it are outlined.

1.3.2.1 Identifying the purpose of the model & its details

The first step of modelling is deciding the purpose of the model and its potential applications, for both of which simplicity and accuracy are key elements. These features help in identifying the central elements in the context of the questions(s) being asked about the system, which the modelling will attempt to answer [5]. For instance, the inclusion of time or space in the model can be necessary depending on the scientific question, but can also increase the complexity of the model and limit its potential. Thus, a precise question to be answered via the model and critical appraisal of what components are vital for the model are both necessary.

The data collection step is also crucial. The more data, knowledge, and expertise is utilised for the formulation of the model and its assumptions, the more accurate and robust will be the resulting model predictions. However, some data may not be available, or of good enough quality. If the missing/low quality data were essential for the model, further adaptations in the modelling approach might be required.

1.3.2.2 Selecting the type of model & designing the model

The type of modelling approach to be applied depends on a multitude of features. Whether time is included in the model affects if the model is static or dynamic. A static model can analyse how two variables affect each other, or how components in a large network are connected excluding any

variations in time. On the other hand, dynamic mathematical models have the ability to describe how the system changes with time. Space is another factor that can be included in the modelling process. A model that does not take space into account, is spatially homogeneous, otherwise, it is spatially heterogeneous (for instance a model describing the diffusion of a protein in a cell in both space and time). Lastly, a model can be either open or closed. The former is true when a system receives an external stimulus or has an output, whereas in the latter such external components are absent.

After deciding the type of the modelling approach that is going to be applied, the entities of the model need to be identified, together with any variables, parameters, and relationships, which can affect the model. The values of the variables and parameters (e.g., rates of reactions and concentrations of proteins) are informative of the properties of the system, so it is essential to obtain them from the literature in the form of data and knowledge or by performing new experiments. The relationships of the entities included in the model can be expressed with simple reactions, based on which the equations (ODEs) describing the system attributes varying in time and/or space and their respective concentrations can be formulated in the next step, using the Law of Mass Action.

1.3.2.3 Solving and testing the model

In the case of dynamic models, the next step is to solve the ODEs of a model and then to study how the variables vary with time. There are two classes of methods for solving a model: analytical solutions that provide explicit solutions and can be calculated with pen and paper, or computer simulations. Since mathematical models are described with equations, it could be expected that an analytical solution is always possible, and lead to the questioning of why a computational solution is needed. A paradigm demonstrating this necessity is located in A3 Section of Appendix A.

As discussed before, ODEs can be formulated using the law of Mass Action, which requires the estimation of kinetic parameters based on quantitative data [125]. A computational solution to the Mass Action ODEs can be obtained with a variety of software, such as tools from MATLAB. Biological mathematical models using such software, have been developed for signalling pathways, such as MAPK [126] and NF-κB [127], providing valuable new insights. In the former, modelling highlighted the cell-autonomous effect of the spatial profile of calcium signalling in the activation of MAPK in epidermis, while in the latter, evidence was produced to support that a regulatory protein named β-TrCP could regulate the transcriptional activity of NF-κB.

Analysing a model is conducted in two ways. Prior to any detailed mathematical analysis, it is informative to solve the model computationally and detect any evident mistakes and drawbacks by comparing the results to either data sourced from the specific or similar system, and/or to the general understanding of the system based on its biology (i.e., external consistency). An example is to test if any concentration values are negative or change beyond an expected range. In addition, the internal consistency of the model can be investigated, which can demonstrate if the components of the model follow the assumptions of the model. If loss of mass is not accounted for in the model, then there should be a mass conservation throughout the experiment. The timing of relative changes of components also needs to make sense biologically, based on the order of events. If there is a serial activation of A, B and C, then the quantity of activated B should precede that of C. Following any adaptations and improvements prompted by model testing, further analyses can be undertaken.

1.3.2.4 Revising/Improving the model and interpreting the model

Based on the findings of the previous step, the model might need to be adjusted or reformulated. Understanding the causes of any failures of the model, will be helpful in resolving them. If the results, for instance, fail to align with knowledge and data of the system being modelled, revising the parametrisation might be insightful. If unsuccessful, then the focus might need to be redirected back to the model assumptions and to the understanding of the system.

If on the other hand, there is no need for further revising, then the model can be explored. Part of the process of exploring a model is studying whether the system reaches a steady state. A steady state is a condition in a system in which the system concentrations do not vary in time. In this state, the system can be dynamic but a balance has been achieved, which is common for biological systems. In order to find steady states, the simulation is performed, and the model variables are examined for reaching a constant value in time.

Sensitivity analysis can also be used to further explore a model. It aims to simulate how a chosen model output may change with respect to parameter variation. Typically, a high sensitivity of a mathematical model is indicative of a mistake, however in biology and especially in signalling pathways, it can be common [5]. There are two main types of sensitivity analyses: global and local. In global sensitivity analysis all parameters are varied with respect to one another simultaneously, while in local sensitivity analysis one parameter is varied at a time, whilst all others are held in their initial value. The effect of parameter variation is then measured by any changes in the steady state

value of a specific entity or in other features, such as amplitudes at a certain time point, or variable trajectories. [128]

Following exploration, validation can be performed by simulating scenarios that have not been used in the design of the model. The results of the simulations can be checked against the existing intuition about the process before any new hypothesis is formed. It is important to note here that one common misconception about validating against numeric data is that the closer the prediction to the data, the better it is, which is not always true, as overfitting a model can decrease its ability to produce new knowledge. However, validation is not always the best practise. In the cases in which there is a lot of available data about the system, it can be valuable to evaluate the model using data that has not been used for the building of the model, otherwise it might be more beneficial to use these data for the formation of the model.

The final model can be used to understand the system, answer questions about it, and test hypotheses or inform the creation of new hypotheses. If there is a lack of or limited new valuable results, then the model can be altered or extended to answer related questions. Furthermore, refinements of a model might be beneficial to increase the quality of the model and thus the competence and significance of its predictions and/or extrapolations. It is notable, however, that the quality and usefulness of a model is not proportional to its complexity. In contrast, simple models are sometimes easier to use and adapt to slightly different conditions, allowing the exploration of a wider variety of scientific questions. Of course, for the model to have any predictive power a certain level of complexity is necessary. Therefore, a balance between simplicity and complexity of a model is ideal and that balance is unique to the specific aim of each model. [5]

1.3.3 Advantages and applications in biology

Mathematical modelling approaches present multiple advantages. There is a lack of limitations based on experimental feasibility. For example, in a wet lab experiment, a specific cell line and reagent might be required, whereas any system could theoretically be modelled, as long as there is enough computer memory and processing capacity. In addition, wet lab reagents can be costly, and projects may require years of work. However, the expense of mathematical modelling is limited to the price of the computer setup and software. Furthermore, after a model has been developed thousands of simulations can be performed in a short period of time. Due to these advantages, mathematical models can be used for a proactive screening of future experiments, in order to prioritise and/or optimise them. Interestingly, modelling is an integral part of drug development in

pharmaceutical companies in both preclinical and clinical stage. For example, pharmacokinetic and pharmacodynamic modelling can be performed prior to phase I clinical trials, in which the safety of a drug is assessed together with its pharmacokinetic profile in humans [129]. The increased appreciation for such applications of mathematical modelling is also evident in the formation of related networks, such as the UK Quantitative Systems Pharmacology Network in which there is a collaboration between academia and companies from the pharmaceutical industry [130].

The usefulness of mathematical models can also be demonstrated by their wide use in the study of human diseases. One example is the study of an animal model of rheumatoid arthritis [131]. The model was developed based on rats, aiming to study the progression of the disease and to simulate the effects of the corticosteroid dexamethasone. It included components in different levels of biological systems: corticosteroid dynamics and inter-regulatory effects of tumour necrosis factor (TNF)-α and Interleukin 1β, up to the level of the organism with the inclusion of disease endpoints, such as paw edema and bone density. The same model was later used to simulate the effect of a proposed treatment for rheumatoid arthritis, dexamethasone, which suggested that the potential mechanism of action of the drug was through the inhibition of interleukins 6 and 1β and that lower dosage might be sufficient [132]. Such insights gained through mathematical modelling regarding effective dosage and the mechanism of action of a drug are valuable and can be applied to the design of preclinical as well as clinical trials.

1.3.4 Limitations of mathematical modelling

Even though mathematical modelling can be insightful, there are limitations in this approach. Parametrisation is one of the most time-consuming steps in the process of building a mathematical model. The level of difficulty of obtaining parameter values can vary substantially, as they can be available in the literature, or required to be estimated from published data, or little data can be available and instead there is some intuition on the system, or finally having no data at all. Usually, most mathematical models are on average somewhere in a medium level of difficulty, which is not necessarily bad, because the ultimate goal of the model is to create new insights, which would have been more difficult if all the parameters and relationships were already studied and known.

Specifically for the concentrations of the entities of the model, they can be difficult to be collected. This is not only because the precise system that is being modelled might not have been studied in enough detail yet but also because there might be technical difficulties in calculating them. For instance, it is technically challenging to measure the exact local concentration of a variety of

components in the specific subcellular localisation of interest, especially when that location is a small region of the cytoplasm that is not encapsulated or divided from other regions in a way that makes it possible to separate them from others. This has resulted in some models avoiding using precise concentrations, but relative quantities instead, and their units to be shown as “dimensionless”, or “item”, or simply as “units”.

In addition, some relationships in systems can also be unknown. Understudied processes and biological entities if included in models can limit the ability of obtaining precise values of rates of reactions and concentrations required for modelling. Fortunately, the amount of biological data collected from the scientific community is increasing daily, similarly to the number and level of organisation of databases that store related data.

Caution is also required in the interpretation of results acquired from mathematical models. A model’s abilities are restricted by the formulated assumptions and the purpose of the model. Furthermore, the model needs to be rigorously explored mathematically and biologically before using it to make predictions and extrapolations. Therefore, modelling can be hindered by its requirement of obtaining insight from multiple scientific fields. However, forming and maintaining intradisciplinary collaborations has become easier in the last decades with the further development of technology and expansion of professional networks.

Creating more powerful mathematical models that predict more accurately the behaviour of proteins of a system is a future aim of the field but requires an abundance of qualitative and quantitative data about the system. The more frequent and widespread use of omics experiments could assist towards this goal, thus offering valuable information to the scientific community. Therefore, the pipelines of model analysis and exploration ought to be developed to accommodate for higher complexity without limiting the predicting power or deviating from the scope of the model. Challenging aspects of such multiscale models include the integration of the effect of the environment in the system, of different timescales, and merging of features of the model with different modalities, such as deterministic and stochastic parameters. [5]

1.4 Neurodegenerative diseases

1.4.1 Introduction

One of the most complicated human organs is the brain, with over 80 billion neurons and a similar number of non-neuronal cells [133]. In each cell the biological processes are regulated by approximately 20,000 genes [18, 134]. The complexity of its processes is beyond doubt and several of their aspects remain elusive. For instance, the brain manages to control our every move and store information in a way that is still not understood.

Unfortunately, problems with the function of the brain and nervous system can lead to a variety of symptoms, such as blindness and slow and slurred speech. Despite significant progress in the understanding of the nervous system in recent decades, some central challenges remain, such as the disease mechanism of neurodegenerative diseases. Neurodegenerative diseases, as the name suggests, are defined by the death of neurons, although recent research has revealed important contributions from a variety of cell types in the brain to the disease process [135]. In these complicated diseases, functions of the brain decline in a progressive manner, lowering of the quality of life of people with neurodegenerative diseases, also affecting their family and carers. These diseases are more frequent in more senior people, so they affect a percentage of the population, which is growing together with today's aging population, highlighting the importance of deciphering the disease mechanism(s) [136].

Although there is an enormous effort from scientists all around the globe to uncover the mechanism of neurodegenerative diseases (over 25,700 related publications have been released in 2021, as retrieved from PubMed on 9th May 2022), there are still very limited disease modifying treatments available for a variety of reasons. One of the contributing factors is the delay of detection from the disease onset. For example, the onset can be decades prior to the diagnosis of a person with Alzheimer's disease [137]. The mechanistic complexity of these diseases is also impeding their deciphering, as a variety of genetic and environmental factors have been implicated throughout the years. In addition, the heterogeneity of the clinical picture is large. One disease is not always clearly separate from another. Patients can present a combination of clinical symptoms that belong in two diseases, thus leading to viewing of some groups of diseases in a form of spectra, analogous to the schizoaffective spectrum [138].

1.4.2 Biological Models

Neurodegenerative diseases have been studied using a variety of approaches, with significant advances deriving from studies of genomics [139]. Variants have been identified to be causative for the development of neurodegenerative diseases or increase the risk of disease development. The mode of inheritance of such variants has been discovered to be autosomal, X-linked or mitochondrial, depending on whether they are located in autosomes (i.e., body chromosomes) allosomes (i.e., sex chromosomes), or mitochondrial DNA. Their identification has been accelerated the last decade with the development of technologies that allow high quality sequencing of DNA.

Together with the existence of variants, changes in the proteome can also be indicative of alterations in the function of the nervous system. Comparison of protein levels between a group of people with a disease *versus* of a healthy group, can provide insight into the importance of certain genes and gene products in the development of a disease. Protein expression levels can be studied in a variety of models, such as isolated cells or brain tissues. An example of an approach that can be adopted for this purpose is the use of antibodies for either the semi-quantitative analysis after electrophoretic separation of extracts of proteins, or the quantitative analysis through ELISA (enzyme-linked immunosorbent assay). For proteins for which high quality antibodies have yet to be manufactured, mass spectrometry can be performed, for semi-quantitative and quantitative analysis based on the biophysical properties of cleaved peptides of the protein(s) of interest. Of note, changes in the gene expression level do not always coincide with the changes in the amount of its product, highlighting the complementarity of proteomic and genomic studies [22]. For instance, a protein might be very unstable, and even though more mRNA is being produced, it fails to lead to an equal increase of the protein level.

In parallel to the level of expression of a protein, studying its function might be indicative of the disease mechanism. The function can either be directly measured, e.g., enzymatic assays, or hypothesised through its structure and a correlation with structures of other proteins whose functions are known. The comparison of the structure of a variant, with that of a wild-type protein, can indicate which domain and function could be pharmacologically altered in the former to obstruct the development of disease.

The medium of study of protein function and level of expression can vary substantially from non-human organisms to human organoids. Organisms that have been used in this field of study include *Caenorhabditis elegans* [140], *Drosophila melanogaster* [141], mouse [142], and rat [143]. Human cell cultures are another classic way to study human proteins in the wet-lab, including NT2 and SH-

SY5Y, as well as neuronal cells derived from iPSCs (induced pluripotent stem cells) of people with a disease [144]. Advancements in cell culture methodologies include mixed cultures, which can have more than one cell type to study their interactions, as well as 3D cultures that can be produced with the help of polymerising material that creates a 3D matrix [145]. In addition, cerebral organoids, which are 3D formations that contain various neuronal cell types [146], are being used in neurodegeneration studies [147]. Tissue from brains can also be studied, with for example patch clamp electrophysiology, which -even though it is a challenging technique- it can inform about properties of neuronal cells or more generally of brain regions of people with a disease, compared with healthy people [148].

Each model has its strengths and limitations. Some models are simpler, but further away from human neuronal cells and others closer to human neuronal cells but more complicated, making the extrapolation of conclusions more challenging. So, even though comparing data from multiple models and conditions can be challenging, it can be quite insightful. In this effort to collect, handle and analyse multiple types of data, systems biology and omics approaches can be a useful approach.

1.4.3 Systems biology approaches for neurodegeneration

Genes associated with the development of a neurodegenerative disease have mainly been studied individually in overexpression models, models with knock-in mutations, and knock-out/down models. The approach of studying in parallel multiple genes is rarely used but can be insightful. When *LRRK2* was studied together with other familial Parkinson's disease genes, it indicated commonalities in the molecular mechanism of disease [149-151]. However, these models are suboptimal for studying multiple genes simultaneously.

Fortunately, systems biology can facilitate this endeavour of understanding neurodegenerative diseases. The two main systems biology approaches that will be used in this Thesis are protein-protein interaction networks analysis and mathematical modelling.

Network analysis has the ability to enhance the functional understanding of a disease and potentially its mechanism, based on genomics and proteinomics. A high number of different genes linked with a neurodegenerative disease might lead someone to conclude that a multitude of cellular functions and/or pathways are involved in the disease development. However, it can not be excluded that there are commonalities between the genes. A holistic view of the genetic factors of

a disease offered by their display in a disease network can assist in the identification of common pathways associated with the development of the disease [103, 152]. Even though this approach requires validation, it provides a resource- and cost-effective analysis to inform wet-lab experimental design [103].

Top-down network approaches have been used to study neurodegenerative diseases by a multitude of research groups. For instance, a study adopted this approach to functionally analyse Alzheimer's disease and predict novel candidate disease genes [153]. A more systematic approach was used to develop a pipeline to identify and study disease modules in a project named DIAMOnD [154]. Alternatively, bottom-up approaches can also provide insight into neurodegeneration. Common and unique pathways were identified through enrichment of disease networks built based on familial frontotemporal dementia (FTD) and Parkinson's disease (PD) genes [104, 155]. A similar network approach was used by a different group to study amyotrophic lateral sclerosis (ALS), which aimed at associating genetic mutations to altered protein interaction properties [156]. Prioritisation of genes associated with ALS have also been conducted using network approaches [157, 158].

Protein-protein interaction networks can be enriched with additional information. For instance, clinical data was incorporated in a research project for ataxia genes, which explored whether genes related to comorbidities were present in the disease network and if so, the manner through which they are interconnected [159]. Additionally, gene expression can enrich disease networks. In a study of FTD, groups of highly co-expressed genes were detected, and potential pathways associated with different brain regions were suggested [160]. Incorporating functional annotations to create a hybrid network, e.g., Gene Ontology and pathway annotations, can also be insightful and has been used to highlight potential mechanisms related to Parkinson's disease [161].

Neurodegenerative diseases have also been studied through other systems biology approaches, such as mathematical modelling. Ouzounoglou et al created an in silico model focused on α-synuclein and its effect on dopaminergic neurons to study the development of PD. Simulated α-synuclein dynamics included overexpression and post translational modifications, but focused mainly on oligomerisation and degradation through the ubiquitin proteasome system, chaperone-mediated autophagy and macroautophagy [162]. Interestingly, experimental data was used to first calibrate and then to validate the model.

A similar project was used investigated by Kuznetsov et al who also aimed to study the onset of PD through modelling of the kinetics of the aggregation of α-synuclein [163]. Healthy neurons were simulated, and multiple processes were included in the model, such as production, misfolding,

aggregation, transport, and degradation of α-synuclein. Various factors that could lead to the aggregation of α-synuclein were studied. Interestingly, an increase in the production of α-synuclein was demonstrated to be incapable of leading to its accumulation. In contrast, the system was more sensitive to the efficacy of its degradation through the proteasome and autophagy.

A larger scale modelling approach was adopted to study Alzheimer's disease by Mizuno et al in 2012. They developed AlzPathway, which is a pathway map that includes multiple signalling pathways related to the disease. It required the manual curation of over 100 review articles and included 1347 molecules and 1070 reactions in multiple cell regions and cell types [164]. This model has since been updated with additional molecules and reactions to include new published data [165].

1.4.4 Hereditary spastic paraplegias

The Hereditary Spastic Paraplegias (HSPs) are a group of neurodegenerative diseases that are clinically and genetically heterogeneous. They are characterised by progressive weakness and spasticity of the lower limbs [166] due to degeneration of the upper-motor neurons [167]. The initial symptoms include frequent falls, cramps, stiff legs and abnormal or unstable gait. Clinical features include bilateral spasticity, leg hypertonicity, positive Babinski sign, muscle weakness, hyperreflexia, bladder dysfunction, loss of vibratory sensation in the ankles and *pes cavus*. If they include additional symptoms, then the form of HSPs is no longer pure, but complicated. These symptoms could include cerebellar ataxia, seizures, cognitive or mental impairment, optical atrophy, and peripheral neuropathy [168] (Table 1-1).

There are more than 70 genetic types of HSPs, which is reflected in the differences of their clinical phenotypes [169]. For instance, the age of onset can vary from early childhood to adulthood, the form of the disease can be pure or complicated, as discussed before, and all the modes of inheritance have been observed (Table 1-1). Examples of associated genes are *SPAST* (SPG4), mutations in which account for most autosomal dominant forms [170], and *SPG11* that when mutated leads to most autosomal recessive cases [171]. This disease affects 0.2-9.6/100,000 people, depending on the mode of inheritance and geographical area [172, 173]. However, only symptomatic treatment can be provided to the patients, as the underlying disease mechanism(s) are still unclear.

Table 1-1. Sources of variation in the clinical phenotype of patients with HSPs	
Age of onset	Early childhood – late adulthood
Form	Pure and complicated
Modes of inheritance	Autosomal recessive, autosomal dominant, X-linked, mitochondrial, and unknown
Usual Symptoms	Bilateral spasticity and weakness of the lower body, leg hypertonicity, positive Babinski sign, muscle weakness, hyperreflexia, bladder dysfunction, loss of vibration sensation in the ankles, and *pes cavus*
Additional symptoms present in some complicated forms	Cerebellar ataxia, seizures, cognitive or mental impairment, cataracts, retinal alteration, optical atrophy, peripheral neuropathy, dystonia, and parkinsonism

1.4.5 Parkinson's disease

1.4.5.1 Historical overview

Parkinson's disease (PD) is named after James Parkinson, who described its symptoms as a neurological disorder in 1817 [174]. Previous short descriptions of PD-like presentations (i.e., parkinsonism) can be found in numerous ancient texts from around the world [175]. Symptoms consistent with parkinsonism were also reported by Zihe Zhang in 1228 [176], Yikvi Sun in 1584 [176] and Ferenc Papai Pariz in 1690 [177]. In 1899 Edouard Brissaud suggested that substantia nigra could be related to the disease [178], which was followed by descriptions of inclusions in that location by Frederick Lewy and Konstantin Tretiakoff [179, 180]. The link of PD with dopamine was described by Oleh Hornykiewicz [181], which led to L-dopa (i.e., levodopa) being identified as an anti-PD drug in 1961 [182]. Additional associated molecules were identified in the following years with parkinsonian effect, such as MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) [183, 184], or anti-parkinsonian effect, like bromocriptine and apomorphine [185, 186]. Other milestones in the history of PD include the identification of its first genetic cause, α-synuclein, in 1997 by Polymeropoulos [187], and the first double-blind clinical trial of a cell-based therapy in 2001 [188].

1.4.5.2 Clinical presentations

The incidence of PD increases with age, with a mean age of onset at 60 years (late onset PD). However, in some rare occasions it can also manifest in younger people (under 40 years young), and in this case it is designated as young onset PD [189].

Currently there is lack of diagnostic biomarkers for PD. Therefore, diagnosing PD is mainly based on the clinical presentation in each person. Detailed inclusion and exclusion criteria have been described to help with this assessment, such as the UK Parkinson's disease Society Brain Bank criteria [190]. For instance, supportive criteria include unilateral onset and rest tremor present, while exclusion criteria include MPTP exposure. Patient history and their response to therapy can also be helpful in the diagnosis.

The main motor symptoms of Parkinson's disease are tremor, bradykinesia and rigidity [191]. Tremor is the first presenting symptom in most people, but it might be absent in others. This involuntary rhythmic oscillatory movement is more prominent when the person is distracted or at rest. The term bradykinesia comes from the greek words "βραδύς" and "κίνηση", meaning slow and movement, respectively. Its existence is necessary for the positive diagnosis of PD. Bradykinesia might manifest in multiple ways, e.g., progressive decreasing of handwriting, and reduced range of facial expressions. Rigidity is expressed as an increase in the tone of the person's muscles, leading their limbs to feel stiff, which can sometimes be accompanied by cramping. Additional motor symptoms can include poor balance and gait instability. Usually, people with PD make small steps, their feet can feel frozen to the floor, and there is limited movement of their arms. Typically, motor symptoms initially manifest unilaterally and then progress to the other side of the body, continuously in an asymmetric manner. [135]

Even though PD is classically perceived as a motor disease, people with PD have a combination of motor and non-motor symptoms with both categories affecting the quality of people's lives. Interestingly, non-motor symptoms can be present prior to motor symptoms and thus might be used in the future to predict the onset of PD. [192] Examples of non-motor symptoms are neuropsychiatric, autonomic, related to sleep, pain, and anosmia. Neuropsychiatric symptoms include depression, anxiety, and apathy, which can be present in the premotor stage of the disease, as well as cognitive decline and dementia, which tend to manifest in later stages. Dysfunction of the autonomic nervous system can lead to constipation, excessive sweating, erectile dysfunction, and increased saliva production [193]. People's sleep can also be disturbed in the pre-motor stage of PD, for instance, through the disorder of the rapid eye movement (REM) stage of sleep [194]. Frequently, people with PD are affected by pain, for example in their neck and shoulders [195].

Most people with PD also present anosmia (i.e., loss of smell) typically in the pre-motor stage [196].

PD is a progressive neurodegenerative disease, meaning that most peoples will manifest a worsening of their symptoms over time. This is indicative of the increased cellular death that is occurring in a molecular level.

1.4.5.3 Pathology

Categorical diagnosis of PD relies upon post-mortem examination. This is due to the requirement of two pathological conditions: the loss of neurons in substantia nigra pars compacta, and the presence of Lewy bodies, which need to be confirmed with post-mortem neuropathological examination [190, 197, 198].

The substantia nigra pars compacta is an area in the basal ganglia that is particularly affected in people with PD. The basal ganglia are involved in multiple functions, such as initiation and timing of movement, action selection, decision making, planning and learning [199], but have also been linked to reward [200], drug addiction, memory, and psychopathology [201]. The substantia nigra is a dopaminergic nucleus of the basal ganglia, which is important for their function and thus is involved in the motor impairments caused by PD. This nucleus is producing dopamine, therefore its malfunction and consequent decrease of dopamine that is observed in PD [202, 203] can affect multiple dopamine-associated functions, such as movement control, and emotional limbic activity [204]. This brain region is visible with the naked eye in tissue sections, as the cell bodies of the dopaminergic neurons that it contains are pigmented with melanin. In contrast, it might not be visible in the brain of people with PD due to the death of dopaminergic neurons. It is noteworthy that alongside neuronal death, other cell types might also be involved in the development of the disease, such as astrocytes and microglia [135].

The second requirement for PD diagnosis is the presence of Lewy bodies, which are intracellular protein aggregates that have formed in homocentric circular structures [197, 205]. One of their main components is α-synuclein, which is encoded by *SNCA,* the first gene to be associated with PD [187, 206]. Overexpression of and mutations in this gene increase protein aggregation and neurodegeneration. Although there is a clear causative link between *SNCA* and PD, whether the formation of the Lewy bodies itself can lead to the development of PD is unclear. Evidence from another neurodegenerative disease, Huntington's disease, support the notion that protein aggregates might be protective [207]. On the other hand, a post-mortem examination of individuals who had undergone cell replacement therapy for PD, revealed that α-synuclein

aggregation can follow prion-like process, spreading the existence of Lewy bodies in healthy neurons [208-210]. Additional evidence has shown that the spread of the pathology in the body of people with PD follows stereotypical patterns that can be used to stage the disease, as they correlate with the clinical presentations (aka Braak stages) [211]. The accuracy of those patterns is debated [212, 213], but the suggestion that PD has a prion-like behaviour is nevertheless intriguing [135].

1.4.5.4 Molecular mechanisms

Historically, PD was considered a nongenetic neurodegenerative disease, however genetic analyses revealed multiple genetic forms. The first gene to be linked to PD was *SNCA* (encoding α-synuclein), with coding mutations causing a higher tendency for α-synuclein aggregation [187]. Since then, more mutations have been causally linked to PD, such as in *PRKN* [214] and *LRRK2* [215, 216]. Overall, the identification of a mutation as the cause of PD is quite rare (<5%). Interestingly, there is a variation of this frequency between countries, with some geographic areas reaching up to 40%, with particular enrichment for mutations in *LRRK2* and *GBA1* [217]. In parallel to causative mutations, there are variants that increase the risk of developing PD, such as variants of *LRRK2*, and *GAK* [218]. Altogether, these genetic data demonstrate some functional commonalities, suggesting certain pathways to be involved in the development of the disease, such as proteostasis (including protein aggregation) [219, 220], mitochondrial health [214, 221, 222], and inflammation [223].

Alongside genetic factors, environmental factors that cause nigral degeneration and parkinsonism have also been identified. The more deeply understood is MPTP, which is the precursor of MPP+, a potent neurotoxin. The toxin is produced in glial cells and enters and accumulates in the dopaminergic neurons of substantia nigra pars compacta through the dopamine transporter. There, it expresses its toxicity by inhibiting complex I, a protein complex of the respiratory chain in the mitochondria [135]. Multiple animal models of PD have been based on this effect of MPTP, allowing the deeper understanding of the link between mitochondrial health and PD [224].

Additional environmental factors have been linked with PD. Epidemiological studies and research using animal models suggest that the pesticide rotenone and the herbicide paraquat can lead to cell death in the substantia nigra pars compacta and a higher risk of developing PD [225, 226]. A more inconclusive case is that of postencephalitic parkinsonism, which is developed after infection with the H1N1 influenza virus and is responsive to levodopa [227].

1.4.5.5 Treatments and therapies

There is a critical need for disease modifying therapies in PD. It is the second most common neurodegenerative disease [228] affecting more than 10 million people around the globe [229]. As mentioned before, this disease affects mainly older people, and since life expectancy has been increasing, the number of people with PD is predicted to double in the next decades to reach 14 million by 2040 [230].

Currently, there are only symptomatic treatments available for PD. One of the approaches for managing PD is the administration of drugs that replace dopamine. Levodopa is the most effective symptomatic oral drug [135]. It can be administrated orally and is able to cross the blood-brain barrier. Therefore, it is able to reach the central nervous system and produce dopamine, as it is its biological precursor [204]. Inhibitors of levodopa's degradation, such as monoamine oxidase inhibitors [231] and Catechol-*O*-methyltransferase inhibitors [232], can be given in parallel to alleviate some of the side effects of its long-term use, including pain and poor balance. Other complications are nausea and gastrointestinal issues but overall levodopa is usually well tolerated [135]. Alternatively, dopamine receptor agonists can be administered, which are able to mimic the action of dopamine. Some examples are apomorphine, ropinirole, and pramipexole, but their duration of action and linked side-effects (e.g., gambling, hypersexuality and compulsive eating [233]) have limited their use [204].

A subgroup of PD patients is eligible for an alternative approach, deep brain stimulation. Electrodes are implanted in brain nuclei of the subcortical region, and are connected to a pacemaker, which is placed in the anterior chest wall. This circuit delivers stimuli to the selected brain region, the identity of which is dependent on the symptoms that are affecting the patient the most. It can long-lastingly alleviate symptoms that normally respond to levodopa, and also tremor. Side-effects of this method include cognitive decline, brain haemorrhage, and infections [135, 234].

The research in this field is continuous and more treatments are being explored. Earlier this year, there are 547 clinical trials in the USA [235] and 39 clinical trials in the UK [236] that are recruiting people to take part (queried on 24/1/2022). Experimental approaches include glucocerebrosidase modulators [237], LRRK2 inhibitors [238], anti-synuclein drugs [239], neuroprotective compounds [240] and cell replacement therapy [241].

1.4.5.6 Systems biology approaches for PD

In the last decades over 60 GWAS have been conducted for PD and have identified over 500 genetic associations [41]. However, the identification of genes associated with a disease through analysing

GWAS data is not straightforward [64]. A holistic approach that combined genetic, proteinic and functional data, led to the prioritization of candidate genes for PD, using protein networks and pathway enrichment [155]. Another study that used a multi omics approach to prioritize candidate genes for PD, combined cell-type specific epigenomic variant annotations, GWAS data and functional annotations [242]. Such systems biology approaches can also suggest the mechanism by which a variant can lead to disease. Indeed, a research project used cell type specific multi-omic data for individual PD GWAS loci (epigenomics and interactomics) and suggested that a PD-linked single nucleotide polymorphism could lead to an altered transcriptional regulation of another PD-linked gene, *STAB1*. Interestingly, this functional connection was indicated to be present only in microglia [243].

It is evident that for system biology approaches collaboration between different research groups and sharing of information are key. A large-scale collaboration among several countries has yielded in the formation of the International Parkinson's Disease Genomics Consortium [244]. In its website several resources are available, including whole exome sequencing data, GWAS data and a genotyping platform. It has more than 180 members and continues to contribute to the understanding of PD by producing research, which has been published in more than 70 articles [244]. An additional international collaborative project focused on PD, is FOUNDIN-PD. Its aim is to create a multi-omic data set by using inducible pluripotent stem cells obtained from people with PD [60]. Such endeavours are creating hope that we are not far away from understanding the mechanism of PD and thus from developing a disease-modifying therapy. A process that is involved with PD, along with multiple other neurodegenerative diseases is macroautophagy, which will be discussed in more detail in the following section [245, 246].

1.5 Macroautophagy

1.5.1 Overview of (macro)autophagy

Homeostasis is the constitutive goal of cells in order to survive. Various stress conditions and the accumulation of damages/alterations in components of the cell (ranging from proteins to organelles) can activate the process of autophagy. This process was first described by Christian de Duve [247] more than 50 years ago and was named autophagy from the greek words: αυτό (self) and φαγία (eating). The term refers to a set of processes that involve lysosomal degradation of cytoplasmic cell components, such as cytosol, protein complexes, and organelles. It is tightly

regulated by autophagy related genes and pathways, which can be grouped as mTOR (mechanistic target of rapamycin, mTOR) dependent and mTOR independent pathways. Such a tight regulation is essential, as autophagy is central for both cell survival and cell death.

Autophagy occurs constitutively in low basal levels in all eukaryotic cells. In the presence of various stress conditions autophagy's levels increase to quickly provide the cell with energy and materials for biosynthesis [248] or to rapidly remove dangerous macromolecules that need to be degraded [249]. Examples of these stress conditions are nutrient deprivation, growth factor starvation, hypoxia, and inflammation. One the other side, the basal level of autophagy is equally important, for example in post-mitotic cells, e.g., neurons [250, 251], that are unable to dilute accumulating toxic compounds by cell division such as damaged proteins and organelles [252-255] and toxic macromolecules [249]. In addition to its role in ensuring cell survival, autophagy can also lead to cell death, though a specific type, type II programmed cell-death, which differs from apoptosis [256, 257].

1.5.2 Types of autophagy

Autophagy is a broad term describing three distinct processes [258]: microautophagy, chaperone-mediated autophagy, and macroautophagy.

Microautophagy: is the process by which the lysosomal membrane invaginates to include cytoplasmic contents. The lysosomal membrane is randomly invaginated and then differentiated into an autophagic tube that encloses parts of the cytosol. At the top of the tube vesicles are formed, which then fuse homotypically and bud into the lumen [259].

Chaperone-mediated autophagy: is a selective process found mainly in mammalian cells. The substrates are proteins with a specific motif [260], the KFERQ pentapeptide, which is recognised by Hsc70 (heat shock-cognate protein of 70KDa, Hsc70) [261]. The complex of the substrate bound with the chaperone is targeted to the surface of the lysosomes, where it interacts with LAMP2A (lysosome-associated membrane protein 2A, LAMP2A) to then cross the membrane and get degraded by the lysosomal hydrolases [262].

Macroautophagy: is the most studied subtype of autophagy, leading to the terms autophagy and macroautophagy being used interchangeably in the literature. During this process intracellular cargoes are engulfed into double membrane structures, named autophagosomes, which later fuse with lysosomes, leading to the degradation of their contents [263]. Amino acids, sugars, fatty acids,

and nucleotides, which result from the degradation, are subsequently released into the cytoplasm for reuse [264].

This Thesis will focus on macroautophagy, for a set of reasons. This type of autophagy is essential for all cell types. In addition, even though it is the most well-studied type of autophagy, the complexity of the process is substantial because there are multiple proteins and pathways involved, including non-canonical autophagy pathways. Furthermore, its role in pathophysiological responses includes energy metabolism, growth regulation, organelle turnover and ageing [245, 246]. Additionally, it contributes to the physiological embryonic development, the removal of apoptotic cells and organelles, and antigen presentation. It is also involved in human diseases, such as neurodegenerative diseases, cancer and cardiovascular disorders [245, 246], as it protects the cell against starvation, toxins, aggregation-prone proteins, and infectious agents [265].

1.5.3 Main events of macroautophagy

Macroautophagy is a complex and incompletely understood process, in which more than 35 autophagy related genes are involved [266]. Briefly, the process of macroautophagy is usually initiated in response to intracellular stress. This signal leads to the formation of the phagophore (also named isolation membrane), a curved double membrane in the cytosol, which then elongates and fuses, forming the autophagosome, as shown in Fig 1-4. The maturation of the autophagosome includes its translocation close to the nucleus and the subsequent fusion with a lysosome, forming an autophagolysosome, also named autolysosome [266]. The contents of the autophagolysosome are degraded by lysosomal enzymes and the produced building blocks are recycled.

Macroautophagy is tightly regulated, as it might lead to unwanted results if not properly controlled, e.g., cell death. The main regulators of macroautophagy are the mTOR pathway, the JNK (c-Jun N-terminal kinase, JNK) pathway, and the AMPK (AMP (adenosine monophosphate)-activated protein kinase, AMPK) pathway, as well as pathways regulated by Ca^{2+}, cAMP (cyclic-AMP, cAMP), calpains and the IP3R (inositol-trisphosphate receptor, IP3R) [267].

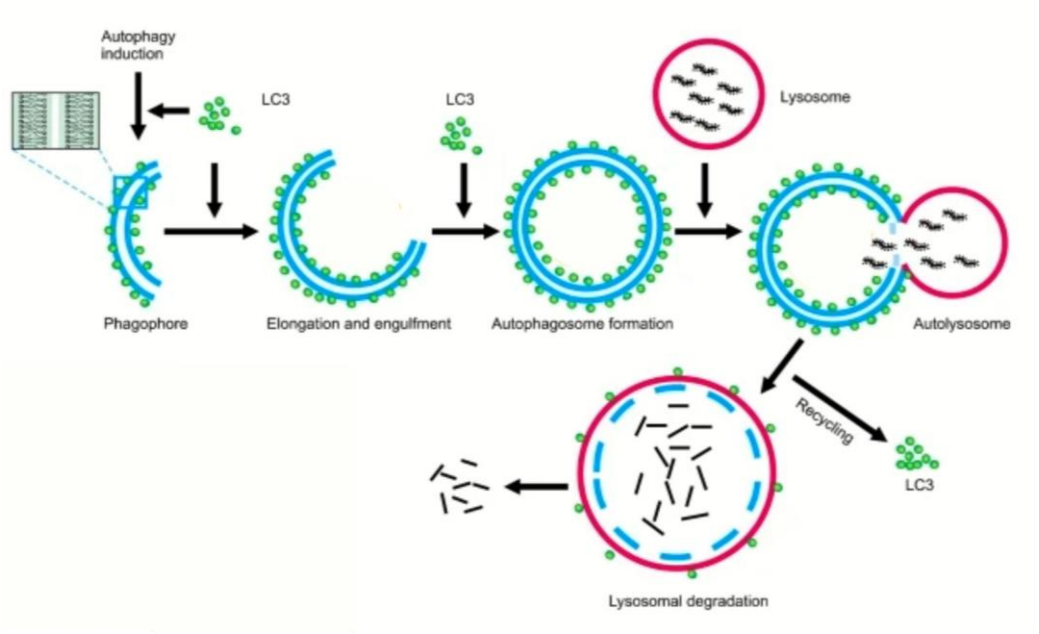


Figure 1-4. Schematic representation of macroautophagy

In macroautophagy, there is initially the formation of the phagophore, then of the, autophagosome and finally of the autolysosome. Adapted from Jing et al [265].

1.5.4 Stages of macroautophagy

1.5.4.1 Initiation and Nucleation of phagophores

1.5.4.1.1 Mechanism

The most upstream complex of macroautophagy is the ULK1 complex [268] that takes part in the production of the phagophore together with the PIK3C3 (Phosphatidylinositol 3-Kinase Catalytic Subunit Type 3, PIK3C3)-containing Beclin-1 complex (also known as PIK3C3-complex 1, or more simply PIK3C3-C1) [269]. The former is composed of ULK1/2 (Unc-51 Like autophagy activating Kinase, ULK), ATG13 (Autophagy-Related Genes, ATG), RB1CC1 (RB1 inducible Coiled-Coil 1, RB1CC1) and ATG101 [270], and the latter of PIK3C3, BECN1 (Beclin1), PIK3R4 (PhosphoInositide-3-Kinase Regulatory subunit 4, PIK3R4) and ATG14 [271, 272]. When an intracellular stress occurs it leads to the inactivation of mTOR that leads to the activation of ULK1 and ULK2 via phosphorylation, and the subsequent formation of the ULK1 complex [273-275]. The ULK1 complex then activates the PIK3C3 complex [276]. These two complexes are translocated in autophagosome nucleation regions, where they stimulate the production of PI3P (PhosphatidylInositol 3-Phosphate, PI3P), which facilitates the elongation of the phagophore and the obtaining of a certain initial curvature [277]. More specifically, PI3P recruits effector proteins required for the formation of the nucleation sites. Examples of proteins recruited by PI3P include DFCP1 (Double FYVE-Containing

Protein 1, DFCP1) and WIPIs (WD-repeat protein Interacting with PhosphoInositides, WIPIs) [266, 278]. The nucleation sites are ER (Endoplasmic Reticulum, ER) associated structures enriched in PI3P, named omegasomes [269].

1.5.4.1.2 Membrane sources

The membrane sources for the nucleation of the phagophore are still unclear. It is hypothesized that parts of the phagophore originates from the ER, MAMs (Mitochondria Associated ER Membranes, MAMs), Golgi, plasma membrane and recycling endosomes [279].

1.5.4.2 Elongation of the phagophore

1.5.4.2.1. Mechanism

The elongation of the phagophore is mainly regulated by two ubiquitin-like conjugation reactions. The first is the conjugation of ATG12 with ATG5 through the E1-like activity of ATG7 and the E2-like activity of ATG10. The conjugated ATG12:ATG5 then interacts with ATG16L1, which is recruited by WIPI2b [280] when the latter detects PI3P [281].

The second important conjugation system is that of LC3I (MAP1LC3 (microtubule-associated protein 1 light chain 3), usually called LC3 or LC3I) [282]. LCE-I gets conjugated with PE (phosphatidylethanolamine, PE) creating the LC3:PE, otherwise known as LCE-II, through a similar process. Here, the E1-like and E2-like activities are expressed by ATG7 and ATG3, respectively, while the ATG12:ATG5:ATG16L1 complex acts as the E3-like enzyme [283, 284]. The high lipophilicity of PE guides the recruitment of LC3II to the autophagosome and its distribution over both sides of the membrane, which is important for the phagophore elongation [285].

LC3II has an extended role in macroautophagy. It acts as a binding receptor that interacts with adaptor proteins involved in selective autophagy, such as p62, which is encoded from *SQSTM1* (Sequestosome-1, *SQSTM1*), and TBC1D25 (TBC1 Domain family member 25, *TBC1D25*) [286]. These are cargo proteins that guide substrates for degradation by directly binding to LC3II while the autophagosome is closing. LC3II possibly has an additional role in the maturation of the phagophore, as it is involved in membrane tethering and hemifusion (i.e., the intermediate stage of membrane fusion, in which the outer membranes have merged, but the inner membranes still remain separate) [287]. Interestingly, the stages of the autophagosome formation can be determined based on the detection of LC3II and ATG12:ATG5:ATG16L1 [288], as shown in Table 1-2. This is because LC3II is produced later in the stage of phagophore formation, but ATG12:ATG5:ATG16L1 is present only in the forming phagophore [289].

Table 1-2. Identification of different stages of phagophores and autophagosomes based on the existence of two conjugates: ATG12:ATG5:ATG16L1, and LC3II.		
Stage	**ATG12:ATG5:ATG16L1**	**LC3II**
Pre-phagophores / early phagophores	+	-
Phagophores	+	+
Mature autophagosomes	-	+

1.5.4.2.2 Membrane sources

Alongside the previously discussed uncertainty of the membrane sources of the phagophore nucleation is the ambiguity for the elongation step. Omegasomes are located near the MAMs, which might indicate that this is a membrane source [290], but other sources might include ER exit sites, ER-Golgi intermediate compartment, Golgi, plasma membrane and recycling endosomes [277].

1.5.4.3 Closure of the phagophore and fusion with endosomes and lysosomes

1.5.4.3.1 Mechanism

The next stage of macroautophagy is the closure of the phagophore and unfortunately this step is poorly understood. The differentiating characteristic of this process compared to other membrane fusion steps is that it is between two narrow tips of phospholipid bilayers [277].

A protein that is regulating the closure of the phagophore is UVRAG (UV radiation resistance associated gene, UVRAG) [291, 292]. It recruits proteins to the phagophore, such as the class-C VPS (vacuolar protein sorting-associated proteins, VPS) that activate Rab7 [292]. The maturation and fusion of autophagosomes also depends on microtubules and dynein in order to be transferred from various locations in the cytoplasm to the microtubule organizing centre [293, 294].

It has been suggested that the ESCRT proteins (Endosomal Sorting Complex Required for Transport, ESCRT) are involved [295]. However, it is unclear whether they enable the phagophore closure and/or the fusion with lysosomes. They might recruit proteins to endosome membranes, such as Rab7 (a small GTPase that recruits more proteins required for fusion) [296, 297] and certain SNAREs (soluble NSF attachment protein receptor, SNARE) [298-302], which also facilitate the fusion with lysosomes [303, 304].

1.5.5 Regulation of macroautophagy

1.5.5.1 Regulation of the initiation of macroautophagy

1.5.5.1.1 mTORC1 dependent regulation

One of the main regulators of macroautophagy is the energy state of the cell. The main complex involved in the regulation of macroautophagy as a consequence of energy status is mTORC1 (mTOR complex 1, mTORC1), which is composed of mTOR, DEPTOR (DEP domain containing mTOR interacting protein, DEPTOR), RAPTOR (Regulatory Associated Protein of mTOR, RAPTOR), AKT1S1 (AKT1 Substrate 1, AKT1S1), mLST8 (mTOR associated protein, LST8 homolog, mLST8) and the Tti1/Telo2 complex (TELO2-Interacting protein 1, Tti1) (Telomere maintenance 2, Telo2) [305]. As an energy biosensor, mTORC1 also regulates the consumption of energy through the proteinosynthesis pathway, e.g., translation and ribosome biogenesis.

More specifically, mTORC1, which is active when the energetic state of the cell is normal or high, inhibits macroautophagy through a direct interaction with the ULK1 complex. When energy is reduced, mTOR is de-activated, leading to a negative regulation of mTORC1 and the activation of the kinase activity of ATG13. ATG13 then activates ULK1/2 and aids the phosphorylation of RB1CC1 by ULK1/2 and so on induces macroautophagy [273-275]. Furthermore, mTORC1 impacts macroautophagy by regulating DAP1 (Death Associated Protein, DAP1), which is a suppressor of macroautophagy [306].

Pathways that activate macroautophagy through decreasing mTORC1 activity are multiple, increasing the complexity of the system. They include those of the limited essential amino acid detection via Rag [307, 308] and of the elevated ratio of AMP/ATP that activates AMPK [309]. Interestingly, AMPK can inhibit mTORC1 via two routes: it phosphorylates RAPTOR [310], and activates TSC1 (Tuberous Sclerosis complex subunit 1)/TSC2 dimer, a negative regulator of mTORC1 [311].

A main protein through which macroautophagy is regulated is BECN1. As noted above, BECN1 is a component of the PIK3C3-containing Beclin-1 complex. It is also an important regulator of the activity of this complex and thus of the initiation of macroautophagy. Some examples of its positive regulators are AMBRA1 (Activating Molecule in BECN1-Regulated Autophagy Protein 1, AMBRA1) [312], UVRAG [313], and SHLB1 [314], and some negative regulators are BCL2, BCL2L1 [315, 316], and the complex of IP3R with BCL2 [317].

An additional protein involved in the regulation of macroautophagy through mTORC1 is MAPK8 (mitogen-activated protein kinase 8, MAPK8) (also known as JNK1). MAPK8 phosphorylates both

BCL2, and BCL2L1 which then unbind from BECN1, leading to the activation of macroautophagy [318]. In parallel, signalling through ULK1 leads to the phosphorylation of AMBRA1 that is in a complex with the BECN1 complex and dynein. This modification releases the rest of the complex from dynein, allowing it to translocate to the autophagy initiation sites [319].

Overall, there are multiple proteins involved in the regulation of macroautophagy through mTORC1 and their pathways are intertwined, increasing the complexity of the system, and highlighting the need of a systems biology approach.

1.5.5.1.2 mTORC1 independent regulation

There are two main regulators of macroautophagy that are mTORC1 independent: AMPK and p53. Interestingly, AMPK can not only modulate macroautophagy in an mTOR-dependent fashion, as it was noted above, but also in an mTOR-independent way: it can directly phosphorylate and activate ULK1 [320].

Notably, p53 has a bidirectional relationship with macroautophagy, as macroautophagy suppresses p53, and p53 regulates macroautophagy [321]. p53 is activated by variety of stressors, including DNA damage, oxidative stress and metabolomic stress, and either facilitates stress adaptation or elimination of cells that are beyond repair [321]. p53 activation induces macroautophagy by directly regulating proteins, such as ULK1, ATG7 [322], Dram (Damage-Regulated Autophagy Modulator) [323], and Isg20L1 [324]. Interestingly, the basal levels of p53 inhibit macroautophagy [325]. The functional association of p53 and macroautophagy is essential for stress responses, metabolism, and cancer.

1.5.5.2 Regulation of phagophore elongation

There is much less information about the regulation of macroautophagy in the stage of phagophore elongation. However, what is known is that starvation regulates macroautophagy, including in this step. It signals through ULK1 for the elongation of phagophore via ATG9. Under starvation conditions, ATG9 translocates from the trans-Golgi and endosomes towards the forming autophagosome [326] and promotes its elongation by delivering vesicles from membrane sources [279].

1.5.6 Non-canonical macroautophagy

The mechanism of macroautophagy described so far consists the canonical pathway. Some of the described steps can be altered and the produced pathways are named non-canonical macroautophagy, or else non-canonical macroautophagy.

The most studied alternative pathway of macroautophagy is that in which BECN1 is absent [327], a protein necessary for the induction of macroautophagy in the canonical pathway. This route of macroautophagy has been observed in multiple studies using cell lines. In a study where Z18 -a compound that binds BCL2 (B cell leukaemia/lymphoma 2, BCL2) and BCL2L1 (BCL2 like 1, BCL2L1) (otherwise known as BCL-XL)- was administrated to HeLa cells, a BECN1-independent autophagosome formation was observed [328]. A similar result was obtained using pro-apoptotic compounds (i.e., staurosporin, MK801 and etoposide) in another study that used primary cortical neurons [329].

Other differences of non-canonical macroautophagy pathways include the absence of requirement of the ULK1 initiation step [330], and of ATG9 and the ATG proteins of the conjugation steps (i.e., ATG5, ATG7 and LC3s) [331]. Therefore, macroautophagy is clearly a biological process of high complexity, which can be better understood using a systems biology approach.

1.6 Aims of Thesis

The overarching goal of this Thesis is to study neurodegenerative diseases (i.e., HSPs, and PD) using holistic approaches to gain insight into the molecular mechanism of disease. This can be broken down to the following aims:

1) Investigate whether any insight into the disease mechanism can be gained for the HSPs, using PPIs.

Since multiple (<70) genes when mutated lead to the same disease, the HSPs, the hypothesis is that there is at least one process or pathway that connects them and is strongly associated with the disease mechanism. Proteins that act in the same process or pathway interact with each other, therefore, building a protein-protein interaction network could highlight the disease mechanism.

This aim will be performed by first identifying the genes that can lead to HSPs, collecting their PPIs, and then building, filtering, and analysing the resultant HSP-PPIN(s). The components and structure

of the HSP-PPIN(s) will provide insight into the mechanistic connectivity between proteins that are encoded by genes that lead to HSPs. Enrichment of associated functions will then be conducted for the components of the network to hint the identity of the biological processes involved in the disease mechanism(s).

2) Explore the potential divergence of disease mechanism(s) among HSPs, based on disease features.

The HSPs showcase a large heterogeneity of clinical presentations, which hint the existence of mechanistic subtypes. Therefore, I hypothesised that supplementing the HSP-PPIN(s) with clinical data for each HSP gene and studying whether certain clinical data are clustered in a part of the network more associated with certain biological processes, could reveal any mechanistic subtypes.

Clinical data will first be incorporated into the HSP-PPIN(s). Parts of the network(s) associated with different modes of inheritance, types of HSPs, or different clinical features will be extracted and analysed through enrichment to detect any differences and similarities on the associated biological processes, and thus the potential subtype disease mechanism(s).

3) Identify connections between macroautophagy and neurodegenerative diseases, using PPINs.

Macroautophagy has been associated with multiple neurodegenerative diseases in which proteins accumulate, which aligns with its biological role in proteostasis. Therefore, it was hypothesised that this association would be expressed by sharing common proteins or common interactors.

The evaluation of the relationship between macroautophagy and each neurodegenerative disease (here Alzheimer's disease, Parkinson's disease, Amyotrophic Lateral Sclerosis, and Frontotemporal dementia) will be performed through creating and overlapping the respective PPINs.

4) Investigate the differential kinetics of the initial stages of macroautophagy in healthy people vs people with PD.

Mathematical modelling is a powerful tool that can create new insight into biological processes. The hypothesis was that simulating macroautophagy in healthy human cells and enriching the model with data regarding how it is affected by a neurodegenerative disease, could provide a mechanistic explanation for their association and therefore highlight therapeutic targets.

For this purpose, a mathematical model of the initial stages of macroautophagy in humans will be created based on the literature, using MATLAB. The simulation will be run for healthy people and for people with PD based on published differential amounts of proteins that are part of the macroautophagy model.

2nd Chapter:

PPIN analysis of HSPs

2. PPIN analysis of HSPs

Main points of this chapter:

- Protein-protein interaction networks were built based on genes that lead to the development of the Hereditary Spastic Paraplegias.

- The protein-protein interactions used to build the networks were collected via PINOT and were experimentally derived human data that were manually curated in primary databases.

- The global and core network created based on the collected interactions revealed that most genes leading to HSPs are interconnected and therefore might be part of the same biological processes.

- Functional enrichment suggested that processes related to protein transport and vesicle dynamics and transport are common themes of the core network and might be related to the disease mechanism.

- The majority of the genes that were suggested to be causing Hereditary Spastic Paraplegias and discovered after the creation of the networks of this Thesis, were part of the networks, supporting the strength of this approach.

2.1 Introduction

Hereditary Spastic Paraplegias (HSPs) is a complicated group of neurodegenerative diseases with a very limited understanding about its mechanistic details. Even though the HSPs were first described 140 years ago [169], the molecular mechanism responsible for disease onset is still unknown. However, few mechanisms have been proposed including alteration of lipid metabolism and endoplasmic reticulum shaping, disruption of mitochondria homeostasis, and dysfunction of intracellular active transport and endolysosomal trafficking [170, 332-334]. All of the above have been suggested based on the multiple functions of each protein associated with HSPs and in vivo and in vitro experiments, whereas no holistic approach solely focused on HSPs has been applied, to the best of my knowledge.

Functional biology classically uses a single-gene approach, where genes are studied in isolation. This approach is powerful, but it allows modelling of only one or very few genes at a time [103]. On the other hand, protein-protein interaction (PPI) networks (PPINs) are a powerful systems biology tool to evaluate the entirety of genes/proteins involved in a disease altogether through a holistic approach. The connections within the PPIN can be mathematically analysed to gain insight into the global relationships among the players of the disease, thus creating an in silico model system to investigate the molecular mechanism associated with those global connections and generate hypotheses to further support functional research and disease modelling.

Network approaches have been previously used to study the HSPs. Some groups focused on the comparison of HSPs to Charcot Marie-Tooth type 2 (a neurodegenerative disease that affects peripheral motor and sensory nerves leading to distal muscle weakness and atrophy) and included predicted PPI data in their analysis [335], or considered that HSPs and Ataxias are two sides of the same spectrum, thus, studying all the implicated genes as one unit [336]. An additional study of HSPs based on whole-exome sequencing of patients, identified new HSP genes based on a network analysis approach [337]. However, the network was constructed based on a list of genes of HSPs and diseases with related phenotype, and it included additional types of interaction not limited to experimentally proven human proteins (e.g., *Drosophila melanogaster* co-citation of proteins). Therefore, this PhD will provide the first study focused on the HSP-PPIN solely based on experimentally proven human PPIs.

2.2 Aims and Objectives

The aim of this chapter is to apply the holistic approach of PPIN analysis to study the HSPs, as well as specific molecular processes involved in this disease to improve our understanding of their underlying molecular mechanism(s), which aids in the identification of potential points of pharmaceutical intervention.

Therefore, the objectives of this chapter are: (i) to create a PPIN using HSP associated genes, and (ii) to analyse it based on network topology and enrichment.

2.3 Methodology

2.3.1 Selection of seeds

In this study, the Mendelian HSP genes are referred to as "seeds" while "nodes" or "interactors" are proteins with which seeds interact. "Edges" are the experimentally demonstrated and manually curated human PPIs that connect the nodes within the PPIN.

The proteins used to build the HSP-PPIN (i.e., seeds) were selected based on their genetic association with HSPs [338]. 66 genes were identified and selected (HSP seeds; Table 2-1). An extended table with the current knowledge on the HSP types and the names of the associated gene and proteins is provided (Table S2-1, Appendix B). Of note, additional 16 seeds, labelled as test-seeds, were included as they have been strongly associated with HSP and/or a mixed phenotype of HSP and other diseases (Table 2-2). The final list of seeds for the HSP-PPIN is composed of 83 seeds (HSP-seeds n=66, test-seeds n=17) and is presented in Table 2-3.

Table 2-1. Hereditary Spastic Paraplegia genes

ALDH18A1, AMPD2, AP4B1, AP4E1, AP4M1, AP4S1, AP5Z1, ARL6IP1, ARSI, ATL1, ATP13A2, B4GALNT1, BSCL2, C12orf65, C19orf12, CAPN1, CPT1C, CYP2U1, CYP7B1, DDHD1, DDHD2, DSTYK, ENTPD1, ERLIN1, ERLIN2, FA2H, FARS2, GBA2, GJC2, HSPD1, IBA57, KIF1A, KIF1C, KIF5A, KLC2, L1CAM, MAG, MARS, NIPA1, NT5C2, PGAP1, PLP1, PNPLA6, RAB3GAP2, REEP1, REEP2, RTN2, SLC16A2, SLC33A1, SPART, SPAST, SPG7, SPG11, SPG21, TECPR2, TFG, TPP1, UBAP1, UCHL1, USP8, VPS37A, WASHC5, WDR48, ZFR, ZFYVE26 and *ZFYVE27*

Table 2-2. Genes related to HSPs

Gene name	**Association with HSP**
ACO2	Homozygous missense mutation associated with complicated HSP [339]. Other associated diseases include Optic atrophy 9 and Infantile cerebellar-retinal degeneration [340, 341].
ALS2	Infantile onset ascending hereditary spastic paraplegia is considered an HSP type [342]. Other associated diseases include amyotrophic lateral sclerosis 2 and Juvenile primary lateral sclerosis [342-345].

Table 2-2. (continued) Genes related to HSPs	
BICD2	Has been proposed to be a rare cause of HSP phenotype [346]. Other associated diseases include Spinal muscular atrophy, lower extremity-predominant, 2B, prenatal onset, autosomal dominant (SMALED2B), and Spinal muscular atrophy, lower extremity-predominant 2A, childhood onset, autosomal dominant (SMALED2A) [347-350].
CCDC50	Reported as a possible HSP gene due to its genetic locus [351].
CCT5	A homozygous missense mutation is considered to lead to a mixed phenotype of HSP and hereditary sensory neuropathy [351, 352].
EXOSC3	Considered to lead to a complicated form of HSP [333, 351, 353]. Other associated diseases include pontocerebellar hypoplasia 1B (PCH1B) [354].
GAD1	A homozygous nonsense mutation is considered by some to lead to autosomal recessive HSP [351, 355] Other associated diseases include Cerebral palsy, spastic quadriplegic 1 (CPSQ1) [356].
HACE1	Involved in a form of Spastic Paraplegia (i.e., Spastic paraplegia and psychomotor retardation with or without seizures, SPPRS) [357, 358] and by some considered a HSP gene [351, 358].
IFIH1	An heterozygous missense mutation was found to be present in family members with HSP [359]. Other associated diseases include Diabetes mellitus, insulin-dependent, 19 (IDDM19) and Aicardi-Goutieres syndrome 7 (AGS7) [360-362].
KCNA2	A mutation has been found in two unrelated families with HSP and has been considered an HSP gene by some [351, 363]. Other associated diseases include Developmental and epileptic encephalopathy 32 (DEE32) [364, 365].
KIDINS220	It is associated with autosomal dominant HSP with a non-classical presentation, including, nystagmus and obesity [351, 366]. Other associated diseases include Spastic paraplegia, intellectual disability, nystagmus, and obesity (SINO) [367].
LYST	A homozygous missense mutation leads to Chediak-Higashi syndrome with spastic paraplegia [368-371].
MT-ATP6	A homoplasmic mutation was identified in several members of a family to cause a late-onset spastic paraplegia-like disorder [372]. Other associated diseases include Neuropathy, ataxia, and retinitis pigmentosa (NARP), Leigh syndrome (LS), Leber hereditary optic neuropathy (LHON), and Ataxia and polyneuropathy, adult-onset (APAO) [373-381].
MT-CO3	A frameshift mutation leads to a childhood onset HSP [382]. Other associated diseases include Leber hereditary optic neuropathy (LHON), and Recurrent myoglobinuria mitochondrial (RM-MT) [383, 384].

Table 2-2. (continued) Genes related to HSPs	
MT-ND4	A heterozygous mutation can lead to adult onset HSP [338, 385]. Other associated diseases include Leber hereditary optic neuropathy (LHON) and Leber hereditary optic neuropathy with dystonia (LDYT) [386-390].
RETREG1	An homozygous mutation is considered to lead to a mixed phenotype of HSP and hereditary sensory neuropathy [391]. Other associated diseases include Neuropathy, hereditary sensory and autonomic, 2B (HSAN2B) [392].
SELENOI	An autosomal recessive mutation can lead to a complicated form of HSP named Spastic paraplegia 81, autosomal recessive, SPG81 [393, 394].

Table 2-3. Seeds for the HSP network	
HSP-seeds (n=66)	*ALDH18A1, AMPD2, AP4B1, AP4E1, AP4M1, AP4S1, AP5Z1, ARL6IP1, ARSI, ATL1, ATP13A2, B4GALNT1, BSCL2, C12orf65, C19orf12, CAPN1, CPT1C, CYP2U1, CYP7B1, DDHD1, DDHD2, DSTYK, ENTPD1, ERLIN1, ERLIN2, FA2H, FARS2, GBA2, GJC2, HSPD1, IBA57, KIF1A, KIF1C, KIF5A, KLC2, L1CAM, MAG, MARS, NIPA1, NT5C2, PGAP1, PLP1, PNPLA6, RAB3GAP2, REEP1, REEP2, RTN2, SLC16A2, SLC33A1, SPART, SPAST, SPG7, SPG11, SPG21, TECPR2, TFG, TPP1, UBAP1, UCHL1, USP8, VPS37A, WASHC5, WDR48, ZFR, ZFYVE26,* and *ZFYVE27*
Test-seeds (n=17)	*ACO2, ALS2, BICD2, CCDC50, CCT5, EXOSC3, GAD1, HACE1, IFIH1, KCNA2, KIDINS220, LYST, MT-ATP6, MT-CO3, MT-ND4, RETREG1* and *SELENOI*

2.3.2 Collection of PPIs and creating the HSP-PPINs

The seeds were used as input to run the Protein Interaction Network Online Tool (PINOT) bioinformatic tool [395]. The interactions provided from PINOT were then screened to include PPIs with a final score above 2 (these interactions were detected in at least two publications or using at least two different methods). The filtered interactions were uploaded in the form of a text file in Cytoscape [106], a network visualisation tool, to obtain the global HSP-PPIN, which was then subjected to topological analysis. Each node of the network was scored based on the number of seeds to which it connected. The nodes interacting with at least one seed, named “inter-interactomes hubs (IIHs)”, were used to extract a subnetwork composed of IIHs and the connected seeds. This subnetwork will be referred to as the “core” HSP-PPIN. Of note, from the core network, similarly with previous publications [104], ubiquitin (UBB and UBC) was excluded to remove

potential non-specific interactions, because it interacts with a large number of proteins in order to lead to their degradation. The interactions for the global HSP and core HSP networks were downloaded on the 09/07/2019, PINOT (beta version; freely available at http://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html).

For the 2nd layer network, the seeds were the proteins of the first layer network. The PPIs were obtained through PINOT on 22/11/2019. The PPIs were filtered, and the core of this network was obtained, following the same process, as previously described for the global and core HSP-PPIN. The phrase HSP-PPIN will be referring to the 1st layer unless specified.

2.3.3 Enrichment

The genes of the core network were analysed for functional enrichment (i.e., functions that are more frequent in this set of genes compared with the annotated human genome [111]), using the GO terms for Biological Processes (GO-BPs), which describe the biological goal of a protein's function (see Section 1.2.3.2.1).

The output of the functional enrichment analysis includes a list of enriched GO terms for Biological Processes and their respective enrichment ratios. The enrichment ratio of each term is a measurement of how many more times a specific GO term was present in the set of genes studied compared to the expected number calculated from its frequency in the annotated human genome. An explanation regarding the calculation of expected values is shown in Table 2-4. Briefly, the expected number of genes in the gene set of interest annotated with a specific GO term is calculated based on the frequency of the GO term annotation in the human genome. The enrichment ratio is the ratio of the actual number of genes of interest annotated with a specific GO term to the expected number. So, for example, if the actual number of genes is double the expected, the enrichment ratio will be equal to two.

So, the enrichment ratio and number of expected genes can be calculated using the following formulas:

$$\text{Enrichment Ratio} = \frac{\text{Number of genes with a GO term in the data}}{\text{Number of expected genes with a GO term in the data}} \quad (2\text{-}1)$$

$$\text{Number of expected genes with a GO term in the data} = \frac{\text{Number of genes in the data} * \text{Number of genes annotated with a GO term in the GO database}}{\text{Total number of annotated genes in the GO database}} \quad (2\text{-}2)$$

Table 2-4. Calculation of expected number of genes having a specific GO term		
	Human genome	**Set of genes under study**
Number of genes characterised by a specific GO term	200	α
Total number of annotated genes	20,000	100
Frequency of the GO term	1/100	α/100 => α=1
Note: The number of genes expected to be characterised by a specific GO term (α) in a set of genes (here n=100) is calculated based on the frequency of that term in the human genome that has been annotated. In this case, for the ratio of 200/20,000 to be equal to α/100, α must be equal to one. So, if two genes out of the 100 that are being analysed have been annotated with this GO term, the enrichment ratio will be 2/1=2. GO: Gene Ontology		

Three independent online tools were used for this analysis to reduce any tool-specific bias. More specifically: g:Profiler [112], Gene Ontology using Panther's tool [110, 113, 396] and WebGestalt [114]. These were running with different algorithms, multiple test correction and/or versions of the GO database, summarised in Table 2-5.

Table 2-5. Summary of the main settings in the enrichment analysis tools				
	Version of tool	**Statistical test**	**Multiple- comparison correction**	**Version of GO database used**
g:Profiler	July 2019	Over-representation enrichment analysis	Bonferroni's corrections	11/07/2019
Gene Ontology using Panther's tool	September/ October 2019	Fisher's exact test	Bonferroni's corrections	03/07/2019
WebGestalt	October 2019	Over-representation enrichment analysis	FDR	14/01/2019

The same process was followed for the localisation enrichment that used Cellular Components (GO-CCs).

Pathway enrichment was performed using the online analysis tool of Reactome (v69& v70 in September and December 2019) [397]. The resulted pathways with p value below 0.05 were

retained and filtered further to remove those with 3 or less proteins involved, as these are not relevant to the analysis that aims to find communal pathways across multiple HSP genes.

Text mining was performed at the merged GO-BPs terms from the 3 tools (see Section 2.3.4). The number of terms related to projections, endosomes, microtubules, membranes, vesicles, and axons were counted based on the presence of the key words “projection*”, “endo*”, “microtubu*”, “membrane”, “vesic*” and “axo*”, respectively, in the name of the GO-BP term. The results were manually quality checked to make sure no unrelated term was included. An enrichment analysis followed based on the frequency of the key words in the resulted GO-BP terms compared to their frequency in the in-house dictionary, which included a wide collection of GO terms, based on the formulas (2-1) & (2-2). In addition, a p value was calculated by comparing the frequency of the presence of the key words in each set of GOs with an equal number of but randomly selected GOs. In more detail, the randomization was performed using an in-house R script of the lab that selected GO terms from the in-house dictionary used for grouping (see Section 2.3.4) and repeated this process 100,000 times. The results were plotted to a distribution, and a p-value was calculated, based on the number of standard deviations that separated the distribution mean and the actual number of the GO terms of the network that had the key words, using pnorm() from R.

2.3.4 Grouping of Gene Ontology Terms for BP and CC

The resulted GO terms for Biological Process were grouped by semantic similarity into semantic classes (level 1 grouping) using in-house developed dictionaries. The semantic classes were further clustered into functional blocks (level 2 grouping).

General GO terms (i.e., semantic classes: metabolism and physiology, functional block: general) were reported but not further analysed in this project, as done in previous publications [104]. Such general terms result in all types of different studies, failing to provide insight into the specific functions that are enriched in individual core networks. In contrast, they dilute the importance of the rest of the terms, by decreasing their ratio compared to the total number of resulted terms.

In order to reduce any tool specific bias, the functional or location blocks enriched in more than 1 enrichment tool (g:Profiler, GO and WebGestalt) were retained for further analysis. For the retained blocks, the union of the semantic classes resulting from each individual tool was analysed. The p-value was adjusted accordingly (p-value=0.05/3).

The number of proteins contributing to the enrichment for each semantic class was calculated using the post-filtering results of g:Profiler and WebGestalt. The exclusion of the results from GO

via Panther was decided because it does not include this information in the downloadable output. In addition, semantic classes in which less than 4 genes were involved, were excluded, because the main interest lies in identifying processes common among multiple HSP genes.

2.3.5 Software and Web-applications

The software and web-applications used for this project are presented in Table 2-6.

Table 2-6. Software and web applications

Name	Version	Website
Cytoscape	3.7.1	https://cytoscape.org/
Gene Ontology	(via PANTHER) 14.1	http://geneontology.org/ and http://pantherdb.org/
g:Profiler	e94_eg41_p11_9f195a1	https://biit.cs.ut.ee/gprofiler/gost
PINOT	1.0; beta version	http://www.reading.ac.uk/bioinf/PINOT/PINOT_form.html
R	3.5.1	https://www.r-project.org/
R studio	1.1.463	https://www.rstudio.com/
Reactome's online analysis tool	v69& v70	https://reactome.org/PathwayBrowser/#TOOL=AT
WebGestalt	fcc27621	http://www.webgestalt.org/

2.4 Results

2.4.1 Modelling parameters

PPINs can be considered static mathematical models. From that perspective, some settings and goals of the modelling can be defined as a first step before any further analysis. The aim of the HSP-PPIN analysis is to study whether the genes in which alterations can cause HSPs are interconnected, and if so, whether they have common biological processes. The data were human PPIs collected through PINOT. For some of the seeds PPI data were not available and potential explanations will be discussed in the next section (Section 2.5). Overall, there were a lot of available PPI data compared to mathematical models, so the HSP-PPIN was built as a first layer network, which

includes the seeds and only their direct interactors. The parameters of the network were the features of the interactions, and especially the final score, as produced by PINOT. No variables or universal constants were included as PPINs are static representations of the interactome of the proteins of interest.

2.4.2 Global network

A PPIN was built using the chosen seeds (Table 2-3) and their experimentally detected human PPIs, which were collected through PINOT [395]. It was visualised through Cytoscape and filtered to include the PPIs with final score over 2 (925 out of 2,967 PPIs) to create the global HSP-PPIN. The filtering was conducted to keep the interactions that have been replicated in the literature, either by more than one method or publication, thus creating more confidence in their existence.

The global HSP-PPIN consisted of 814 nodes, connected through 925 edges. It included 57 HSP seeds, 11 test seeds, and 746 direct protein interactors of seeds (first layer interactors). Most of the network components were interconnected forming a main graph, which included the majority of seeds (n=53, 77.9%) and total nodes (n=755, 92.8%). However, 59 nodes (8.2%) formed 14 disconnected smaller graphs, as seen in Fig 2-1. Interestingly, a similar percentage of the two types of seeds were part of the separated nodes: 12 out of the 66 HSP seeds (18.2%), and 3 out of 17 (17.6%) of test seeds. Additional seeds were not part of the global network as there were no PPIs or their PPIs did not pass the filtering process: 9 out of the 66 HSP seeds (13.6%), and 6 out of 17 (35.2%) of test seeds.

After selecting the seeds for the protein network analysis of HSPs, a gene was found to be causative for HSPs, *RNF170* [398]. It was noticed that this gene was present in the global HSP-PPIN, connected directly with *ERLIN2*. This is highlighting the potential of PPIN analysis as a tool to identify and prioritise candidate genes from genetic analysis of HSP patients and showcasing the value of network analysis for human diseases. A more in-depth analysis of the predictive power of PPINs will be explored in a later section (Section 2.4.6).

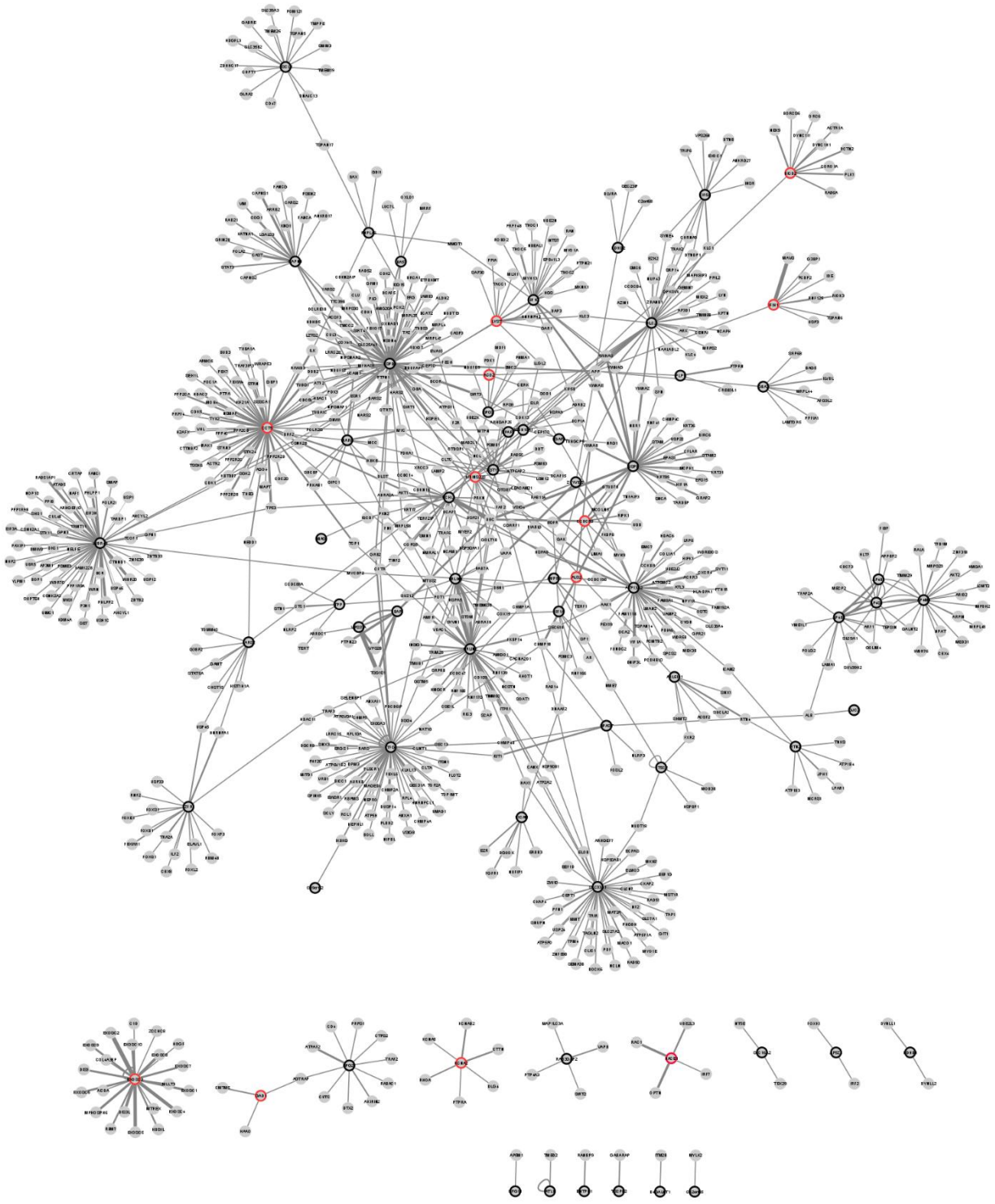


Figure 2-1. The global HSP-PPIN

The HSP global network is the visualisation of all the binary interactions of the seeds that were collected through the online tool PINOT following filtering based on the final score. The nodes corresponding to the HSP-seeds have a black border, while the test-seeds have a red border. The thickness of each edge positively correlates with its final score as calculated by PINOT. The network was visualised using Cytoscape v3.7.2.

2.4.3 Core network

Each interactor of the main graph of the global network was scored based on the number of seeds to which it connected, and the degree distribution was plotted (Fig S2-1, Appendix B). All interactors of at least two seeds (IIHs) (111 out of 746, 14.9%), and the seeds to which they are connected were selected and used to extract the core HSP network (Fig 2-2). The nodes that fail to connect two or more seeds were considered of less importance because one aim of the PPIN analysis was to discover communal links among the seeds in order to explain how mutations in a large number of different genes can all lead to the same disease.

The core network was composed of 164 proteins and 275 edges (Fig 2-2). As expected, there is a bias in retaining seeds *versus* interactors in the network based on the rationale of the filtering process. 68.2% of HSP seeds, 47.1% of test seeds, and 14.9% of direct interactors (45, 8, and 111, respectively) were included in the core network, compared with their initial amounts (66, 17, and 746, respectively). The full list of genes that are part of the core network is shown in Table 2-7.

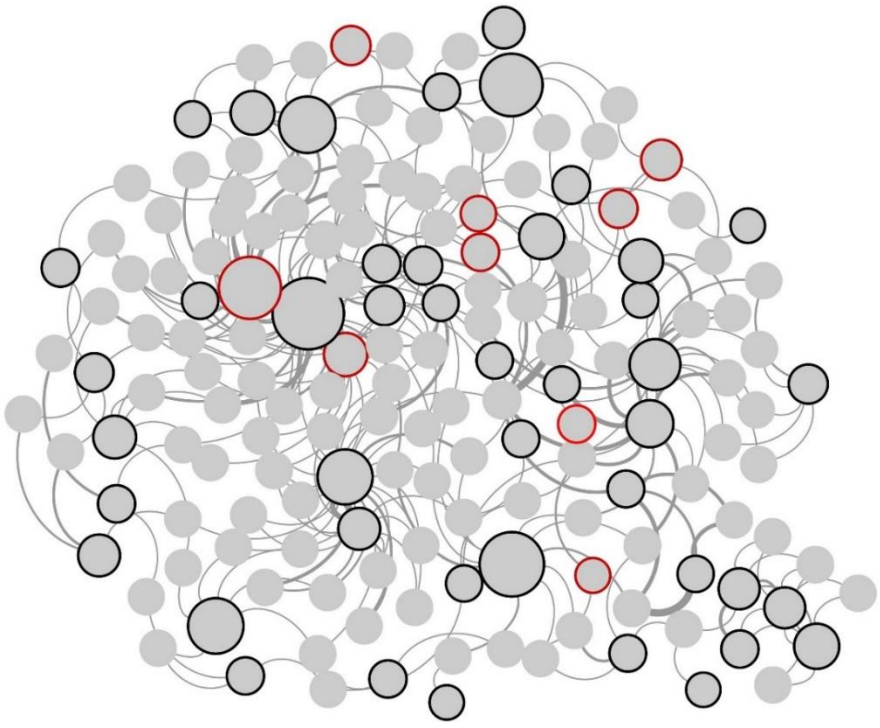


Figure 2-2. The core HSP-PPIN

The global network was filtered based on the connectivity of nodes with the seeds to produce the core network. HSP seeds are the nodes with black border, while the test seeds those with red border. The size of each node positively correlates with its number of connections (i.e., node degree). The thickness of each edge positively correlates with its final score as calculated by PINOT. The network diagram was produced using Cytoscape (version 3.7.1).

Interestingly, *CCDC50*, a test seed included in this study based on its chromosomal location being within the genetic loci of SPG14, was part of the core network, suggesting it could indeed be the responsible gene for this HSP type.

In the core HSP-PPIN the presence of some additional genes that have been previously linked with HSP and its clinical phenotype were investigated. Their identity and link to HSPs are summarised in Table 2-8. None of them were present in the core network. However, one of these genes was present in the global network, *AFG3L2*, which was an interactor of one seed, *GBA2*.

Table 2-7. The genes of the core protein-protein interaction network for the Hereditary Spastic Paraplegias

HSP seeds (n=45, 68.2%)	*ALDH18A1, AMPD2, AP4B1, AP4E1, AP4M1, AP4S1, ARL6IP1, ATP13A2, BSCL2, C19orf12, CAPN1, DDHD2, DSTYK, ERLIN1, ERLIN2, FARS2, GBA2, GJC2, HSPD1, IBA57, KIF1A, KIF1C, KIF5A, KLC2, L1CAM, MAG, MARS, NT5C2, PGAP1, PLP1, PNPLA6, RTN2, SLC33A1, SPART, SPAST, SPG7, TFG, TPP1, UBAP1, UCHL1, USP8, VPS37A, WDR48, ZFR,* and *ZFYVE27*
Test seeds (n= 8, 47.1%)	*ACO2, ALS2, BICD2, CCDC50, CCT5, IFIH1, KIDINS220,* and *LYST*
Interactors (n=111, 14.9%)	*ADRA1D, ALB, AMFR, APP, ATF2, ATP2A2, ATP5F1, CANX, CDC5L, CERK, CFTR, CHMP1A, CHMP1B, CHMP4B, CHRNA9, CLTC, COX15, CSNK2B, DCAF7, DDB2, DIABLO, DLST, DNAAF2, DTNBP1, EGFR, ESR1, EVA1C, F2R, FKBP8, FN1, FXR2, GAK, GALNT2, GOLIM4, GOLT1B, GRB2, HAX1, HDAC1, HDAC11, HNRNPA1, HSP90B1, HSPA5, HSPA8, HSPE1, ICAM2, IKBKE, IKBKG, ILK, IQCB1, ISLR, IST1, KIF5B, KLC1, KLC3, LLGL2, MCOLN3, MMGT1, MRPL58, MTUS2, MYC, MYCBP2, MYEF2, MYH9, NEDD1, NIPSNAP1, NME7, NUDT18, PDHA1, PDK3, PKN2, POLR2G, POT1, PTPN1, RAB11A, RAB14, RAB5C, RAB7A, RFWD3, RTN4, SEC61B, SFN, SIRT3, SNW1, SOD1, STOM, SYNE4, SYVN1, TCP1, TEPSIN, TIMM29, TINF2, TMEM63B, TP53, TRAK2, TSG101, TSPAN17, TUBA1C, TUBG1, UBC, USP45, VAPA, VCAM1, VDAC1, VPS28, XRCC3, YWHAB, YWHAE, YWHAG, YWHAQ, YWHAZ,* and *ZRANB1*

Table 2-8. Additional genes related to HSPs

Gene names	Reason for studying
AFG3L2	It is part of the suggested spectrum of ataxias and spastic paraplegias [336]. It leads to Spinocerebellar ataxia 28 (SCA28) and Spastic ataxia 5, autosomal recessive (SPAX5).
FTL	Neurodegeneration with brain iron accumulation 3 (NBIA3)
FXN	It is part of the suggested spectrum of ataxias and spastic paraplegias [336]. It leads to Friedreich ataxia (FRDA).

Table 2-8. (continued) Additional genes related to HSPs	
MARS2	It is part of the suggested spectrum of ataxias and spastic paraplegias [336]. It leads to Spastic ataxia 2, autosomal recessive (SPAX3), which is otherwise named Autosomal recessive spastic ataxia with leukoencephalopathy (ARSAL).
MTHFR	Methylenetetrahydrofolate reductase deficiency (MTHFRD) can lead to adolescent and adult-onset symptoms include spastic paraparesis. It is one of the conditions that can be mistakenly diagnosed for HSP due to their similarities [338]. Therefore, this gene was included as a negative control.
PLA2G6	It is part of the suggested spectrum of ataxias and spastic paraplegias [336]. It leads to Neurodegeneration with brain iron accumulation 2A and 2B (NBIA2A and NBIA2B), and Parkinson's disease 14 (PARK14).
SACS	It is part of the suggested spectrum of ataxias and spastic paraplegias [336]. It leads to Autosomal recessive spastic ataxia Charlevois-Saguenay (ARSACS).
SETX	It is part of the suggested spectrum of ataxias and spastic paraplegias [336]. It leads to Spinocerebellar ataxia, autosomal recessive 1 (SCAR1) and autosomal dominant amyotrophic lateral sclerosis (ALS4).
SLC2A1	It is part of the suggested spectrum of ataxias and spastic paraplegias [336]. It leads to GLUT1 deficiency syndrome 1 (GLUT1DS1), whose phenotype includes spasticity, seizures, and motor incoordination.
WDR45	Involved in macroautophagy [399, 400], which could be linked to HSPs due to the association of vesicle trafficking as a potential mechanism of disease [170].

2.4.4 Functional enrichment

The core-HSP network represents the most interconnected part of the graph and contains all the protein interactors that are communal to at least 2 seeds. Therefore, it can be used to explore functionalities shared across different HSP genes.

The genes of the core network were therefore analysed to detect functions that were more frequent in the core network, compared to their frequency in the whole human genome (i.e., functional enrichment analysis [111]). There is a variety of available tools for functional enrichment (termed biological processes in these tools) with each resulting in potentially different results depending on the type and specific parameters in the statistical analysis [111]. Thus, three different tools were chosen to reduce the effect of the bias of each individual tool: g:Profiler, Gene Ontology, and WebGestalt.

2.4.4.1 Specific enriched GO terms

Significantly enriched terms resulted via g:Profiler are shown in Fig 2-3. In this Figure, only those with enrichment ratio higher than 2 are depicted for ease of visualisation due to their high number (n=109). The enrichment results via Gene Ontology and WebGestalt were 162 and 108, respectively. The vast number of obtained GO terms hinders the effort of result interpretation, so a part of the specific GO terms had to be selected. Regarding the criterion of this selection, p-values are corrected differently in the different tools, however the enrichment ratio is calculated through the same formula (see Section 2.3.3). Therefore, the top 5 terms were selected from each tool, based on the enrichment ratio (Table 2-9). Of note, all of these GO terms had an enrichment ratio higher than 20. Even though the number, identity and the enrichment ratio of the specific terms obtained from the analysis via the three tools varied, multiple similarities can be observed in the results. Firstly, 80% of the top 5 most enriched terms of Table 2-9 were present in all three tools. Secondly, the identity and order of decreasing enrichment ratio of 2 of these tools were identical (g:Profiler and Gene Ontology). Thirdly, all of the top 5 terms were associated with the same themes: endomembrane system and the transport of proteins, which is indicated to be important for the HSPs through these results. These results were in accordance with the overlap of the totality of GO-BP terms resulted from each tool, as 63.8% of terms resulted from at least 2 tools (Fig S2-2, Appendix B).

Figure 2-3. Enrichment ratio of GO-BP terms as resulted from the analysis of the genes of the core HSP-PPIN, using g:Profiler

The analysis was performed on 19/04/2019 and the results include only the significantly enriched GO-BP terms. The colour-coding corresponds to different level 2 GO groups (see Section 2.4.4.2). GO: Gene Ontology, BP: Biological Process

Table 2-9. The top 5 GO terms of each functional enrichment tool, based on enrichment ratio			
	g:Profiler	**Gene Ontology**	**WebGestalt**
1.	Endoplasmic reticulum tubular network formation	Endoplasmic reticulum tubular network formation	Protein folding in endoplasmic reticulum
2.	Viral budding via host ESCRT complex	Viral budding via host ESCRT complex	Viral budding via host ESCRT complex
3.	Viral budding	Viral budding	Multivesicular body assembly
4.	Multivesicular body assembly	Multivesicular body assembly	Multivesicular body organization
5.	Multivesicular body organization	Multivesicular body organization	Viral budding

2.4.4.2 Grouped enriched GO terms

All significantly enriched GO-BPs were grouped by semantic similarity into semantic classes (level 1 grouping) and semantic classes were further organised into functional blocks (level 2 grouping) thus allowing an easier result interpretation (see Section 2.3.4 and [104, 155, 401, 402]).

GO terms can be of various levels of depth (the more in depth the GO term is, the more specific/detailed it is) and are organised hierarchically. Broad GO terms are not useful in functional enrichment analysis and are being ignored or even removed from the results [104]. This is because the aim of such analysis is to detect functions that could be underlying a disease, thus the more specific the GO term, the more insightful it can be. For this reason, general terms were grouped in the functional block “general”, noted but not analysed any further.

Interestingly, the results of all three tools based on the level 1 and level 2 grouping were similar (Table 2-10) in accordance with the results of the previous section. More specifically, 80% of the level 2 groups of GO terms were identical between all three tools. For the rest 20%, the level 2 groups resulted from 2 out of 3 tools. Regarding the level 1 groups, 40% resulted from all the tools and 64% from at least 2. This demonstrates that even though the extracted specific GO terms were slightly different among the tools, the main findings that are discussed below are supported by all of them. So, from now and on, the results of the three enrichment tools will be merged to increase their coverage. To correct for this merging, only the functional groups that are present in at least 2 tools will be included in the analysis and the p-values will be adjusted accordingly (p=0.05/3).

Table 2-10. The grouping of the semantic classes into functional blocks and their overlap in the three functional enrichment tools

Semantic Class (level 1 grouping)	Overlap among the tools		Functional block (level 2 grouping)
Cell cycle	1/3	3/3	Cell cycle
Cell cycle - segregation/cytokinesis	3/3		
Cell cycle-cytoskeleton	1/3		
Cell death	1/3	2/3	Cell death
Cell death - mitochondria	1/3		
Development - brain	2/3	2/3	Development
Development - neuronal	2/3		
Intracellular organisation	2/3	3/3	Intracellular organisation
Intracellular organisation - cytoskeleton - cell projections	1/3		
Intracellular organisation - membrane	3/3		
Intracellular organisation - organelle	1/3		
Intracellular organisation - organelle - endosome	3/3		
Intracellular organisation - organelle - ER	2/3		
Intracellular organisation - organelle - mitochondria	2/3		
Intracellular organisation - vesicle	3/3		
Physiology - host	3/3	3/3	Physiology - virus
Physiology - virus	3/3		
Protein metabolism	2/3	2/3	Protein metabolism
Protein metabolism - folding	1/3		
Protein localisation	3/3	3/3	Protein localisation
Protein localisation - membrane	1/3		
Protein localisation - mitochondria	1/3		
Response to stimulus - signalling - ERBB	1/3	3/3	Response to stimulus
Response to stimulus - signalling - growth factor	2/3		
Response to stimulus - stress - ER stress	1/3		

Table 2-10. (continued) The grouping of the semantic classes into functional blocks and their overlap in the three functional enrichment tools			
Transport	3/3	3/3	Transport
Transport - intracellular	3/3		
Transport - intracellular - endocytosis	3/3		
Transport - intracellular - mitochondria	1/3		
Transport - intracellular - vesicle	3/3		
Waste disposal - ubiquitin-proteasome - ER	3/3	3/3	Waste disposal

The GO terms of all three tools were associated with a range of themes varying from axonal transport to cell differentiation. The majority of GO terms were associated with the functional blocks “Intracellular organisation” (25.1%), “Transport” (23.6%), “Protein localisation” (15.9%), and “Cell death” (8.9%) (Fig 2-4). These are in accordance with the results obtained after manual evaluation of the top-5 enriched terms and suggesting that these processes have a role in the molecular mechanism of HSPs. A more detailed view of the results with examples of semantic classes can be seen in Fig 2-5.

Interestingly, there were words in the GO-BP terms that were common across functional blocks, such as neuron and axon. Therefore, further analysis was conducted to find common themes across functional blocks and reduce any effect of the organisation of the functional blocks on the results. Text mining was performed for single words within all the GO-BP terms and detected significant enrichment for “axon” (n=17/347, 4.8% [10.4 fold enrichment] p-value=2.72×10^{-19} after 100,000 random simulation), “endosomes” (n=7/347, 2.1% [6.7 fold enrichment] p-value=1.37×10^{-43}), “membrane” (n=55/347, 15.8% [6.4 fold enrichment], p-value=2.48×10^{-65}), “projection” (n=14/347, 4.1% [6.4 fold enrichment], p-value=4.95×10^{-8}), and “vesicles” (n=24/347, 6.8% [5.3 fold enrichment], p-value=4.64×10^{-44}).

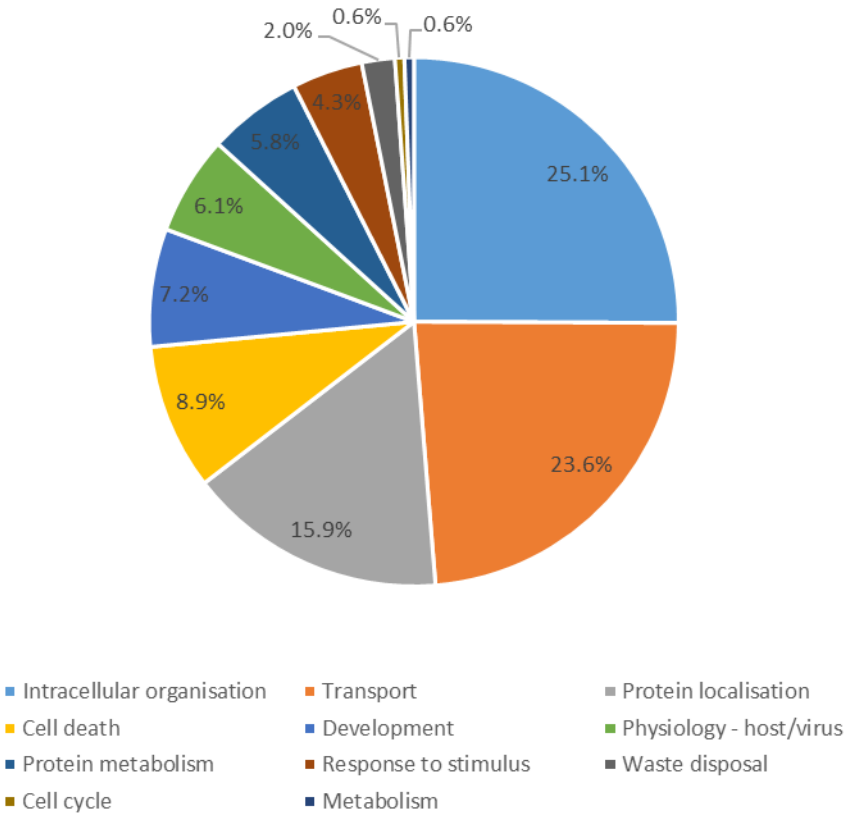


Figure 2-4. Distribution of the number of GO-BP terms resulted from the functional enrichment of the core HSP-PPIN in functional blocks

Functional enrichment was performed for the components of the core HSP-PPIN. The resulted GO-BP terms (n=347) were grouped into semantic classes using in-house R script and then into functional blocks (n=11, excluding "general"). The number of terms of each functional block was calculated from g:Profiler, WebGestalt, and GO (through PANTHER) using the terms of semantic classes that were present in at least two tools.

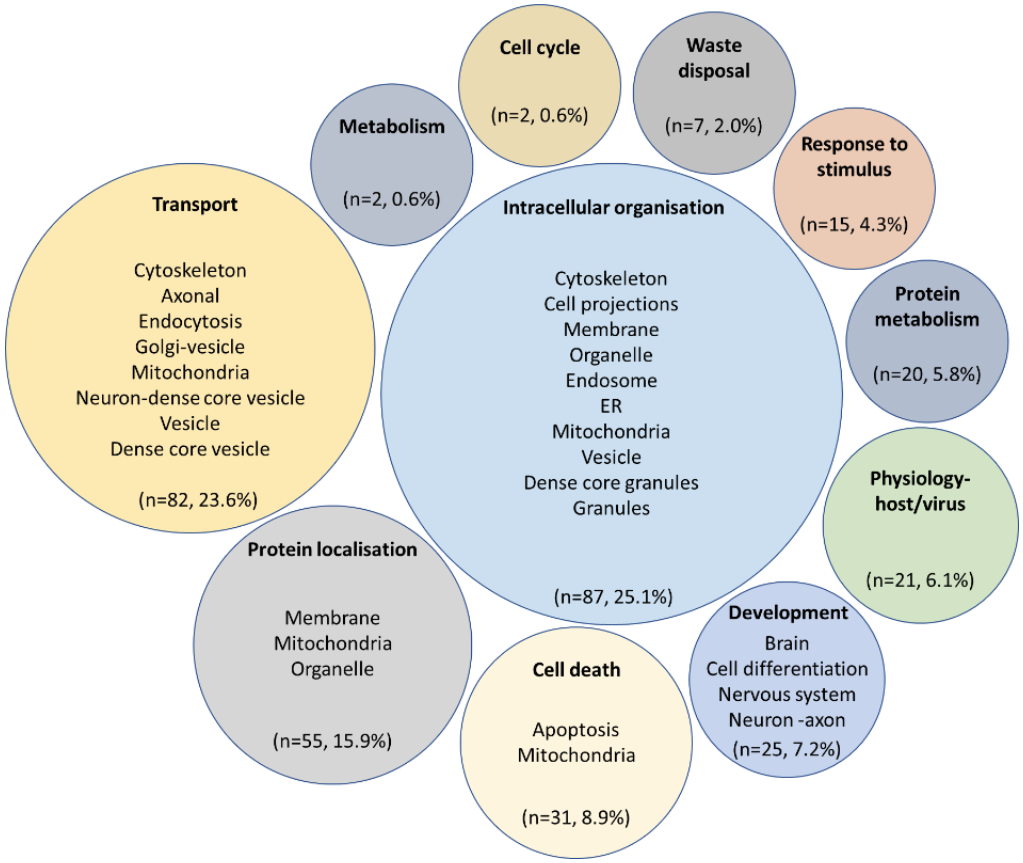


Figure 2-5. Graphical representation of the functional enrichment results of the core HSP-PPIN

Functional enrichment was performed for the components of the core HSP-PPIN. The resulted GO-BP terms (n=347) were grouped into semantic classes (n=54; excluding "metabolism") using in-house R script and then into functional blocks (n=11; excluding "general"). The number of terms of each functional block was calculated from g:Profiler, WebGestalt, and GO (through PANTHER) using the terms of semantic classes that were present in at least two tools. A more detailed version is shown in Fig S2-3 (Appendix B).

Additional information can be extracted by the number of genes of each level 2 category (Table 2-11), as it is important to distinguish if some enriched functions are a property of several or few genes of the network, which can lead to different conclusions. g:Profiler and WebGestalt provide the list of proteins in the submission query that are responsible for each enriched GO-BP. Similarly to previously, the functional groups that resulted from only one tool were excluded and so did the genes that contributed to that enrichment, leaving the total number of enriched genes to 143.

The highest number of proteins belong in the functional block of "Intracellular organisation" (n=125, 88%), followed by "Transport" (n=91, 64%), "Protein metabolism" (n=84, 59%), and "Protein localisation" (n=64, 45%) (Table 2-11). Of note, a minimum of 45% of the proteins of the

HSP-core network contributed to the enrichment of at least one of these groups, underlining their significance in HSPs in accordance with previous results of this project.

Table 2-11. Distribution of genes of the core network in each level 2 Gene Ontology category

Level 2 grouping of GO terms	Number of genes	Percentage of genes
Intracellular organisation	125	87.4%
Transport	91	63.6%
Protein metabolism	84	58.7%
Protein localisation	64	44.8%
Response to stimulus	52	36.4%
Cell death	44	30.8%
Physiology - host/virus	34	23.8%
Waste disposal	7	4.9%

Note: The distribution of enriched genes of the core network (n=143) that belong in each Gene Ontology level 2 category was analysed. The number of genes and their percentage in each category was calculated using the results of g:Profiler and WebGestalt.

The enrichment of each relevant functional block was then mapped on the HSP-core network (Fig 2-6). From this Figure, it is evident that most genes are involved in a variety of functions categorised in multiple functional blocks, while the minority of nodes did not contribute to any enrichment (nodes coloured grey).

The functional blocks of GO terms are visualised in the core HSP-PPIN through the colour of each node for the results of g:Profiler and WebGestalt. The colour-function correspondence is located on the right of the image, while the grey nodes are those who had no enriched functions. The network was visualised using Cytoscape.

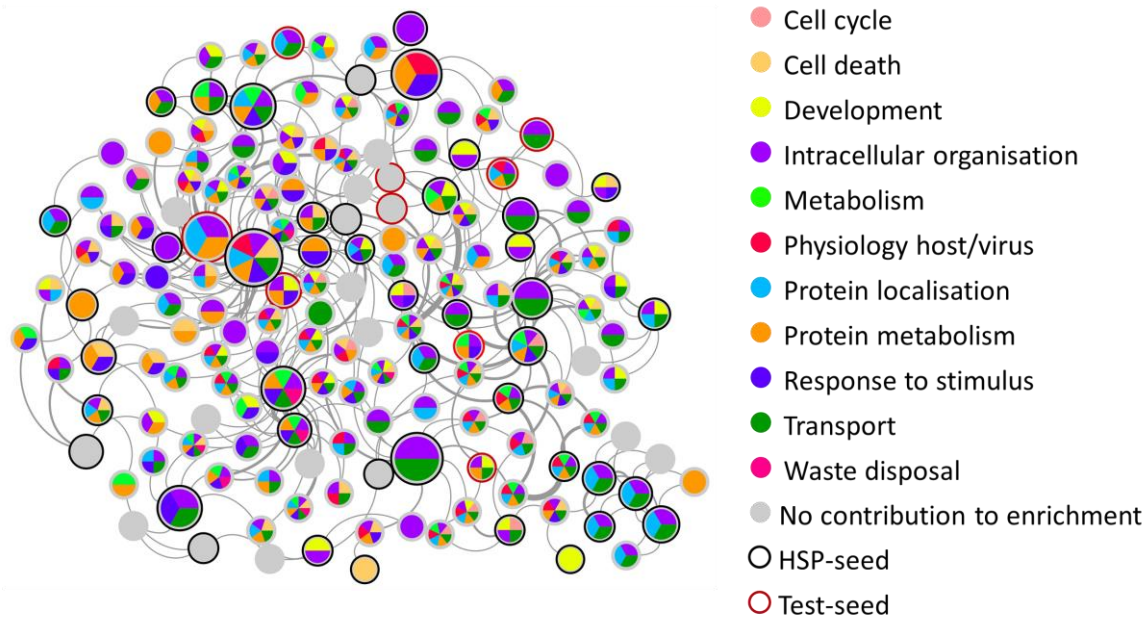


Figure 2-6. Visualisation of the functional enrichment of the core HSP-PPIN

The core network is the interconnected part of the global network that includes the proteins that interact with at least two seeds, and the connected seeds. The nodes corresponding to the HSP-seeds have a black border, while the test-seeds have a red border (ACO2, ALS2, BICD2, CCDC50, CCT5, IFIH1, KIDINS220, LYST). The size of each node positively correlates with its number of connections (i.e., node degree). The thickness of each edge positively correlates with its final score as calculated by PINOT. Adapted from [403].

2.4.5 Pathway enrichment

Interestingly, the independent analysis of the proteins of the core-HSP network through Reactome, suggested enrichment in similar pathways (Fig 2-7). The two most significantly enriched pathways were: vesicle-mediated transport (REA identifier: R-HSA-5653656, p-value: 1.1E-16) and membrane trafficking (REA identifier: R-HSA-199991, p-value: 1.1E-16), with 49 and 47 associated proteins, respectively. These constitute the 29.9% and 28.7% of the total proteins of the network, respectively, and the 37.1% and 35.6% of the enriched proteins of the network, which is indicative of the significance of these pathways in the disease mechanism of HSP.

The pathway enrichment was visualised in the core HSP-PPIN through the colour of each node for the results of Reactome. The colour-function correspondence is located on the top right of the image, while the grey nodes are those who had no enriched functions. The network was visualised using Cytoscape.

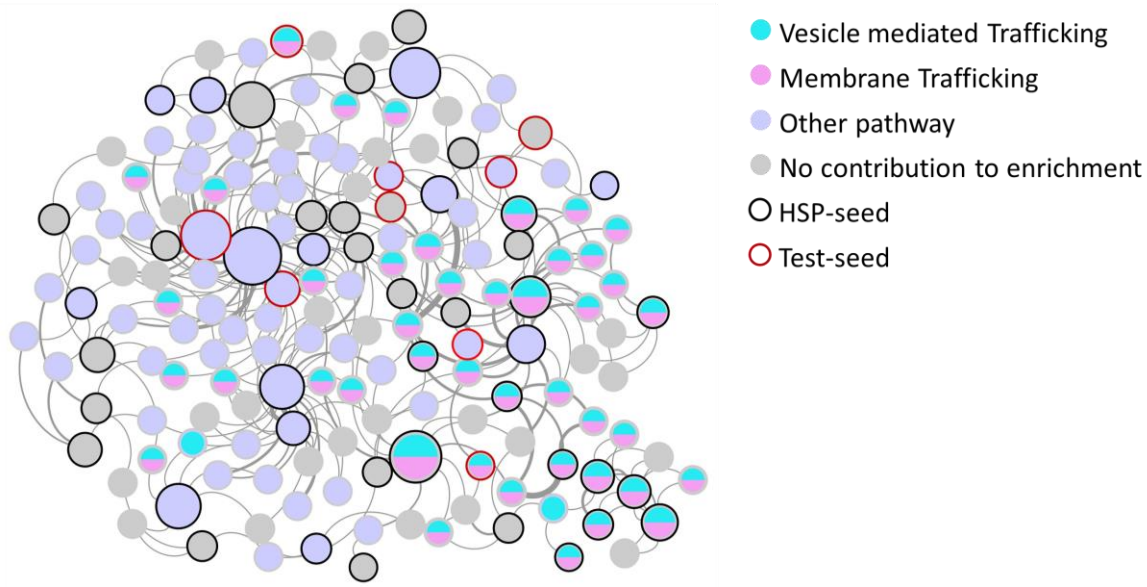


Figure 2-7. Visualisation of the pathway enrichment of the core HSP-PPIN

The core network is the interconnected part of the global network that includes the proteins that interact with at least two seeds, and the connected seeds. The nodes corresponding to the HSP-seeds have a black border, while the test-seeds have a red border (ACO2, ALS2, BICD2, CCDC50, CCT5, IFIH1, KIDINS220, LYST). The size of each node positively correlates with its number of connections (i.e., node degree). The thickness of each edge positively correlates with its final score as calculated by PINOT. Adapted from [403].

2.4.6 Predictive power of HSP-PPIN

In order to test for the presence of more genes that since the selection of the seeds have been causally linked with HSP, the 2nd layer of HSP was created, following the same steps as for the 1st layer. The core of the 2nd layer network consisted of 7,783 proteins and 47,399 interactions.

The new genes reported to be linked to HSPs in 2021, whose presence in the HSP network was tested were: *RNF170* [398], *SPTAN1* [404], *ADAR* [405], *VRK1* [406], *SACS*, *ENTPD1* and *CWF19L1* [407], and *GLRX5* and *ELP2* [408]. The results are presented in Table 2-12. Briefly, *RNF170* was directly linked to 3 HSP seeds, and indirectly to 6 seeds in total. *SPTAN1, ADAR, VRK1, ELP2, GLRX5*, and *SACS* connected with the HSP seeds only indirectly, and *ENTPD1* and *CWF19L1* are not present in the network. Interestingly, even though *SPTAN1, ADAR,* and *VRK1* only connected with HSP seeds through another protein, they were interacting with a high percentage of the core of the 2nd layer HSP-PPIN (15.5-20.7%).

Table 2-12. Exploration of the presence of new HSP genes in the core of the 2nd layer HSP-PPIN

Gene	N_{seeds} direct connection	N_{seeds} indirect connection	$N_{interactors}$ (% of total)	Reference
RNF170	ERLIN1, ERLNI2, and ATP13A2 (16.7%, 3/18 direct interactors)	+ RTN2, SPAST, and TGF (1.1%)	527 (6.8%)	[398]
SPTAN1	0	SLC33A1, KIF5A, CAPN1, ERLIN1, HSPD1, ATP13A2, and WASHC5 (0.4%)	1,608 (20.7%)	[404]
ADAR	0	KIF1C, KIF1A, FARS2, ZFR, KLC2, ALDH18A1, ERLIN2, ERLIN1, HSPD1, and PNPLA6 (0.5%)	1,599 (20.5%)	[405]
VRK1	0	HSPD1, WDR48, IBA57, KIF1C, and TFG (0.4%)	1,209 (15.5%)	[406]
ENTPD1	0	0	0 (0%)	[407]
CWF19L1	0	0	0 (0%)	[407]
ELP2	0	CAPN1, and KLC2 (1.5%)	131 (1.7%)	[408]
GLRX5	0	GBA2, HSPD1, ZFYVE27, KLC2, DDHD2, and PLP1 (0.8%)	738 (0.9%)	[408]
SACS	0	KLC2, AP4E1, and SPAST (1.2%)	257 (3.3%)	[407]

Note: $N_{interactors}$ (col 4) refers to the total number of direct and indirect (through one node) interactors of the protein of interest. The total number of proteins of the core 2nd layer network is 7,783.

2.5 Discussion

In this project the Hereditary Spastic Paraplegias (HSPs) were studied using the holistic approach of network analysis. The HSPs are neurodegenerative diseases with considerable amount of genetic and clinical heterogeneity [169, 170], rendering them suitable to be studied using a protein-protein interaction network approach. This work focused on the study of HSPs, considering HSPs a single disease and not including all the genes associated with a disease spectrum in which HSP is involved (e.g. HSP-ataxia spectrum [409]) or genes with related phenotype, in contrast to prior studies [336, 337]. Another novel feature of this analysis is the exclusion of PPI data that were predicted or identified in species other than *H. sapiens* [335, 337]. This was achieved by using the bioinformatic tool PINOT for the collection of the PPIs of the seeds [395].

PINOT produces a list of experimentally demonstrated and manually curated binary interactions containing unique, human PPI data obtained by merging and processing PPI data from seven databases: BioGrid [410], InnateDB [90], IntAct [89], MBInfo [411], MINT [412], UniProt [413] and bhf-ucl. It gains access to the information that these primary databases hold in real time through PSICQUIC [102] and then applies multiple filters in order to provide the user with a list of interactions for which there is accurate information in these PPI databases. Through PINOT, interactions are also scored taking into consideration the number of publications and different methods used for their detection. Therefore, PPIs accurately curated and characterised by a confidence score were collected for the genes listed in Table 2-3. PPIs that had been detected with at least two different detection methods or published in at least two research papers were then selected, thus decreasing the number of false positive PPI data [395].

2.5.1 Proteins of the global network

In this work the global HSP-PPIN was first created. It included 57 HSP seeds and 11 test seeds. The missing seeds had no PPIs collected through PINOT, or their interactions were filtered out based on the quality control process of the tool.

The global network was constituted of two parts: 1 interconnected graph, and 14 smaller graphs. Most of the seeds that were part of the smaller graphs had a low number of interactors: 1 for 6 seeds (*ATL1, B4GALNT1, C12orf65, ENTPD1, SPG11,* and *TECPR2*; HSP seeds), 2 for 3 (*AP5Z1, DDHD1,* and *SLC16A2;* HSP seeds), 3 for 1 (*GAD1*, test seed), 4 for 2 (*HACE1*, test seed; *RAB3GAP2*, HSP seeds), and 6 for 1 (*KCNA2*, test seed). Such a low number of interactors could explain why these were not connected with the rest of the graph. The average number of seeds' interactors

from the filtered PPIs was 36 and the median 21, which shows that they had fewer chances on being connected with a node that also connects to another seed. However, the HSP seed *SPG21* and the test seed *EXOSC3* had 10, and 21 interactors, respectively, and still were not connected to the rest of the graph. Interestingly, *SPG21* connected through one interactor with another seed (i.e., *GAD1*, test seed) but they failed to connect with the rest of the seeds.

There are three hypotheses for the presence of the 14 small unconnected graphs: (i) either the interactions linking some of the seeds have not been discovered or curated in PPI databases by the time of the analysis (possibly explaining the absence of understudied proteins with limited number of interactors), or (ii) these links are not direct, but instead mediated by two or more interactors, or (iii) they are truly not connected and thus potentially mutations in these genes cause HSPs through a different mechanism/pathway.

Through the exploration of the 2nd layer HSP network, it was unveiled that the second hypothesis was indeed able to explain all the disconnected test seeds (i.e., *GAD1, HACE1,* and *KCNA2*), and most of disconnected HSP seeds (i.e., ATL1, *AP5Z1, B4GALNT1, DDHD1, EXOSC3 RAB3GAP2, SLC16A2, SPG11, SPG21, TECPR2*), including *SPG21* and *EXOSC3.* In fact, the addition of the 2nd layer interactome was able to connect nearly all of the seeds in a unique graph. However, *C12orf65,* and *ENTPD1* were still not part of the 2nd layer network. Therefore, their absence from the global network would need to be explained by one of the three hypotheses mentioned above: the seeds are understudied, 2 or more interactors (in this case more than 2) mediate the connection with other seeds, or they are actually not connected to the other HSP seeds. The 3rd hypothesis would indeed explain the existence of a unique clinical feature for people with mutations in *ENTPD1* (i.e., aggressive behaviour [351]). No unique clinical feature was identified for people with mutations in *C12orf65* to the best of my knowledge, however, it produces a mitochondrial protein, so it could be hypothesised that HSP is developed through a different pathway. This is further supported by the absence of most other mitochondrial proteins from the global network.

2.5.2 Proteins of the core network

The global network was then filtered to include the interconnected part of the network, in which interactors link at least two proteins derived from HSP genes (i.e., seeds). This part of the HSP-PPIN is the core network and its analysis can lead to important observations.

Out of the initially selected 83 seeds, 53 were retained in the core HSP-PPIN (Fig 2-2) and 68 were part of the core network of the 2nd layer HSP-PPIN. This indicates that they are functionally associated and therefore convergent molecular mechanisms can lead to the development of HSPs.

In this study test seeds were included, which are genes that have been controversially linked to HSPs with no general consensus. Eight test seeds were present in the core HSP-PPIN, providing in silico evidence to suggest their importance in the HSP system. The presence of the test-seeds *BICD2, CCT5, KIDINS220, ACO2, LYST* and *IFIH1* in the network is in alignment with the biological processes and cellular components suggested to play a role in HSPs from the present and previous research, namely vesicle-mediated transport, protein folding, cell death, metabolism, and antiviral responses [414-421]. The latter process' link to HSP is supported by the findings of this work but has yet to be strongly associated with neurodegeneration. However, the implication of *IFIH1* in the Aicardi-Goutières syndrome is suggestive, as it is disorder affecting the brain, immune system and skin and its symptoms include spasticity, dystonic posturing, and other neurological dysfunctions [415]. The incorporation of *ALS2* in the core HSP-PPIN is unsurprising, because it is accepted as an HSP gene from many clinicians and researchers [170, 338, 355]. However, there was no previous indication that *CCDC50* is linked to HSPs except based on its chromosomal location. *CCDC50* is in the 3q28 genetic locus, while the locus containing the generic risk association detected for the HSP subtype SPG14 is 3q27-28 [422]. *CCDC50* directly interacts with two proteins that are shared interactors of six HSP-seeds (which is more than 95.5% of the proteins of the global and 74.5% of the proteins of the core HSP-PPIN). This result is an in silico prediction (based on the analysis of the CCDC50 interactome) that alterations in *CCDC50* could be causing the SPG14 HSP type. Therefore, based on the physical association of the proteins derived from these test seeds with other proteins whose mutations lead to HSPs, it is suggested to screen for mutations in these genes for the HSP cases with unidentified genetic cause.

Of note, genes that were later found to be associated with HSPs were present in the HSP network either in the first or second layer. Two genes that were found to be causative for HSPs after the data analysis for the current work, *RNF170* and *SPTAN1* [398, 404], with the former being present in the global HSP network (1st layer) and directly connected to *ERLIN2*. The latter was present in the 2nd layer of the core HSP network and was connected to 11 HSP seeds through one protein. Out of the total 9 proteins associated with HSP after the seed selection for this project, 7 were found either in the 1st or 2nd layer core HSP-PPIN (78%) (Table 2-12). These results are showcasing the clinical and biological relevance of the HSP-PPIN analysis as a tool to identify and prioritise candidate genes from genetic analysis of HSP patients, and to hint about key processes involved in the disease mechanism.

In conclusion, PPIN analysis is a valuable tool that can indicate mechanistic links between different genes implicated in a disease. Furthermore, PPIN may be applied as a tool to prioritise candidate genes for explaining the genetic association between disease and risk loci.

2.5.3 Functional enrichment

Analysing a disease-focused PPIN based on **functional** annotation provides an opportunity to increase our understanding of the underlying mechanism(s) using a holistic approach. Therefore, enrichment was performed for the components of the core HSP network using three independent tools, g:Profiler, Gene Ontology, and WebGestalt for functions, and using Reactome for pathways to unveil commonalities across network components.

These tools resulted in similar but not identical lists of GO terms, Enrichment Ratios, or p-values, even though whenever given the opportunity the same (or the most similar) options among tools were chosen. This is due to three main reasons. Firstly, the tools were using different versions of the GO data (28/12/2018, 02/02/2019, and 01/14/2019, respectively for g:Profiler, Gene Ontology and WebGestalt). Secondly, the statistical tests they were using were slightly different (Overrepresentation enrichment analysis, Fisher's exact test, and Overrepresentation enrichment analysis, respectively). Thirdly, the specific parameters within the statistical tests were not identical (significance threshold: p-value<0.05, p-value<0.05 and FDR<0.05, respectively). All the aforementioned differences are acceptable alternatives to analyse functional enrichment [111] but the specifics of the results could change across tools. Therefore, there was no attempt in this project to compare the enrichment ratios or p-values among the three tools; the identity of the enriched GO terms and the functional classes of GO terms were compared, instead, to evaluate the consistency of the results across enrichment tools. Overall, using more than one tool for functional enrichment -even though it is challenging- is beneficial, as it removes the bias associated with specific choices in a statistical test or its parameters, and increases the confidence on any drawn conclusions [111].

Even though the number of specific terms resulted from the functional enrichment was high (n=347), several common themes arose. After level 2 grouping of the GO terms, the categories with the highest number of enriched GO terms "Intracellular organisation" (25.1%), "Transport" (23.6%), and "Protein localisation" (15.9%), and "Cell death" (8.9%) (Fig 2-4 and 2-5). The groups with the highest number of associated genes were "Intracellular organisation" (87.4%), "Transport" (63.6%), "Protein metabolism" (58.7%) and "Protein localisation" (44.8%) (Table 2-11), which is not identical

but in accordance with the previous results. In addition, most of the specific terms describe functions related to protein transport (e.g., protein targeting), regulation of protein level (e.g., protein metabolic process, ERAD pathway) and vesicle dynamics and transport (e.g., multivesicular body organisation, vesicle organisation, vesicle mediated transport). Of note, the notion that protein transport and vesicle dynamics are in the centre of the functions implicated in the development of HSPs is supported by several previous publications. Particularly, much evidence supports the implication of intracellular active transport, endolysosomal trafficking pathway and ER shaping in HSPs [170, 332-334]. Interestingly, functional publications were not used for the generation of the HSP-PPIN, therefore the conclusions obtained here based on PPIs represent a further validation of some of the published functional analyses.

Intracellular active transport has been strongly associated with HSPs. Three genes have been shown to have such roles (i.e., *KIF5A* [423], *KIF1A* and *KIF1C* [424], involved in HSP type SPG10, SPG30 and SPG58 respectively) and indications exist for numerous others (e.g. *SPAST* [425, 426], *ATL1* [427, 428], *SPART [429]*, and *NIPA1* [430], involved in HSP type SPG4, SPG3, SPG20 and SPG6 respectively). For example, mutations in the former genes all lead to lower affinity for microtubules [423, 424, 431] and more specifically in *KIF5A* this has been shown to lead to a decrease in the speed of microtubule-dependent anterograde axonal transport [423]. Even though only two of the three proteins strongly linked to intracellular active transport and half of the indicated ones were in the core network, the functional enrichment analysis showed transport and intracellular organisation to belong in the most enriched categories in agreement with the aforementioned published data.

Endolysosomal trafficking pathway has recently been reported to be dysfunctional in HSP models. More specifically, a recent study showed that in SPG48 patient fibroblasts there was accumulation of membrane material in endolysosomes, and in a mouse model where *AP5Z1* (i.e., the gene responsible for SPG48) was mutated there was dysfunction in the vesicular-mediated trafficking of cargoes [432]. Additionally, the loss of function of the proteins of the HSP types SPG11 and SPG15 (i.e., spatacsin, and zinc finger FYVE domain-containing protein 26, respectively), leads to an impairment of the autophagic lysosome reformation process [433]. Alternatively, it has been suggested that the lysosomal dysfunctions observed in HSP models could be caused by a decrease in the degradation capacity [170]. Both hypotheses are in accordance with this project's findings, as terms associated with vesicles, endosomes, and transport, as well as proteolysis and protein metabolism were amongst the enriched ones.

ER shaping has been suggested to be a key pathway linked to the aetiology of the HSPs, as numerous proteins related to HSPs have been associated with this process, including *RAB3GAP2, ARL6IP1, REEP2,* and *ATL1* [332, 434] (involved with HSP type SPG69, SPG61, SPG72 and SPG3, respectively). For example, *ATL1*, *SPAST* and *REEP1* encode proteins with hairpin loop domains, which they insert into the tubular ER membrane, contributing to its shaping [170]. In more detail, *ATL1* is involved in the creation of the three-way junctions between ER tubules [435-437], *SPAST* in the linkage of the ER with the microtubules [438] and *REEP1* in the formation and stabilisation of the ER tubular network [439]. Based on the present analysis, the most enriched GO term using g:Profiler (Fig 2-3) was "endoplasmic reticulum tubular network formation", suggesting that this indeed could be a mechanism leading to HSPs. However, this term resulted from the presence of three out of only four proteins with this function, so whereas it could explain its association with some subtypes, it does not seem to explain the underlying mechanism of most HSPs.

Evidence suggests that **lipid metabolism** is implicated in the mechanism of some HSPs [440-442]. This might be true for specific types of HSPs, such as SPG26 (*B4GALNT1*) [443, 444], SPG46 (*GBA2*) [445-447], and SPG5A (*CYP7B1*), which encode enzymes directly involved in the synthesis of the ganglioside [444], the hydrolysis of glucosylceramide [447] and the metabolism of cholesterol [448]. However, this idea was not supported by the functional enrichment analysis of the core network that included most of the genes associated with the disease, as no related GO term was obtained.

The role of **mitochondria** has also been hypothesised to be involved in HSPs [449-453], for example in HSP type SPG13 (*HSPD1*) [449] and SPG7 (*SPG7*) [450], through the disruption of mitochondrial quality control and the axonal accumulation of mitochondria, respectively. Even though some related GO terms resulted from the analysis of this project, most of them are describing functions related to protein transport and membrane organisation, which have been found to be enriched in the rest of the organelles and cell as well. So, there is no evidence from this analysis supporting that mitochondria are particularly important for most HSPs. However, it has to be noted that most mitochondrial-related seeds were not part of the network, as not enough PPIs, or of not good enough quality were detected. Therefore, it can not be excluded that mitochondria could play an important role in the development of HSPs, perhaps through a different mechanism.

Interestingly, the specific GO term **substantia nigra development** was obtained. This is in accordance with the clinical features of HSPs but problems in its development have not previously been suggested to be one related to the disease mechanism of HSPs to the best of my knowledge.

Such a result needs to be further analysed and evaluated using in vivo and in vitro experiments as well as genetic analysis of HSP patients with unidentified causative genes.

There have been multiple reviews suggesting the main disease mechanism(s) of HSPs. The results of this chapter support the involvement of some of these suggestions. Out of the ten mechanisms suggested by Lo Guidice et al [355] those supported by the results of this work were mainly endosome membrane trafficking and vesicle formation, abnormal membrane trafficking and organelle shaping, dysfunction of axonal transport, but also, axon development. Regarding the main HSP mechanisms prioritised by de Souza et al [338], out of the total 5, those in accordance with this work were membrane trafficking and organelle shaping, and axonal transport. Our results are in most agreement with the suggestion from Blackstone [333] that the key biological processes are the organelle shaping and biogenesis and the membrane cargo and trafficking, further supporting the hypothesis that the HSPs could be considered transportopathies [454]. These results highlight the potential of the approach of PPINs coupled with functional enrichment that can identify the most relevant functions among the genes of interest associated with a complex disease, which is essential for discovering disease modifying targets and interventions.

A previous study that analysed HSPs together with CMT2 proposes some different pathways to be involved in the disease mechanism of HSPs (i.e. Epstein-Barr virus infection, Herpes simplex infection and Antigen processing and presentation) [335]. The discrepancy between this study and the results of this project can be explained by differences in the methodology. Specifically, the authors used a list of 95 HSP genes as seeds to build a network, which included several genes that have not been associated with HSPs in the literature. Then the network was expanded using the DIAMOnD algorithm, which aims to identify additional disease proteins based on network analysis [154]. After this step, the network was analysed for functional enrichment in its entirety. This decision contradicts the suggestions of the creators of the algorithm, as the resulted HSP-candidate genes were not filtered based on topology or functional similarity before the subsequent analysis, which could lead to the incorporation of numerous false positive results [154]. Lastly, their data (DAVID v.6.8; HIPPIE v2.0) were of 2016 (e.g., HIPPIE v2.0 had 100,000 less interaction than its 2019 version) and also included predicted interactions. On the other hand, in this study through the use of PINOT, only the manually curated, and experimentally demonstrated with at least two methods or publications PPIs were used. In addition, the seeds were carefully chosen to include only those that have been linked to HSPs through genetic studies on HSP patients. Furthermore, the core network consists of only the seeds that connect directly or through one node, and the nodes that connect at least two seeds. Lastly, the PPIs were collected from primary databases through PSICQUIC, which ensures the usage of the latest version of each database. Thus, in this PhD project

only the most relevant proteins were included in the HSP-PPIN and the higher number of up-to-date PPIs were filtered based on their quality.

2.5.4 Association of HSP with related diseases

Some genes related to disorders of the spastic-ataxia spectrum or with similar clinical phenotypes were either main HSP genes (HSP seeds), or among those selected to be part of the additional seed list for the generation of the HSP network (test seeds) (Table 2-3). Interestingly, some of these genes were indeed present in the core network:

- *ALS2* was part of the core network, suggesting an association of **Infantile onset ascending spastic paraplegia** with HSPs. Mutations in *ALS2* can cause Infantile onset ascending spastic paraplegia [343], Juvenile primary lateral sclerosis [455] and Amyotrophic Lateral Sclerosis 2 [345]. Infantile onset ascending spastic paraplegia's only known genetic cause so far is *ALS2* [456]. Most researchers agree that this disease belongs to HSPs, so its presence in the core network was not surprising, but instead it could provide additional validation of that notion.

- *SPG7*, and *PNPLA6* are part of the core network and are also associated with **Hereditary cerebellar ataxias** (HCAs) [457, 458]. Interestingly, *SPG7* is a very common gene leading to cerebellar ataxia [459]. Mixed phenotypes of HSPs and HCAs can be caused by *GBA2*, and *KIF1C*, which were also part of the core HSP-PPIN. Briefly, *GBA2* leads to cerebellar ataxia with spasticity [445], while *KIF1C* causes predominant cerebellar ataxia with spasticity in the lower limbs [460]. Additional genes are involved in both diseases, such as *F2H* [461-463], *SYNE1* [464-467], *PLA2G6* [468], and *KCNA2* [363-365], supporting the notion that HSPs and HCAs could be part of the same spectrum and the use of the term spastic ataxia [336, 409, 469]. Even though a handful of genes are implicated in both diseases, or show mixed clinical phenotypes, through this study only 4 were found to be in the core network.

- *PDK3* was also part of the network and is associated with **Charcot Marie Tooth**. It leads to an X-linked form of Charcot Marie Tooth, CMTX6 [470]. Interestingly, it was a direct interactor of two seeds, and through one protein connected with 20.7% of the core HSP-PPIN, including 4 seeds. This disease is linked with Hereditary Spastic paraplegias, as spastic paraplegia can be present in people with Charcot Marie Tooth [471] and there are also commonalities in the associated biological processes. Mutations in motor proteins, such as *KIF5A* and *KIF1C* can

lead to HSPs, while cargo transport dysfunction (e.g., *DYNC1H1*, *DNM2*) can lead to Charcot Marie Tooth [471, 472].

Of note, based on the previously discussed results regarding the associated biological processes linked to HSPs, several similarities can be observed with the ones linked to Parkinson's disease. In more detail, it has been associated with disruptions in the endolysosomal trafficking pathway [473-476], lipid metabolism [477] and the function of the substantia nigra [198]. Interestingly, a previous study using network analysis supports this observation, as it demonstrated a significant functional link between genes of HSPs and Parkinson's disease [337]. This suggests that different neurodegenerative diseases could have more common functional pathways than could be indicated based on their clinical manifestations. However, no supporting data for this hypothesis were found though this project.

The presence of genes in the core HSP-PPIN that are also implicated in other diseases could hint about functional connections between them. However, a more dedicated study in which two networks, one for HSPs and one for the other disease (e.g., HCAs), are built and compared, could provide a stronger indication about their possible overlap and differential biological mechanisms.

3rd Chapter:

Further analysis of HSPs based on clinical features

3. Further analysis of HSPs based on clinical features

Main points of this chapter:

- The Hereditary Spastic Paraplegias are clinically heterogeneous, so a number of disease features were studied for potential mechanistic discrepancies
- The analysis of the mode of inheritance and the type of Hereditary Spastic Paraplegias failed to suggest any differences, in contrast to clinical presentations.
- A combinatorial analysis based on enrichment followed by application of machine learning tools suggested 2 clinical subgroups of Hereditary Spastic Paraplegias:
    - TS cluster (Thinning of corpus callosum & Seizures)
    - EPOD cluster (Early onset, Peripheral neuropathy, Optical atrophy, and Dementia/mental retardation)
- Enrichment suggested protein quality control and degradation potentially through the ERAD pathway to be especially important for the TS cluster, while for the regulation of protein localisation and transport in neurons via vesicles to be more tightly associated with the EPOD cluster.

3.1 Introduction

The Hereditary Spastic Paraplegias (HSPs) are a complicated group of heterogeneous neurodegenerative diseases characterised by progressive spasticity and weakness of the lower limbs [166] accompanied by degeneration of the upper-motor neurons [167]. The heterogeneity and complexity of HSPs derive from the clinical presentations as well as the underlying genetic causes. The age of onset can vary from early childhood to late adulthood, all modes of inheritance can be observed, and the form of the disease can be pure or complicated. Complicated forms of HSPs are defined by the co-occurrence of additional symptoms, including peripheral neuropathy, seizures, dementia or mental impairment and optic atrophy [168] (Table 1-1; Section 1.4.4). Regarding the genetic complexity, mutations in over 70 genes [169] have been associated with HSPs, which is one of the highest numbers of causative genes associated with a single Mendelian disease [333]. In such a complex scenario of presentations with a vast array of genetic components

involved in disease pathogenesis, it is not quite clear if all these disease presentations, albeit being classified under the same term of HSPs, can be referred to the same molecular alteration. This missing piece of knowledge is extremely important to support clinical classification and therapeutic investigation.

This work is the first study in which PPINs solely based on experimentally detected and manually curated human PPIs of HSP genes are applied to the investigation of HSPs to identify biological processes involved in disease subtypes following stratification based on the presentation of specific clinical features.

3.2 Aims and Objectives

The aim of this chapter is to investigate whether clinical features tend to aggregate in some groups of proteins of the HSP-PPIN and thus provide a mechanistic link between genetic mutations and clinical presentations.

The objectives of this chapter is to: (i) layer clinical data related to (1) the form of HSP (pure or complicated), (2) the mode of inheritance (autosomal dominant, autosomal recessive, mitochondrial, and X-linked), and (3) a list of clinical presentations in people with HSP (such as thinning of corpus callosum and seizures), (ii) investigate whether there are any clinical clusters in the network, and (iii) if any clusters are present, perform further analysis based on enrichment.

3.3 Methodology

3.3.1 Source of clinical data

After consultation with a collaborator and an expert in HSPs, Prof Henry Houlden, the source of clinical data was chosen to be extracted from the database of the Neuromuscular Disease Centre of Washington University in St Louis, MO, USA [351].

3.3.2 Enrichment

Enrichment for biological processes was performed as described in Chapter 2. The same process was also followed for enrichment of cellular localization using the GO category Cellular Component,

which were grouped in semantic classes (level 1 grouping), and those further grouped in location blocks (level 2 grouping).

Pathway enrichment was performed as described in Chapter 2 using Reactome's online analysis tool (v69& v70 in September and December 2019) [397].

3.3.3 Principal component analysis & Hierarchical clustering

Principal Component Analysis (PCA) was conducted through R (v. 3.6.1) using the prcomp() function of the stats package to compare functional enrichment profiles. The analysis of both the number and percentage of GO terms in each functional block was considered essential because there was a substantial difference in the number of resulted GO terms of the 6 groups, whose functional enrichment profiles were compared ($19<n<114$). The same process was followed for the comparison of the cellular localisation profiles of the clinical clusters.

Hierarchical clustering was performed using the hclust() function (R stats package) based on the distance of the groups in the PCA plot (i.e., Euclidean distance) producing the cluster dendrogram. However, one unit of distance in the x-axis of the PCA plot is more important than on the y-axis, as PC1 (x-axis) explains more variation than PC2 (y-axis). Therefore, the coordinates of each clinical group were transformed, by multiplying them with the explained variation, leading the distance between points to have the same significance in any direction and can thus to be suitable for the subsequent analysis with hierarchical clustering. Choosing the number of clusters derived from the hierarchical clustering that best fit the data was based on the Multiscale bootstrap resampling method, using the R package pvclust [478] that assigns p-values to the branches of the dendrogram. In this package, the pvclust p-value shows that confidence in each result (e.g., pvclust=0.95 shows 95% confidence). After identifying the clusters, the overlap of protein components for a pair of subnetworks, was compared for subnetworks that belonged in the same *versus* different clusters. This was performed using two tailed t-test with unequal distribution.

3.3.4 Distance index and Pearson's correlation

The number of genes displaying a pair of clinical features was compared with the expected number based on their frequencies. The enrichment ratio for each clinical feature (fx) compared to the frequency of a feature (fa) was calculated ($R_{fx\text{-}a}$) (see Enrichment Ratio formula) and then

normalised with the enrichment ratio of the clinical feature to itself ($R_{fa\text{-}a}$) by calculating the following:

$$R_{Difx\_a} = \frac{R_{fx} - R_{fa}}{R_{fa}}$$

Then, the difference of the normalised ratio of each clinical feature with the rest is calculated. The sum of the absolute differences between all the ratios of two features was calculated and named distance index.

Pearson's correlation (r) was performed between all the normalised enrichment ratios (R_{Difxa}) of two features. The p-value was calculated using the formula:

$$p - value = \frac{\mathrm{r} * \sqrt{\mathrm{n} - 2}}{\sqrt{(1 - \mathrm{r}^2)}},$$

with n being the number of comparisons (n=5, as the $R_{fa\text{-}a}$ were removed)

3.4 Results

3.4.1 HSP analysis based on mode of inheritance

It was investigated whether the mode of inheritance of different HSPs is clustered in different parts of the HSP network, which could indicate that there are mechanistic subtypes of HSP. The clinical data regarding the mode of inheritance were visualised in the core HSP-PPIN (Fig 3-1A), and the cores of these networks were extracted (Fig 3-1B and C). Even though HSPs can have autosomal dominant, autosomal recessive, mitochondrial and X-linked inheritance, only the core network of the first two will be studied. This is because there were no genes with mitochondrial inheritance in the core HSP-PPIN, and those with X-linked inheritance were only 2 and connected through more than one interactor, so no core network could be extracted.

There are 41 autosomal recessive genes in the network, and 14 autosomal dominant genes, out of which 5 have both modes of inheritance, which corresponds to 12.2% and 35.7% of each respective group. The autosomal dominant-network consists of 83 and the autosomal recessive-network of 143 proteins, with 79 common proteins (95.2%, and 55.2%, respectively). Functional enrichment of these networks resulted in 36 and 43 semantic classes, and 9 and 10 functional blocks respectively. A comparison of the results of the two networks is shown in Table 3-1. The autosomal dominant-

network had 1 unique semantic class (1/36, 2.8%), and no unique functional blocks (0/9, 0%), while the autosomal recessive-network had 8 unique semantic classes (8/43, 18.6%) and one unique functional block (1/10, 10%).

These results are hard to interpret due to the low number of unique semantic classes for the two categories and their identity. If these results are studied in isolation, it could be claimed that the unique semantic classes of autosomal recessive- differentiate it from the autosomal dominant-HSP network. For example, “Protein localisation – mitochondria” and “Transport - intracellular – axonal” can lead to the conclusion that mitochondria and neurons play a more important role in the autosomal recessive-network. This is -however- not the case because of other shared semantic classes, such as Transport - intracellular - mitochondria” and “Transport - intracellular - neuron - dense core vesicle”. However, it could be claimed that the autosomal dominant-network has a preference to multivesicular body sorting pathway, while the autosomal recessive in autophagy, and that maybe cytoskeleton is more important for the autosomal recessive -network, as there are no semantic classes related to cytoskeleton for the autosomal dominant-network. Nonetheless these are not substantial differences, and in combination with the large overlap of the two networks, no strong conclusions can be drawn from this analysis.

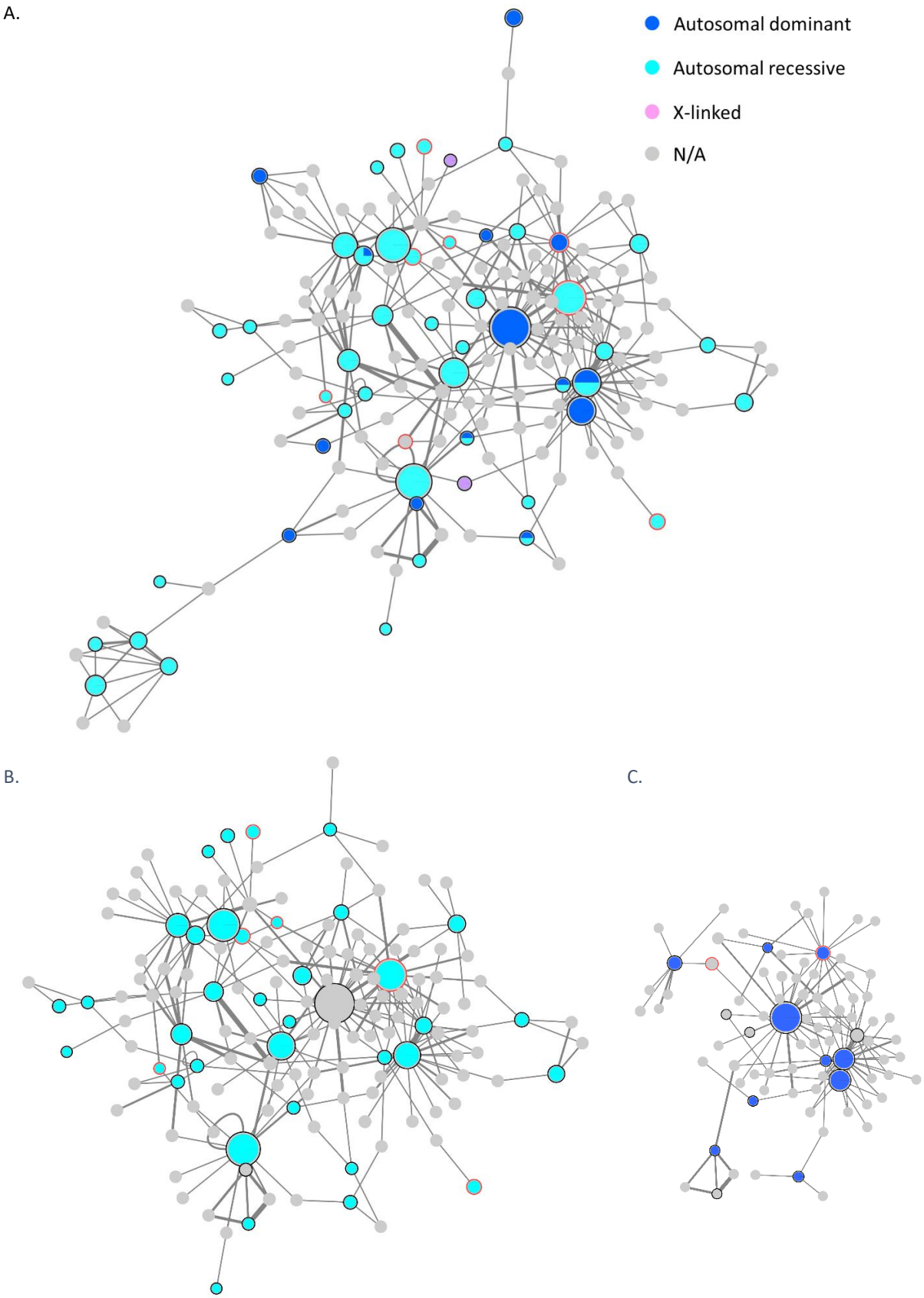


Figure 3-1. Analysis of the HSP mode of inheritance in the core HSP-PPIN

*(**A**) The mapping of the mode of inheritance in the core HSP-PPIN. (**B**) The autosomal recessive HSP network. (**C**) The autosomal dominant HSP network.*

Table 3-1. Unique semantic classes and functional groups for the AD- and AR- HSP networks

Mode of inheritance	Unique semantic classes	Unique functional groups
Autosomal dominant	1. Waste disposal - ubiquitin-multivesicular body sorting pathway	-
Autosomal recessive	1. Cell cycle 2. Intracellular organisation - cytoskeleton 3. Intracellular organisation – vesicle 4. Protein localisation – mitochondria 5. Protein metabolism - protein complex organisation 6. Transport - intracellular – axonal 7. Transport - intracellular - cytoskeletal 8. Waste disposal – autophagy	1. Cell cycle

3.4.2 HSP analysis based on form of HSPs

It was investigated whether the form of HSP (complicated or pure) is clustered in different parts of the HSP network, which could indicate that there are mechanistic subtypes of HSP. The clinical data regarding the form of HSPs were added in the core HSP-PPIN (Fig 3-2A), and the core of these networks was extracted (Fig 3-2B, and C).

There are 38 complicated genes in the core HSP-PPIN, and 12 pure, out of which 8 have both forms. The complicated-network consists of 145 and the pure-network of 56 proteins, with 49 common proteins (which corresponds to 33.8%, and 87.5% of the total number of proteins of the respective networks). Functional enrichment of these networks resulted in 49 and 11 semantic classes, grouped in 11 and 5 functional blocks respectively. A comparison of the results of the two networks is shown in Fig 3-3 and Fig S3-1 (Appendix C), based on the percentage and number of the GO terms, respectively. The complicated HSP network had 38 unique semantic class (38/49, 77.6%), and 7 unique functional blocks (7/11, 63.63%), while the pure HSP network had no unique semantic classes nor any unique functional blocks.

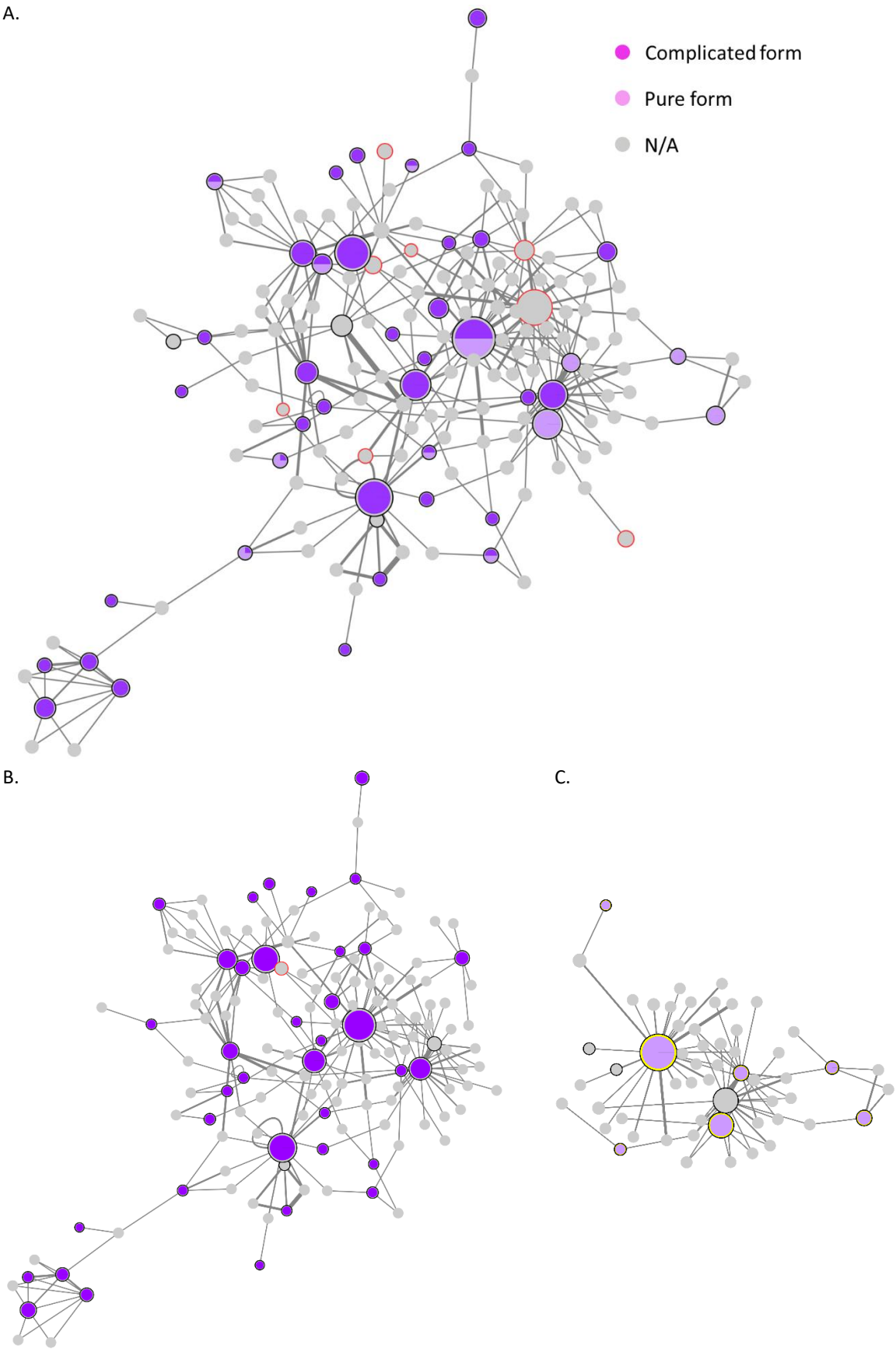
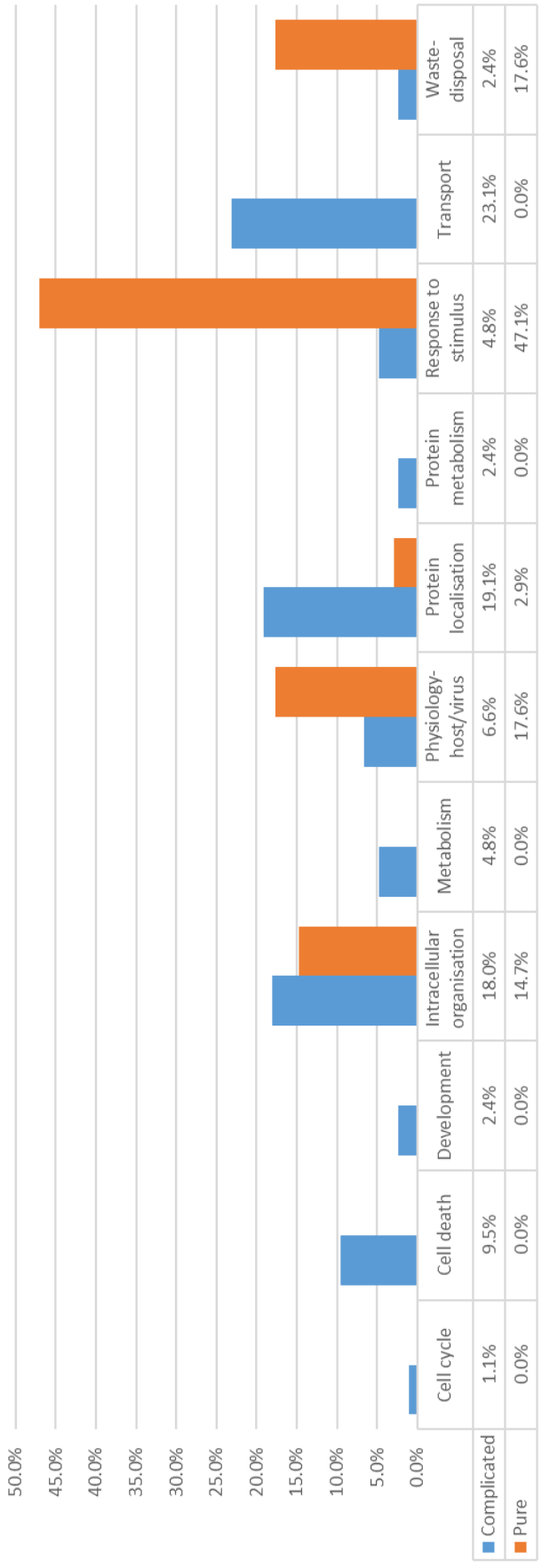


Figure 3-2. Analysis of the HSP form in the core HSP-PPIN

*(**A**) The mapping of the form of HSP in the core HSP-PPIN. (**B**) The complicated HSP network. (**C**) The pure HSP network.*



	Cell cycle	Cell death	Development	Intracellular organisation	Metabolism	Physiology-host/virus	Protein localisation	Protein metabolism	Response to stimulus	Transport	Waste-disposal
Complicated	1.1%	9.5%	2.4%	18.0%	4.8%	6.6%	19.1%	2.4%	4.8%	23.1%	2.4%
Pure	0.0%	0.0%	0.0%	14.7%	0.0%	17.6%	2.9%	0.0%	47.1%	0.0%	17.6%

Figure 3-3. Comparison of GO-BP terms between complicated and pure form of HSPs based on their percentage

Hypotheses can be formed by observing the distribution of the GO terms. Based on their percentage (Fig 3-3), it can be claimed that the pure network has a specificity to the functional groups “Physiology host-virus”, “Response to stimulus” and “Waste disposal”, as these are around 3, 10 and 7 times higher compared with the complicated network. However, when taking into account their numbers, this claim seems weak, as they are similar or lower, with 6, 16, and 6, compared to 25, 18, and 9 for the pure *versus* complicated HSP network, respectively.

3.4.3 HSP analysis based on a set of clinical features

3.4.3.1 HSP clinical subtypes’ specific networks

The typical HSP phenotype can be present with a wide group of clinical features in different individuals. These include peripheral neuropathy (P), thinning of the corpus callosum (T), seizures (S), dementia or mental retardation (D), and optic atrophy (O). Some patients also manifest early disease onset (E). Medical reports and case studies sometimes declare the presence of the above clinical features in cases with mutations of HSP genes. Therefore, the seeds of the HSP core network were colour-coded based on the clinical features to which they can lead (Fig 3-4). Of note, some seeds are associated with just 1 feature (n=9, 16%) while others with 2 (n=18, 32%), 3 (n=12, 21%), or 4 (n=7, 12%). This incorporation of clinical information in the seeds of the network allowed the extraction of 6 subnetworks from the core-HSP PPIN, each of them collecting the interconnected seeds (and their interactors) that were associated with each specific clinical feature (Fig S3-2, Appendix C).

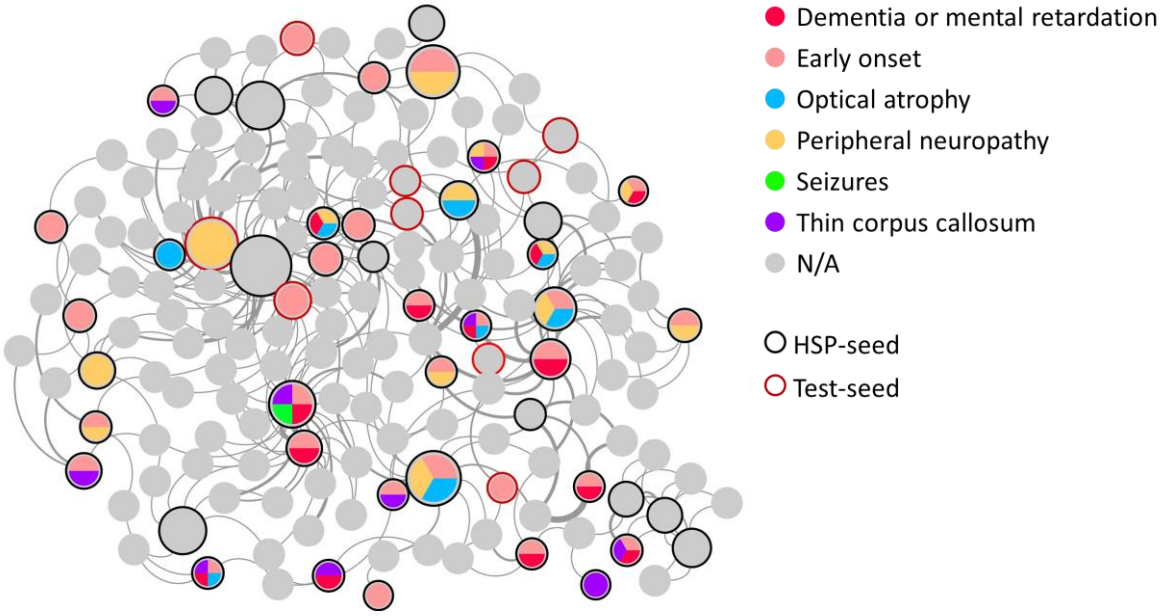


Figure 3-4. Visualisation of clinical characteristics of HSPs caused by each HSP gene in the core HSP-PPIN

The presence of clinical characteristics in HSPs is visualised in the core HSP-PPIN by the colour of each node. The colour correspondence is located on the top right of the image. The nodes corresponding to the HSP-seeds have a black border, while the test-seeds have a red border. The size of each node positively correlates with its degree. The thickness of each edge positively correlates with its final score calculated by PINOT. The network was visualised using Cytoscape. Adapted from [403].

Enrichment of biological processes (GO-BPs) was performed for each clinical subnetwork, using g:Profiler, Gene Ontology and WebGestalt, as previously described. The percentage and numbers of GO terms within each functional block were measured to weight their relevance to the HSPs (Fig 3-5 and Fig S3-3, Appendix C).

The number of functional groups (13) and of the clinical subnetworks (6) required a more sophisticated method for their comparison. Principal components analysis (PCA) was therefore utilised to reduce the complexity of these functional enrichment results to 2 principal components, allowing the comparison of the 6 clinical subnetworks (Fig 3-6A). The chosen principal components were PC1 and PC2 due to their high explained variation (Fig 3-6B). In the PCA graph, each clinical subnetwork is represented by a single point and interestingly some were clustered together (Fig 3-6A). More specifically, the functional enrichment profiles of the clinical features peripheral neuropathy, optic atrophy, dementia or mental retardation, and early onset displayed a similar reduction to PC1 and PC2. Of note, this result was obtained with PCA performed on both the percentage of the GO terms in each functional block (Fig 3-6) and their number (Fig S3-4).

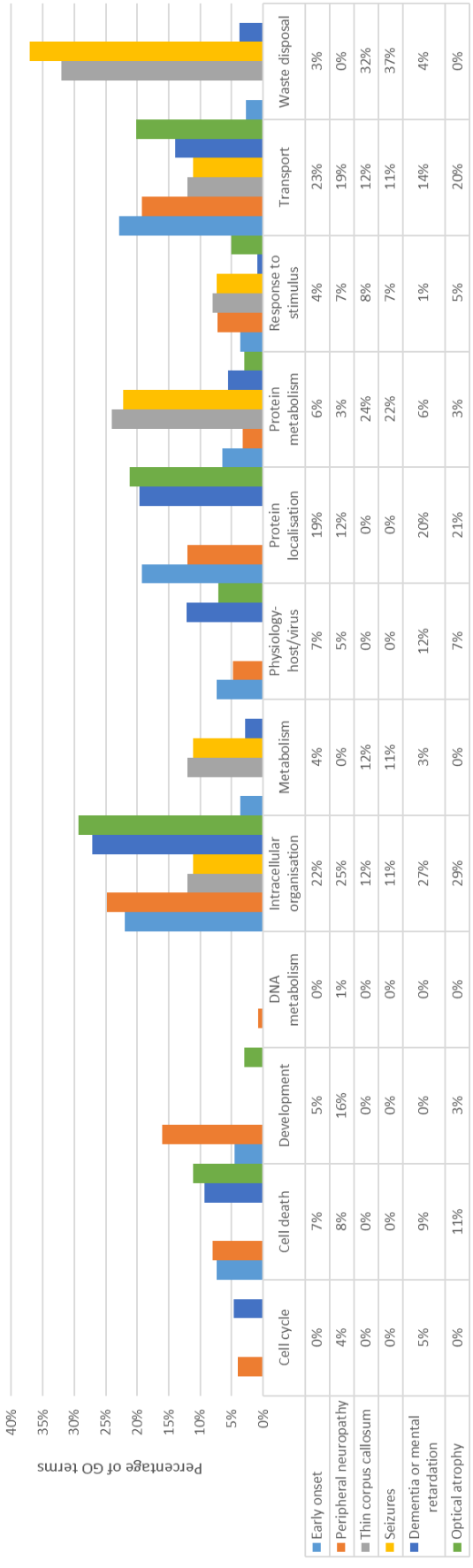
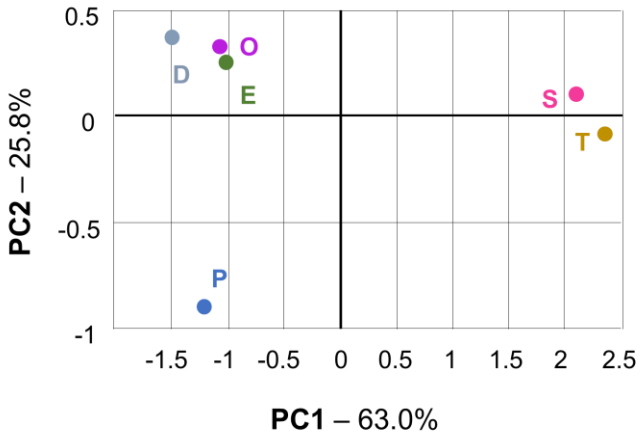


Figure 3-5. Comparison of the functional enrichment results of the clinical subnetworks using the percentage of the GO terms of each functional group

A.



B.

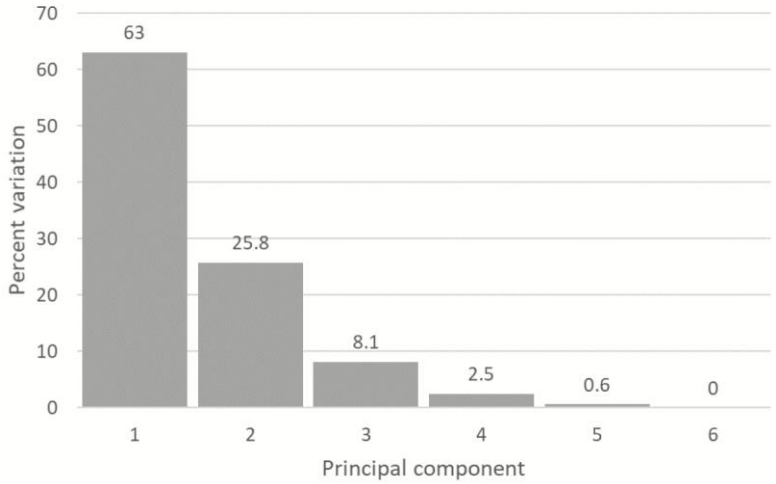


Figure 3-6. Principal Component Analysis performed for the comparison of the clinical subnetworks *(**A**) The percentage of GO-BP terms for each functional group was analysed with PCA through R, resulting in the PCA graph, which shows the distribution of the gene groups in the PC1 and PC2 axes. (**B**) The explained variation from each PC axis. (The results for the analysis based on the number of GO-BP terms in shown in Fig S3-4.) Adapted from* [403].

Even though the PCA plot can provide an initial visual insight into potential clustering, the more dedicated method, hierarchical clustering (HC), was used to determine the clusters within the PCA graph. The results of the HC were plotted into a cluster dendrogram (Fig 3-7 & Fig S3-5). In order to find the number of clusters that best fits the data, the Multiscale bootstrap resampling method was used [478] (Fig S3-6). This method suggested the existence of 2 clusters: Cluster A, and Cluster B, as seen in Fig 3-7, with a 99%, and 91% confidence (pvclust p-value =0.99 and 0.91 respectively). This method also led to the same result using the number of GO-BP terms (Fig S3-5). Cluster A is composed of two clinical groups: thin corpus callosum and seizures (from now on named TS), while cluster B is composed of four groups: early onset, peripheral neuropathy, optic atrophy and dementia or mental retardation (from now on named EPOD).

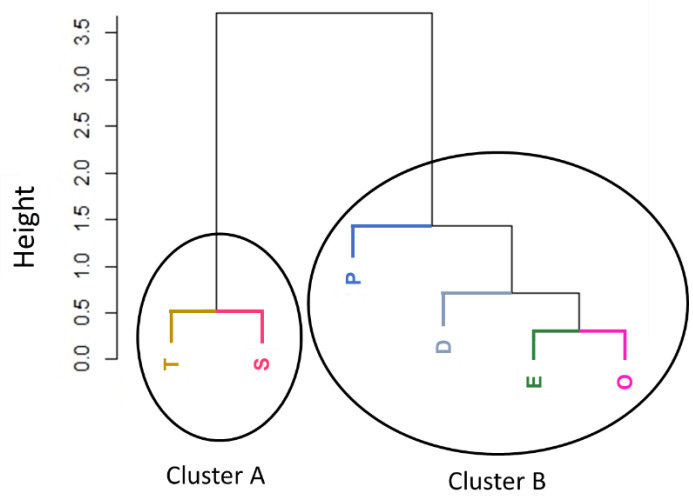


Figure 3-7. Hierarchical clustering performed for the comparison of the clinical subnetworks

Cluster dendrogram produced based on hierarchical clustering of the gene groups as analysed in Fig 3-6. (The results for the analysis based on the number of GO-BP terms in shown in Fig S3-5) Adapted from [403].

The co-clustering of the T and S is not surprising as they shared 23 proteins, with their only difference being that the former had 5 extra proteins. The common proteins corresponded to 100% of the proteins of S and to 82.1% of the proteins of T. Interestingly, another pair of datapoints that clustered together based on hierarchical clustering consists of E and O, which have shared components. Their common proteins are 33, which is 29.8% of the proteins of E and 81.0% of the proteins of O. These results prompted the exploration of additional overlaps, shown in Table 3-2.

Table 3-2. Overlap of protein composition among clinical subnetworks

	D	E	O	P	S	T
D	100%	48.2%	45.2%	28.8%	100%	89.3%
E	96.5%	100%	81.0%	60.3%	100%	96.4%
O	33.3%	29.8%	100%	53.4%	8.7%	7.1%
P	36.8%	38.6%	92.9%	100%	21.7%	17.9%
S	40.4%	20.2%	4.8%	6.8%	100%	82.1%
T	43.9%	23.7%	4.8%	6.8%	100%	100%

Note: The overlap is calculated as a percentage of the total proteins of the subnetwork of the corresponding column. With orange colour are the comparisons within the same cluster, while in green those between clusters. Adapted from [403].

To explore whether the similarity of protein composition drives the results of HC, the percentage of overlapping protein identity between gene groups within the same cluster was calculated and compared with the percentage of overlapping protein identity between different clusters (Fig 3-8). The comparison of overlapping protein identity within the same cluster and between clusters was not significant (p=0.07). This result indicates that an increased overlap of proteins between two clinical subnetworks is insufficient to drive their co-clustering in the PCA graph, as determined by HC.

3.4.3.2 Matching with clinical data

The PCA and HC analyses indicate that the functions related with the HSP genes could be distinct between the two clusters. Such a functional difference could indicate that the features of each cluster would tend to co-occur in patients as they might be caused through the same mechanism. Interestingly, the analysis of clinical data that included the presence of these features in patients with mutations in various HSP genes, showed that in most cases the order that the clinical features tend to co-occur in patients follows the order of similarity of functional profiles, as suggested by the cluster dendrogram (Table S3-1).

When the results were analysed through Pearson's correlation (Table S3-2), the TS cluster significantly positively correlated (T with S, p-value 0.0003). In addition, a pair of the EPOD cluster, also positively correlated (P with O, p-value 0.0010). This result further supports the previous findings supporting a functional connection of the components within the clusters TS and EPOD.

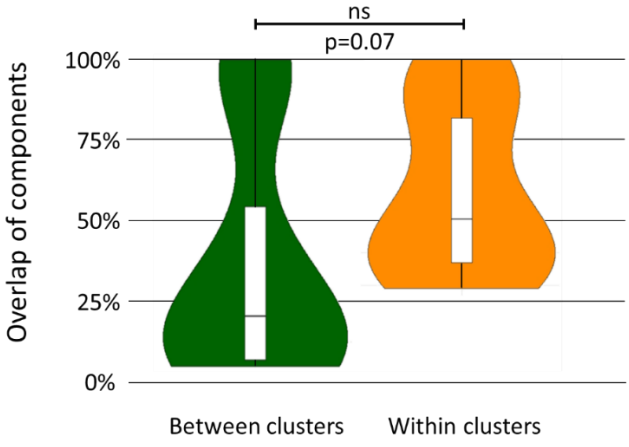


Figure 3-8. Comparison of the overlap of proteins between two subnetworks that belong in different (left) or the same (right) cluster(s). *Adapted from* [403].

3.4.3.3 Exploring potential differences between the clusters

An initial functional insight into the clustering of the PCA data was provided by the loading score of each functional block (Fig 3-9 and Fig S3-4C). More specifically, the overlap of the PCA graph and

the loading score graph can provide an explanation as to which functional blocks guide the localisation of each clinical subnetwork in a particular position in the graph. For example, thin corpus callosum and seizures (T & S) were clustered together mainly based on higher similarities in the functional blocks “metabolism”, “protein metabolism”, and “waste disposal” (from the comparison of Fig 3-6A and Fig 3-9).

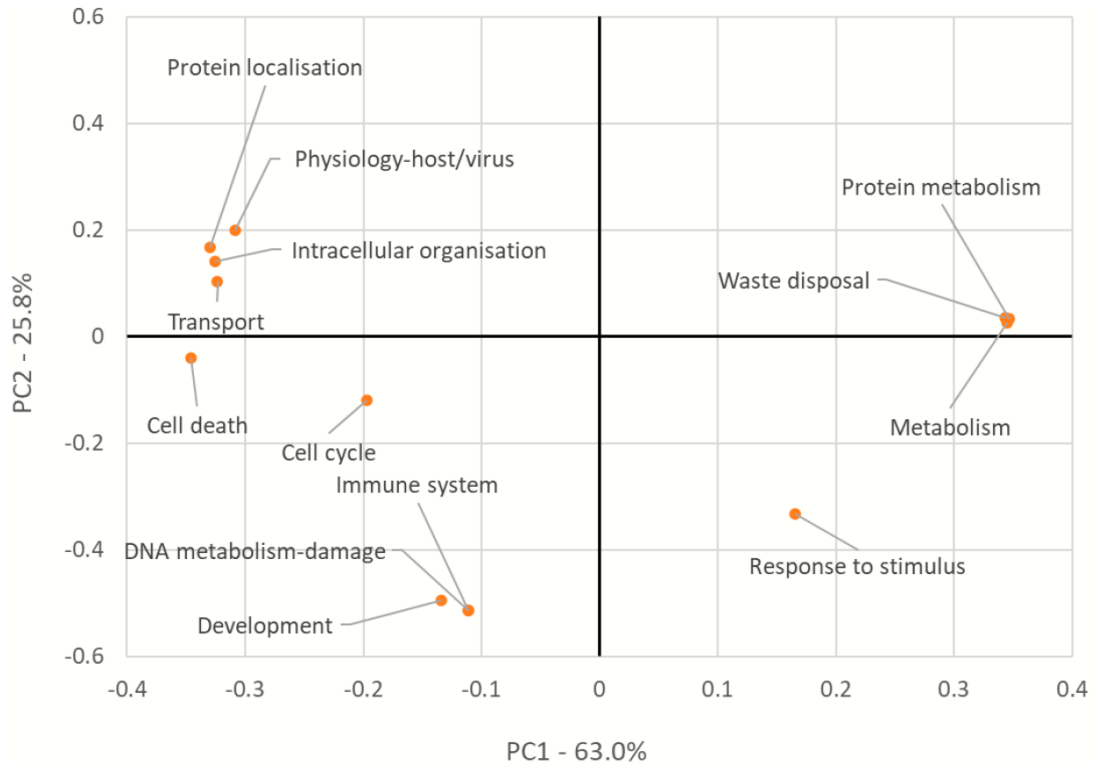


Figure 3-9. Loading score of PCA for the analysis based on the percentage of GO terms

The putative differences of the two clusters were further explored with enrichment analysis in two levels: biological processes, and cellular components. Comparing the functional profiles of the two clusters (see Fig 3-10, Fig S3-7) suggested some differences in the distribution of the enriched GO terms. The TS cluster resulted in 25 terms, 12 semantic classes and 5 functional groups, while EPOD in 158 terms, 56 semantic classes and 10 functional groups. There was an overlap in the identity of the terms (n= 14, which is 56.0% of those of TS, and 8.9% of those of EPOD), semantic classes (n=10, which is 83.3% of those of TS, and 17.9% of those of EPOD) and of the functional groups (n=4, which is 80% of those of TS, and 40% of EPOD).

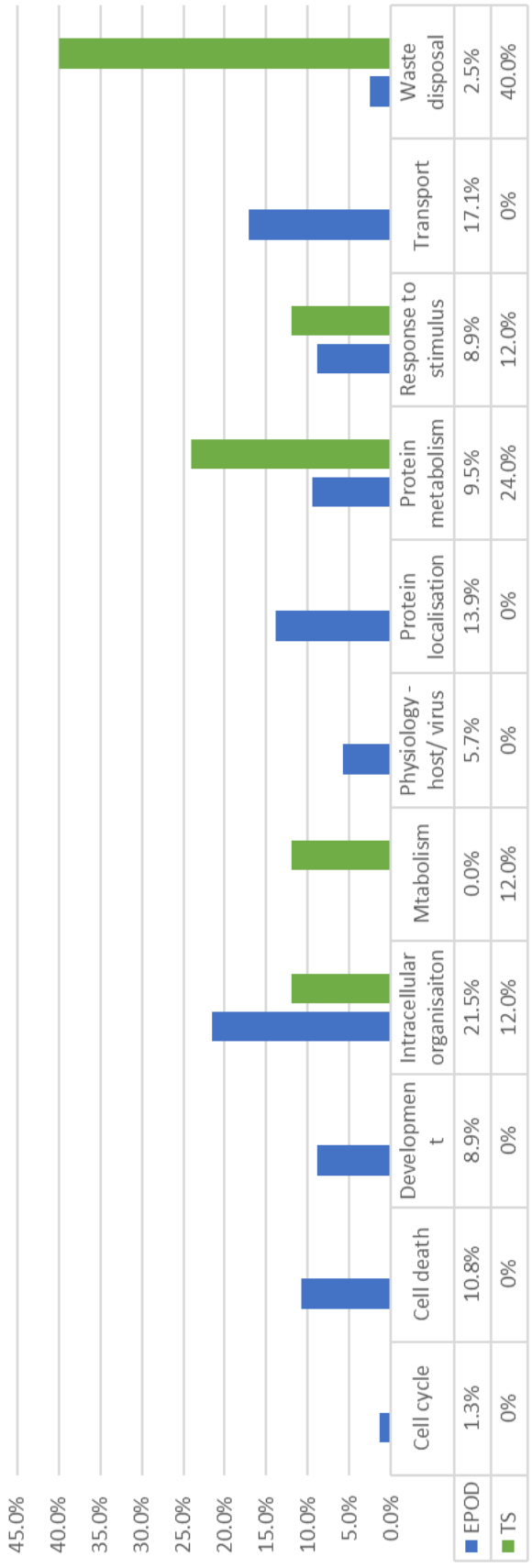


Figure 3-10. Comparison of the functional enrichment profiles of EPOD and TS using the percentage of the GO-BP terms in each functional group

However, there was a differential distribution of the GO terms in each functional block. The functional groups “Waste disposal” (n=10, 40.0%) and “Protein metabolism” (n=6, 24.0%) were suggested to be some of the most important processes in the cluster TS, based on the percentages of their GO terms. In accordance with these results, it should be mentioned that the unique functional group of TS was “Metabolism”, which included 3 semantic classes related to the catabolic process (e.g., macromolecule catabolic process (GO:0009057)). In contrast, for EPOD the most important functional groups were “Intracellular localisation” (n=34, 21.5%), followed by “Transport” (n=27, 17.1%), and Protein localisation (n=22, 13.9%). These results indicate that transport and localisation of proteins is a more important process in the cluster EPOD than in cluster TS, for which the catabolic processes of the cell are of higher relevance.

Interestingly, the GO terms “proteasomal protein catabolic process” and “proteasome-mediated ubiquitin-dependent protein catabolic process” are unique to TS. These terms are 2 among the 25 of TS (8%), while no term relevant to proteasome can be found among the 158 of the EPOD cluster, even though they are 6-fold higher (148 vs 25), highlighting the importance of the presence of this semantic class uniquely to cluster TS. This result is in accordance with the previous observations about the higher significance of catabolic processes for the TS cluster.

The potential differences between the two clusters were further investigated using enrichment focused on cellular components, using Gene Ontology Cellular Component (GO-CC) terms (see Fig 3-11, Fig S3-8). The enrichment resulted in 10 GO-CC terms, grouped in 7 semantic classes and 6 location groups for TS, and 84 terms, in 46 semantic classes and 17 location groups for EPOD. There was a substantial overlap between the two clusters. There was no unique GO-CC term for TS, and therefore nor semantic classes and location groups. However, their distribution in the categories was differential.

Figure 3-11. Comparison of the location enrichment profiles of EPOD and TS using the percentage of the GO-CC terms in each location group

Interestingly, a higher percentage of GO terms for the TS was related to the ER (30.0% vs 6.0%), melanosomes (10.0% vs 1.2%), and three forms of membranes (i.e., membrane networks, 10.0% vs 1.2%; membranes of organelles, 20.0% vs 9.5%; and other membrane terms 20.0% vs 9.5%). These were 5-fold, 8-fold, and 2- to 8-fold more increased in the TS cluster compared to EPOD. The profile of the latter included more location groups. The functional group with the highest percentage of terms was “Projection” with 12 terms (n=12/84, 14.3%). This category included terms, such as “distal axon” and “neuron projection”. The second largest group was “Vesicle”. Examples include “multivesicular body”, “late endosome”, “early endosome”, and “pigment granule”. Of note, EPOD included various neuron related terms, for instance “myelin sheath”, “neuronal cell body”, and “dendrite cytoplasm”. These results suggest that the membrane system and melanosomes are of higher relevance to the TS cluster, while vesicles and cellular projections, such as those of neurons, are more related to the EPOD cluster. These results should be further investigated using an in vitro and/or in vivo experimental setup.

3.5 Discussion

In order to explore whether the clinical heterogeneity of the HSPs might reflect subtypes with differences in the molecular mechanism of disease, the functional profile of parts of the core network associated with different features was explored. The mode of inheritance was studied first, which resulted in the formation of 2 subgroups, one for the autosomal dominant part of the network and another for the autosomal recessive part. Functional enrichment indicated the possible differential association of the two networks with cell cycle, autophagy, and cytoskeleton for the autosomal recessive HSP-network, and multivesicular body sorting pathway for the autosomal dominant HSP-network. However, the high number of common genes between these two groups (95.2% of the autosomal dominant network was part of the autosomal recessive network), the low number of unique results from the functional enrichment (2.8% of the semantic classes of the autosomal dominant network were unique) and their identity, hindered the drawing of any strong conclusions.

Potential mechanistic subtypes associated with the two different forms of HSPs, complicated and pure, were also explored. The results showed a trend for the pure network to be more specialised in “Physiology host-virus”, “Response to stimulus” and “Waste disposal”. However, the percentage of overlap of the pure and complicated networks was high (87.5% of the pure network was also part of the complicated network). In addition, there was a lack of unique semantic classes and

functional blocks for the pure network. Therefore, the interpretation of the results requires caution, and no strong conclusions can be drawn from this analysis.

More promising data were generated during the exploration of a set of clinical features, when they were layered in the HSP-PPIN. Functional enrichment analysis in combination with machine learning tools were utilised, namely PCA analysis and hierarchical clustering. The results pointed towards the existence of at least two main subtypes of HSPs based on the associated functions: (i) thin corpus callosum and seizures (i.e., TS cluster), and (ii) early onset, peripheral neuropathy, dementia or mental retardation and optical atrophy (i.e., EPOD cluster). Enrichment analysis of biological processes for the two clusters proposed that waste disposal, protein metabolism and the function of proteasome are more associated with the TS cluster, which all fall into the same theme of protein quality control and degradation. The EPOD cluster was more related to protein localisation, and transport. The two HSP clusters also differed in their association with cellular components, as TS showed a higher enrichment in the ER, different types of membranes, and melanosomes, whereas neuron-related cellular localisations and vesicles were more strongly associated with EPOD. Overall, similar conclusions can be drawn from the results from enrichment in the level of functions, and cellular components, namely that the protein quality control and degradation maybe through the ERAD pathway were especially important for the TS cluster, while the tight regulation of protein localisation and transport in neurons via vesicles could be more tightly associated with the EPOD cluster.

These findings provide an indication of separate implicated functions, pathways and potentially disease mechanisms in groups of genes that lead to the development of different clinical features. Based on the molecular mechanism of disease, it is proposed that HSPs are subdivided into at least 2 major groups. These results indicate that not all the clinical manifestations of HSPs refer to the same disease at a molecular level and that it is indeed possible to stratify HSPs patients based on the putative molecular mechanisms of disease. These results require further validation, but they suggest that when aiming at drug discovery for the HSPs and when designing clinical trials, the molecular heterogeneity of disease would need to be taken into consideration.

Even though the interactome analysis can provide us with useful knowledge, the conclusions from such an in silico analysis should be further explored, as they require direct functional validation. With the discovery of more protein-protein interactions, the human interactome will become more complete and might be able to help us understand better the connecting processes of large groups of genes and potentially point towards the disease mechanism. More specifically, future work focusing on proteins related to HSP as suggested by this study and especially those that have not

been studied yet in as much depth as others, might fill in the gap of the clinical and functional data required to unveil which pathways and processes are more related to the disease mechanism and might bring the scientific and clinical communities closer to a treatment or even cure. To aid towards that direction, people from various scientific communities and sectors could collaborate to set up a database that includes a detailed list of symptoms and mutations in genes/loci, as agreed by most experts, and that is accessible to the other researchers for further exploration.

4th Chapter:

Network analysis of macroautophagy and its interplay with neurodegenerative diseases

4. Network analysis of macroautophagy and its interplay with neurodegenerative diseases

Main points of this chapter:

- Finding the key players of macroautophagy: no general consensus
- So, multiple sources were used:
  - Autophagy Regulatory Network (ARN),
  - Gene Ontology,
  - REACTOME,
  - KEGG, and
  - UniProt
- 511 proteins were filtered to 156 main macroautophagy proteins, which were used as seeds to built the macroautophagy network
- Most neurodegenerative genes of Alzheimer's disease, Parkinson's disease, Amyotrophic lateral sclerosis, and Frontotemporal dementia were present in the macroautophagy network
- Large overlap of the networks of each neurodegenerative disease with the macroautophagy network

4.1 Introduction

4.1.1 Association of macroautophagy with neurodegenerative diseases

Macroautophagy is one of the main systems for the degradation of cell components. It is especially important for post mitotic cells, including neurons in which it is involved in the removal of a misfolded proteins that are prone to aggregate [249]. Deficiencies in this process can lead to the formation of protein aggregates, which is the hallmark of multiple neurodegenerative diseases. Indeed, mice deficient of neuronal ATG5 and ATG7 -two proteins that regulate the elongation of the phagophore- showed neuronal accumulations of cytoplasmic proteins, as well as progressive

deficits in motor functions [250, 251]. These results were observed in the absence of mutations leading to higher propensity of aggregation. Therefore, in the presence of such mutations an efficient macroautophagic machinery might be even more essential for maintaining correct cell proteostasis, which then reflects in a healthy cellular environment.

4.1.1.1 Alzheimer's disease and autophagy

Alzheimer's disease (AD) is the most common neurodegenerative disease mainly affecting people over the age of 65 [479]. One of its main symptoms is dementia but others include impairment in language and visuospatial abilities, depression, and anxiety. The pathological hallmark of Alzheimer's disease is the accumulation of protein aggregates consisted of amyloid-β and hyper-phosphorylated tau, forming amyloid plaques and neurofibrillary tangles, respectively [480, 481].

There is evidence linking dysfunction of macroautophagy and AD in both animal models and humans. Heterozygous deletion of *BECN1* (a key macroautophagy protein) in a mouse model of AD led to an increase of the accumulation of amyloid-β and neurodegeneration [482]. In people with AD, there is reduction in Beclin-1 protein levels in the affected brain regions [482]. This protein was detected in neurons that had survived until the time of the sample collection, suggesting a potential advantage of those cells compared to others that had already died. However, AD is a progressive disease that continues to lead to neuronal death, so one of its mechanisms could be through the reduction of the level of Beclin-1 and of the efficacy of macroautophagy.

Furthermore, mutations in *PS1* (Presilinin-1, *PS1*), the most common cause of early-onset familial AD [483], were reported to impair macroautophagy. More specifically, *PS1* was showed to be required for the targeting of v-ATPase in lysosomes [484] that leads to the autophagolysosomal acidification [485], which is essential for the degradation of its contents. The impairment of macroautophagy as suggested by experiments in mice was confirmed in human fibroblasts from people with AD [484].

4.1.1.2 Parkinson's disease and autophagy

Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease. Degeneration of the neurons in the substantia nigra lead to a deficit of dopamine and consequent alterations of the dopaminergic circuits that are responsible for the control of movements. The main symptoms for individuals with this disease are motor (rigidity, bradykinesia, and tremor) [191], but also include disturbances in sleep, loss of smell, constipation, depression,

anxiety, and dementia [192-194, 196]. The pathological hallmark of PD is the accumulation of aggregated proteins in cellular inclusions, named Lewy Bodies, whose main component is α-synuclein [197, 198, 205].

Macroautophagy has been strongly linked to Parkinson's disease. α-synuclein's gene (*SNCA*) is one of the genes most strongly associated with PD. Aggregation of α-synuclein is leading to the generation of Lewy bodies and mutations in its gene (e.g., gene triplication, leading to protein overexpression) can cause a familial form of PD [486]. Interestingly, α-synuclein has been shown to inhibit the formation of autophagosomes [487]. Furthermore, α-synuclein is cleared by chaperone mediated autophagy, and it was demonstrated that mutated α-synuclein can block this autophagic pathway [488]. These studies suggest the existence of a positive feedback loop whereby overproduction of α-synuclein or mutated α-synuclein might impair clearance based on autophagy, this leads to increased levels of α-synuclein in the cell, finally leading to misfolding and aggregation in Lewy bodies. Another PD gene suggested to affect autophagy is DJ-1, which regulates it through multiple pathways, such as the JNK/Beclin1 pathway [489].

The association between macroautophagy and PD is further strengthened by evidence that link a modulation of the autophagy regulators and the levels of aggregates. The administration of rapamycin (i.e., negative regulator of mTOR) increased the clearance of aggregates in animal models of PD through the upregulation of macroautophagy [490-492]. Additionally, there is evidence suggesting an increased downregulation of two genes of autophagy in brains of people with PD, *LAMP2A* and *Hsc70* [493]. More specifically, there were increased levels of miRNAs in the substantia nigra pars compacta and the amygdala, which were negatively regulating the autophagy genes hinting to a role of the process in the development of the disease.

4.1.1.3 Frontotemporal dementia and autophagy

Frontotemporal dementia (FTD) is a clinically, pathologically and neuroanatomically heterogeneous group of progressive neurodegenerative diseases, targeting the frontal and temporal lobes [494]. FTD is accompanied by neuronal loss, gliosis, and microvacuolar changes in multiple brain regions [495], such as the frontal lobes. It can affect a broad range of brain functions such as social behaviour and language, and it is the second most common cause of young onset dementia [494]. Regarding its genetic basis, genetic cases of FTD account for around 10% of patients, and most of which (60%) are caused by mutations in *C9ORF72*, *MAPT*, and *GRN* [496, 497]. Other implicated genes include *TARBP*, *FUS*, *VCP*, *CHMP2B*, and *TBK1* [495, 498-500]. Intracellular inclusions that include aggregated proteins have been identified in brain tissue of people with FTD. Mutations in

C9ORF72 can lead to the development of FTD and in the formation of intraneuronal inclusions of its protein product (a dipeptide with repeat expansion) [501]. In addition, inclusions of TDP-43 (encoded by *TADBP)* and of FUS, have also been identified in people with FTD [502, 503].

Multiple evidence has been reported linking FTD with autophagy. Tau accumulation, which is present in some cases of people with FTD [504], can disrupt axonal vesicle transport, leading to an accumulation of autophagosomes and an increased tau-induced toxicity [505]. In addition, a growing number of drugs that affect autophagy can ameliorate the disease phenotype of FTD. Such examples are methylene blue [506-509], trehalose [510], and rapamycin [511]. Furthermore, the suggested mechanism of neurodegeneration in FTD by mutations in *VCP* and *CHMP2B* includes the dysfunction in the autophagic protein degradation pathway [495]. Another FTD gene, *TBK1*, has also been linked with autophagic degradation of protein aggregates [512]. More specifically, it directly phosphorylates the autophagy receptors optineurin and p62 that target cargo to the forming autophagosome [513], and is involved in the autophagosome-lysosome fusion [514].

4.1.1.4 Amyotrophic Lateral Sclerosis and autophagy

Amyotrophic Lateral Sclerosis (ALS) is a degenerative disorder in which motor neurons are progressively lost from the central nervous system. One of its hallmarks is that both upper and lower motor neurons are affected, resulting in muscle weakness and loss of autonomic functions, such as mobility and lung function. Respiratory failure is the most common cause of death and occurs usually from 2 to 4 years after the onset of symptoms [190]. Intracellular inclusions have been detected in tissue from people with ALS. The majority of people with ALS present intracellular inclusions containing TDP-43 [503]. Additional inclusions have been detected in people with mutations in *C9ORF72*, *FUS* and *SOD-1*, which contained aggregates of the respective protein products [501, 502, 515].

The link between ALS and macroautophagy is supported by multiple studies. p62 is essential for guiding cargo to autophagosomes. Interestingly, it is encoded by the *SQSTM1* gene, whose mutations are linked to familial forms of ALS [516]. Optineurin is also involved in directing cargos to autophagosomes and mutations in its gene, *OPTN*, are linked with familial ALS, as well [517]. TBK1 regulates the activity of optineurin [517] and is similarly linked to ALS, probably through the same pathway [214]. Guanine nucleotide exchange C9orf72, a protein strongly associated with familial ALS, is suggested to regulate macroautophagy through its GDP/GTP exchange factor activity [231-233]. Finally, a study in ALS suggested a progressive neuronal accumulation of ubiquitin-positive

protein aggregates and a dysfunction of ESCRT-III [518, 519]. This finding combined with the discovery that ESCRT-III in cortical neurons is important for the fusion of autophagosome with lysosomes [298, 299, 301], is another piece of evidence suggesting a link between macroautophagy and ALS.

4.1.2 Autophagy through systems biology approaches

Macroautophagy is a process characterised by high complexity, rendering it an ideal process to be studied with a systems biology approach to unveil its mechanistic details. For instance, a proteomic approach was used to prioritise candidate proteins that were implicated in starvation-induced autophagy in *Drosophila melanogaster*. Further experiments in which the identified candidate genes were mutated, confirmed the association of a candidate, *Desat1*, demonstrating the predictive strength of combining proteomic and genetic analyses [520]. An epistatic analysis of pairs of genes related to autophagy and apoptosis was scaled up to a systems level through the development of a RNAi based methodology. Processing of the produced results using PPI data suggested a novel pathway connecting *CASP3* (Caspase-3) and *ATG5* (Autophagy protein 5) [521], which has since been confirmed [522]. A different study used genome-wide siRNA screening to investigate the mechanism of basal autophagy, identifying 236 genes that regulates autophagy and the convergence of multiple signalling pathways that regulate autophagy to the inhibition of PIK3C3 [523]. A noteworthy project created a detailed map of the mTOR signalling pathway [524]. It contained 946 components and 777 relationships, including a part that was related to autophagy. This comprehensive signalling map was made available via a collaborative web service platform that allowed its further development, through curation from other researchers and facilitated system-level analyses.

Protein interaction network analysis is an eligible systems biology tool which has been utilised for the analysis of the autophagic process. For example, an aging model based on the fungus *Podospora anserina* was studied by building a PPIN of autophagy using published and new PPI data produced by the study [525]. Transcriptome data was incorporated in the network, and 7 modules were identified based on network topology, suggesting different associated pathways. Novel human candidate proteins that could be implicated in autophagy were identified in study based on PPIN analysis in yeast [526]. Mapping of conserved interactions between homologs of human and yeast proteins allowed these predictions, some of which were experimentally verified [526]. A disease protein interaction network was built around autophagy in another study in which its

relationship with other diseases was investigated [527]. Interestingly, the study suggests a strong functional link between autophagy and cancer, and that autophagy could be a connecting factor between cancer and the immune system, with MAPK1 potentially playing an important role.

The link between neurodegeneration and autophagy was similarly investigated via systems biology approaches. One study analysed 416 genes and focused on the transcriptional and microRNA-based post-transcriptional regulation of autophagy and lysosomal function, and their role in neurodegeneration [528]. The genes linking the two processes were enriched for phenotypes, such as motor-related, abnormal brain morphology, brain size, and gliosis. This study further supported the contribution of autophagic genes to AD and PD, but also suggested their involvement in other diseases that affect the brain, namely, tuberous sclerosis, and neuronal ceroid lipofuscinoses. In this study, some PPIs were used for visualisation purposes only, and the collection of PPIs was not described in the methods nor the supplementary material. Within the larger field of systems biology, this highlights the lack of PPIs studies with a focus on autophagy in the context of neurodegeneration. In the case of AD, PPIN analysis has been used for investigating the processes linked with disease and the association of AD with other conditions, such as type 2 diabetes [529-531]. Of note, recent PPIN studies focused on the molecular mechanism of AD, consistently included autophagy in the processes associated to the disease mechanism [529, 530, 532-534]. However, there is only one PPI study (to the best of my knowledge) in which autophagy was found to be important for dementia, more specifically for tauopathies via an analysis of an interaction network of disease proteins and drug target [535].

Similarly, PPIN analysis of PD has been used to study its association with conditions such as COVID-19 [536], and mitochondrial dysfunction [537], and its mechanism, with autophagy indeed confirmed relevant [155, 536-540]. However, PPIN has yet to be applied to specifically analyse the potential molecular overlaps of the pathway of autophagy with PD.

Therefore, in this chapter the overlap of the autophagy pathway with neurodegenerative diseases will be studied using the systems biology approach of PPIN analysis. Since, systems biology approaches facilitate a deeper understanding of complicated processes, like macroautophagy, such a study could suggest potential mechanism(s) through which it associates to neurodegenerative diseases. Identifying causal links between the two processes, and potential disease-modifying targets could improve the quality of life of people with neurodegenerative diseases, such as AD, PD, FTD and ALS.

4.1.3 Resources and bioinformatic tools for autophagy

There is a multiplicity of resources and databases that allow a systems biology study of macroautophagy. Several web-based resources are available for the prediction of the interaction of the protein of interest with the ATG8 family, a protein family essential for the elongation and maturation of the phagophore. They are based on the screening of an interaction motif, known as LCE-interacting region or ATG8-interacting motif or LC3 recognition sequence (LIR, AIM or LRS, respectively) in the protein sequence. Examples are the iLIR server [541], the high-fidelity AIM system [542], the Eukaryotic Linear Motif [543], and the ShortLinearMotif(SLiM)Search [544].

Additionally, there are databases that hold information about autophagy-related genes. Some examples are: iLIR Database, iLIR@viral [545], THANATOS Database [546], Human Autophagy Database (http://autophagy.lu/), Autophagy Database [547], ncRNA-Associated Cell Death Database [548], AutomiRDB [549] and the Gerontology-Autophagic-MicroRNA Database [550].

A systems biology approach was adopted in the creation of the Autophagy Regulatory Network (ARN), a bioinformatic tool for the study of the mechanism and the regulation of autophagy [551]. This database includes the proteins of the autophagy machinery, four sets of their regulators (direct regulators, transcriptional regulators, post-transcriptional regulators, and miRNA regulators and their transcription factors) and related signalling pathways. ARN allows much flexibility as the user is able to select the type of proteins of interest (as grouped above), but also the type of interaction data (manually curated and/or predicted).

4.2 Aims and Objectives

The aim of the rese4arch reported in this chapter is to understand whether neurodegenerative diseases and macroautophagy are interconnected and if so, the degree of association. To achieve this goal, PPINs of 4 neurodegenerative diseases, AD, PD, FTD and ALS will be produced, together with the PPIN of macroautophagy. Finally, the potential overlap of those networks will be evaluated to determine whether with an unbiased systems approach, an enrichment of autophagy proteins can be found in the protein interactome of neurodegenerative diseases.

4.3 Methodology

4.3.1 Selection of seeds for neurodegenerative diseases

The list of seeds for the neurodegenerative networks were kindly provided from Dr James Tomkins, who at the time was working in a comparative functional study of neurodegenerative diseases through a PPIN approach. The list of seeds was obtained by manual curation of the literature followed by expert evaluation of the findings. The seed list is presented in Table 4-1.

Table 4-1. Seeds of ND networks

Neurodegenerative disease	Gene name	SwissProtID	Gene name	SwissProtID
Alzheimer's disease (n=4)	*ADAM17* *APP*	P78536 P05067	*PSEN1* *PSEN2*	P49768 P49810
Amyotrophic lateral sclerosis (n=29)	*FUS* *SOD1* *ALS2* *SETX* *SPG11* *ANG* *FIG4* *ATXN2* *VCP* *UBQLN2* *SIGMAR1* *CHMP2B* *PFN1* *HNRNPA1* *MATR3*	P35637 P00441 Q96Q42 Q7Z333 Q96JI7 P03950 Q92562 Q99700 P55072 Q9UHD9 Q99720 Q9UQN3 P07737 P09651 P43243	*TUBA4A* *CHCHD10* *C9orf72* *NEFH* *PRPH2* *DCTN1* *DAO* *SQSTM1* *TBK1* *GLE1* *GLT8D1* *NEK1* *CFAP410* *KIF5A*	P68366 Q8WYQ3 Q96LT7 P12036 P23942 Q14203 P14920 Q13501 Q9UHD2 Q53GS7 Q68CQ7 Q96PY6 O43822 Q12840
Parkinson's disease (n=19)	*PARK7* *DNAJC6* *DNAJC13* *FBXO7* *GBA* *LRRK2* *PINK1* *PRKN* *RAB39B* *SMPD1*	Q99497 O75061 O75165 Q9Y3I1 P04062 Q5S007 Q9BXM7 O60260 Q96DA2 P17405	*SNCA* *SYNJ1* *ATP13A2* *PRKRA* *WDR45* *KANSL1* *KAT8* *GAK* *RAB29*	P37840 O43426 Q9NQ11 O75569 Q9Y484 Q7Z3B3 Q9H7Z6 O14976 O14966

Table 4-1. (continued) Seeds of ND networks				
Frontotemporal dementia (n=17)	*MAPT*	P10636	*DCTN1*	Q14203
	GRN	P28799	*FUS*	P35637
	C9orf72	Q96LT7	*TARDBP*	Q13148
	VCP	P55072	*TBK1*	Q9UHD2
	SQSTM1	Q13501	*TIA1*	P31483
	UBQLN2	Q9UHD9	*RAB38*	P57729
	IFT74	Q96LB3	*HLA-DRA*	P01903
	OPTN	Q96CV9	*TMEM106B*	Q9NUM4
	CHCHD10	Q8WYQ3		

4.3.2 Selection of seeds for the macroautophagy network

The selection of MA seeds was a multistep process, briefly described in Fig 4-1. The names of macroautophagy related proteins were collected from multiple sources (Reactome, KEGG, UniProt, Autophagy Regulatory Network, and Gene Ontology) and merged in a unique list of 511 genes that was then filtered, retaining 156 genes. Each gene had a score confidence depending on the confidence of its association with macroautophagy.

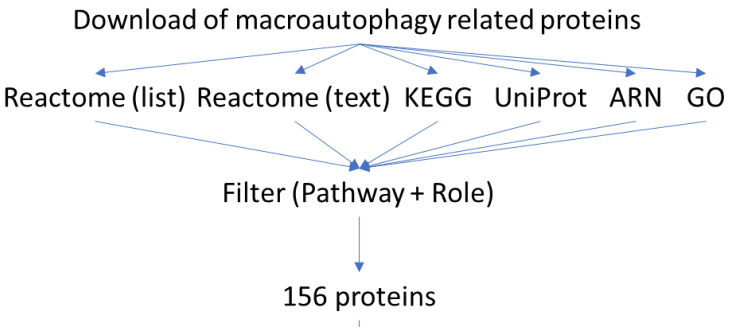


Figure 4-1. Brief overview of the seed selection process for the macroautophagy network

Sources included Reactome (both the list of components and the description of the process in text), KEGG, UniProt, Autophagy Regulatory Network (ARN), and Gene Ontology GO). The lists were brought together and filtered based on their replication across databases, the stage of macroautophagy in which they are involved, and their role.

In more detail, from the ARN website 233 genes were extracted on 14/01/2020, as they were labelled as core macroautophagy proteins (n=37) and direct regulators of core proteins (n=196) (score=1).

Information was downloaded from KEGG on 24/01/2020. The human autophagy pathway (hsa04140) consisted of 137 genes (score=1).

Gene Ontology contains lists of proteins involved with a series of processes labelled with GO-BP terms. The GO-BP term macroautophagy (GO:0016236) contained a range of related terms, the most relevant for this project being autophagosome (GO:005776), and regulation of macroautophagy (GO:0016241). The three groups of genes were extracted on 27/02/2020 and consisted of 192 genes (score=1).

Reactome had two main sources of information, a text describing the process, and a list of proteins. Within the text, two groups of proteins were identified, those that were considered to be part of macroautophagy and those that maybe were part of macroautophagy, so a score of 1 (n=72) and 0.5 (n=25) was attributed to them, respectively. The list of proteins from Reactome was first downloaded on 2/11/2018 and also on 9/1/2020, and the information from both versions was used. A higher confidence score was given to the proteins present in both versions (score=1, n=68), compared to those added in the second version (score=0.75, n=64). There were no proteins present in the first and not the second version. The total number of unique proteins from Reactome were 176.

UniProt is another database that contains a range of information about the processes in which genes are involved. Examples include the Description of function, Gene Ontology, and Location. The list of genes identified from the above sources was queried to extract information about a potential link with macroautophagy based on manual text mining from the Description of function section. The key used was “phag” and the results were evaluated to exclude any non-autophagy related terms. Information was retrieved from 130 genes (March 2020), with some having a clear relationship with macroautophagy (score=1, n=119), while for others the link was inferred by similarity with other macroautophagy associated proteins (score=0.5, n=11).

The 511 unique genes were then reduced further based on two criteria, namely confidence score filter (filter S), and functional contribution to the pathway of macroautophagy (filter P). In more detail, genes with a confidence score higher than 1.5 (filter S, n=189) were kept and further evaluated based on their contribution to the pathway (filter P). The “start” of macroautophagy was arbitrary considered to be the activated ULK1 complex and the “end” of macroautophagy was considered to be right before the fusion with lysosomes. Genes prior to the “start” or after the “end” of macroautophagy were removed. The direct regulators of proteins involved in the macroautophagy pathway were retained, while regulators of regulators were removed. This filtering process led to a final list of 156 genes (156/511, 30.5%) presented in Table 4-2.

Table 4-2. Seeds of the macroautophagy network

Gene	SwissProtID	Score	Gene	SwissProtID	Score
BECN2	A8MW95	2	*RAB23*	Q9ULC3	
GRAMD1A	Q96CP6	2	*STBD1*	O95210	2
MAPK10	P53779	2	*TOMM20*	Q15388	1.75
PRKACA	P17612	2	*TOMM22*	Q9NS69	1.75
RAB39B	Q96DA2	2	*TOMM40*	O96008	1.75
RAB8B	P61006	2	*TOMM5*	Q8N4H5	1.75
RUBCN	Q92622	4	*TOMM6*	Q96B49	1.75
RUFY4	Q6ZNE9	2	*TOMM7*	Q9P0U1	1.75
TMEM74	Q96NL1	2	*TOMM70*	O94826	1.75
TP53INP1	Q96A56	2	*VDAC1*	P21796	1.75
TP53INP2	Q8IXH6	3	*CAPN1*	P07384	2
TRIM21	P19474	2	*CAPNS1*	P04632	2
UBQLN4	Q9NRR5	2	*CLTC*	Q00610	2
DYNC1H1	Q14204	1.75	*FYCO1*	Q9BQS8	2
DYNC1I1	O14576	1.75	*MAP1B*	P46821	2
DYNC1I2	Q13409	1.75	*MAPK15*	Q8TD08	2
DYNC1LI1	Q9Y6G9	1.75	*MTMR14*	Q8NCE2	3
DYNC1LI2	O43237	1.75	*MTMR3*	Q13615	3
HDAC10	Q969S8	2	*PARK7*	Q99497	1.75
HDAC6	Q9UBN7	2.75	*RAB1A*	P62820	2
HTT	P42858	2	*SH3GLB1*	Q9Y371	4
MAPK8	P45983	4	*CAMKK2*	Q96RR4	2
TBC1D14	Q9P2M4	2	*STK11*	Q15831	2
TBC1D17	Q9HA65	2	*BCL2*	P10415	3
TBC1D25	Q3MII6	2	*BCL2L1*	Q07817	3
TBC1D5	Q92609	2	*CISD2*	Q8N5K1	2
TECPR2	O15040	2	*DYNLL1*	P63167	4
TMEM173	Q86WV6	2	*DYNLL2*	Q96FJ2	4
UBQLN1	Q9UMX0	2	*HMGB1*	P09429	3
UBQLN2	Q9UHD9	2	*NEDD4*	P46934	2
DAPK1	P53355	3	*NRBF2*	Q96F24	4
DAPK2	Q9UIK4	2	*PIK3CB*	P42338	2
DAPK3	O43293	2	*PIK3R2*	O00459	2
ITPR1	Q14643	2	*USP10*	Q14694	2
HSPA8	P11142	1.75	*LAMP2*	P13473	2
AMBRA1	Q9C0C7	5.5	*RAB33B*	Q9H082	2
ATG2A	Q2TAZ0	3	*RAB7A*	P51149	4
ATG2B	Q96BY7	3	*RAB7B*	Q96AH8	2
ATG9B	Q674R7	4.5	*TECPR1*	Q7Z6L1	2
MAP1LC3A	Q9H492	5	*CHMP2A*	O43633	2.5
MAP1LC3B	Q9GZQ8	5	*CHMP2B*	Q9UQN3	2.5
NBR1	Q14596	3	*CHMP3*	Q9Y3E7	2.5
OPTN	Q96CV9	2	*CHMP4A*	Q9BY43	2.5
PLAA	Q9Y263	2	*CHMP4B*	Q9H444	2.5
SQSTM1	Q13501	4.75	*CHMP4C*	Q96CF2	2.5
UBXN6	Q9BZV1	2	*CHMP6*	Q96FZ7	2.5
ULK2	Q8IYT8	4	*SNAP29*	O95721	3
VCP	P55072	2.75	*STX17*	P56962	3
VMP1	Q96GC9	3	*VAMP8*	Q9BV40	2
WDFY3	Q8IZQ1	2	*VPS11*	Q9H270	3
WDR45	Q9Y484	3.5	*VPS16*	Q9H269	3
WDR45B	Q5MNZ6	2.5	*VPS18*	Q9P253	4
YOD1	Q5VVQ6	2	*VPS33A*	Q96AX1	4

Table 4-2. (continued) Seeds of the macroautophagy network					
ZFYVE1	Q9HBF4	2	*ATG10*	Q9H0Y0	6
PRKAA1	Q13131	6	*ATG101*	Q9BSB4	6
PRKAA2	P54646	6	*ATG12*	O94817	6
PRKAB1	Q9Y478	3	*ATG13*	O75143	6
PRKAB2	O43741	4	*ATG14*	Q6ZNE5	6
PRKAG1	P54619	4	*ATG16L1*	Q676U5	6
PRKAG2	Q9UGJ0	4	*ATG3*	Q9NT62	6
PRKAG3	Q9UGI9	3	*ATG4A*	Q8WYN0	5
ATG16L2	Q8NAA4	2.5	*ATG4B*	Q9Y4P1	6
GABARAPL3	Q9BY60	2.5	*ATG4C*	Q96DT6	5
MAP1LC3B2	A6NCE7	2.5	*ATG4D*	Q86TL0	5
MAPK9	P45984	3	*ATG5*	Q9H1Y0	6
CALCOCO2	Q13137	2	*ATG7*	O95352	6
CSNK2A1	P68400	1.75	*ATG9A*	Q7Z3C6	6
CSNK2A2	P19784	1.75	*BECN1*	Q14457	6
CSNK2B	P67870	1.75	*GABARAP*	O95166	5
FUNDC1	Q8IVP5	2.75	*GABARAPL1*	Q9H0R8	5
KEAP1	Q14145	2	*GABARAPL2*	P60520	5
MAP1LC3C	Q9BXW4	5	*PIK3C3*	Q8NEB9	6
MFN1	Q8IWA4	1.75	*PIK3R4*	Q99570	6
MFN2	O95140	2.75	*RB1CC1*	Q8TDY2	6
MTERF3	Q96E29	1.75	*ULK1*	O75385	6
PGAM5	Q96HS1-2	1.75	*UVRAG*	Q9P2Y5	5
PINK1	Q9BXM7	2.75	*WIPI1*	Q5MNZ9	5.5
PRKN	O60260	2.75	*WIPI2*	Q9Y4P8	6

4.3.3 Building the first and second layer of the networks PINOT

Using PINOT and the list of seeds, the PPIs were collected (21/03/2020) and then filtered to only include those with a final score higher than 2, in order to increase the confidence in the results. The interactors of the first layer were then used as seeds for another round of PPI collection through PINOT (21&23/03/2020) to create the 2nd layer interactome following the same process. The networks were then visualised in Cytoscape, as described in previous chapters.

4.3.4 Statistical analysis

The statistical analysis was performed though R, similarly to the analysis of word enrichment of described in Section 2.3.3.

4.4 Results

4.4.1 MA network

The macroautophagy (MA) network was built based on PPI data collected through PINOT for the identified seeds. It is composed of 151/156 seeds (96.8%, 746 PPIs), 2,832 direct interactors (1st layer; 56,982 PPIs) and 9,377 indirect interactors (2nd layer; 52,792 PPIs), bringing the total number of proteins to 12,360 (110,520 PPIs).

The 5 missing seeds from the MA network were *ATG16L2*, *GABARAPL3*, *TMEM74*, *TOMM6* and *TOMM7*. For *GABARAPL3* there were not any PPIs resulted from PINOT, while for the rest of the MA seeds some PPIs were collected, 19, 2, 4, and 13 respectively. However, the PPIs had a score of 2, meaning that were identified in a single publication with a single method, so there were excluded from further analysis.

In the MA network, the 10 genes with the highest centralities were extracted. For closeness centrality the list in a descending order is: *CUL3*, *COPS5*, *TP53*, *EGFR*, *HSPA8* (MA seed), *CUL1*, *CDC5L*, *SNW1*, *HSP90AA1*, and *UBC*. For betweenness centrality: *CUL3*, *EGFR*, *APP*, *COPS5*, *HSPA8* (MA seed), *TP53*, *FN1, CDC5L*, *MYC*, and *ESR1*. For node degree: *CUL3*, *COPS5*, *CAND1*, *CUL1*, *CDC5L, TP53, EGFR, ESR1, SNW1*, and *FN1*. The percentages of MA seeds in these lists are: 1%, 1% and 0%, respectively. The degree of most nodes is below 500 (99.9%), but one node, CUL3, reaches as many as 1,429 interactors. The distributions of the node degree, betweenness centrality and closeness centrality are shown Fig S4-1.

4.4.2 ND networks

The neurodegenerative networks (ND networks) were created in a similar manner for AD, PD, ALS, and FTD. The AD network was produced using 4 seeds (see Section 4.3.3) all of which were present in the network. The 1st layer interactors were 466 and the 2nd layer interactors were 6,419. The seed with the highest node degree was APP, which had 415 interactors, followed by PSEN1 with 45 interactors.

The ALS network was created based on 29 seeds. The seed *PRPH2* did not result in any PPIs produced by PINOT. The remaining 28 seeds, however, resulted in 1,102 1st layer interactors and

9,095 2nd layer interactors. The seed with the highest node degree was *SOD1*, followed by HNRNPA1 with 223, and 197 interactors, respectively.

The FTD network was produced using 17 seeds all of which gave PPIs that contributed to the network. The 1st layer interactors were 702 and the 2nd layer interactors 8,540. The seeds with the most interactors were *VCP* and *SQSTM1* with 142, and 126 interactors, respectively.

The PD network was created using 19 seeds, all of which produced PPIs kept in the core PD network. The network consisted of 791 1st layer interactors and 8,583 2nd layer interactors. The seeds with the highest node degree were *LRRK2* and *PRKN* with 227 and 142 interactors respectively.

4.4.3 Overlapping the ND seeds in the MA network

The MA network was explored to investigate whether the ND seeds where part of it, and if so in which layer. The results show that most seeds were part of the MA network and are presented in Table 4-3 and Fig 4-2A. All the seeds of AD were part of the MA network, with 3 out of the 4 being in the 1st layer interactors and 1 being in the 2nd layer (75.0%, and 25.0%, respectively), meaning that they either interact directly or through one protein with main MA proteins. Similarly in PD, all seeds were part of the MA network with 5, 8 and 6 in the seeds, 1st, and 2nd layer interactome respectively (corresponding to 26.3%, 42.1%, and 31.6%, respectively). Conversely, some seeds from ALS and FTD were absent from the MA network (n_{ALS}=3, 10.3%; n_{FTD}= 1, 5.9%). The rest were present in all three layers of the network, seeds, 1st layer and 2nd layer interactors. More specifically, 4, 10 and 12 ALS seeds were part of the respective layers, while for FTD they were more evenly distributed with 4, 6 and 6, respectively.

A statistical analysis was performed to assess whether the overlap of the ND seeds in the MA network was significant. Briefly, a group of random genes equal in number to the list of neurodegenerative genes under analysis was selected and it was investigated whether they were part of the MA network. The number of random matches to the MA network was recorded and the process was repeated 100,000 times, thus obtaining their distribution. The p-value corresponding to the observed number of matches of the ND seeds with the MA network was calculated based on the random distribution. The overlap between AD and the MA network was not significant (p=0.059), in contrast to the rest of the neurodegenerative diseases, including the overall group of ND seeds (see Table 4-3, and Fig S4-2, Appendix D).

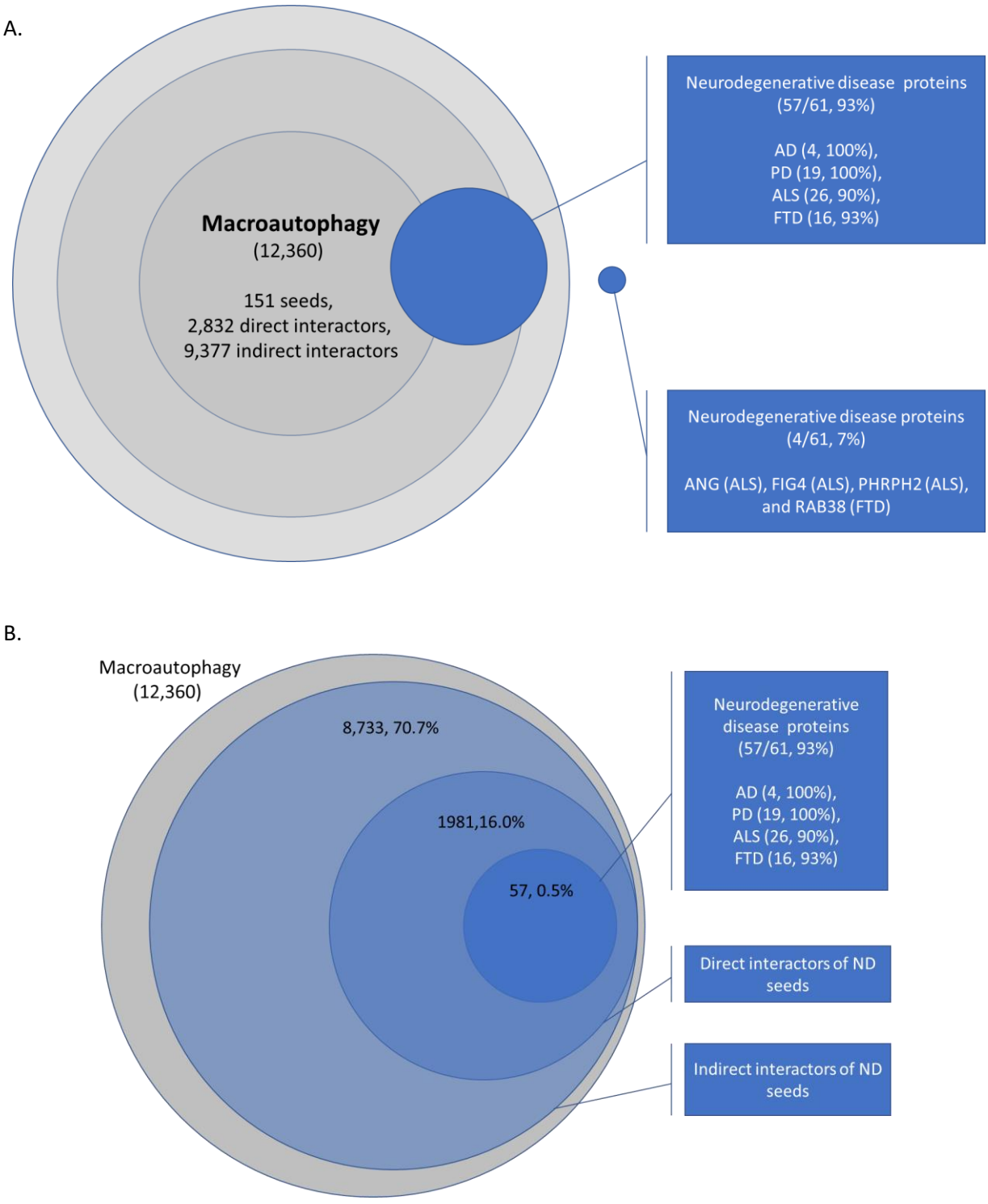


Figure 4-2 Overlap of ND seeds with the MA network

*(**A**) ND seeds for AD, PD, ALS and FTD present and absent from the MA network. (**B**) ND seeds and their direct and indirect interactors within the MA network*

Table 4-3. ND seeds present in the MA network

Neurodegenerative disease	Present in the MA network	Percentage	p-value
Alzheimer's disease (n=4)	4	100%	5.86×10^{-2}
Parkinson's disease (n=19)	19	100%	3.16×10^{-4}
Amyotrophic lateral sclerosis (n=29)	26	89.7%	1.20×10^{-5}
Frontotemporal dementia (n=17)	16	94.1%	5.90×10^{-4}
Total (n=61)	57	93.4%	1.2×10^{-12}

Interestingly, the ND seeds were directly interacting with 1,981 proteins, which constitute 16.0% of the MA network and indirectly (through one protein) interacting with 8,733 proteins (70.7%) (Fig 4-2B). Overall, the ND seeds and their direct and indirect interactors (n= 10,771) constitute 87.1% of the MA network, demonstrating the strength of the association between neurodegeneration and macroautophagy.

A more detailed analysis followed to identify in which layer of the MA network the ND seeds were located, and the results are shown in Table 4-4.

The four seeds that were not present in the MA network were *ANG* (ALS seed), *FIG4* (ALS seed), *PRPH2* (ALS seed), and *RAB38* (FTD seed). The third gene was expected to be absent from the network, as it failed to result in any PPIs through PINOT, as previously mentioned. Interestingly, the rest of the missing seeds had a high percentage of their interactors being present in the 2nd layer of the MA network (see Table 4-5, and for more details see Table S4-1). For instance, *RAB38* had a single direct interactor (*RAB32*), which was in the second layer interactome of 22 MA seeds. Therefore, the overlap of the ND networks with the MA network was also investigated.

Table 4-4. ND seeds present in each layer of the MA network

Neurodegenerative disease	MA seed	1st layer	2nd layer
Alzheimer's disease (n=4)	0 (0%)	3 (75.0%)	1 (25.0%)
Parkinson's disease (n=19)	5 (26.3%)	8 (42.1%)	6 (31.6%)
Amyotrophic lateral sclerosis (n=26)	4 (15.4%)	10 (38.4%)	12 (46.2%)
Frontotemporal dementia (n=16)	4 (25.0%)	6 (37.5%)	6 (37.5%)
Total (n=57)	10 (17.5%)	24 (42.1%)	23 (40.4%)

Table 4-5. Connectivity of direct interactors of the missing ND seeds with the MA network

ND seed	Direct interactor	N MA seeds connected through two proteins
ANG (n=6/7, 85.7%)	*ACTN2*	9
	ANXA2	55
	FST	0
	PCNA	70
	PLAUR	13
	RNH1	24
	S100A10	22
FIG4 (n=4/4, 100%)	*ANK1*	7
	PIKFYVE	2
	SNX27	12
	VAC14	9
PRPH2	*0*	-
RAB38 (n=1/1, 100%)	*RAB32*	22

Note: For more details see Table S4-1 (Appendix D)

4.4.4 Overlapping the ND networks in the MA network

It was investigated whether and if so in what degree each ND network was overlapping with the MA network. The analysis was performed by counting the number of the components of each layer of the ND network were part of each of the 3 layers of the MA network (Tables 4-6 and Table S4-2, Appendix D). For instance, while none of the seeds of AD were part of the seed list of MA, its 1st layer had 155 proteins (33%) in the 1st layer of MA, and its 2nd layer had 6,274 (97.7%) in the 2nd layer of MA. Interestingly, all studied NDs had a high overlap of their 2nd layer with the respective 2nd layer of MA (ranging from 94-98%), while the overlap of the 1st layers was more limited (ranging from 33-68%).

Table 4-6. Distribution of the ND network in the MA networks (percentages)

	MA			
		Seeds	1st layer	2nd layer
AD	Seeds	0.0%	75.0%	100.0%
	1st layer	2.1%	33.3%	99.8%
	2nd layer	1.3%	32.5%	97.7%
ALS	Seeds	14.3%	50.0%	92.9%
	1st layer	3.9%	54.7%	99.1%
	2nd layer	1.0%	24.4%	97.1%
FTD	Seeds	23.5%	58.8%	94.1%
	1st layer	5.0%	67.8%	98.4%
	2nd layer	1.1%	26.6%	98.4%
PD	Seeds	26.3%	68.4%	100.0%
	1st layer	4.2%	63.1%	99.5%
	2nd layer	1.1%	25.9%	97.9%

Note: For the absolute values of the distribution of the ND network in the MA networks see Table S4-2 (Appendix D)

4.5 Discussion

In this chapter the association between neurodegenerative diseases and macroautophagy was studied. Investigating such an association in which a large number of genes are involved is a task for which systems biology approaches can be utilised, such as PPIN analysis.

4.5.1 Selection of seeds

The process of creating of a PPIN initiates with choosing the genes that will be its seeds. This step was more complicated than expected for the macroautophagy network (MA network), as there was very little agreement of the identity of the main macroautophagy proteins across databases. More specifically, the sources used were ARN, KEGG, Reactome, GO and UniProt, which resulted in 233, 137, 176, 192 and 130 genes, respectively. These had limited overlap, as the unique list of the merged genes was consisted of 511 genes. This is unfortunately a common problem in biology, as identifying a main protein of a process is prone to unconscious biases based on the interests and background of each researcher, but also sensitive to the choices of the borders of the biological process, the level of detail, and the level of regulation to be included in the process. To limit these biases, multiple sources were used, and filters were applied, which led to a list of 156 seeds based on which the MA network was constructed, using PPIs collected through PINOT.

4.5.2. MA network and ND networks

The MA network consisted of 12,360 components, which were 151 seeds, 2,832 direct interactors, and 9,377 interactors of the direct interactors. The missing seeds were *ATG16L2, GABARAPL3, TMEM74, TOMM6* and *TOMM7*. The main identified reason for their absence was the limited data of PPIs, as they had 19, 0, 2, 4, and 13 PPIs with a final score lower than 3. Further studies on these proteins aimed at identifying new or verifying previously suggested PPIs could aid in the endeavour of understanding their connection with macroautophagy through PPIN analysis.

In the MA network a few proteins had a high number of connections with the rest of the network. *CUL3* had 1,429 interactors, while 99.9% genes had less than 500. This is a consequence of its biological role, as it is involved in multiple processes, including the ubiquitination of proteins targeted for proteasomal degradation [552-554]. *COPS5* is part of the COP9 signalosome complex, which is involved in multiple cellular and developmental processes, such as apoptosis, and the

regulation of the JNK signalling pathway [555, 556]. As par the complex, it interacts directly with a large number of proteins that are regulated by it, so its high node degree is not surprising [556-563]. Similarly, the following two proteins with the highest node degree, CAND1 and CUL1, are associated with the ubiquitin system, explaining this result [554, 564]. CDC5L was the 5th protein with the highest node degree. It is a DNA binding protein involved in cell cycle control, while it may also have a transcription activation role [565-573]. Interestingly, the only disease it has been strongly associated with is multicystic renal dysplasia [574], which has not been clearly linked to macroautophagy or neurodegeneration. A potential link of this disease to macroautophagy can be deduced by its association with BCL2 [575], which regulates a variety of processes including macroautophagy [318, 576, 577]. This potential link of CDC5L with macroautophagy could be further studied by researchers to elucidate this probable relationship.

Networks of NDs were also constructed for ALS, AD, FTD, and PD, due to their previously suggested association with macroautophagy [484, 487, 491, 511, 514]. Most seeds were present in the networks, except for PRPH2, which was a seed for ALS. Querying this gene in PINOT did not result in any PPIs, which can be explained by the protein being understudied. However, this is surprising because PRPH2has been linked with ALS for almost 20 years [578, 579]. Further studies around this gene and its product could lead to a deeper understanding of its function and potentially the mechanism through which it is implicated in complicated human diseases, such as ALS.

The rest of the ND seeds resulted in PPIs that were included in the respective ND networks, which were consisted of between 6,400 and 9,100 proteins. The size of the networks is unproportionate to the number of seeds: 4-19 seeds led to a network made of 6,400-9,100 proteins, while 156 seeds to a network of 12,360 proteins. This can be explained by the focus of the scientific community in studying human diseases rather than biological processes, resulting in more PPIs detected for genes that lead to neurodegeneration rather than macroautophagy.

4.5.3 Overlaps

The overlap of macroautophagy and neurodegeneration was investigated with two ways: (i) identifying the presence of ND seeds in the MA network, and (ii) studying the intersection of each ND network with the MA network.

The first approach demonstrated that most ND seeds were part of the MA network. All the seeds of AD, and PD were part of the network, while just a few seeds were missing for ALS and FTD (n_{ALS}=3, 10.3%; n_{FTD}= 1, 5.9%). The presence of these seeds was statistically significant for all NDs except for

AD, demonstrating the strength of the association of neurodegenerative diseases with macroautophagy. Interestingly, the distribution of the seeds in the 3 layers of the network (i.e., seeds, 1st layer interactome, 2nd layer interactome) varied among the diseases. In AD and PD most were in the 1st layer; in ALS most were part of the 2nd layer, while in FTD there was an even distribution between 1st and 2nd layer. In addition, the interactions of the ND seeds within the MA network were explored, which demonstrated that they directly interact with 16% of the MA network, further supporting the association between MA and NDs.

Some ND seeds were absent from the MA network. However, most of them had a direct interactor, which was simultaneously a 2nd layer interactor in the MA network. Interestingly, the average number of MA seeds to which they connected was 19 (ranged from 0 to 70). This result led to the analysis of the overlap of the MA and ND networks.

Of note, most ND networks were almost completely embedded within the MA network. For instance, 99.1% of the first layer interactome and 97.1% of the 2nd layer interactome of ALS were part of the 2nd layer interactome of MA. It can be hypothesised that this is the case because some seeds of ALS were part of the seed list of MA. However, the same result can be observed for AD, none of whose seeds are also MA seeds, whereas its 2nd layer is almost completely embedded in the 2nd layer of MA (97.7%).

These results further support previously published literature demonstrating the association of neurodegeneration with macroautophagy through a systems approach. A noteworthy study used a PPIN approach was used to study the association of macroautophagy with multiple diseases including a category of neurological diseases [527]. In this study 770 autophagy genes were used, which exceed the total number of 511 macroautophagy related genes identified with our methodology prior to the filtering process. Their list of autophagy genes included predicted human autophagy genes, for instance, based on the association of their homologs with autophagy. Furthermore, the overlap of autophagy with diseases was measured only though identifying autophagy genes that were also recognized as causative for human diseases. This resulted in detecting only 9 genes that link autophagy and “neurological conditions”, with a 1.52 enrichment ratio. Therefore, the full potential of PPINs was not used in this study. However, interesting results included the suggestion that autophagy genes are linking different categories of diseases, such as cardiovascular and respiratory diseases, which is in accordance with the results of the current thesis, as autophagy seems to link with multiple neurodegenerative diseases. This is also supported by another study that demonstrated that two genes that can lead to PD are linked through macroautophagy, as parkin leads to the degradation of UCH-L1 by the autophagy-lysosome

pathway [580]. Several other PD genes have been linked with autophagy, including *SNCA* and *DJ-1* [486-488, 490-493]. However, the precise mechanism is yet to be deciphered, as a study showed that in a mouse model of PD, dysfunction of autophagy led to improved motor ability, even though progressive neuron loss was observed in substantia nigra pars compacta [581].

Overall, the current study created more evidence to support that neurodegeneration and macroautophagy are strongly linked, using the systems biology approach of PPIN analysis. In contrast to the initial expectation, the overlap of neurodegenerative disease could not be localised in a small part of the macroautophagy network. Therefore, a different approach should be used to study the potential mechanism through which mutations in genes that lead to the development of neurodegenerative diseases affect macroautophagy, such as mathematical modelling.

5th Chapter:

A process diagram of macroautophagy

5. A process diagram of macroautophagy

Main points of this chapter:

- The main aim of this chapter is to design a diagram of the process of macroautophagy that will allow the building of a mathematical model in Chapter 6.
- Basic data sources: REACTOME, KEGG, and recent literature
- Borders of model:
    - Stage of macroautophagy: Initiation to Phagophore elongation
    - Biological role of proteins: Key players of macroautophagy and their direct regulators
- The study of the association of macroautophagy with neurodegeneration was narrowed to its link with Parkinson's disease and specifically LRRK2.
- The effect of LRRK2 on macroautophagy will be simulated using data showing alterations in its protein interactions when it LRRK2 is mutated.
- The final main diagram of this chapter is composed of 23 components. 10 additional components link the diagram with LRRK2.

5.1 Introduction

5.1.1 Mathematical models of components of macroautophagy

The systems biology approach of mathematical modelling has been previously applied to investigate macroautophagy and related proteins. A model of mammalian basal and induced macroautophagy was created by Martin et al to describe autophagic vesicle dynamics in single cells [582]. Data collected from live-cell microscopy were used to build and refine the model. The first step of the model was the stimulus dependent activation of PIK3C3-C1 and its final step was the turnover of autophagic vesicles. The model accurately predicted the change in vesicle synthesis rate due to a depletion in ATG9, as confirmed by in vivo experiments.

Sakai and colleagues focused on the membrane dynamics of phagophore formation [583]. Components of this model included proteins, complexes and lipids that affect the curvature of the phagophore membrane (named curvature generators) as well as the bending energy of the membrane. The stabilisation of intermediate structures (disk-shaped and cup-shaped) was shown to facilitate the formation of the spherical phagophore, which was further supported by in vivo experimental data from mouse embryonic fibroblasts, used by the authors to estimate the membrane area and bending angle of autophagosomes.

These models demonstrate the value of investigating complicated biological processes -such as macroautophagy- using mathematical modelling, as this approach can reveal novel insight to be used as guidance for further experimental research.

5.1.2 Mathematical models of neurodegenerative diseases

Multiple mathematical models have been developed to investigate neurodegenerative diseases. For instance, a model developed to investigate the pathogenesis of Alzheimer's disease described the cross-talk between the amyloidogenic and non-amyloidogenic pathway after administration of secretase inhibitors to predict the Aβ production rate and plasma levels [584]. Their predictions were validated by measuring plasma $A\beta_{40}$ levels from human volunteers. Another model focused on the role $A\beta_{42}$ in developing Alzheimer's disease and the potential benefit of immunotherapy [585]. Through stochastic modelling, the $A\beta_{42}$ turnover and accumulation were simulated and demonstrated that antibodies could delay AD onset by reducing the level of $A\beta_{42}$ dimers.

Although a number of models have been distinctly developed to study neurodegenerative diseases or macroautophagy, as previously discussed, only few models have investigated their relationship. One example is the modelling of α-synuclein dynamics based on biomolecular reactions, describing its overexpression, post-translational modification, oligomerization, and degradation [162]. The effect of chaperone-mediated autophagy, macroautophagy and the proteasome system were included in the simulated degradation process of α-synuclein. Predictions of the model were validated based on experimental data from human cell lines.

The increasing body of evidence supporting an interplay between autophagy and neurodegeneration [586-596] highlights the need for the development of more sophisticated mathematical models that study the potential causality between these two processes. Therefore, as part of this Thesis, a mathematical model of autophagy and more specifically macroautophagy

will be built. Based upon data linking *LRRK2* to macroautophagy [597, 598], this investigation will focus upon and test how mutations associated with Parkinson's disease in this gene impact upon macroautophagy.

5.1.3 LRRK2 in Parkinson's disease

LRRK2 is a gene containing 51 exons, encoding a large protein with a size of around 280kDa. It has two enzymatic domains, a kinase and a GTPase domain, and multiple PPI domains [599]. It has been hypothesized that LRRK2 is involved (among other functions) with the regulation of protein trafficking through the endosomal pathway, macroautophagy, and the function of lysosomes and synapses [215, 216, 600-602].

Mutations in *LRRK2* were first identified as causative for PD in 2004 [215, 216] and are one of the most common genetic causes of this disorder [603]. There is variability related with both their frequency in populations and effect size. Regarding the former, some populations have a high frequency of *LRRK2* mutations that cause PD, such as the Ashkenazi Jews, and the Imazighen and Timazighin (previously named North African Berber) [603]. Regarding the latter, out of 1,697 variants of *LRRK2* (as resulted from the Genome Aggregation Database, also known as gnomAD [604], on 27/5/2022), only a few have a clear causal relationship with PD (e.g., G2019S, R1441C) [215, 216, 605, 606]. The majority of pathogenic *LRRK2* mutations are located in its catalytic core, affecting its enzymatic activity [607-609]. Interestingly, a protective LRRK2 variant (R1398H) has also been identified in multiple populations [610, 611].

The clinical presentation of PD associated with LRRK2 is similar to that of idiopathic PD but with some distinctions. Even though most people with mutations in *LRRK2* have Lewy body pathology, some have either neurofibrillary tangles [215, 216, 612-614] or TDP-43 deposition [615], which are more closely linked to other neurodegenerative diseases, such as frontotemporal dementia [616, 617]. There are still a lot of unanswered questions regarding LRRK2's function, and the exact mechanism through which it leads to PD is yet to be deciphered.

5.2 Aim and Objectives

The aim of this part of the research is to create an accurate but comprehensive diagram of macroautophagy that encapsulates the core components and events in humans, providing the

foundations for the construction of a dynamic model (to be presented in Chapter 6). This model will be used to test the impact of LRRK2 upon macroautophagy in a healthy *versus* disease state. The latter will be built based on alterations in bidirectional interactions of LRRK2 with other proteins, when LRRK2 is mutated.

The first objective for this endeavour is to assess the availability of relevant data, which will inform the decision of defining the boundaries of pathway to be modelled and other features of the model and therefore of the diagram. The second objective is to construct a diagram that represents macroautophagy, using data from databases, literature, and previous data from Chapter 4. Data from LRRK2 interactors of interest will also be incorporated in the diagram. Finally, the diagram will be evaluated and simplified to only retain essential components and relationships.

5.3 Methodology

5.3.1 Comparison of PD genes

For a comparison of data availability among genes leading to PD (which were previously identified in Section 4.3.1), the number of results in PubMed (a), number of PPIs (b), and number of mutational data affecting their PPIs (c) were collected.

(a) The literature search through PubMed was performed by querying the official gene name (11/2/2022).

(b) The comparison of PPI data was performed using the downloaded data of Chapter 4 (see Section 4.3.3) by querying the official gene name in the unfiltered data set and counting the total number of PPIs for each PD gene.

(c) The search of mutational data was performed using the data set from IntAct (queried on 12/07/2021) based on the official gene name and by counting the total number of results.

5.3.2 Construction and simplification of macroautophagy diagram

5.3.2.1 Basis of the diagram

The basis of the macroautophagy diagram was the two equivalent diagrams from Reactome (autophagy; *Homo sapiens*; R-HSA-9612973) and KEGG (autophagy; animal; Homo sapiens; hsa04140) downloaded on 24 & 30/1/2020, respectively (Fig S5-1 & S5-2, Appendix E). The

diagrams were filtered and simplified, as shown in Fig 5-1. The filter was based on the borders of the process, as defined in Section 4.3.2. Where appropriate, the diagrams were also simplified to remove intermediate steps that add limited value to the model; such simplifications were performed only in the diagram from Reactome, as the diagram from KEGG lacked similar intermediate steps. For instance, the creation of LC3II is depicted as the result of 6 reactions in Reactome, but was simplified to 2: (i) ATG4 cleaves pro-LC3 to LC3I, and (ii) ATG7, ATG3, and ATG12:ATG5:ATG16L1 produce LC3II from LC3I. The filtered and simplified diagrams from the two sources were then merged by keeping all the available information.

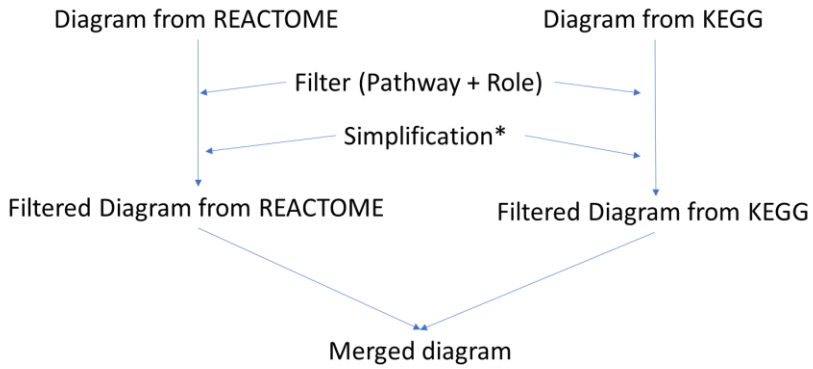


Figure 5-1. Processing of the diagrams from Reactome and KEGG

5.3.2.2 List of references

A list of references was used to enrich the diagram. The list was acquired on 24/05/2021 through PubMed, using the keyword "macroautophagy" and the following criteria: 2018-2020, Review/Systematic review, Free full text, Humans, and English. Any papers that were exclusively about one protein or complex were removed, while papers that were specifically about macroautophagy (excluding Chaperone Mediated Autophagy and Microautophagy) were retained, resulting in 5 papers. The consensus document "Guidelines for the use and interpretation of assays for monitoring autophagy" [34] was also added to the reference list leading it to be consisted of 6 papers [34, 618-622].

5.3.2.3 Simplifications

The simplifications of the macroautophagy diagrams were based on 2 main steps. Firstly, entities of the diagram were omitted if they had only one outbound and one inbound relationship, and their contribution to the biological process was retained in the diagram, by incorporating an equivalent relationship between the directly upstream and directly downstream entities (bypassing the

intermediate entity). Secondly, branches of the diagram through which information did not flow (i.e., nodes with only unidirectional relationships: outbound or inbound) were also removed. The exception in this step was the activated ULK1 complex (as it is the first component of the pathway for this project), LC3II (as it is the last entity of the model and diagram), and the LRRK2 interactors (as no entities upstream of LRRK2 will be included in the model and diagram).

5.3.3 Analysis based on LRRK2

5.3.3.1 Processing of LRRK2 data

After LRRK2 was chosen as the PD gene of interest, the collected PPIs and mutational data were further analysed. The PPIs of LRRK2 were filtered to only include those with final score >3 (thus only retaining interactions that have been replicated at least twice in peer-reviewed literature) to increase the confidence in the results. Regarding the mutational data analysis for LRRK2, multiple filters were applied in this analysis. Only mutational data that affected a binary interaction between two human proteins were included. If the effect of the mutation was labelled with general terms (i.e., “Mutation with no effect” MI:2226; and “Mutation” MI:0118), then that piece of information was excluded from further analysis. Interactions that included GTP or GDP were retained due to the function of LRRK2, but interactions in which other chemical substances were present, were excluded from further analysis. After filtering the mutational data set, a unique list of affected interactions was created.

Data regarding protein-protein interactions affected by mutations in a protein are available in IntAct. LRRK2 interaction data were extracted from IntAct on 12/07/2021 and subsequently filtered. The criteria for exclusion included: (i) interactions of non-human proteins, (ii) more than 2 protein participants, and (iii) mutations of the category “Mutation with no effect” (MI:2226), and “Mutations” (MI:0118).

5.3.3.2 Identifying associations with macroautophagy

The most relevant interactors of LRRK2 where those with associations with the macroautophagy pathway. For the mining of information of associations between LRRK2 and macroautophagy, or between LRRK2 interactors and macroautophagy, the reference list used to build the macroautophagy diagram [34, 618-622] (see Section 5.3.2.2) was mined for data. Data supporting associations with macroautophagy were investigated firstly by reading in detail the papers (excluding the “Guidelines for the use and interpretation of assays for monitoring autophagy” [34]),

and also by querying the official gene names of *LRRK2* and the *LRRK2* interactors, and then manually extracting any relevant information. For the mining of information from "Guidelines for the use and interpretation of assays for monitoring autophagy" [34] only the latter step was performed due to the number of pages of the document (i.e., 549 pages). The LRRK2 interactors were grouped in 3 categories based on the collected data: "yes" group (evidence of association with macroautophagy), "?" group (possible association with macroautophagy), and "no" group (no data supporting an association with macroautophagy). For the "?" group, their description in UniProt was also consulted, leading to their categorising as either "yes", or "no", depending on the presence or absence of data supporting their association with macroautophagy.

A common problem in biology is the existence of alternative names for genes and proteins [26]. Therefore, it was necessary to exclude the possibility of the lack of supporting data for the genes of the "no" group being due to the use of an alternative gene name, a protein name, or the name of a complex in which they take part. In parallel, it was investigated whether their UniProt page had macroautophagy related terms in their description or in the associated biological processes. Text mining for the term "phag" was performed and the results were manually checked for macroautophagy related terms (e.g., positive selection of "regulation of macroautophagy", GO:0016241; negative selection of "phagocytosis", GO:0006909). If data supporting the association were obtained, then the gene could be moved to the "yes" group.

In some cases, the information collected from the UniProt pages and reference list hinted a link with macroautophagy but not in a specific enough way to be able to add it in the diagram. For those, the additional step of searching for information in the references of either UniProt or the Guidelines was performed. If that did not lead to a conclusion, they were added in the "?" category.

The proteins in the "?" group were investigated further in the literature, leading to none being included in the diagram. There were 4 main reasons: (i) absence of association with any other component of Enriched Diagram B, (ii) absence of evidence in humans, (iii) lack of specific information as to how they are linked to macroautophagy, or (iv) existence of too many complicated links with macroautophagy. The last category was populated by proteins in the proteasome and tubulins that have multiple protein-protein interactions and cross-talks with macroautophagy proteins. Including proteins of this category would drastically increase the number of proteins of the macroautophagy model and its complexity, which contradicts the need of including only the most necessary entities and relationships in the current diagram.

5.3.4 Software and databases

The software used in this chapter to create the diagram of macroautophagy was Cell Collective (online version for researchers, accessed on September of 2021) [623]. Information about macroautophagy that was retrieved from databases for Chapter 4 was also used here (see Sections 4.3.2 and 4.3.3).

5.4. Results

The results of creating the macroautophagy diagram will be structured following the steps of modelling. The first step of modelling is outlining the purpose of the model, which has been completed (see Section 5.2), which guides the decision of selecting the type of the model and the classes of its components. The available data will then be explored, based on which the scale and other features of the model will be defined. Next, the focus will be in designing the diagram in detail. This step will be followed by enriching the diagram with data related to LRRK2 interactors, adjusting the diagram based on feedback from an expert in autophagy, and filtering the data to only include those that contribute to the flow of information. The end result will be the final diagram of macroautophagy, which will be used as a basis for the macroautophagy model to be created in Chapter 6.

5.4.1 Selecting the type of the model type and the classes of its components

The type of the model for a series of attributes was chosen and is reported in Table 5-1. Regarding the first choice, an explanatory model was considered to be more suited, as it can capture relationships between components and provide explanations, which is aligned with the objectives of this model. The model was also chosen to be dynamic, as a part of the signalling pathway of macroautophagy needed to be simulated in the model. In this dynamic model, the timescale was chosen to be continuous in order to allow detection of even small temporary concentration changes. For simplicity, no outputs or input signals were part of the model, instead the timepoint start with the most upstream entity being in an activated state. Finally, no randomness is incorporated in the model, so it is deterministic. Selecting the features of the macroautophagy model to be created in Chapter 6 also guides the creation of the macroautophagy diagram to be formulated in the current chapter.

Table 5-1. Type of model based on a series of features	
Attribute	**Choice**
Correlative vs Explanatory	Explanatory
Dynamic vs Static	Dynamic
Continuous or Discrete timescale	Continuous
Open vs Closed	Closed
Deterministic vs Stochastic	Deterministic

Another important decision is related to the classes of components that will be part of the diagram and model. Proteins and complexes will be represented in a specific state of activation as single variables. An additional variable will be representing any other activation state of those components and the transformation from one state to the other will depend on another entity of the model, or time. Lipids will not have an activation state, but they can transform to another entity and their concentrations can change over time (i.e., variables). This multivariate model will also include two independent variables, time and the mutated LRRK2, which will be affecting other entities, but itself will be stable throughout the experiment.

Regarding the types of reactions of the model connecting the variables, there will be of 3 main types: activation/inhibition, formation/disassociation, and production/degradation. The production or degradation will refer to the creation of a component in the model without the need for a substrate, or the decrease of its amount, respectively. In this case, the removal of the need for a substrate is not biologically accurate but it reduces the complexity of the model. In all types of relationships another component of the model can be a modifier or co-factor, meaning that it can affect the rate of these reactions directly or by adding a parallel reaction and therefore change the overall rate of the process.

Parameters will also be included in the model. These will be mainly the rates of the aforementioned reactions, as well as the initial quantities of the components of the model (i.e., initial conditions). No universal constants will be part of this model. As for the equations describing the model, they will be deterministic ODEs.

5.4.2 Exploring available data to define the model

5.4.2.1 Available data for macroautophagy

There are multiple sources of data regarding macroautophagy. The primary source of information is the peer-review literature. In PubMed, the query for macroautophagy resulted in 105 publications in 2020, 60 in 2021 and 6 in 2022 (on 9/2/2022). "Autophagy", frequently used as an interchangeable term with "macroautophagy" [624], resulted in 2,070, 2,198, and 188 publications respectively (on 9/2/2022).

In addition, there are databases in which diagrams and information regarding biological processes, such as macroautophagy, can be found. Two of those sources that hold a brief description of macroautophagy together with an informative diagram are Reactome and KEGG (Fig S5-1 & S5-2, Appendix E). There are 150 molecules, including 137 proteins in the diagram from Reactome, while in KEGG there are 147 components, including 141 genes. These databases are informative; however, they are updated semi-regularly. The entry of macroautophagy from Reactome was last revised on 31st of October 2019 (at the time of writing this paragraph, April 2022), while that information from KEGG is not available. However, the most recent article that was used to create the diagram of KEGG was published in August 2015; since then > 8,500 publications related to autophagy have been released (based on PubMed queried on April 2022 using the term "autophagy").

A more dedicated resource for autophagy is the Autophagy Regulatory Network [551]. It is a manually curated database focused on autophagy components and their direct interactions. The incorporated data spans from lists of main autophagy proteins to direct regulators, transcription factors, miRNAs, and pathways. Predicted autophagy regulators are also included to prompt their further investigation by the autophagy community. There are a total of 397,764 interactions stored in the database (on 9/2/2022).

Regarding existing models, there are multiple models in a database called BIOMODELS [625]. It was established in 2005 from the EBI [626] and has become a core resource for modelling. There are currently (as of April 2022) 1,041 manually curated, 1,309 non-curated, and 833 automatically generated models, as well as 542,964 model parameters, which is rendering this platform attractive for modellers. More specifically, 33 autophagy-related models resulted when querying the database in April 2022, each of which can be a valuable source for missing data for this project.

Therefore, there is a multiplicity of primary and secondary sources, including platforms and databases, that can aid in the collection of data for creating a model of macroautophagy.

Based on the available data, the dimensions of time and space of the mathematical model of macroautophagy had to be defined. The majority of data relating to macroautophagy are referring to the initiation step and the phagophore formation, while the later steps are less studied [620]. Therefore, the model focused on the initiation of macroautophagy until the phagophore elongation, and more specifically the start of the macroautophagy pathway was the activated ULK1 complex and the end of the pathway the production of LC3II (Fig 5-2A).

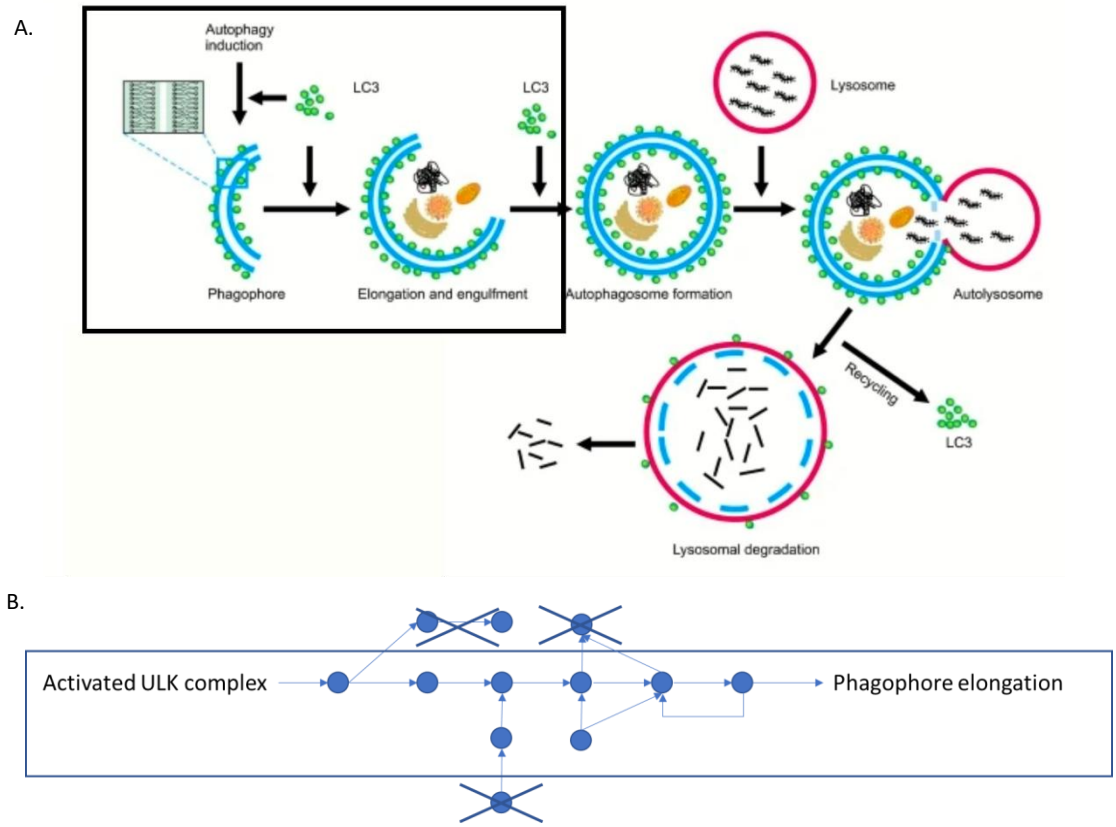


Figure 5-2. Defining the borders of macroautophagy

*(**A**) The stages of macroautophagy to be modelled in this Thesis were decided to be from the initiation of the ULK1 complex until before the autophagosome formation, adapted from Jing et al [265]. (**B**) The main macroautophagy proteins and their direct regulators were included in the model, while components through which information did not flow were excluded.*

Regarding the location of the model, it was the phagophore forming site of the cell, which was considered spatially homogenous. Adding more locations in the model would have increased the complexity of the model, so instead of modelling proteins being transferred from other cellular regions, they were modelled as activated. For instance, while ATG16L1 is recruited to the phagophore assembly site by WIPI2, in this model it was assumed that it is already there but instead that it is activated by WIPI2. The parameters of the reaction were adjusted to mirror the difference between such a case and a real activation step.

Furthermore, the level of distance of included components from the main players that contribute to most of the flow of the process needed to be decided. An additional problem surfaced during this step: the identification of the main players of autophagy. For this purpose, the scoring of macroautophagy components developed in Chapter 4 was used to resolve this issue. The main components were identified as those with a score of 4 or higher. Then, the main autophagy proteins and their direct regulators were retained in the model. Due to the large amount of research around macroautophagy as seen in previous paragraphs of this section, incorporating additional indirect regulators would have resulted in a diagram with a larger size than acceptable for this project. In addition, it was decided that components through which information did not flow would also be removed, as shown on Fig 5-2B.

5.4.2.2 Available data for the overlap of macroautophagy with neurodegenerative diseases

The second part of the modelling for which the availability of data was required to be assessed, was related to the overlap of macroautophagy and neurodegenerative diseases. There are multiple types of data that can be used for this endeavour, such as PPIs by creating and overlapping PPINs. In Chapter 5, in which this approach was adopted, it became evident that there is a major overlap of macroautophagy with neurodegenerative diseases, such as Parkinson's disease, Alzheimer's disease, Frontotemporal dementia, and Amyotrophic Lateral Sclerosis. If the overlap had been limited, then that small part of the PPIN would have been modelled in this chapter for each of the individual 4 diseases. However, the amount of overlap rendered this option more complicated than aimed in this project. Therefore, one disease was selected, Parkinson's disease, which had one of the largest overlaps with the macroautophagy network. For instance, in the level of seeds, 100% of the PD seeds (n=19/19) were present in the macroautophagy network, compared to 100% (4/4), 89.7% (26/29), and 94.1% (16/16) for AD, ALS, and FTD, respectively. Of note, between the AD and PD seeds, all of which were present in the MA network, only those of PD had a significant p-value (3.16×10^{-4}), pointing towards PD having a wider and more significant overlap with macroautophagy. Therefore, the association of PD with macroautophagy was chosen to be studied in more detail.

The overlap between the PD and MA networks was extensive. As shown on Chapter 4 (Sections 4.4.3 and 4.4.4), there was a 63.1% overlap of the first layers, and a 97.9% overlap of the second layers. Therefore, the relationship between a single gene of PD with macroautophagy was going to be modelled. The ideal candidate for this study needed to have much related literature, a large number of PPIs, and much mutational data affecting its PPIs. These criteria were fulfilled by LRRK2. In more detail, querying the name of each gene of the PD network in PubMed (on 11/2/2022), the

most publications were for *LRRK2* (631) followed by *PINK1*, *SNCA*, and *GBA*, with 523, 508, and 492 results, respectively (Table 5-2). In addition, *LRRK2* had the highest number of interactors as collected by PINOT (with final score>2; 222). This was 63% higher than the second ranked gene, *PRKN* (136 interactors), which was followed by *SNCA* with 106 interactors (Table 5-3). Regarding data for mutations that affect their interactions with other proteins, the highest number of data entries was for *SNCA* (1,969 interactions), followed closely by *LRRK2* (1,717 interactions), with the third gene being *KANSL1* with just 54 entries (Table 5-4). Therefore, the association of LRRK2 with macroautophagy was chosen to be studied in more detail.

Table 5-2. Amount of literature for each PD gene

Gene	N results in PubMed	Gene	N results in PubMed
LRRK2	631	*SMPD1*	24
PINK1	523	*PRKRA*	23
SNCA	508	*SYNJ1*	20
GBA	492	*DNAJC6*	20
PARK7	169	*KAT8*	17
ATP13A2	94	*RAB39B*	17
PRKN	51	*DNAJC13*	9
GAK	44	*KANSL1*	7
FBXO7	31	*RAB29*	7
WDR45	28		

Table 5-3. Number of interactors of each PD gene based on PINOT			
Gene	N interactors	Gene	N interactors
LRRK2	222	*PINK1*	23
PRKN	136	*DNAJC13*	22
SNCA	106	*KAT8*	20
PARK7	100	*RAB39B*	20
GAK	74	*RAB29*	13
ATP13A2	59	*SYNJ1*	10
FBXO7	55	*WDR45*	9
KANSL1	28	*SMPD1*	6
GBA	26	*DNAJC6*	4
PRKRA	24		

Table 5-4. Amount of mutational data that affect interactions with other proteins in IntAct			
Gene	N entries	Gene	N entries
SNCA	1969	*RAB39B*	10
LRRK2	1717	*KAT8*	6
KANSL1	54	*WDR45*	1
PARK7	43	*SYNJ1*	0
PINK1	42	*DNAJC6*	0
PRKN	41	*DNAJC13*	0
PRKRA	25	*GBA*	0
RAB29	21	*SMPD1*	0
FBXO7	19	*ATP13A2*	0
GAK	18		

5.4.3 Initial designing of the macroautophagy diagram

To design the pathway, it was decided that the base would be the diagrams of Reactome and KEGG identified in Section 5.4.2.1 (Fig S5-1 & S5-2, Appendix E). The diagrams were filtered, simplified, and merged, leading to the creation of the Merged diagram (see details in Section 5.3.2).

In this stage, it was essential to investigate whether core macroautophagy proteins were already part of the diagram, or whether they had to be incorporated. As decided in Section 5.4.2.1, the analysis of Chapter 4 was used to define the core proteins. The 46 proteins associated with

macroautophagy with a score of 4 or higher were crossed checked against the Merged diagram. Most proteins were already present in the diagram (n=34/46, 73.9%), while the rest of the proteins (n=12/46, 26.1%) failed to pass the filters of inclusion (Filter S and Filter P, discussed in Section 4.3.2) and were not added in the diagram. For instance, VSP18 and VSP33A were not incorporated in the diagram as they are involved in the fusion of the phagophore membrane, which occurs in a later stage of macroautophagy not studied in this project (Section 4.3.2). The resulted diagram was named Enriched Diagram A and it was composed of 29 components and 5 descriptive terms, which were linked with 45 relationships.

The diagrams used as the model's basis are representing the main steps of macroautophagy, however, they might not include the most recent literature. Reactome was last updated in October 2019 and the latest paper used for KEGG was published in August 2015. Therefore, a list of publications [34, 618-622], was used to mine information and enrich the Enriched Diagram A, producing the Enriched Diagram B (see Section 5.3.2).

5.4.4 Associations of LRRK2 with macroautophagy

The direct and indirect associations of LRRK2 with macroautophagy were investigated and the extracted data were used to enrich the diagram. The direct associations were explored through mining data from the reference list, and UniProt. Then, the associations of protein interactors of LRRK2 with macroautophagy were investigated (i.e., indirect associations). Briefly, this was conducted by obtaining the LRRK2 interactors, retaining those with associations with macroautophagy, and with data supporting that their interactions with LRRK2 are affected by mutations in LRRK2.

5.4.4.1 Obtaining LRRK2 interactors

PINOT was used to collect the human PPIs of LRRK2, which resulted in 1,434 interactors. The interactors that have been identified with multiple distinct methods and/or published in multiple papers were selected by filtering based on the final score being higher than 3 (n=173/1,434, 12.1%). Therefore, this study focused on the LRRK2 interactors for which there is higher confidence.

5.4.4.2 Filtering of the LRRK2 interactors based on links with macroautophagy

Multiple alternative approaches, described in more detail in Section 5.3.3, were used to select the LRRK2 interactors associated with macroautophagy. Firstly, the reference list of papers was consulted, resulting in 14/173 LRRK2 interactors to be linked to macroautophagy (i.e., "yes" group; 8.1%), 6 to be potentially linked (i.e., "?" group; 3.5%) and 153 for which no data were available (i.e., "no" group; 88.4%). Further investigation led to 17 LRRK2 interactors belonging in the "yes" group and 156 in the "no" group. Therefore, as a result of this analysis the diagram was enriched with 17 proteins to produce the Enriched Diagram C and excluded the rest of the LRRK2 interactors (n=156) (see Table 5-5). Then, Enriched diagram C was simplified to exclude any proteins that did not contribute to the flow of information for macroautophagy or from LRRK2 to macroautophagy.

Table 5-5. Filtering of LRRK2 interactors based on their association with macroautophagy

ABCE1	*OPG2*	***HSP90AA1***	*MOGS*	***PPP2R2A***	*RPL13*	*SEC16A*	*TUBA1C*
ACTG1	*COQ8A*	*HSP90AB1*	*MRPL19*	*PRDX3*	*RPL19*	*SENP3*	*TUBB*
ACTR2	*CSE1L*	*HSPA4*	*MSH2*	***PRKACA***	*RPL23*	*SF3B2*	*TUBB2A*
AGO1	*CYFIP1*	*HSPA8*	*MSN*	***PRKDC***	*RPL24*	*SF3B3*	*TUBB4A*
AGO2	***DDB1***	*HSPA9*	*MTHFD2*	*PRPF6*	*RPL3*	*SFN*	*TUBB4B*
AHCYL1	*DIS3*	*KIF2A*	*MYL6*	*PSMD11*	*RPL30*	*SFXN1*	*TUBB6*
AIFM1	*DNAJA1*	*KPNB1*	*MYL9*	*PSMD2*	*RPL34*	*SH3GL1*	*TUBG1*
ARFGAP1	*DNM1*	*LARP4*	*MYO1B*	*PSMD6*	*RPS11*	*SLC25A11*	***TUFM***
ARHGEF7	*DNM1L*	*LARP7*	*MYO1C*	*PTCD3*	*RPS13*	*SLC25A22*	*USP39*
ARPC1B	*DVL1*	*LAS1L*	*MYO1D*	*PYGB*	*RPS14*	*SLC25A4*	*VIM*
ARPC2	*DVL2*	*LDHA*	*NCL*	*RAB10*	*RPS15*	*SLC25A5*	*VPS4A*
ATP5MG	*DVL3*	*LRP6*	*NCLN*	***RAB1B***	*RPS15A*	*SLC25A6*	*WSB1*
ATP5PO	*DYNC1H1*	*LRRC47*	*NUP107*	*RAB29*	*RPS16*	*SNAPIN*	*YWHAB*
BAG5	*ECHS1*	*LRRK1*	*NUP133*	*RAB32*	*RPS18*	***SNCA***	*YWHAE*
CCT3	*EEF1A2*	***MAP1B***	*NUP160*	***RAB5B***	*RPS2*	*SPTLC1*	*YWHAG*
CDC37	*EEF1G*	*MAP2K3*	*OPA1*	*RAC1*	*RPS20*	***SQSTM1***	*YWHAH*
CDC42	*EEF2*	*MAP2K6*	*PCNA*	*RBM39*	*RPS27*	*SSR4*	*YWHAQ*
CFAP20	*EPRS1*	*MAP2K7*	*PFKP*	*RGS2*	*RPS3*	*STUB1*	***YWHAZ***
CHD1L	*GAK*	*MAPT*	*PKM*	*RO60*	*RPS3A*	*TCF25*	*ZRANB2*
CHGB	*GNAI2*	*MBP*	*PLEC*	*RPL10A*	*RPS7*	***TP53***	
CKAP5	***GSK3B***	*MDN1*	***PPP1CA***	*RPL11*	*RPS8*	*TTC27*	
CNP	*HACD3*	*MMS19*	***PPP1R8***	*RPL12*	*SAMHD1*	*TUBA1A*	

Note: The LRRK2 interactors with an identified link with macroautophagy are in bold.

5.4.4.3 Which of these do we have mutational data for?

The potential manner through which mutations in LRRK2 affect macroautophagy through its protein interactors was identified as an aim of this model. Therefore, it was required to explore for which of the 17 LRRK2 interactors associated with macroautophagy there were data showing a change in the interaction when *LRRK2* is mutated. IntAct holds such a data set, so the LRRK2 interaction data were extracted and filtered (see Section 5.3.3). There were relevant data for 8 LRRK2 interactors (n=8/17, 47.1%). The gene name and SwissProt ID of these LRRK2 interactors are shown in Table 5-6, together with the *LRRK2* mutation that affects each interaction. Interestingly, the G2019S mutation and D1994A mutation affects all the interactions in a negative manner (6/6, and 4/4, respectively), while others can have an opposite effect on different interactions. For instance, R1441G leads to an increased strength of interaction with PPP1CA, whereas to a disrupted rate with PRKACA and YWHAZ (Table S5-1; Appendix G).

Table 5-6. Genes encoding proteins whose interactions with LRRK2 are affected by *LRRK2* mutations

Gene	SwissProt ID	*LRRK2* mutations
PRKACA	P17612	Q5S007:p.Ser1444Ala, Q5S007:p.Arg1441Cys, Q5S007:p.Arg1441Gly, Q5S007:p.Ser1443_Ser1444delinsAlaAla, Q5S007:p.Arg1441His
SNCA	P37840	Q5S007:p.Gly2019Ser
GSK3B	P49841	Q5S007:p.Asp1994Ala, Q5S007:p.Gly2019Ser
RAB5B	P61020	Q5S007:p.Gly2385Arg, Q5S007:p.Arg1441Cys, Q5S007:p.Asp1994Ala, Q5S007:p.Ile2020Thr, Q5S007:p.Gly2019Ser
PPP1CA	P62136	Q5S007:p.Asn1437His, Q5S007:p.Arg1441Gly, Q5S007:p.Tyr1699Cys, Q5S007:p.Ser910Ala, Q5S007:p.Ser935Ala, Q5S007:p.Ser955Ala, Q5S007:p.Ser973Ala
YWHAZ	P63104	Q5S007:p.Arg1441Gly, Q5S007:p.Ser1444Ala
SQSTM1	Q13501	Q5S007:p.Gly2385Arg, Q5S007:p.Asp1994Ala, Q5S007:p.Gly2019Ser
CDC37	Q16543	Q5S007:p.Gly2385Arg

In the final model of macroautophagy of Chapter 6, the aim is to simulate the effect of different *LRRK2* mutations on individual interactions. Therefore, the mutations of *LRRK2* were prioritised to include those for which there was more information. As seen in Table 5-7, most mutations affected only 1 interaction (n=9/15, 60%), while 1 mutation (i.e., Q5S007:p.Gly2019Ser aka G2019S) affected 6 interactions. The mutations affecting at least 3 interactions will be studied further, leading to the

inclusion of 5 mutations that affected 33% of the total interactions (R1441C, R1441G, D1994A, G2019S, and G2385R, corresponding to Q5S007:p.Arg1441Cys, Q5S007:p.Arg1441Gly, Q5S007:p.Asp1994Ala, Q5S007:p.Gly2019Ser, and Q5S007:p.Gly2385Arg). This did not result in the exclusion of any LRRK2 interactor, meaning that all interactors were affected by at least one of the mutations that will be simulated in a later stage.

Table 5-7. Mutations of *LRRK2* and the number of proteins with whom the interactions are affected

Feature short label	Feature range(s)	Original sequence	Resulting sequence	N affected interactions
Q5S007:p.Asn1437His	1437-1437	N	H	1
Q5S007:p.Arg1441Cys	1441-1441	R	C	3
Q5S007:p.Arg1441Gly	1441-1441	R	G	3
Q5S007:p.Arg1441His	1441-1441	R	H	1
Q5S007:p.Ser1443_Ser1444delinsAlaAla	1443-1444	SS	AA	1
Q5S007:p.Ser1444Ala	1444-1444	S	A	2
Q5S007:p.Tyr1699Cys	1699-1699	Y	C	1
Q5S007:p.Asp1994Ala	1994-1994	D	A	4
Q5S007:p.Gly2019Ser	2019-2019	G	S	6
Q5S007:p.Ile2020Thr	2020-2020	I	T	1
Q5S007:p.Gly2385Arg	2385-2385	G	R	3
Q5S007:p.Ser910Ala	910-910	S	A	1
Q5S007:p.Ser935Ala	935-935	S	A	1
Q5S007:p.Ser955Ala	955-955	S	A	1
Q5S007:p.Ser973Ala	973-973	S	A	1

Therefore, the 8 filtered LRRK2 interactors (i.e., *SQSTM1*, *CDC37*, *GSK3B*, *PPP1CA*, *PRKACA*, *RAB5B*, *SNCA*, and *YWHAZ*) were retained in the macroautophagy diagram, while the other LRRK2 interactors were removed. An additional gene had to be retained in the diagram due to the way it functions. *CCDC37* mainly affects macroautophagy through the complex of its protein product with HSP90AA1, so the latter was also included in the diagram. The resulted diagram is displayed for transparency, but it is too complicated to be studied in detail (Enriched diagram D, Fig S5-3,

Appendix E). More simplification was required and performed in a later stage, described in Section 5.4.5.

5.4.5 Further adjustments

5.4.5.1 Expert curation of the macroautophagy network

The macroautophagy pathway is complex and the subject of extensive ongoing research, and so the opinion of an expert in the field of macroautophagy, Dr Sharon Tooze, was sought to provide targeted curation of the diagram of macroautophagy. Prior to this step, however, some of her recent publications were studied in more detail [627-630] to incorporate any missing information or correct any potential mistakes that could have been made by misinterpreting the literature or using published data either disproven by or not convincing to field experts. Relationships were added in the diagram (Enriched Diagram E, Fig S5-4, Appendix E) based on the extracted information, including the positive regulation of the ULK1 complex through LC3II and C9orf72 [628]. Interestingly, data were obtained supporting the association of RAC1 with macroautophagy [630]. RAC1 was a LRRK2 interactor that was previously excluded from the analysis due to lack of data supporting its association with macroautophagy (Section 5.4.4.2). Therefore, the LRRK2 interactor list was updated to include RAC1 and the analysis that followed this step (i.e., filtering of the LRRK2 interactors based on the existence of mutational data) was repeated. RAC1 was retained after this step and therefore Tables 5-6 and 5-7 were updated and their new versions are presented in Tables 5-8 and 5-9.

After updating the model based on recent publications of Dr Sharon Tooze, her advice was sought for a list of topics, some of them controversial, as gathered from the literature. For instance, there is growing evidence that the ATG9 trafficking pathway can affect macroautophagy [631-633]. It was questioned whether such information was essential to be added in the diagram, taking into account the limit of components that can be included in the model (20-30 in this case). The questions and decisions made based on the feedback are summarised in Table S5-2 and the resulted diagram named Enriched diagram E is presented in Figure S5-4 (Appendix E).

Table 5-8. Genes encoding proteins whose interactions with LRRK2 are affected by *LRRK2* mutations (updated to include RAC1)

Gene	SwissProt ID	LRRK2 mutations
PRKACA	P17612	Q5S007:p.Ser1444Ala, Q5S007:p.Arg1441Cys, Q5S007:p.Arg1441Gly, Q5S007:p.Ser1443_Ser1444delinsAlaAla, Q5S007:p.Arg1441His
SNCA	P37840	Q5S007:p.Gly2019Ser
GSK3B	P49841	Q5S007:p.Asp1994Ala, Q5S007:p.Gly2019Ser
RAB5B	P61020	Q5S007:p.Gly2385Arg, Q5S007:p.Arg1441Cys, Q5S007:p.Asp1994Ala, Q5S007:p.Ile2020Thr, Q5S007:p.Gly2019Ser
PPP1CA	P62136	Q5S007:p.Asn1437His, Q5S007:p.Arg1441Gly, Q5S007:p.Tyr1699Cys, Q5S007:p.Ser910Ala, Q5S007:p.Ser935Ala, Q5S007:p.Ser955Ala, Q5S007:p.Ser973Ala
YWHAZ	P63104	Q5S007:p.Arg1441Gly, Q5S007:p.Ser1444Ala
SQSTM1	Q13501	Q5S007:p.Gly2385Arg, Q5S007:p.Asp1994Ala, Q5S007:p.Gly2019Ser
CDC37	Q16543	Q5S007:p.Gly2385Arg
RAC1	P63000	Q5S007:p.Gly2019Ser, Q5S007:p.Lys1906Met, Q5S007:p.Arg1441Cys, Q5S007:p.Ile2020Thr, Q5S007:p.Tyr1699Cys

Table 5-9. Mutations of *LRRK2* and the number of proteins with whom the interactions are affected (updated to include RAC1)

Feature short label	Feature range(s)	Original sequence	Resulting sequence	N affected interactions
Q5S007:p.Asn1437His	1437-1437	N	H	1
Q5S007:p.Arg1441Cys	1441-1441	R	C	4
Q5S007:p.Arg1441Gly	1441-1441	R	G	3
Q5S007:p.Arg1441His	1441-1441	R	H	1
Q5S007:p.Ser1443_Ser1444delinsAlaAla	1443-1444	SS	AA	1
Q5S007:p.Ser1444Ala	1444-1444	S	A	2
Q5S007:p.Tyr1699Cys	1699-1699	Y	C	2
Q5S007:p.Lys1906Met	1906-1906	K	M	1
Q5S007:p.Asp1994Ala	1994-1994	D	A	4
Q5S007:p.Gly2019Ser	2019-2019	G	S	7
Q5S007:p.Ile2020Thr	2020-2020	I	T	2
Q5S007:p.Gly2385Arg	2385-2385	G	R	3
Q5S007:p.Ser910Ala	910-910	S	A	1
Q5S007:p. Ser935Ala	935-935	S	A	1
Q5S007:p.Ser955Ala	955-955	S	A	1
Q5S007:p.Ser973Ala	973-973	S	A	1

5.4.5.2 Additional adjustments and final diagram

In mathematical modelling a balance between simplicity and complexity is key [5]. Even though, descriptive terms, such as “Initiation of macroautophagy and nucleation of isolation membrane”, were helpful in building the model and retaining as much information as possible in an accurate manner, they were removed at this stage. Similarly, relationships that expressed regulation without being specific about whether it is positive of negative were also excluded. Additionally, components through which information did not flow were also removed, with the exception of the initial and last components of the model (ULK1 complex and LC3II) and the LRRK2 interactors. Therefore, the main diagram that is going to be modelled was created. For ease of visualisation purposes, the macroautophagy diagram is presented without the connections with LRRK2 interactors in Fig 5-3.

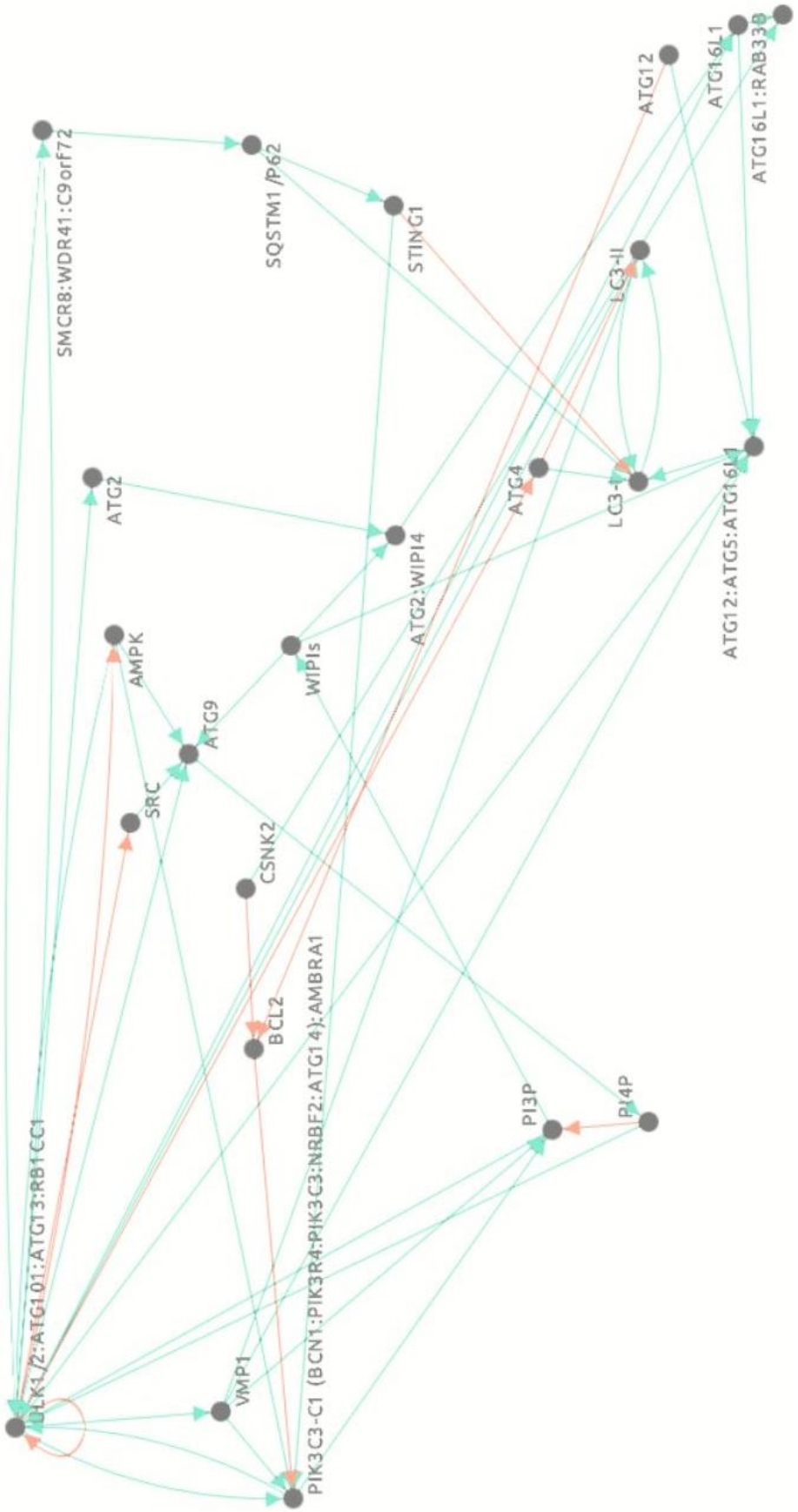


Figure 5-3. Updated simplified process diagram of macroautophagy

5.5 Discussion

Creating a mathematical model of macroautophagy is a complicated multi-step process that requires much information to be collected and several decisions to be made. For the former, the vast number of produced publications relating to macroautophagy is a valuable source of information. A substantial total of 728 publications resulted from the query of "macroautophagy" in PubMed and of 13,459 publications from using the term "autophagy", as of the 9th of February of 2022. The amount of available literature is accompanied by a multiplicity of other resources, such as Reactome, KEGG, Autophagy Regulatory Network and BIOMODELS that are valuable in building a model. These demonstrate the international effort in understanding macroautophagy and its connection with human diseases.

A number of decisions were made based on the exploration of the available data regarding macroautophagy. The vast amount of research around this topic led to incorporating only the core macroautophagy proteins and their direct interactors to limit the entities of the model and diagram. The location of the model and diagram was defined to be the site of phagophore formation. It was also decided to focus on the initial stages of the process (i.e., initiation of macroautophagy until the elongation of the phagophore).

The rationale of the choice to start the modelled part of macroautophagy at the activated ULK1 complex was partly due to the stages prior to the activated ULK1 complex, which includes a variety of separate pathways depending on the signal that leads to the activation of macroautophagy, as discussed in Chapter 4. Briefly, examples include starvation and low energy levels, which induce macroautophagy through mTOR and JNK, and AMPK, respectively [318, 634, 635]. However, the aim was to model macroautophagy in a non-signal specific manner, so anything prior to the activation of the ULK1 complex was excluded. An additional reason for this decision was mTORC1 being upstream of the ULK1 complex. mTORC1 is a central complex that regulates cell growth and survival, and macroautophagy, responding to growth factors and amino acids [636-639]. Its inclusion in the model would have increased the complexity of the model to a level higher than aimed for this project. Regarding the decision to end the modelled pathway at the step of elongation, there were two main reasons: (i) multiplicity of downstream pathways, and (ii) lack of specificity of the associated proteins to macroautophagy. The former relates to the ability of the autophagosome to fuse with a variety of vesicles, including early or late endosomes, multivesicular bodies, lysosomes, or even the plasma membrane [640, 641]. Regarding the latter, two examples will be used, CHMP2B, and CHMP3. CHMP2B is associated not only with the fusion of autophagosome with the lysosome, but also with the multivesicular body sorting pathway, nucleus

organisation and viral budding [642-645]. CHMP3 is also associated with mid body abscission and the regulation of early endosome to late endosome transport [645, 646]. Therefore, the fusion of the phagophore was not included in the current model.

As seen from the differences in the diagrams of Reactome and KEGG (Fig S5-1 & S5-2, Appendix E), neither the start, the end, nor the number of regulators matches between these databases. In Reactome, the ULK1 complex signifies the beginning, and LC3II the end of the pathway, while a limited number of regulators are included. In KEGG, the pathway is broader. Different stimuli that lead to the activation of macroautophagy are included in its diagram, who's last step is the degradation of the components of the autophagolysosome. In the diagram created in this chapter the borders are of higher similarity to those of Reactome, as the beginning and the end match. However, additional regulators were incorporated to allow for a more accurate representation of the current knowledge of macroautophagy. In addition, some steps were simplified to limit the complexity of the diagram and of the mathematical model to be created from this diagram. This led to the main diagram of macroautophagy (Fig 5-3) been composed of 23 components (proteins, protein complexes and lipids), which is more restricted compared to the 150 from Reactome and the 147 from KEGG.

Since the model of Chapter 6 will focus on simulating how mutations that lead to neurodegeneration can affect macroautophagy, the availability of data supporting the overlap between macroautophagy and neurodegeneration was also assessed. Together with the literature, the PPIN analysis of the previous chapter supports a substantial intersection between the two processes. Therefore, the aim of the model was adjusted to simulating the overlap of macroautophagy with a single protein that leads to Parkinson's disease, LRRK2. This decision was based on analyses of the number or amount of available publications, protein interactors, and mutational data that affect protein interactions, which were available through IntAct.

Of note, a large volume of data does not necessarily correlate with experimental value. Even though the connection of two LRRK2 interactors, *MAPT* and *SEC16A*, with macroautophagy was mentioned in one of papers in the reference list [34], there was a lack specific information to allow their incorporation in the diagram. In another case, that of *SSR4*, there was no referenced human evidence supporting the involvement of the protein in macroautophagy [34, 647-650]. Therefore, more research is required to fill in these gaps of knowledge and improve our confidence in the associations of certain proteins with macroautophagy in humans.

An additional source of data for the association of LRRK2 with macroautophagy could have been the Autophagy Regulatory Network (ARN). Interestingly, querying LRRK2 in the ARN website (9th

February 2022) [551], resulted in zero direct interactions with autophagy proteins, or autophagy regulators, which is in disagreement with the results from this project. Through the current study, the relationship of *LRRK2* and *SQSTM1* (producing p62) was collected from the literature and incorporated in the diagram of macroautophagy. However, ARN included 110 post-translational regulators, 25 transcriptional regulations, and 2 pathways for LRRK2. Among the post-translational regulators, 6/9 LRRK2 interactors of the final diagram were present. Therefore, an alternative approach could have been to incorporate these data from ARN in the diagram. That would, however, require manual evaluation of the data of each interaction, and filtering based on the level of distance from components of the diagram. Nonetheless, 3/9 total LRRK2 interactors (33%) were uniquely identified by using the methodology of this Thesis and were absent from the ARN data set. This result demonstrates that even though valuable information can be extracted from specialised databases, the approach of this study resulted in the inclusion of 3 extra connections that could affect the behaviour of the model and thus lead to stronger predictions about macroautophagy and its link with neurodegeneration.

The association of multiple LRRK2 interactors with macroautophagy proteins as identified in the current study is not surprising. Overexpression of LRRK2 with the G2019S mutation has been suggested to play a role in the regulation of macroautophagy [651] due to accumulation of autophagic vesicles in neurons and increased neuronal death observed in rats [652]. Evidence from human cells have also been produced. In human neuroblastoma cells, which overexpressed the LRRK2 G2019S mutation, the induction of macroautophagy resulted in the further accumulation of autophagosomes and shortening of neurites [653]. A more recent analysis, studied the overexpression of LRRK2 G2019S, which was demonstrated to induce macroautophagy in the neuromuscular junction [654]. In a study with human neuroglioma cells that avoided using an overexpression system, which represent a less natural cellular state, the inhibition of the kinase activity of LRRK2 resulted in an induction of macroautophagy [588]. A follow up study of the researchers revealed that this stimulation was dependent on PIK3C3-C1 and independent of mTOR and ULK1 complex [586]. In addition, increased basal macroautophagy has been reported in fibroblasts from patients with PD, carrying the LRRK2 G2019S mutation [655]. It has been suggested that LRRK2 affects macroautophagy through its impact in vesicular dynamics [602].These findings are highlighting the association of LRRK2, and thus Parkinson's disease, with macroautophagy, supporting the results of this chapter.

6th Chapter:

A mathematical model of macroautophagy

6. A mathematical model of macroautophagy

Main points of this chapter:

- The main aim of this chapter is to formulate, parametrise and solve an ordinary differential equation (ODE) mathematical model of the initial stages of macroautophagy (initiation and elongation of phagophore)
- The model is based on the macroautophagy process diagram detailed in Fig 6-1 and is informed and tested using literature data
- The model is composed of 26 entities, and 45 relationships, each described by an ODE and was solved using the SimBiology toolbox of MATLAB
- The model was investigated for both healthy and Parkinson's disease states, based on differential concentrations/amounts of its entities
- The LC3 ratio (i.e., LC3II/LC3I) was used as an output to compare macroautophagy in the healthy vs the Parkinson's disease state
- Macroautophagy in PD is suggested by the model to start earlier but reach slightly lower levels as described by the LC3 ratio compared to healthy controls

6.1 Introduction

In people living with neurodegenerative diseases, such as Parkinson's disease (PD), specific brain functions are altered, including control of movement, cognition, sleep, and smell [191-194, 196]. These effects are also reflected in alterations in cell and protein biology in animal models of PD and in post-mortem analyses of human brain specimens of people with PD. For instance, a post-mortem study of cerebellar tissue of people with PD carrying LRRK2 mutations, demonstrated that there was a decrease in the protein levels of GBA and VSP35, and by implication of their activity [656]. LAMP2A has also been found to be decreased in Parkinson's disease [657] including in people with LRRK2 G2019S mutations [658]. In mice expressing the same mutation, the opposite trend was observed for ATP13A2, as it was upregulated [659]. Interestingly, components of the macroautophagy pathway and their regulators have also been found to be differentially expressed in people with PD. For example, a higher expression of negative regulators of LAMP2A and Hsc70

was found in the substantia nigra and amygdala of people with PD [493]. Macroautophagy has been discussed in detail previously (Section 1.5), but a brief summary of the process is presented in Fig 1-4 also copied in this section as an *aide memoire*.

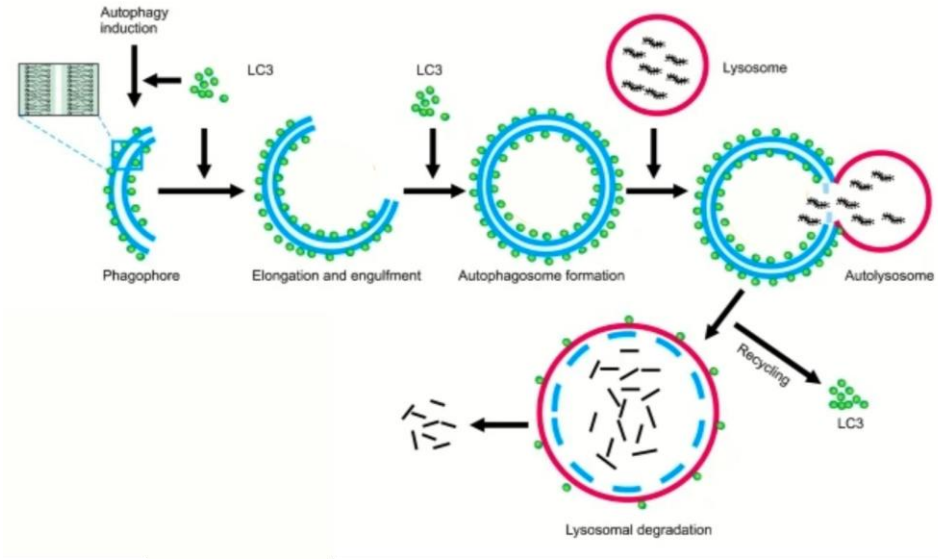


Figure 1-4. Schematic representation of macroautophagy.

In macroautophagy, there is initially the formation of the phagophore, then of the, autophagosome and finally of the autolysosome. Adapted from Jing et al [265].

Furthermore, the effect of PD pathology is more prominent in some brain regions. This can be explained in part by the spreading of the disease according to Braak staging [211]. At any given time-point, an area that has been affected for a longer period of time (e.g., brainstem *versus* neocortex) could have a more severe PD pathology. However, the differential features of distinct areas of the brain can also play a role in their susceptibility to disease, e.g., level of vascularity [660], basic metabolism [661], and differential expression of proteins [662]. The molecular basis of this differential impact of PD pathology in diverse brain regions has yet to be deciphered.

6.2 Aims and objectives

The aim of this chapter is to formulate, parametrise and solve a mathematical model of macroautophagy based on the macroautophagy process diagram (Fig 5-3) that focused on its initial stages (initiation of macroautophagy and phagophore elongation) in a human cell. Considering that

the components of the macroautophagy pathway and their regulators have been found to be differently expressed in people with PD, the model will be used to investigate whether there is a distinction in how macroautophagy functions when comparing brains of healthy people vs people with PD.

The first objective for this endeavour is to reduce the complexity of the macroautophagy diagram reported in Fig 5-3, so that a minimal mathematical model capturing key behaviours can be created. The model will be formulated using ordinary differential equations (ODEs). The second objective is to collate reaction rates for macroautophagy, and also protein concentration/amount data from healthy people vs people with PD. Finally, the model will be solved, to explore whether the macroautophagy process differs in by PD, and if so, in what manner.

6.3 Methodology

6.3.1 Model reduction

The macroautophagy process diagram that was produced in Chapter 5 (Fig 5-3) focused on the initial stages of macroautophagy (initiation of macroautophagy and phagophore elongation) in a single human cell. In the current chapter, Fig 5-3 was further simplified to develop a mathematical model of macroautophagy, mainly by removing intermediate entities in serial reactions of the diagram, such as STING1, ATG4, and VMP1. Therefore, the total number of entities was reduced from 23 to 14, as shown in the simplified process diagram of Fig 6-1, produced in Cytoscape (v3.7.1). In addition, the “regulations” (black arrows) not incorporated in Fig 6-1, due to the lower confidence on the details of the relationships (i.e., positive or negative). The connections of the LRRK2 interactors (CDC37, GSK3B, PPP1C1, PRKACA, SNCA, SQSTM1, RAC1, and YWHAZ) with this diagram were also incorporated and are presented in Fig S6-1. Of note, the LRRK2 interactor RAB5B was excluded from further analysis due to lack of high confidence regarding to connections with other proteins of the model.

6.3.2 Model formulation

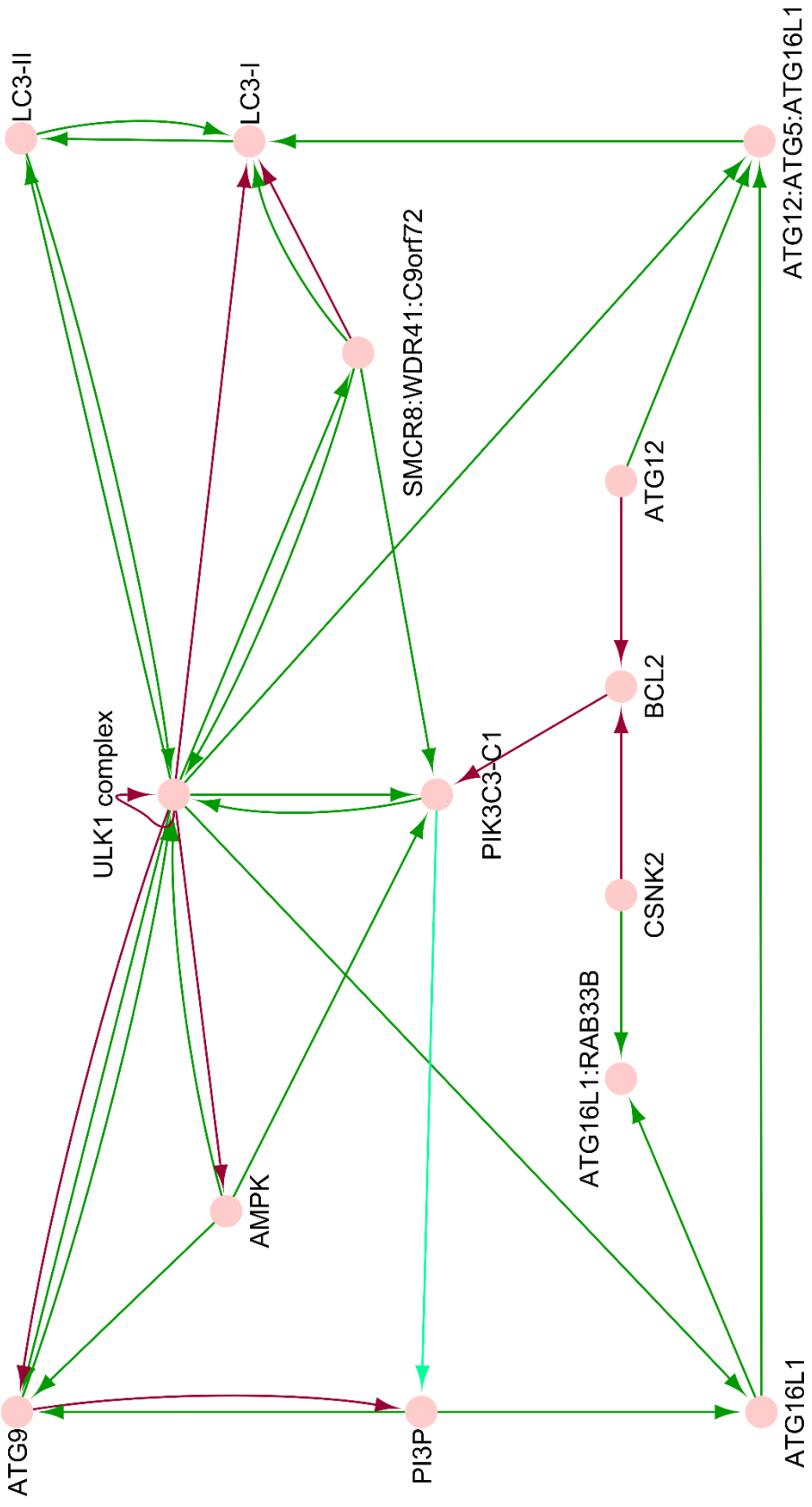


Figure 6-1. Simplified diagram of the macroautophagy process to be used as a basis for the mathematical modelling

It is based on Fig 5-3 and excludes the connections of LRRK2 interactors, which are included in Fig S6-1. In green are the activations, in red the inhibitions, and in cyan the productions. This picture was created through Cytoscape v.3.7.1

Initial simulation of this model showed PI3P increasing unbounded, which does not match the biology of this process [663]. Thus, a relationship was added to represent its natural decrease over time (PI3P → PI, k= - [PI3P] t). Similar relationships were added for the rest of the model entities for consistency.

The formulated rate reaction equations that describe the mathematical model of macroautophagy are presented in Table S6-1.

6.3.3 Parameterisation

For the parametrisation of the mathematical model of macroautophagy, the relationships describing the model (i.e., rate reaction equations of Table S6-1) were analysed into individual reactions and the type of each reaction was labelled. The labels used were sensing, binding, complex formation, phosphorylation, lipidation, ubiquitylation, truncation, GTP exchange, recruitment, natural de-activation, stabilising in membrane, and unknown. To retrieve reaction rate values for each type of reaction, the BioModels database was queried on 9th March 2022, for "autophagy" and the 4 resulted models published after 2010 were used as data sources (their unique identifiers in BioModels are: BIOMD0000000640, BIOMD0000000560, BIOMD0000000559, BIOMD0000000105) [162, 664-666]. The reaction rate values were plotted per type of reaction and outliers were removed (outliers were defined as values with a distance of more than 2 magnitudes to the others; see Fig S6-2). The average reaction rate value per type of reaction across the models (e.g., "phosphorylation") was calculated and used. Briefly, the resultant values were as follows: "binding" was set equal to 1.73×10^{-4} (molecule sec)$^{-1}$, "truncation" to 8.60×10^{-5} (molecule sec)$^{-1}$, "inactivation" to 5.00×10^{-10} sec^{-1}, and "ubiquitination" to 1.00×10^{-3} (molecule sec)$^{-1}$. Data were unavailable for the remaining categories. Based on their biological similarities, the reaction rate constants for "sensing", "binding", "complex formation" and "GTP exchange" were considered equal. In addition, "phosphorylation", "lipidation", and "ubiquitylation" were also considered to have an equal value. "Recruitment" was considered to be half as fast as "sensing", as it requires transport from another region of the cell. "Stabilising in membrane", and "unknown" types of reaction were assumed to be half as fast as "recruitment". In most cases the rate constant was stated for 2 reactants. In cases where more than 2 reactants were involved, the reaction rate was adjusted with respect to the amount of the additional reactant species, for example, if a "binding" reaction was between 3 components instead of 2.

For relationships in the model composed of multiple individual reactions, the assumptions were: (i) If a relationship is composed of more than one reaction in a series, then the rate of this relationship is equal to the slowest reaction, minus a tenth of its value for each additional reaction; and (ii) If a relationship is composed of more than one reaction in parallel, then the reaction rate constant is assumed to be equal to the sum of the individual reaction rate constants, to limit their number.

When the reaction rates were tested by running a simulation in the preliminary model, around 5,000h were needed for LC3II to have a higher amount than LC3I, instead of approximately 1h that is commonly reported in the literature [586, 667]. Since, the reaction rate constants were calculated using average values from previously published models of autophagy-related processes, instead of data from targeted experiments in macroautophagy, they were adjusted accordingly to reflect the biology of the process. More specifically, all the rates of reactions were multiplied by 5,000.

For the calculations of the amount of each model entity, there was no single source providing all the necessary data for the entities in the model. Therefore, data from two sources were collected and combined. Amounts of proteins were retrieved from a previously published model [582]. In the published model, the concentrations of PIK3C3, WIPIs, ATG9, and LC3I were set equal to 0.55nM, while the concentration of PIs was 5.54nM. In parallel, the readings from a proteomic published analysis were used, as they included relative amounts of multiple proteins in brains of healthy people vs people with PD [668]. The averages of the readings from brains of healthy people and people with PD were calculated for each protein. Then, the retrieved data from the 2 sources were combined in the following manner. Since data for WIPIs and PIs were unavailable in the proteomic analysis, the average of the averages of the relative reading for PIK3C3, ATG9, and LC3I was equated to 0.55nM. Based on this equation (2039.258 equates to 0.55nM), the concentrations of each protein were calculated from their reading. The concentrations (nM) were then converted to amounts (number of molecules) using the volume of the cell from the model [582] ($V=3\times10^{-12}$ L). For genes with multiple isoforms and for protein complexes, their minimum amount was used. There was no data in the proteomic analysis for BCL2, so its number of molecules was set equal to the average value of the entities in the model.

Similarly, the amounts of the components of the model were calculated for different brain regions using data from BRAINEAC (http://www.braineac.org/; [669]). The average of the relative amounts of PIK3C3, ATG9, and LC3I (PIs and WIPIs had no data in BRAINEAC) was calculated and equated to 0.55nM [582]. Based on this reading, the amounts of the model entities were calculated and converted to numbers of molecules, as described above.

The amounts of the model entities and the values of the reaction rate constants are presented in Table S6-2, and Table S6-3, respectively.

6.3.4 Model solution

The reaction rate equations (Table S6-1) were used as input into the SimBiology toolbox of MATLAB to create the mathematical model, using the settings “Mass action” to describe the type of relationships of the model and “odes15” as the solver. The ODEs describing the model were derived by the toolbox and were then manually examined for their accuracy (Table S6-4; Table S6-5).

The model as produced by the toolbox is presented in Fig 6-2. In this figure, the boxes are representing the entities of the model (i.e., proteins, complexes, and lipids), while the arrows are representing their relationships (see Section 6.3.1 and Appendix F for the full description). The solid lines are linking the main components of the reaction (identified by the number inside the circle), while the reaction can be also connected with an entity of the model that modifies a reaction through a dashed line. This figure deviates from the classic representation of models in which for each reaction the parameters are shown, due to the synonymity of the reaction identifier (numbers between 0 and 54), to the index of the parameters. For instance, reaction 1 has k_1 as its parameter.

The larger number of components compared to the macroautophagy process diagram of Fig 6-1 is due to the nature of modelling (n=26 vs 14), as both the active and inactive states of some components needed to be represented in the model (e.g., LC3I and LC3I_act). Other components were present only in one state, either because including their activation steps was beyond the scope of this model (e.g., CSNK2_act) or because they represented lipids that are either present or absent (lack of need of activation).

Some activations within the model of Fig 6-2 are delayed. To account for this, two more adjustments were applied. There are 3 entities that activate ATG16L1: PI3P, the activated ULK1 complex, and CSNK2 (as seen in equation 11 of Table S6-4). However, these activations are distinct in their timings. ATG16L1 needs to first be recruited to the location of the formation of the phagophore through PI3P. This is represented by reaction 13 in Fig 6-2. Then, ATG16L1 can be further activated by ULK1 complex and CSNK2 (reactions 14 and 50, respectively). Therefore, this prioritisation of reactions was incorporated into the model by inserting a “trigger event” in SimBiology. Briefly, “trigger events” allow the user to program changes in the values of parameters

at a specific timepoint during the simulation. The selected timepoint for switching the reactions 14 and 50 from inactive to active was chosen to be that at which ATG16L1 reached 10% of its maximum activation level (trigger event 1). Similarly, LC3II can be produced only after ATG16L1's activation and the formation of the complex with ATG12 and ATG5, as seen in equation 18 of Table S6-4. Therefore, a second trigger event was incorporated when LC3II reached its 10% of maximum activation level, before which the reactions of LC3II production by the ULK1 complex, CSNK2, and SMCR8:WDR41:C9ORF72 were inactive (reactions 24, 26, and 34, respectively) (trigger event 2).

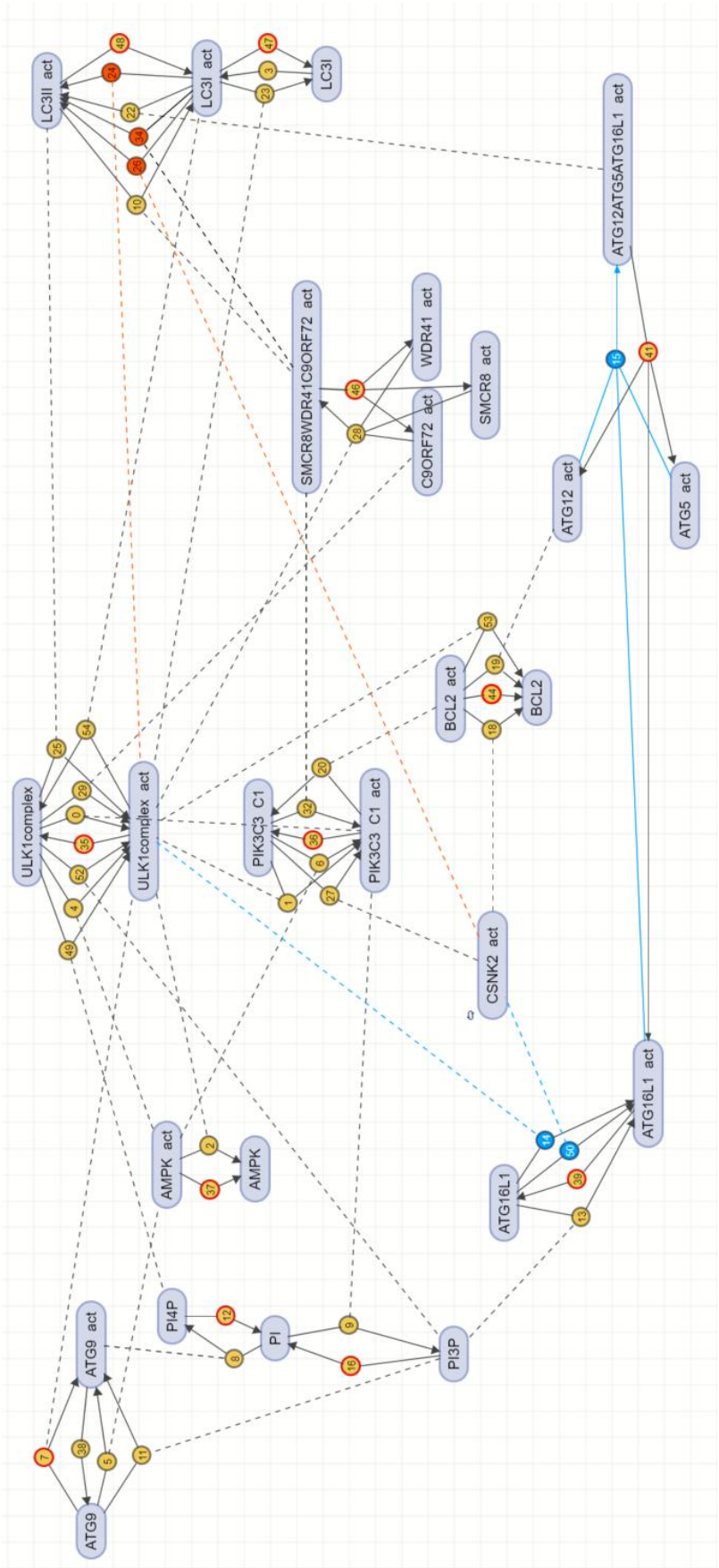


Figure 6-2. Visual representation of the mathematical model of macroautophagy

SimBiology was used to create a model of the initial stages of macroautophagy as described in this chapter. SimBiology also enables the precise visualisation of the model, which is shown here. The entities of the model (i.e., protein, protein complexes and lipids) are depicted as purple boxes, while the reactions that connect them are represented with numbered solid lines. In the cases in which the reaction has a ??, this is connected with the rest of the reaction with a dashed line. Reactions with a blue fill represent those associated with event 1, and with red those with event 2. The border of the name of each reaction is by default black, unless it is representing the natural de-activation, in which case it is red. The numbers refer to the respective reaction numbers as stated in Table S6-1. The image was produced through SimBiology, MATLAB.

6.3.5 Local sensitivity analysis

The chosen output for the local sensitivity analysis (LSA) was the LC3II/LC3I ratio (hereafter referred to as the LC3 ratio), which was calculated through SimBiology, based on setting the following "observable": LC3activation = (LC3II_act ./ LC3I_act).

The LSA was performed manually, by recording the effect of parameter variation on the LC3 ratio. The LSA function within SimBiology could not be used given it does not apply to models that have trigger events, such as those incorporated in the present model. The outputs of the LSA for increased (x10) and decreased (x1/10) reaction rate constants compared to their initial values were assessed visually by plotting and comparing the two graphs of the LC3 ratio. This variation in reaction rate constants was chosen based on literature [664]. The LSA was performed in two ways: based on the value of the LC3 ratio at 4,000s, and based on the value of the ratio for the timeframe $0 \leq t \leq 4{,}000s$.

For the LSA at 4000s, the parameters (i.e., rates of reactions) were categorised in 2 groups based on their impact on the output: those with or without an impact. For the LSA in $0 \leq t \leq 4{,}000s$, the rates of reactions were categorised to 3 groups: no impact, slight impact, high impact.

6.3.6 Visualisation of results

The model was visualised with the following software in different stages of creating the model: Cytoscape (v3.7.1) [106], and MATLAB (version 2021a; MathWorks, Natick, MA) using the SimBiology Toolbox (version 6.1).

6.3 Results

6.3.1 Model simulation

The first step after creating the mathematical model of macroautophagy for a healthy state was to explore the results by solving it in SimBiology. When plotting the molecule levels of all the components of the model together over time (Fig S6-3), the produced picture is complicated, but it can be observed that the model reaches its steady state at around 10,000s (approximately 2h and 45min). However, the model has been built based on the events of macroautophagy initiation and

phagophore elongation. Therefore, the analysis will focus on the events up to approximately 1h (4,000s) (Fig S6-4) [586, 667].

Plots of the molecule levels of smaller groups of model entities for ease of visualisation over time (t=4,000s) are presented in Fig S6-5. It can be observed that the events and the order in which they occur align with the biological understanding of macroautophagy. For the comparison of rates of activation/production of entities of the model, the time point at which the activated and inactivated form of each entity have equal number of molecules as seen from the plots, will usually be used unless otherwise stated. Firstly, the ULK1 complex remains activated throughout the experiment with a small temporary decrease at around 900s (Fig S6-5A&B). The PIK3C3-C1 is activated rapidly by the ULK1 complex. The active and inactive forms of PIK3C3-C1 are equal at 123s (Fig S6-5B). Another 3 components that are directly regulated by ULK1 are ATG9, BCL2, and AMPK. Their active and inactive forms are equal at 407s, 461s, and 1565s, respectively (Fig S6-5C). The dynamics of PIs are also noteworthy (Fig S6-5D). PI4P is produced mainly through ATG9, which is reflected by the timing of its production. It reaches the same levels as PI at 437s and keeps increasing until around 1,000s, when it reaches its steady state. PI3P increases with a smaller rate ($t_{equal,\ PI3P\text{-}PI}$ = 650s) and reaches its steady state at around the same time (t=900s). After the production of PI3P, ATG16L1 is activated, and reaches its 10% of total activated state at 433s. The activated ATG16L1 leads to the formation of the ATG16L1:ATG12:ATG5 complex, which then forms LC3II from LC3I (Fig S6-5E). In parallel, the SMCR8:WDR41:C9ORF72 complex is forming from its components at reaches levels equal to those of ULK1 complex at 148s (Fig S6-5F).

The dynamics of LC3 forms are of particular importance and are shown in Fig 6-3. Briefly, LC3I is slowly decreasing to produce the activated form of LC3I (i.e., LC3I_act), which is then decreasing as it is used to produce LC3II (i.e., LC3II_act). The rate of production of LC3II is increased further after t=890s. This is due to the events triggered after 890s, meaning the activation of the reactions through which the ULK1 complex, CNSK2, and also the SMCR8:WDR41:C9ORF72 complex lead to the formation of LC3II.

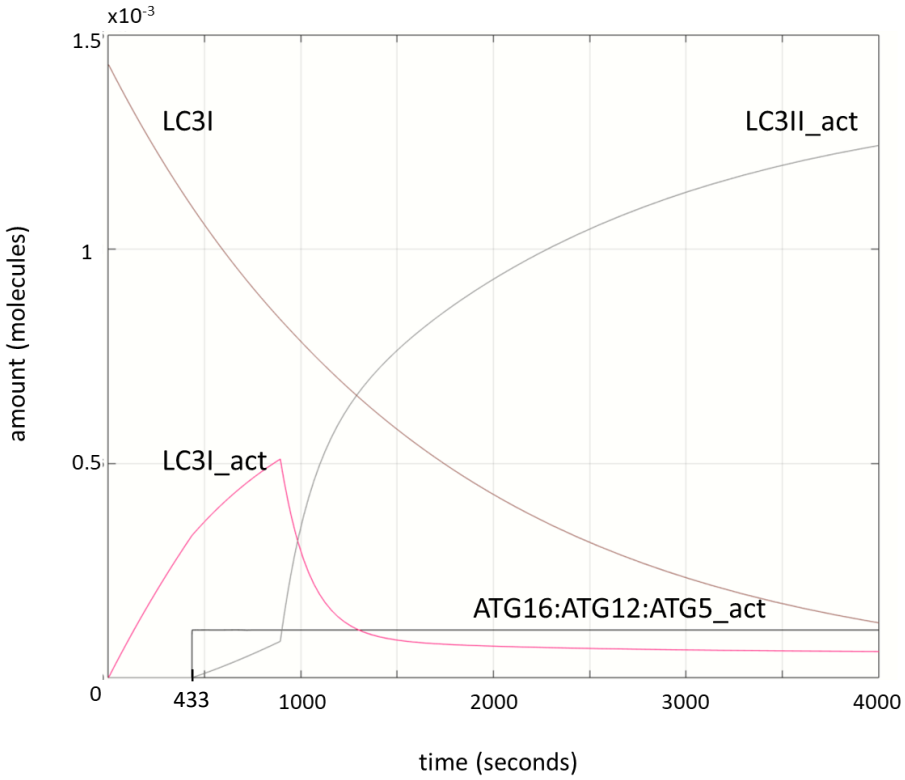


Figure 6-3. The dynamics of the LC3s in relation to the formation of the ATG16:ATG12:ATG5 complex as simulated by the macroautophagy model

The image was produced through SimBiology, MATLAB. ATG16 represents ATG16L1. At t=0s, all of the ULK1 complex of the system is fully activated and all other entities are fully inactivated (except from the negative regulators which are fully activated). At t=433s, ATG16:ATG12:ATG5 complex is formed and leads to the formation of LC3-II_act from LC3-I_act.

6.3.2 Local sensitivity analysis

The mathematical model of macroautophagy was then analysed using a local sensitivity analysis (LSA) to identify the parameters that affect the model more prominently than others. The output for this analysis was chosen to be the ratio of LC3II and LC3I (i.e., LC3 ratio), due to being one of the most widely used cellular marker for macroautophagy, whose increase can indicate an induction of the process [34].

Most parameters in the model had no effect on the output at 4,000s (n=35/45, 77.8%). The effect on the LC3 ratio was minimal (<10%) for 3 parameters (i.e., kf_23, kf_46, and kf_48), while the remaining 10 parameters (kf_3, kf_10, kf_22, kf_24, kf_26, kf_28, and kf_34) did show a discernible difference in the output (Fig 6-4) (see Table S6-4, and Table S6-5 for details of each reaction). Interestingly, the two parameters with the most prominent effect on the LC3 ratio are linked to the

regulation of LC3s by the SMCR8:WDR41:C9ORF72 complex and CSNK2. More specifically, a 10-fold increase of the reaction rate constant kf_10, which describes the inactivation of LC3II (LC3II -> LC3I) by SMCR8:WDR41:C9ORF72 complex, led to an 88% decrease in the LC3 ratio, whilst a 10-fold decrease led to a 243% increase in the LC3 ratio (331% range of effect). Regarding to CSNK2, a 10-fold increase of the reaction rate constant kf_26, which describes the activation of LC3II (LC3I -> LC3II) by CSNK2, led to a 758% increase in the LC3 ratio, whilst a 10-fold decrease led to an 80% decrease in the LC3 ratio (kf_26: -80% in x1/10, +758% in x10 (838% range of effect). Conversely, other entities of the model that are established regulators of macroautophagy had less an effect on the ratio. For instance, the parameter describing the formation of LC3II from LC3I by the ATG16L1:ATG12:ATG5 complex, kf_22, affected the system comparatively mildly. A 10-fold increase of the reaction rate constant led to a 62% increase in the LC3 ratio, whilst a 10-fold decrease of the reaction rate constant led to a 7% decrease in the LC3 ratio (69% range of effect, which is approximately a fifth of the range of kf_10, and a twelfth of the range of kf_26).

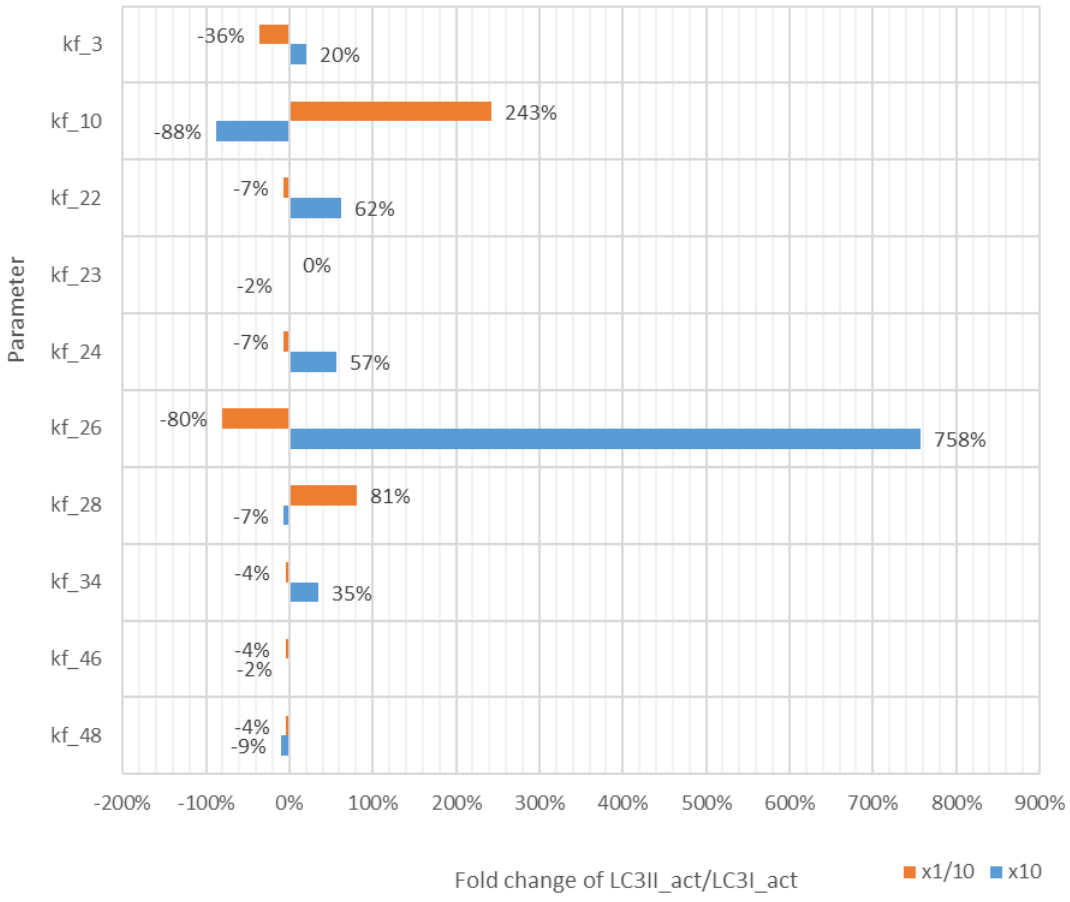


Figure 6-4. Local sensitivity analysis of the macroautophagy model at t=4,000s

Parameters that did not result in any discernible change of the output (LC3 ratio) are not included in this graph (n=35/45, 77.8%).

Furthermore, it can be observed that most parameters with an impact on the LC3 ratio, affected the output differently when they were increased compared to when they were decreased, usually having a greater effect on increasing the ratio than decreasing it. For example, the top two percentages of greater increase in the LC3 ratio were 243% and 758%, which were caused by a 10-fold decrease of kf_10, and a 10-fold increase of kf_26, respectively. On the other hand, the two percentages of greater decrease in the ratio were 80% and 88%, which were caused by a 10-fold increase of kf_10, and a 10-fold decrease of kf_26, respectively. Some exceptions in the asymmetry between positive and negative effect can also be detected. For instance, kf_3 (i.e., the rate of the activation of LC3I) affected the output more in the negative side than in the positive side (-36% in x1/10, and +20% in x10) (Fig 6-4). Overall, from this analysis it is suggested that the LC3 ratio, and thus possibly the level of macroautophagy, is more prone to be increased than decreased with perturbations of the kinetics of the system.

During the analysis of the model through LSA, it was observed that the effect of the parameters on the LC3 ratio varied throughout time. Therefore, LSA for a period of time was also performed. Similarly to the previous analysis, it was performed manually with the LC3 ratio as an output and the timeframe chosen was 0≤t≤4,000s. There was no noticeable effect for 22 parameters (n=22/45, 48.9%); there was a small effect for 17 parameters (n=17/45, 37.8%); while 6 parameters had a large effect in the behaviour of the LC3 ratio. The 6 parameters were: kf_3, kf_10, kf_22, kf_26, kf_28, and kf_34 (see Table S6-4, and Table S6-5 for details of each reaction). As expected, these parameters belonged to those that had an effect on the output during the previous LSA.

An example of a parameter with a large effect on the readout (LC3 ratio) of this analysis is kf_3, which is shown in Fig 6-5. Multiple differences can be observed in the LC3 ratio when comparing panels A and C of Fig 6.5, which correspond to a higher (x10) and lower (x1/10) rate, respectively. Firstly, when the rate is increased, the maximum value of the LC3 ratio is also increased to around 29 compared to 13 (55.17% increase) (Fig 6-5A vs C). Secondly, the dynamics of the ratio differ, as when kf_3 is increased, it first increases very slowly until 890s, then rises rapidly until it reaches its maximum value at around 1,900s, and then drops slowly for the remaining 2,100s to a value of approximately 24 (Fig 6-5A). The behaviour of the LC3 ratio when kf_3 is decreased is similar until 890s, but differs largely thereafter, as it increases constantly until 4,000s without reaching a maximum beforehand. When the LC3 are all plotted in a logarithmic graph (Fig 6-5C & D), it can be observed that even though both LC3I and LC3II values and dynamics differ, a main difference can be identified in the dynamics of LC3II, which with a higher kf_3 rate, it increases for the first 500s of the simulation, retains a high level until 890s and then drops to lower levels, compared to the case in which kf_3 is decreased.

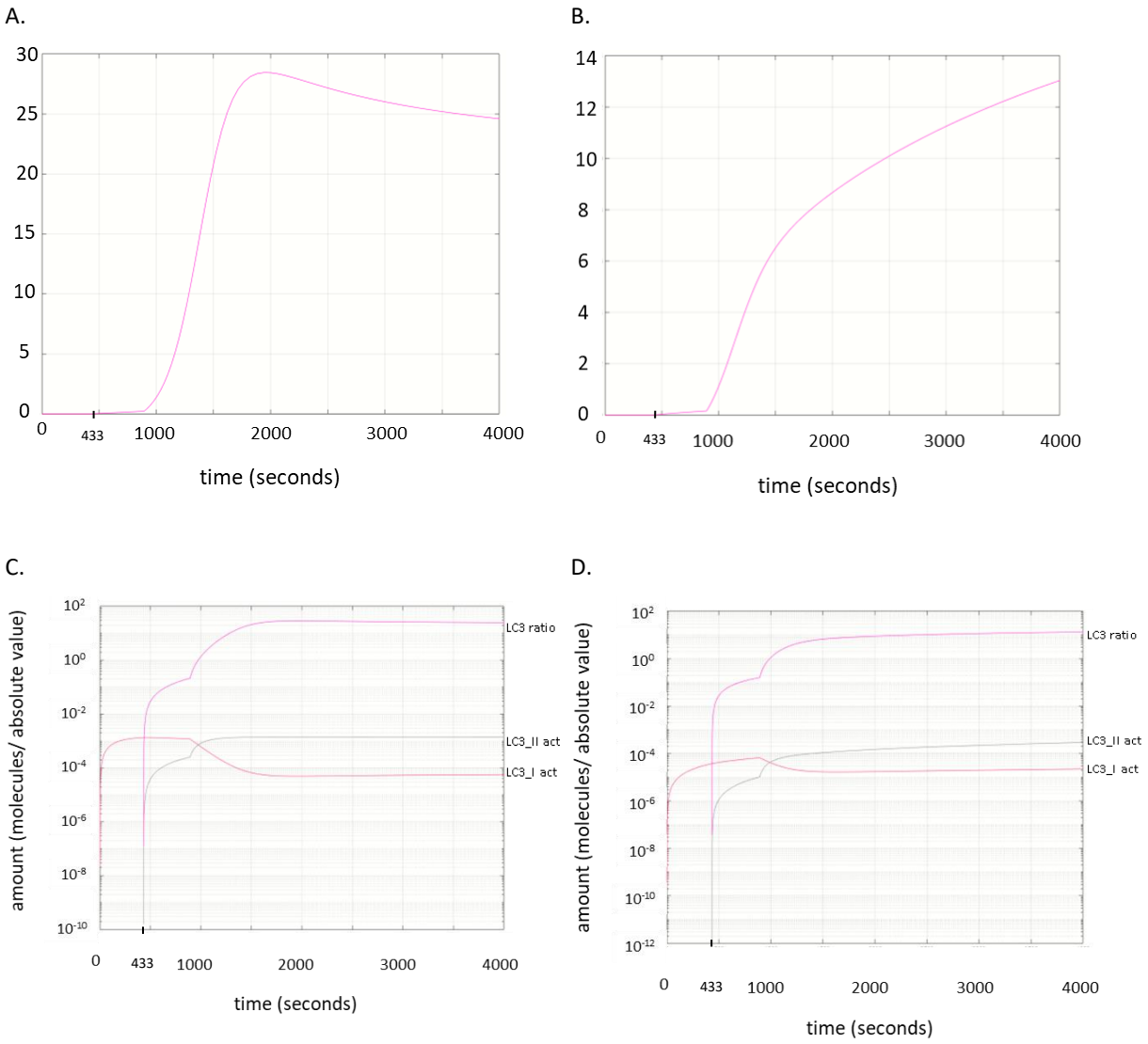


Figure 6-5. An example of the behaviour of LC3 ratio with the variation of a rate of reaction constant

*The LC3 ratio (LC3II_act/ LC3I_act) as resulted from the increase (x10, **A**&**C**) and decrease (x1/10, **B**&**D**) of a parameter, here kf_3. LC3 ratio: light pink, LC3II_act: grey, LC3I_act: dark pink. At t=0s, all of the ULK1 complex of the system is fully activated and all other entities are fully inactivated (except from the negative regulators which are fully activated). At t=433s, ATG16:ATG12:ATG5 complex is formed and leads to the formation of LC3-II_act from LC3-I_act. From the same timepoint and on the LC3 ratio is calculated.*

6.3.3 Comparison of model behaviour in healthy vs PD states

Here we have developed a mathematical model accounting for the initial stages of macroautophagy for a healthy individual. As discussed previously (Sections 4.1.1.2, and 6.1), macroautophagy is altered in brains of people with PD [487, 489, 652-655]. This is reflected in the differential expression of genes and protein amounts [668]. In this section, potential differential amounts of entities of the model in PD will be taken into consideration to simulate macroautophagy in someone with PD. The number of molecules of entities of the model for people

with PD were calculated as described in Section 6.3.3 and are presented in Table S6-6 (Appendix G). The amounts of the entities in the macroautophagy model demonstrated a mean difference of -0.47%, meaning that amounts in the PD were usually lower than in the healthy state (Fig S6-6). The most prominent difference was observed for PIK3C3-C1 (10.74% lower in PD), followed by a decrease in AMPK (-9.71%), and an increase in ATG16L1 and SMCR8 (+8.18%, and +7.56%, respectively). Smaller differences were observed for LC3I and ATG12 (-5.59%, and +5.41%), while the rest of the model entities had differences of less than 5%. The macroautophagy model was then solved using these amounts to simulate the PD state.

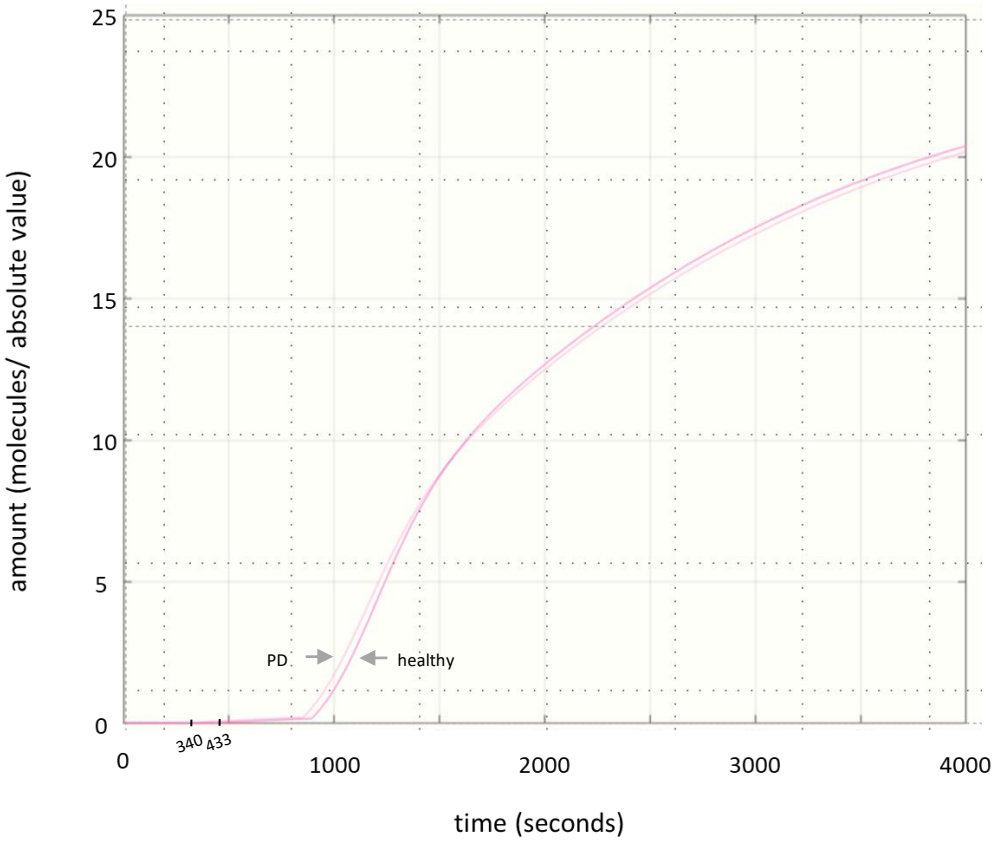
The overall dynamical behaviour of the model in the PD state is generally similar to that of the healthy state, as expected, but presents differences in the timings of some events. The results from this comparison are summarised in Table 6-1. For the majority of the entities, their activation is delayed in the PD compared to the healthy state. However, due to the faster activation of ATG16L1 (reaches its 10% of maximum activation at 340s vs 433s), the ATG16L1:ATG12:ATG5 complex is formed earlier (by 93s) leading to an earlier formation of LC3II from LC3I. These results are presented in greater detail in Fig S6-7, where small groups of entities are depicted (to be compared with Fig 6-3, Fig S6-4, and Fig S6-5, which present the same groups for the healthy state).

Table 6-1. Comparison of key points of the model simulation of a healthy vs PD state

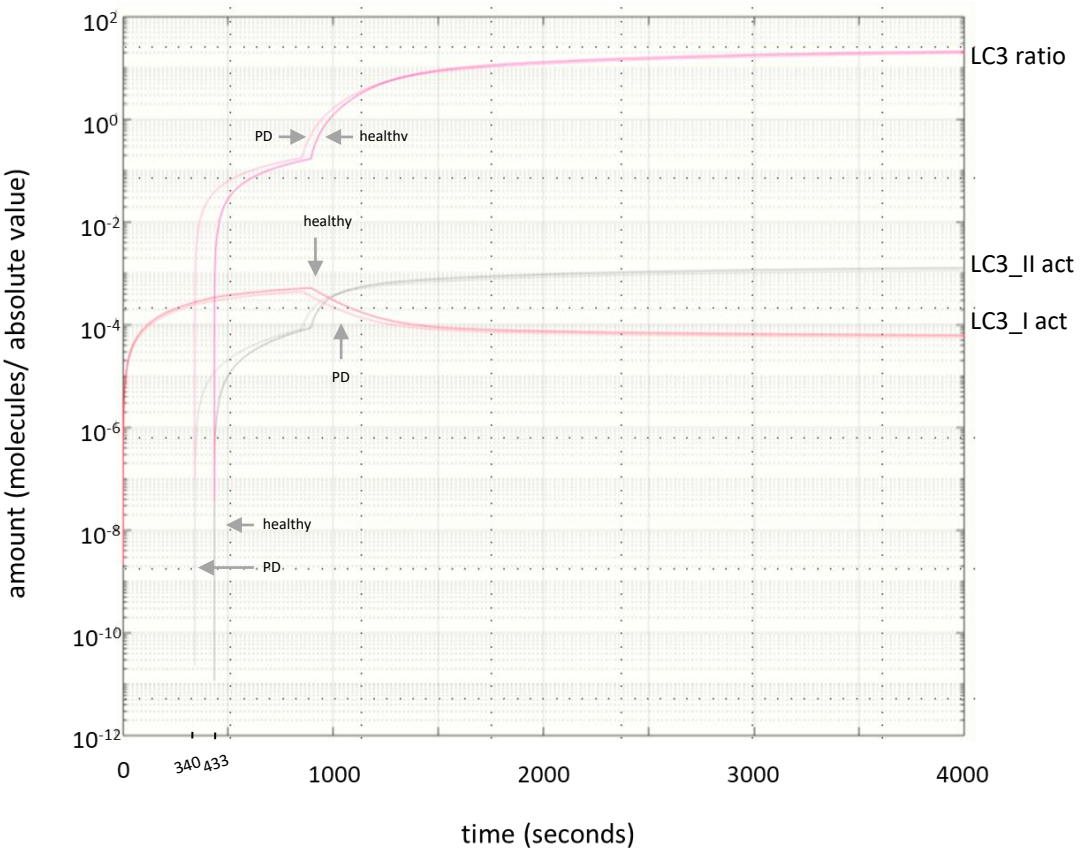
Components	Healthy	PD	Comparison
PIK3C3_C1 (active = inactive)	123s	127s	+4s
ATG9 (active = inactive)	407s	426s	+19s
BCL2 (active = inactive)	461s	477s	+16s
AMPK (active = inactive)	1565s	1576s	+11s
PI-PI4P (active = inactive)	436s	438s	+2s
PI-PI3P (active = inactive)	650s	678s	+28s
ATG16L1:ATG12:ATG5 (formation)	433s	340s	-93s
LC3I = LC3II_act	1288s	1280s	-8s
LC3I_act = LC3II_act	982s	940s	-42s
SMCR8:WDR41:C9ORF72 = ULK1complex_act	148s	144s	-4s

The LC3 ratio was also used to compare the two cases to explore whether PD would lead to noticeable differences. Interestingly, there was indeed a slight difference. In the PD simulation, the LC3 ratio increased earlier (t=850.7s vs t= 891.9s) but at a slower rate, leading to a slightly lower value at t=4,000s, compared with the simulation of the healthy state (Fig 6-6A). More details can be observed when LC3II_act and LC3I_act are plotted alongside the LC3 ratio using a logarithmic scale for the y-axis (Fig 6-6B). The results graphed in Fig 6-6B show that the LC3 ratio appears earlier in the PD case than the healthy state, which is due to the earlier production of LC3II_act (depicted in grey) (340s vs 433s). However, by 850s they reach similar levels, as LC3II has a slower rate of production in the PD case. After 850s an additional factor that drives the differences in the LC3 ratio between the healthy and PD case is LC3I_act (dark pink), which until that timepoint was similar in both states. LC3I_act starts decreasing more quickly and with a delay of a few seconds in the healthy state, whilst the production of LC3II_act is delayed but accumulates more quickly. At 4,000s LC3II_act has a lower value in the PD case compared to the healthy state by 8%. Therefore, even though the value of the LC3 ratio at 4,000s are similar in the two states (20.4 in the healthy state and 20.2 in the PD state; 0.98% difference), their dynamics differ, with macroautophagy (as expressed by the LC3 ratio) in PD starting earlier but being accelerated slower than in the healthy state, reaching slightly lower levels, as expressed by the LC3 ratio. Therefore, these results point towards different dynamics of macroautophagy in the simulation of PD, solely based on differences in the amounts of the model entities.

A.


amount (molecules/ absolute value)
25
20
15
10
5
0
PD
healthy
340 433
0
1000
2000
3000
4000
time (seconds)

B.


amount (molecules/ absolute value)
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
PD
healthv
healthy
PD
healthy
PD
LC3 ratio
LC3_II act
LC3_I act
0
340 433
1000
2000
3000
4000
time (seconds)

C.

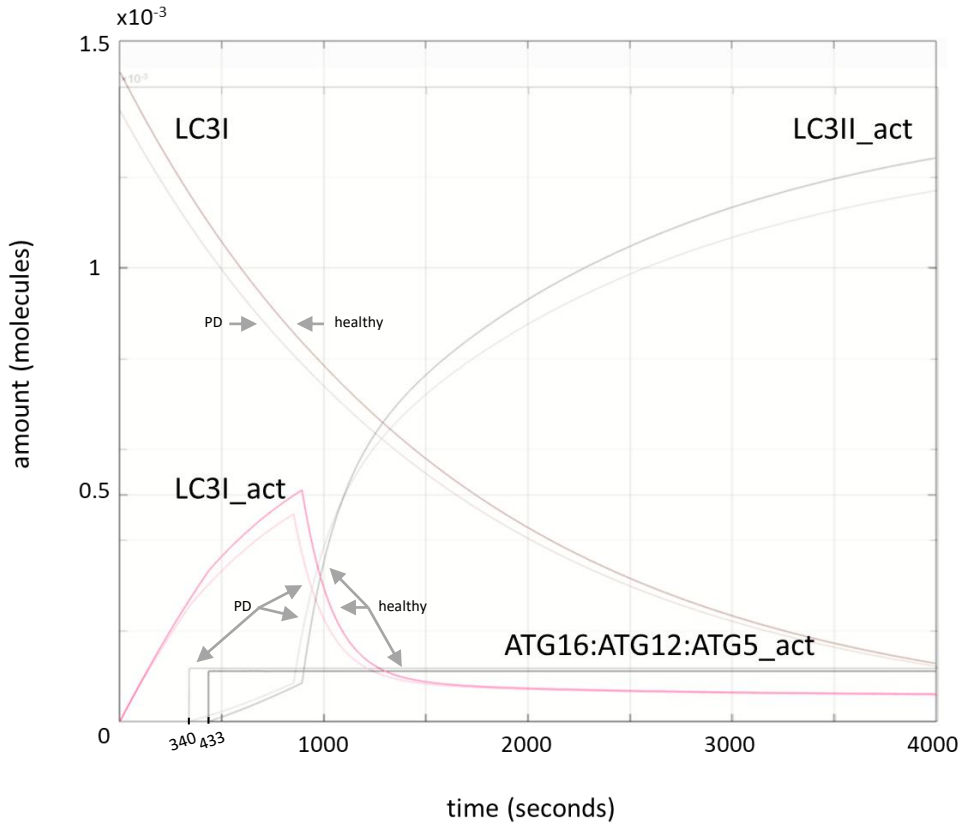


Figure 6-6. Comparing the behaviour of LC3s in the macroautophagy model in healthy vs Parkinson's disease states

(A) The ratio of LC3II_act and LC3I_act is compared between the healthy (darker hues) and Parkinson's disease (lighter hues) states for 4,000s. (B) The ratio of LC3II_act and LC3I_act is shown together with its individual components for healthy (darker hues) and Parkinson's disease (lighter hues) conditions for 0≤t≤4,000s in a logarithmic scale for the y-axis. (C) The different forms of LC3 are compared between the healthy and PD state (darker and lighter hues, respectively). At t=0s, all of the ULK1 complex of the system is fully activated and all other entities are fully inactivated (except from the negative regulators which are fully activated). In the healthy state, at t=433s, ATG16:ATG12:ATG5 complex is formed and leads to the formation of LC3-II_act from LC3-I_act. From the same timepoint and on the LC3 ratio is calculated. The equivalent timepoint in the PD state is 340s. LC3I (brown), LC3 ratio (pink), LC3I_act (dark pink), and LC3II_act (grey). ATG16 represents ATG16L1.

6.4 Discussion

In this chapter a mathematical model of the initial stages of macroautophagy signalling in humans was created. The model was formulated and parametrised by consulting the relevant literature, including published mathematical models. The macroautophagy model (Fig 6-2) was constructed based on the simplification of the macroautophagy process diagram (Fig 6-1). This reduction step led to a process diagram consisting of 26 entities connected by 45 relationships. Some of the components of the final model are considered key players of macroautophagy, such as the ULK1 complex, PIK3C3-C1, and BCL2, while the contribution of others is less strongly established. More

specifically, the role of CSNK2, and of C9orf72, SMRC8, and WDR41 individually and of their complex in macroautophagy is less clearly understood.

The model was solved and analysed using the SimBiology toolbox of MATLAB. Model results showed the initial events of macroautophagy to occur within an hour, which is in accordance with the literature [586, 667]. These events included the activation of PIK3C3-C1 by the ULK1 complex, the subsequent production of PI3P, activation of ATG16L1, formation of ATG16L1:ATG12:ATG5, and production of LC3II from LC3I, which are considered the main initial events of macroautophagy [279, 281, 670]. In parallel, a positive feedback loop was observed that was composed of ATG9, which is activated by the ULK1 complex, leading to an increase in PI4P levels, which in turn is a positive regulator of the ULK1 complex. A negative feedback loop was also formed, which included the ULK1 complex, AMPK, and ATG9. In more detail, the ULK1 complex inhibits AMPK, which normally activates the positive regulator of macroautophagy, ATG9. These loops lead the ULK1 complex to be retained in an activated state, with only a temporary small decrease of 1% at approximately 900s. Of note, as this model is designed to study the initial stages of macroautophagy, events of the next stages of macroautophagy were excluded from this work, leading the system to reach a steady state at around 10,000s (~3h).

Local sensitivity analysis was performed to reveal the main parameters (rates of reaction constants), that affect the behaviour of the model. The selected output for this analysis was the ratio of LC3II and LC3I (i.e., LC3 ratio), as it is a typical marker of macroautophagy progression [34]. LSA at 4,000s and for $0 \leq t \leq 4{,}000$s, suggest similar lists of rates of reaction constants to be the most influential in macroautophagy. Those parameters having the most influence included the activation of LC3I (kf_3), regulation of LC3II levels from SMCR8:WDR41:C9ORF72 (kf_10 and kf_34), formation of LC3II by ATG16L1:ATG12:ATG5 (kf_22), by ULK1 complex (kf_24), and by CSNK2 (kf_26), and the formation of SMCR8:WDR41:C9ORF72 (kf_28). Interestingly, the ULK1 complex had an impact on the LSA at 4,000s but less so for the $0 \leq t \leq 4{,}000$s, in contrast to the rest of the components that had an impact in both analyses. Surprisingly, the two entities less robustly associated with macroautophagy by the literature, CSNK2 and SMCR8:WDR41:C9ORF72, presented the strongest effect on the system. In the LSA at 4,000s, kf_26 (the rate by which CSNK activates LC3II) had the greatest effect on the LC3 ratio (-80% to +758%; range of 838%), followed by kf_10 and kf_28 (the rate by which SMCR8:WDR41:C9ORF72 inhibits LC3II and forms itself, respectively; -88% to 243% and -7% to +81%; range of 331% and 88%, respectively). Of note, kf_22, which is the rate of the reaction by which LC3II is formed by ATG16L1:ATG12:ATG5 that is considered its main activator, had a smaller effect on the output than the aforementioned entities (-7% to +62%; range of 69%). These results need to be interpreted with caution, as the mechanism through which CSNK2 and

SMCR8:WDR41:C9ORF72 affect macroautophagy is not fully understood, and therefore their representation in this model, might not be an accurate reflection of their physiological role. Nevertheless, this analysis highlights the need of further studies focused on the role of CNSK2 and SMCR8:WDR41:C9ORF72 in macroautophagy, as they seem to have a strong effect on the level and dynamics of macroautophagy.

Subsequently, the model of macroautophagy was used to compare two states: healthy *versus* PD. This was performed using amounts of entities of the model, obtained for healthy people and those with PD (data extracted from the literature [668]). At first the amounts of the entities were compared between the two states, revealing that on average most were lower in PD (including PIK3C3-C1, and AMPK), but interestingly ATG16L1, ATG12, and SMCR8 had higher levels in PD compared to the healthy state. Then the model was solved using the PD amounts and the simulation demonstrated that a number of entities of the model formed more slowly compared to the healthy state. For instance, the evaluation of PIK3C3-C1, ATG9, BCL2, and AMPK showed a delay by 4s, 19s, 16s, and 11s, respectively in the PD compared to the healthy state (time point at which the activated and non-activated states of each component have equal amounts). In contrast, ATG16L1 was activated 93s faster (time point at which it reaches its 10% of maximum activation), leading LC3II to be formed 42s faster (timepoint when the amounts of LC3II_act and LC3I_act are equal) in PD compared to the healthy state. Even though the dynamics of the process differ between the two states, the value of the LC3 ratio at 4,000s is similar (0.98% difference). However, the amount of LC3II was lower in the PD state compared to the healthy state by 8%. Biological systems are sensitive and potentially this small change could lead to a lower efficacy of macroautophagy, which could be magnified by an overload of the autophagic machinery due to a higher amount of misfolded proteins and aggregates, as occurs in PD [219, 220]. Of note, these results were solely based on differences in the amounts of model entities between healthy people and people with PD, as described in literature. Additional alterations in the macroautophagy process as a result of the cause of the disease (e.g., mutations in LRRK2) were not included. This will be discussed further in Section 7.2.2.

7th Chapter:

General discussion

7. General discussion

The work presented in this Thesis is centred around systems biology approaches applied to the investigation of mechanisms resulting in neurodegenerative diseases, with a particular focus on the Hereditary Spastic Paraplegias (HSPs), Parkinson's disease (PD), and macroautophagy (MA). The results of this Thesis have been discussed in detail in each chapter. Therefore, in this section, I will summarise the key findings, discuss the limitations, and suggest directions for future work.

7.1 Key findings

7.1.1 Systems biology analysis of HSPs

The HSPs are a group of complicated neurodegenerative diseases, first described over a century ago. There are more than 70 genetic types of HSPs, and various clinical presentations [169], creating a genetic and clinical heterogeneity that is hindering our understanding of the disease mechanism(s). A systems biology approach, and more specifically a network approach, was adopted to holistically study the genes that when mutated lead to the development of the HSPs. The uniqueness of this work, compared to previous studies of HSPs, lies upon multiple steps of creating the network. The seeds of the network were genes that lead to HSPs, excluding genes with related phenotypes. The PPIs of the seeds were collected through PINOT [395], and those with a "final score" over 2 were retained (higher the final score, the higher the confidence on the interaction). Therefore, only quality controlled, experimentally proven, and manually curated PPIs between human proteins were included in the analysis. This approach aids in the exclusion of false positive data, a particularly important consideration when handling large data sets, such as PPIs.

The global HSP network was visualised based on the filtered PPIs and consisted of 1 interconnected graph that included the majority of the HSP seeds and 14 smaller graphs. From the interconnected graph of the global HSP-PPIN, the core network was extracted by including protein interactors common to at least 2 seeds, as the aim of this analysis was to identify commonalities between seeds. 53 HSP seeds were retained in the core network. The 2nd layer interactome showed that nearly all seeds were connected in a graph and suggested that the rest of the seeds were excluded due to limited availability of PPI data. The connectivity of the seeds in the 1st layer core network indicated their functional association and hinted that they might be involved in shared biological processes and pathways.

Indeed, further analysis revealed that terms associated with the functional blocks “Transport” and “Protein localisation” were enriched. More specifically, terms were related to protein transport, regulation of protein levels, and vesicle dynamics and transport (e.g., “protein targeting”, “ERAD pathway”, “multivesicular body organisation”, “vesicle mediated transport”), which help in prioritising some of the multiple previously suggested mechanisms for HSPs, such as membrane cargo and trafficking [170, 332-334, 454]. Of note, the GO term “substantia nigra development” was amongst the results, which aligns with clinical features of HSPs that lead to involuntary movements and loss of fine movement control [166].

Interestingly, most genes associated with the HSPs in the literature that were identified after the creation of the network [398, 404-408] were part of the HSP networks (either the global network or the core of the 2nd layer network), which highlights the ability of PPINs to be used as predictive tools for detecting genes, and a potential strategy for accounting for the missing heritability of diseases [103]. On that note, a HSP candidate gene with limited data supporting a causative association with HSP, *CCDC50* [422], was part of the core network, which suggest that it could be the gene responsible for SPG14. Clinical and experimental validation is required to test this hypothesis.

The clinical diversity of HSPs prompted an interest in exploring the existence of mechanistic subtypes of HSPs. This analysis was performed by mapping clinical data (mode of inheritance, type of HSP, and clinical features) in the core HSP-PPIN and analysing the respective parts of the network via enrichment. There was no evidence of mechanistic subtypes based on mode of inheritance or type of HSP. Conversely, machine learning tools suggested the functional clustering of clinical features in two groups, TS and EPOD (thinning of corpus callosum and seizures; early onset, peripheral neuropathy, optical atrophy, and dementia/mental retardation). Enrichment of the clusters suggested protein quality control and degradation to be especially important for the TS cluster, while regulation of protein localisation and transport in neurons via vesicles to be more tightly associated with the EPOD cluster.

7.1.2 Overlap of macroautophagy and neurodegenerative diseases

Macroautophagy (MA) is associated with multiple neurodegenerative diseases (NDs) including Alzheimer’s disease (AD), PD, Amyotrophic lateral sclerosis (ALS) and Frontotemporal dementia (FTD) [250, 251, 482, 484, 487, 489, 505, 512, 516, 517]. This relationship was studied by assessing the overlap of the PPINs built around macroautophagy seeds and disease seeds.

The MA network consisted of 12,360 components (151 seeds, 2,832 direct interactors, and 9,377 interactors of the direct interactors). The limited availability of PPI data for 5 seeds led to lack of connections with other components of the network and their exclusion from further analyses. On the other hand, proteins related to ubiquitylation of substrates were amongst those with the largest number of connections within the network (i.e., hubs), which was expected due to their biological role.

To investigate the relationship of macroautophagy with neurodegenerative diseases, the presence of the ND seeds in the MA network was investigated. Most seeds were part of the MA network, and this result was significant for ALS, PD, and FTD, but not for AD. Interestingly, most of the missing seeds interacted directly with a 2nd layer interactor of the MA network. Therefore, an additional analysis was performed, in which the overlap of the ND networks with the MA network was assessed. Interestingly, the ND networks were almost completely embedded within the MA network further supporting the strong association between the studied diseases and macroautophagy. This is in agreement with literature and also with the biological role of macroautophagy in disposing overexpressed and misfolded proteins [249, 252-255, 490-492].

7.1.3 Mathematical model of macroautophagy

In order to gain a deeper understanding of the mechanisms and events underlying these associations, the interplay of macroautophagy with neurodegeneration was investigated further. Since the overlap with all tested NDs was extensive, a single disease, PD, was selected to be studied in more detail. In order to simulate macroautophagy in a healthy person vs a person with PD, testing for any differences, a mathematical model of macroautophagy was created. The model was firstly based on a physiological state (“healthy” state) and focused on the initial stages of macroautophagy, including key macroautophagy regulators and their direct interactors. The model was solved using the SimBiology toolbox of MATLAB, and the behaviour of the model was found to be in accordance with the literature. Interestingly, local sensitivity analysis showed that the highest effects on the system were induced by CSNK2 and SMCR8:WDR41:C9ORF72, which are the entities of the model the most understudied in relation to their precise mechanistic link with macroautophagy.

CSNK2 is part of the CK2 complex which catalyses the phosphorylation of more than 300 substrates thereby constituting one of the most pleiotropic members of the human kinome [671, 672]. For instance, it has been linked to neurite outgrowth [673], and suggested to have a neuroprotective

effect on the brain through modulating the NADPH oxidase after ischemic brain injury [674]. CK2 has been associated with multiple diseases, including diabetes [675], multiple types of cancers [676] and neurodegenerative diseases, such as Huntington's disease [677], Alzheimer's disease [678], as well as Parkinson's disease [677]. The suggested mechanism for the associations with Parkinson's disease varies. For instance, CK2 was proposed to phosphorylate α-synuclein [679]. However, while this result was performed in vitro, further in vivo studies failed to support it [677, 680]. Therefore, the mechanistic link with Parkinson's disease remains elusive [677]. On the other hand, data linking CSK2 with autophagy are more limited (PubMed resulted in just 27 papers for "CK2 kinase autophagy", while for LRRK2 and autophagy, and ULK1 and autophagy the respective results were 282, and 1,795, respectively when queried on 13th June 2022). They include the inhibition of autophagic degradation of Keap1 [681] and an effect on autophagy in astroglial cells through a pathway with at least 4 intermediate steps [682], or by multiple parallel pathways under calorie restriction conditions [683], all of which are upstream of the ULK1 complex. However, in this study we show that CK2 has a direct effect of high impact on macroautophagy, as simulated in the mathematical model.

Repeat expansions in the C9orf72 gene cause familial forms of Amyotrophic lateral sclerosis and Frontotemporal dementia [684-686]. The resultant protein is required fort normal lysosome homeostasis [687-692]. However, C9ORF72 is unstable and functions as part of the SMCR8:WDR41:C9ORF72 complex [691-695], which was first identified in 2016 [692, 695]. Little is known about it and its exact functions still need to be elucidated [696]. Research has revealed that it acts as GAP (GTPase activating protein) for RAB GTPase proteins, such as RAB8a and RAB39b to control macroautophagy [695], and as a GEF for other RABs (e.g., RAB5, RAB7, and RAB11) [697]. In regards to macroautophagy, it associates with the ULK1 complex [692] to regulate its expression and activity [698] and also affects later stages of macroautophagy by regulating the expression of lysosomal enzymes [698]. Its localisation is consistent with these roles, as it is detected in the cytoplasmic surface of lysosomes [691, 699]. Interestingly, it is recruited to lysosomes when amino acids are scarce through an interaction with a lysosomal cationic amino acid transporter PQLC2, with this interaction being negatively regulated by amino acids levels [700, 701]. Interestingly, studying its association with PD, revealed that repeat expansion of C9ORF72 was present in a patient with early onset and familial PD [702]. In addition, C9ORF72 was found to be differentially expressed during the progression of PD [703]. The main role of the complex in regulating RAB GTPases could be the way through which is it associated with PD, as these proteins have been strongly linked to the disease mechanism [704].

Data from the literature were then used to simulate a PD state, by creating a second list of amounts for the entities of the model based on literature. The simulation suggested that in the disease conditions, the LC3 ratio (i.e., LC3II/LC3I) increased earlier but reached slightly lower levels (by 8%) at around 1h after the activation of ULK1 complex. The kinetics of LC3I were identical in both conditions, so this difference in the LC3 ratio was solely due to higher levels of LC3II in PD compared with the healthy state. This difference was marginal (8%) however biological systems are sensitive and this small change could lead to larger discrepancies in the kinetics of macroautophagy, which could be amplified further by a higher concentration of misfolded proteins and aggregates, as it occurs in PD [219, 220] that could overload the autophagic machinery. Furthermore, the only modelling parameters that differed between the two states (healthy vs PD) were the amounts of the entities of the model. Therefore, incorporating additional alterations in the autophagic process due to the cause of the disease (e.g., LRRK2 mutations) is hypothesised to lead to stronger effects on macroautophagy in people with PD compared with healthy people.

7.2 Limitations and Future directions of my projects

7.2.1 PPIN analysis and clinical data availability

Even though the potential of PPIN analysis is substantial, its drawbacks also need to be discussed. The results derived from PPIN analyses require cautious interpretation depending on the PPI detection method (e.g., experiment *versus* prediction). Some of the limitations of PPIN analysis, as discussed in Section 1.2.5 were addressed by using PINOT, due to its filtering and scoring system, and exclusion of predicted data. However, the PPI results from PINOT were still affected by the ascertainment bias (discussed in greater detail in Section 1.2.3), increasing the risk of including false positive data. In addition, this approach represents all protein products of a gene with a single node, leading to potential loss of valuable information regarding the system's mode of function (e.g., isoform specific information is excluded). Moreover, the dynamic alteration of protein functionality through the regulation of gene expression, protein level, and activity is not taken into consideration in PPIN analyses, limiting the ability to represent the system's fine mechanistic details. Finally, similarly to all other in silico analyses, the results have to be experimentally validated.

Upon analysis, a bias to disease-related proteins was observed in the PPIN results. Even though it is a natural consequence of the community and researchers being interested in treating and curing diseases, resulting in more funding (and hence more experiments) investigating disease implicated

proteins, it also skews analyses towards disease genes. For instance, out of the 173 LRRK2 interactors that had a final score higher than 3 in Chapter 5, most were associated with at least 2 human diseases (85.0%), while 50% were associated with 6 or more diseases (data from UniProt, 16/3/2022).

Over time, as more protein-protein interactions are detected and as high-quality PPI data accumulate, the confidence in the results of PPIN analyses will increase. Identifying gaps of knowledge by studies such as the current one, can aid in this endeavour. More specifically, future research could focus on the missing seeds from the core HSP-PPIN of Chapter 2 (i.e., *ATL1*, *B4GALNT1*, *C12orf65*, *ENTPD1*, *SPG11*, and *TECPR2*), such as the study of Cogo et al [705], and from the study of the interplay of macroautophagy and neurodegeneration of Chapter 4 (i.e., *TMEM74*, *TOMM6*, *TOMM7*, and *PRPH2*). In addition, more PPI experiments should be conducted for the LRRK2 interactors, which were excluded from further analysis due to lack of data in Chapter 5 (i.e., *BCL2*, *CASP8*, *DNM1*, *EIF2A*, *ERLIN1*, *KIF21A*, *KLC2*, *RAB12*). The issue of data availability is evident in the latter analysis, as 1,261 proteins failed to be confirmed as LRRK2 interactors (detected with less than 2 publications or two methods), which corresponds to 87.9% (n=1,261/1,434) of its total interactors. An example of an approach that could be adopted for the detection of new and validation of previously suggested protein interactions is a Mass Spectrometry analysis of the interactome of the protein of interest, which was previously extracted through tagging followed by a pull-down assay. In addition, multiplexing (analysing multiple samples in a single experiment) in Mass Spectrometry is becoming more common [706] which increases the speed of such experiments.

The data availability problem was also related to the clinical features of people with HSPs. Therefore, to aid towards this direction, people from various scientific communities and sectors could collaborate to set up a database that includes a detailed and regularly updated list of symptoms and mutations in genes/loci, as agreed by most experts, and that is accessible to the other researchers for further exploration.

Conversely, only a subsection of published PPI data might be appropriate for an experiment. Generating and publishing PPI data from specific tissues and cell types can allow the selection of PPIs that are occurring in the biological context relevant to the scope of each experiment (e.g., brain or neuron specific analyses). Such endeavours have recently blossomed, as multiple research groups are generating single cell omic data [66].

Such future work might fill in the gap of the PPI and clinical data, which would render PPIN analyses more powerful. PPIN analyses are already capable of guiding researchers towards detecting the

underlying pathways and processes shared by multiple genes involved in complicated diseases, thus hinting about the disease mechanism(s) and accelerating the development of targeted therapies. Furthermore, due to PPIN analysis' ability to identify groups of proteins connected with disease genes, it can be developed as a predictive tool to prioritise candidate proteins/genes, which then can be genetically evaluated in the search for causes of missing heritability. Finally, by analysing PPINs created using as seeds genes associated with diseases that belong to the same spectrum, indications can be provided regarding mechanistical associations among them.

7.2.2 Mathematical modelling of macroautophagy and its relationship with neurodegeneration

The investigation of associations between LRRK2 interactors and macroautophagy in Chapter 5, revealed two substantial issues. The first was the bias introduced in the study based on the selection of papers and sources used to extract information. Although efforts were made to make this decision as objective as possible, a different list of criteria would have resulted in a different outcome. This became evident when some of the recent papers of Dr Sharon Tooze were included in the analysis, and an additional LRRK2 interactor was found to be linked with macroautophagy (1/8, 12.5% increase in the total number of interactors). Unfortunately, there is no apparent solution for this problem, as any selection of papers would lead to a different result. On the other hand, incorporating all of the papers with the term "autophagy" could only be performed with automated text-mining, due to the volume of published work. However, this solution has its drawbacks compared to manual evaluation; for instance, the confidence in the quality of each piece of collected information would be reduced.

The second issue was the multiplicity of methods that had to be utilised for this analysis. In addition to manual text-mining for the list of official gene names, alternative gene names, official and alternative protein names, as well as names of complexes in which the gene product takes part had to be performed. This process was time consuming, intrinsically error-prone (due to possible misinterpretation of the mined information), and difficult to automate. Alternatively, a researcher can decide to use data already curated in databases (such as Reactome), which however might not include some of the latest research findings.

After selecting the LRRK2 interactors associated with macroautophagy, mutational data affecting individual interactions were required. Such data have been collected by and made available through IntAct for a multitude of human proteins and were explored to collect interactions affected by mutations in LRRK2. Human data regarding bidirectional interactions with proteins of

macroautophagy were found for 9 out of 17 LRRK2 interactors (53%). The completion of this valuable data set of IntAct by primary data generation will require resources and expertise but will be beneficial in the endeavour of understanding the disease mechanism of mutated proteins, such as LRRK2.

The basis of a mathematical model for the initial stages of macroautophagy was developed in Chapter 6 lies in the known relationships between the components and their amounts and parameters. For the former, even though the process of macroautophagy has been studied for a long time by various research groups around the world, several of its aspects are still unclear [34]. This relates to all steps in the macroautophagy process but especially to the later ones, as the initiation step has been studied in more detail in numerous organisms. In addition, many different models have been used to study macroautophagy, so simply compiling the available data would be unfavourable for modelling, as the number and identity of autophagic components, and the rates of reactions, can vary in different biological models [34]. Therefore, further research of the autophagic process is required to elucidate the role of understudied events, as well as effectors, like CSNK2, and SMCR8:WDR41:C9ORF72, whose importance is suggested by this Thesis.

Additionally, the simplification of the model might have introduced a bias to the model. Even small perturbations of the amounts of the model entities, affect the behaviour of the macroautophagy readout, as it was observed from the comparison of the healthy state vs the PD state. Therefore, a slightly diverse set of rules for the simplification step might have led to a macroautophagy model with a different list of entities that could be present in different amounts, leading the model to demonstrate a distinct behaviour.

In parallel, instead of modelling an “average” cell, modelling a specific cell type that is strongly associated with the aim of each study could be beneficial. However, such data sets are usually sparse and incomplete. Therefore, there is a need for generating data about the relative reaction rates and amounts of components of macroautophagy in a cell-type specific manner. Even though such experiments would be costly and time consuming, the benefits from being able to model macroautophagy in specific brain cell types with higher accuracy could be considerable, as they might elucidate the mechanism(s) of the differential response and vulnerability of certain cell types to the same disease [707]. It is of note, however, that the absence of accurate numerical data informing model parameters does not prevent analysis of this system, as qualitative data can still provide valuable constrains that the model needs to satisfy to make accurate predictions, as shown in previous studies [78].

The unavailability of the rates of reactions and amounts of the proteins, complexes and lipids in the model created the need of their calculation from data in the literature in Chapter 6. This step required certain assumptions to be made. For example, it was assumed that some types of reactions occurred at the same rates (e.g., phosphorylation and ubiquitylation). It was also assumed that the quantities of the entities of the model calculated based on a proteomic analysis of specific regions human brains were accurate and indicative of their local concentration in the region of the cell in which the phagophore was forming. These assumptions were necessary due to lack of data, but they might not fully reflect the physiological state within human cells. For example, averages of protein amounts were used for each brain region. In reality, they might differ amongst the different cell types composing the brain region. Therefore, a more granular analysis of protein amounts needs to be performed.

Regarding the protein concentrations, their accurate measurement is a technically challenging task. There are techniques available for quantifying protein concentrations, which are for instance based on ELISA, but they are costly and available only for specific proteins for which an adequate antibody has been developed. In addition, the concentration of a protein within a cell might not be indicative of the local concentration in the region of interest, which is a much more challenging task for experimental biologists. The limitations in accurate protein concentration measurements have led researchers to opt for using transcriptomic data instead, which is -however- also problematic, as mRNA levels might not necessarily be indicative of protein levels [708]. Recent advances in proteinomics methods have enabled the conducting of quantitative experiments in the level of protein that analyse multiple samples in a single experiment [709, 710], detecting the interactome in the single cell level [711], or even in a subcellular region manner [712].

The current limitations in this field should fail to discourage the scientific community from the endeavour of investigating biological processes. As more research is being performed around macroautophagy and macroautophagy related proteins, the gaps of knowledge will narrow, and the produced mathematical models, such as the one produced in this Thesis can be optimised and updated to reflect the new understanding of the process.

Optimisations performed in the macroautophagy model of Chapter 6 could initially focus on parameterisation. In vitro experiments describing the kinetics of key reactions of the model and thus providing more accurate rates of reactions, would further strengthen the predictive power of the model. In parallel, it would be useful to retrieve or create a data set of concentrations from human iPSCs derived from healthy people and people diagnosed with PD. Such data from iPSCs of patients could be insightful, as the effects of the disease on the cell physiology have occurred for

less time compared to cells derived from post-mortem tissues. Therefore, this approach allows the study of the initial stages of disease and the pathobiological changes, and limits the extent of inclusion of changes due to the activation of protective mechanisms against the disease, the impact of agonal state or post-mortem delays in tissue processing [713]. Additionally, iPSCs would produce homogenous population of cells thus allowing the study of the macroautophagy dynamics in a single cell type.

Currently, an increasing number of research groups are producing omic data for specific cell types [66]. This could be beneficial for fitting the macroautophagy model to a cell type of interest, (e.g., neurons or astrocytes). Such an approach would allow the modelling of potential cell-type specific mechanisms of macroautophagy and thus identification of cell-type specific drug targets. Interestingly, Filippi-Chiela et al [714] found that phenotypes can be masked when a mix of cell populations is studied. They demonstrated an absence of correlation between senescence and macroautophagy in human glioma cells, in contrast to a negative correlation when more cell types were included in the experiment. An advance towards the direction of single cell analysis in macroautophagy has been made by Xu et al [715] and Hu et al [716]. The former investigated the dynamics between macroautophagy and apoptosis using single cell sequencing in tandem with live microscopy imaging, whilst the latter performed single cell RNA sequencing to study the role of macroautophagy in the cell-renewal and differentiation of hematopoietic stem cells.

In this Thesis macroautophagy was modelled in a general manner. Differences between the signalling processes that occur with different stimuli that activate macroautophagy were not taken into consideration. In addition, differences in pathways can be noted based on the specific tissue or cell type under study. Therefore, future development of this model could be targeted at developing different versions depending on the selected stimulus, tissue, and cell-type of interest. The scarcity of data for such specific models could be addressed by performing wet-lab experiments that aim to fill that gap of knowledge or test the model, as previously achieved by other research teams [582, 717, 718].

In parallel to optimizations, more experiments can be performed with the current version of the model of macroautophagy for studying Parkinson's disease. Data regarding LRRK2 mutations and their effect on interactions with other proteins that have direct associations with entities of the model have been collected for this study and are described in Appendix G and summarised in Table S6-7. Simulating this effect for each mutation could hint whether the altered strength of protein-protein interactions could explain in part the differential severity of various LRRK2 mutations in human health.

Interestingly, there are differences in the PD pathology among brain regions [211]. A source of information for relative amounts of mRNAs (which is used as a proxy for the amounts of proteins) in post-mortem analyses of multiple brain regions from healthy people is Braineac (http://www.braineac.org/; [669]). Therefore, Braineac data were used to extract relative amounts of entities that are part of the mathematical model of macroautophagy, which were then converted into amounts (see Table S6-8), as previously described in Section 6.3. Future work in this project can utilised these amounts and explore whether the developed model predicts a change in macroautophagy, for instance in the LC3 ratio, in a brain region specific manner in healthy people, and whether macroautophagy is affected differently by PD, depending on the brain region considered.

In addition, the predictive power of the mathematical model of macroautophagy could be explored. Decreasing the quantity of a component or the rate of a reaction could simulate the effect of an inhibitor and predict which reactions and components would affect the model in the desired manner (e.g., rescuing the effect of a mutation). Some of these results could be validated through experiments with inhibitors or knock-downs in cultures of human cell lines. A validated mathematical model of macroautophagy could be used as a tool to screen the effect of various pharmacological components targeting macroautophagy in a fast and cost-effective manner.

7.3 Conclusions

In this Thesis, systems biology approaches were used to shed light on neurodegenerative diseases and their relationships with the complicated biological process of macroautophagy. Disease mechanisms were prioritised against others for the HSPs, and using machine learning tools, two clinical subgroups of HSPs were suggested to exist with potential disease mechanistic discrepancies. Then, macroautophagy and its association with four neurodegenerative diseases was assessed and hinted to be extensive. Therefore, I focused on further investigating its relationship with one disease, Parkinson's disease. For this purpose, a mathematical model of macroautophagy was created, which suggested that macroautophagy in the disease state has differential kinetics compared to the healthy state. Further optimisations and explorations of the model predictions could highlight potential interventions in the macroautophagy pathway that could ameliorate the pathobiology observed in people with Parkinson's disease. A deeper insight into macroautophagy and its link with neurodegenerative diseases through in vivo, in vitro and in silico analysis could lead to the development of disease-modifying therapies for a variety of illnesses, such as cancer, cardiovascular disorders, and neurodegenerative diseases [245, 246].

As demonstrated in this Thesis, systems biology approaches can shed light on complicated processes and human diseases. Specifically, PPIN analysis is a powerful tool that can extract, capture, and combine a large volume of knowledge in a relatively quick and easy manner, creating a comprehensive picture that can summarise state-of-the-art knowledge, helping to confirm existing theories, as well as facilitate the identification of uncertain areas that require further investigation. In parallel, mathematical modelling can simulate complex biological processes, increase our understanding of their mechanisms and suggest potential therapeutic interventions. Overall, the potential of systems biology approaches is vast and could accelerate the discovery of the several missing links within neurodegenerative diseases and in their relationships with macroautophagy, as well as the identification of promising sites for disease-modifying interventions.

References

1. Merriam-Webster, I. *Merriam-Webster Dictionary*. [webpage] 2021 14/12/2021 [cited 2021 15/12/2021]; Available from: https://www.merriam-webster.com/dictionary/system.
2. Trewavas, A., *A brief history of systems biology. "Every object that biology studies is a system of systems." Francois Jacob (1974).* Plant Cell, 2006. **18**(10): p. 2420-30.
3. Administration-NASA, N.A.a.S. *Mars helicopter status updates*. 2022 19/01/2022 [cited 2022 25/01]; Available from: https://mars.nasa.gov/technology/helicopter/status/.
4. Clapeyron, B.t.P.E.m. and K.E. Schreber, *Abhandlung über die bewegende Kraft der Wärme*. Ostwald's Klassiker der exacten Wissenschaften,. 1926, Leipzig,: Akademische Verlagsgesellschaft m.b.H. 2 p.l., 46 p.
5. Voit, E.O., *A First Course in Systems Biology*. 2012: Taylor & Francis Group.
6. Bertalanffy, A.R., et al., *L. von Bertalanffy, General System Theory*. 1968: New York: George Braziller.
7. von Bertalanffy, L., *Der Organismus als physikalisches System betrachtet.* Naturwissenschaften, 1940. **28**(33): p. 521-531.
8. Mesarović, M.D., *Systems theory and biology—view of a theoretician*, in *Systems theory and biology*. 1968, Springer. p. 59-87.
9. Smith, H.H., *Effects of Genome Balance, Polyploidy, and Single Extra Chromosomes on Size in Nicotiana.* Genetics, 1943. **28**(3): p. 227-36.
10. Evans, G.A., *Designer science and the "omic" revolution.* Nat Biotechnol, 2000. **18**(2): p. 127.
11. Health, N.I.o. *Human Genome Project Results*. 2018 [cited 2022 25/01/2022]; Available from: https://www.genome.gov/human-genome-project/results.
12. McPherson, J.D., et al., *A physical map of the human genome.* Nature, 2001. **409**(6822): p. 934-41.
13. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
14. Venter, J.C., H.O. Smith, and M.D. Adams, *The Sequence of the Human Genome.* Clin Chem, 2015. **61**(9): p. 1207-8.
15. International HapMap, C., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
16. Fairley, S., et al., *The International Genome Sample Resource (IGSR) collection of open human genomic variation resources.* Nucleic Acids Res, 2020. **48**(D1): p. D941-D947.
17. Kim, Y.A. and T.M. Przytycka, *Bridging the Gap between Genotype and Phenotype via Network Approaches.* Front Genet, 2012. **3**: p. 227.
18. Willyard, C., *New human gene tally reignites debate.* Nature, 2018. **558**(7710): p. 354-355.
19. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
20. Williams, T.N., *Human red blood cell polymorphisms and malaria.* Curr Opin Microbiol, 2006. **9**(4): p. 388-94.
21. Pearson, T.A. and T.A. Manolio, *How to interpret a genome-wide association study.* JAMA, 2008. **299**(11): p. 1335-44.
22. Munoz, J. and A.J. Heck, *From the human genome to the human proteome.* Angew Chem Int Ed Engl, 2014. **53**(41): p. 10864-6.
23. Jensen, O.N., *Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry.* Curr Opin Chem Biol, 2004. **8**(1): p. 33-41.
24. Lodish, H., et al., *Molecular cell biology*. 2008: Macmillan.
25. Roberts, G.C. and C.W. Smith, *Alternative splicing: combinatorial output from the genome.* Curr Opin Chem Biol, 2002. **6**(3): p. 375-83.

26. Manzoni, C., et al., *Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences.* Brief Bioinform, 2018. **19**(2): p. 286-302.
27. Martens, L., *Proteomics databases and repositories.* Methods Mol Biol, 2011. **694**: p. 213-27.
28. Cox, J. and M. Mann, *Quantitative, high-resolution proteomics for data-driven systems biology.* Annu Rev Biochem, 2011. **80**: p. 273-99.
29. Lam, Y.W., et al., *Proteomics analysis of the nucleolus in adenovirus-infected cells.* Mol Cell Proteomics, 2010. **9**(1): p. 117-30.
30. Boisvert, F.M., et al., *A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage.* Mol Cell Proteomics, 2010. **9**(3): p. 457-70.
31. Andersen, J.S., et al., *Nucleolar proteome dynamics.* Nature, 2005. **433**(7021): p. 77-83.
32. Blagoev, B., et al., *Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.* Nat Biotechnol, 2004. **22**(9): p. 1139-45.
33. Breda A, V.N., Norberto de Souza O, et al., *Protein Structure, Modelling and Applications*, in *Bioinformatics in Tropical Disease Research: A Practical and Case-Study Approach*, D.A. Gruber A, Huynh C, et al., Editor. 2006 Bethesda (MD): National Center for Biotechnology Information.
34. Klionsky, D.J., et al., *Guidelines for the use and interpretation of assays for monitoring autophagy (4th edition)(1).* Autophagy, 2021. **17**(1): p. 1-382.
35. Waanders, L.F., et al., *Quantitative proteomic analysis of single pancreatic islets.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 18902-7.
36. Cha, S., et al., *In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology.* Mol Cell Proteomics, 2010. **9**(11): p. 2529-44.
37. Wisniewski, J.R., et al., *Universal sample preparation method for proteome analysis.* Nat Methods, 2009. **6**(5): p. 359-62.
38. Müller, A., et al., *3D FIB-SEM reconstruction of microtubule–organelle interaction in whole primary mouse β cells.* Journal of Cell Biology, 2020. **220**(2).
39. Trebichalská, Z., et al., *High-Resolution 3D Reconstruction of Human Oocytes Using Focused Ion Beam Scanning Electron Microscopy.* Frontiers in Cell and Developmental Biology, 2021. **9**(3083).
40. Marx, V., *The DNA of a nation.* Nature, 2015. **524**(7566): p. 503-5.
41. Buniello, A., et al., *The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D1005-D1012.
42. House, T.W. *Precision Meicine Initiative-White House-President Barack Obama*. 2016 [cited 2022 26/01/2022]; Available from: https://obamawhitehouse.archives.gov/precision-medicine.
43. Ginsburg, G.S. and K.A. Phillips, *Precision Medicine: From Science To Value.* Health Aff (Millwood), 2018. **37**(5): p. 694-701.
44. Orchard, S., et al., *Managing the data explosion. A report on the HUPO-PSI Workshop. August 2008, Amsterdam, The Netherlands.* Proteomics, 2009. **9**(3): p. 499-501.
45. Martens, L., et al., *mzML--a community standard for mass spectrometry data.* Mol Cell Proteomics, 2011. **10**(1): p. R110 000133.
46. Vizcaino, J.A., et al., *ProteomeXchange provides globally coordinated proteomics data submission and dissemination.* Nat Biotechnol, 2014. **32**(3): p. 223-6.
47. Perez-Riverol, Y., et al., *Making proteomics data accessible and reusable: current state of proteomics databases and repositories.* Proteomics, 2015. **15**(5-6): p. 930-49.
48. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
49. *Crystallography: Protein Data Bank.* Nature New Biology, 1971. **233**(42): p. 223-223.

50. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
51. Martens, L., et al., *Human Proteome Organization Proteomics Standards Initiative: data standardization, a view on developments and policy.* Mol Cell Proteomics, 2007. **6**(9): p. 1666-7.
52. Orchard, S., et al., *Protein interaction data curation: the International Molecular Exchange (IMEx) consortium.* Nat Methods, 2012. **9**(4): p. 345-50.
53. Duvaud, S., et al., *Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users.* Nucleic Acids Res, 2021. **49**(W1): p. W216-W227.
54. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80.
55. Shi, Z., et al., *A systems biology analysis of autophagy in cancer therapy.* Cancer Lett, 2013. **337**(2): p. 149-60.
56. Wanichthanarak, K., J.F. Fahrmann, and D. Grapov, *Genomic, Proteomic, and Metabolomic Data Integration Strategies.* Biomark Insights, 2015. **10**(Suppl 4): p. 1-6.
57. Kristensen, V.N., et al., *Principles and methods of integrative genomic analyses in cancer.* Nat Rev Cancer, 2014. **14**(5): p. 299-313.
58. Kannan, L., et al., *Public data and open source tools for multi-assay genomic investigation of disease.* Brief Bioinform, 2016. **17**(4): p. 603-15.
59. Woodruff, M.C., et al., *Extrafollicular B cell responses correlate with neutralizing antibodies and morbidity in COVID-19.* Nat Immunol, 2020. **21**(12): p. 1506-1516.
60. Bressan, E., et al., *The Foundational data initiative for Parkinson's disease (FOUNDIN-PD): enabling efficient translation from genetic maps to mechanism.* bioRxiv, 2021.
61. Goulielmos, G.N., et al., *Endometriosis research in the -omics era.* Gene, 2020. **741**: p. 144545.
62. Borah, K., Y. Xu, and J. McFadden, *Dissecting Host-Pathogen Interactions in TB Using Systems-Based Omic Approaches.* Front Immunol, 2021. **12**: p. 762315.
63. Usova, E.I., et al., *Integrative Analysis of Multi-Omics and Genetic Approaches-A New Level in Atherosclerotic Cardiovascular Risk Prediction.* Biomolecules, 2021. **11**(11).
64. Schilder, B.M., E. Navarro, and T. Raj, *Multi-omic insights into Parkinson's Disease: From genetic associations to functional mechanisms.* Neurobiol Dis, 2022. **163**: p. 105580.
65. Ben-Ari Fuchs, S., et al., *GeneAnalytics: An Integrative Gene Set Analysis Tool for Next Generation Sequencing, RNAseq and Microarray Data.* OMICS, 2016. **20**(3): p. 139-51.
66. Hasin, Y., M. Seldin, and A. Lusis, *Multi-omics approaches to disease.* Genome Biol, 2017. **18**(1): p. 83.
67. Bonate, P.L., *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. 2011: Springer US.
68. Mukherji, S. and A. van Oudenaarden, *Synthetic biology: understanding biological design from synthetic circuits.* Nat Rev Genet, 2009. **10**(12): p. 859-71.
69. Galdzicki, M., et al., *Standard biological parts knowledgebase.* PLoS One, 2011. **6**(2): p. e17005.
70. Marner, W.D., 2nd, *Practical application of synthetic biology principles.* Biotechnol J, 2009. **4**(10): p. 1406-19.
71. Keasling, J.D., *Synthetic biology for synthetic chemistry.* ACS Chem Biol, 2008. **3**(1): p. 64-76.
72. Drubin, D.A., J.C. Way, and P.A. Silver, *Designing biological systems.* Genes Dev, 2007. **21**(3): p. 242-54.
73. Hodgkin, A.L. and A.F. Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve.* J Physiol, 1952. **117**(4): p. 500-44.
74. Nakano, T., et al., *A kinetic model of dopamine- and calcium-dependent striatal synaptic plasticity.* PLoS Comput Biol, 2010. **6**(2): p. e1000670.

75. Kikuchi, S., et al., *Kinetic simulation of signal transduction system in hippocampal long-term potentiation with dynamic modeling of protein phosphatase 2A.* Neural Netw, 2003. **16**(9): p. 1389-98.
76. Gleeson, P., et al., *NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail.* PLoS Comput Biol, 2010. **6**(6): p. e1000815.
77. Qi, Z., G.W. Miller, and E.O. Voit, *The internal state of medium spiny neurons varies in response to different input signals.* BMC Syst Biol, 2010. **4**: p. 26.
78. Qi, Z., G.W. Miller, and E.O. Voit, *Computational systems analysis of dopamine metabolism.* PLoS One, 2008. **3**(6): p. e2444.
79. Tretter, F., *Mental illness, synapses and the brain--behavioral disorders by a system of molecules within a system of neurons?* Pharmacopsychiatry, 2010. **43 Suppl 1**: p. S9-S20.
80. (EPLF), É.p.f.d.L. *Blue Brain Project*. 2022 [cited 2022 18/01/2022]; Available from: https://www.epfl.ch/research/domains/bluebrain/.
81. Ero, C., et al., *A Cell Atlas for the Mouse Brain.* Front Neuroinform, 2018. **12**: p. 84.
82. Zisis, E., et al., *Digital Reconstruction of the Neuro-Glia-Vascular Architecture.* Cereb Cortex, 2021. **31**(12): p. 5686-5703.
83. Snider, J., et al., *Fundamentals of protein interaction network mapping.* Mol Syst Biol, 2015. **11**(12): p. 848.
84. Oliver, S., *Guilt-by-association goes global.* Nature, 2000. **403**(6770): p. 601-3.
85. Hishigaki, H., et al., *Assessment of prediction accuracy of protein function from protein-protein interaction data.* Yeast, 2001. **18**(6): p. 523-531.
86. Pino, L. and B. Schilling, *Proximity labeling and other novel mass spectrometric approaches for spatiotemporal protein dynamics.* Expert Rev Proteomics, 2021. **18**(9): p. 757-765.
87. Xenarios, I., et al., *DIP: the database of interacting proteins.* Nucleic Acids Res, 2000. **28**(1): p. 289-91.
88. Koh, G.C., et al., *Analyzing protein-protein interaction networks.* J Proteome Res, 2012. **11**(4): p. 2014-31.
89. Orchard, S., et al., *The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases.* Nucleic Acids Res, 2014. **42**(Database issue): p. D358-63.
90. Breuer, K., et al., *InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1228-33.
91. Orchard, S., et al., *The minimum information required for reporting a molecular interaction experiment (MIMIx).* Nat Biotechnol, 2007. **25**(8): p. 894-8.
92. Orchard, S., et al., *Submit your interaction data the IMEx way: a step by step guide to trouble-free deposition.* Proteomics, 2007. **7 Suppl 1**: p. 28-34.
93. Oughtred, R., et al., *The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions.* Protein Sci, 2021. **30**(1): p. 187-200.
94. Turinsky, A.L., et al., *Literature curation of protein interactions: measuring agreement across major public databases.* Database (Oxford), 2010. **2010**: p. baq026.
95. Barbosa-Silva, A., et al., *PESCADOR, a web-based tool to assist text-mining of biointeractions extracted from PubMed queries.* BMC Bioinformatics, 2011. **12**: p. 435.
96. Fernandez, J.M., R. Hoffmann, and A. Valencia, *iHOP web services.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W21-6.
97. Szklarczyk, D., et al., *The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible.* Nucleic Acids Res, 2017. **45**(D1): p. D362-D368.
98. Hu, Y., et al., *Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data.* Nucleic Acids Res, 2018. **46**(D1): p. D567-D574.

99. Alanis-Lobato, G., M.A. Andrade-Navarro, and M.H. Schaefer, *HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks.* Nucleic Acids Res, 2017. **45**(D1): p. D408-D414.
100. Hermjakob, H., et al., *The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data.* Nat Biotechnol, 2004. **22**(2): p. 177-83.
101. Orchard, S., et al., *Preparing to work with big data in proteomics - a report on the HUPO-PSI Spring Workshop: April 15-17, 2013, Liverpool, UK.* Proteomics, 2013. **13**(20): p. 2931-7.
102. Aranda, B., et al., *PSICQUIC and PSISCORE: accessing and scoring molecular interactions.* Nat Methods, 2011. **8**(7): p. 528-9.
103. Manzoni, C., P.A. Lewis, and R. Ferrari, *Network Analysis for Complex Neurodegenerative Diseases.* Current Genetic Medicine Reports, 2020. **8**(1): p. 17-25.
104. Ferrari, R., et al., *Weighted Protein Interaction Network Analysis of Frontotemporal Dementia.* J Proteome Res, 2017. **16**(2): p. 999-1013.
105. *Chapter 5 - Centrality and Hubs*, in *Fundamentals of Brain Network Analysis*, A. Fornito, A. Zalesky, and E.T. Bullmore, Editors. 2016, Academic Press: San Diego. p. 137-161.
106. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
107. Brown, K.R., et al., *NAViGaTOR: Network Analysis, Visualization and Graphing Toronto.* Bioinformatics, 2009. **25**(24): p. 3327-9.
108. Team, R.C., *R Core Team (2020).* 2020.
109. Katsura, I., *Assembly systems in molecular biology: a graph theory of genetic complementation between two related species.* J Theor Biol, 1981. **88**(3): p. 503-12.
110. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
111. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.
112. Reimand, J., et al., *g:Profiler-a web server for functional interpretation of gene lists (2016 update).* Nucleic Acids Res, 2016. **44**(W1): p. W83-9.
113. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic Acids Res, 2017. **45**(D1): p. D183-D189.
114. Wang, J., et al., *WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit.* Nucleic Acids Res, 2017. **45**(W1): p. W130-W137.
115. Rhee, S.Y., et al., *Use and misuse of the gene ontology annotations.* Nat Rev Genet, 2008. **9**(7): p. 509-15.
116. Camacho, D.M., et al., *Next-Generation Machine Learning for Biological Networks.* Cell, 2018. **173**(7): p. 1581-1592.
117. Xu, C. and S.A. Jackson, *Machine learning and complex biological data.* Genome Biol, 2019. **20**(1): p. 76.
118. Oyelade, J., et al., *Clustering Algorithms: Their Application to Gene Expression Data.* Bioinform Biol Insights, 2016. **10**: p. 237-253.
119. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
120. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W214-20.
121. Rabanal-Ruiz, Y., E.G. Otten, and V.I. Korolchuk, *mTORC1 as the main gateway to autophagy.* Essays Biochem, 2017. **61**(6): p. 565-584.
122. Vander Haar, E., et al., *Insulin signalling to mTOR mediated by the Akt/PKB substrate PRAS40.* Nat Cell Biol, 2007. **9**(3): p. 316-23.

123. Waage, P. and C. Guldberg, *Studier over affiniteten.* Forhandlinger i Videnskabs-selskabet i Christiania, 1864. **1**: p. 35-45.
124. van't Hoff, J.H., *Die Grenzebene, ein Beitrag zur Kenntniss der Esterbildung.* Berichte der deutschen chemischen Gesellschaft, 1877. **10**(1): p. 669-678.
125. Petit, C.S., A. Roczniak-Ferguson, and S.M. Ferguson, *Recruitment of folliculin to lysosomes supports the amino acid-dependent activation of Rag GTPases.* J Cell Biol, 2013. **202**(7): p. 1107-22.
126. Cursons, J., et al., *Regulation of ERK-MAPK signaling in human epidermis.* BMC Syst Biol, 2015. **9**: p. 41.
127. Benary, U. and J. Wolf, *Controlling Nuclear NF-kappaB Dynamics by beta-TrCP-Insights from a Computational Model.* Biomedicines, 2019. **7**(2).
128. Ye, M. and M.C. Hill, *Chapter 10 - Global Sensitivity Analysis for Uncertain Parameters, Models, and Scenarios*, in *Sensitivity Analysis in Earth Observation Modelling*, G.P. Petropoulos and P.K. Srivastava, Editors. 2017, Elsevier. p. 177-210.
129. Agoram, B.M. and O. Demin, *Integration not isolation: arguing the case for quantitative and systems pharmacology in drug discovery and development.* Drug Discov Today, 2011. **16**(23-24): p. 1031-6.
130. Sorger, P.K., et al. *Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms*. in *An NIH white paper by the QSP workshop group*. 2011. NIH Bethesda Bethesda, MD.
131. Earp, J.C., et al., *Modeling corticosteroid effects in a rat model of rheumatoid arthritis I: mechanistic disease progression model for the time course of collagen-induced arthritis in Lewis rats.* Journal of Pharmacology and Experimental Therapeutics, 2008. **326**(2): p. 532-545.
132. Svensson, B., et al., *Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in patients with early active rheumatoid arthritis reduces joint destruction and increases the remission rate: a two-year randomized trial.* Arthritis & Rheumatism, 2005. **52**(11): p. 3360-3370.
133. Azevedo, F.A., et al., *Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain.* Journal of Comparative Neurology, 2009. **513**(5): p. 532-541.
134. Salzberg, S.L., *Open questions: How many genes do we have?* BMC Biology, 2018. **16**(1): p. 94.
135. Lewis, P.A. and J.E. Spillane, *The Molecular and Clinical Pathology of Neurodegenerative Disease*. 2018: Elsevier Science.
136. Collaborators, G.B.D.D.F., *Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019.* Lancet Public Health, 2022. **7**(2): p. e105-e125.
137. Gale, S.A., D. Acar, and K.R. Daffner, *Dementia.* Am J Med, 2018. **131**(10): p. 1161-1169.
138. Peralta, V. and M.J. Cuesta, *Exploring the borders of the schizoaffective spectrum: a categorical and dimensional approach.* J Affect Disord, 2008. **108**(1-2): p. 71-86.
139. Diaz-Ortiz, M.E. and A.S. Chen-Plotkin, *Omics in Neurodegenerative Disease: Hope or Hype?* Trends Genet, 2020. **36**(3): p. 152-159.
140. Chandler, R.J., et al., *Modelling the functional genomics of Parkinson's disease in Caenorhabditis elegans: LRRK2 and beyond.* Biosci Rep, 2021. **41**(9).
141. Aryal, B. and Y. Lee, *Disease model organism for Parkinson disease: Drosophila melanogaster.* BMB Rep, 2019. **52**(4): p. 250-258.
142. Esquerda-Canals, G., et al., *Mouse Models of Alzheimer's Disease.* J Alzheimers Dis, 2017. **57**(4): p. 1171-1183.
143. Sosulina, L., et al., *Hippocampal hyperactivity in a rat model of Alzheimer's disease.* J Neurochem, 2021. **157**(6): p. 2128-2144.

144. Gordon, J., S. Amini, and M.K. White, *General overview of neuronal cell culture.* Methods Mol Biol, 2013. **1078**: p. 1-8.
145. Sheard, J.J., et al., *Optically Transparent Anionic Nanofibrillar Cellulose Is Cytocompatible with Human Adipose Tissue-Derived Stem Cells and Allows Simple Imaging in 3D.* Stem Cells Int, 2019. **2019**: p. 3106929.
146. Lancaster, M.A. and J.A. Knoblich, *Generation of cerebral organoids from human pluripotent stem cells.* Nat Protoc, 2014. **9**(10): p. 2329-40.
147. Raja, W.K., et al., *Self-Organizing 3D Human Neural Tissue Derived from Induced Pluripotent Stem Cells Recapitulate Alzheimer's Disease Phenotypes.* PLoS One, 2016. **11**(9): p. e0161969.
148. Tamagnini, F., *Nucleated, Outside-Out, Somatic, Macropatch Recordings in Native Neurons.* Methods Mol Biol, 2021. **2188**: p. 229-242.
149. Hawe, J.S., F.J. Theis, and M. Heinig, *Inferring Interaction Networks From Multi-Omics Data.* Front Genet, 2019. **10**: p. 535.
150. Barbosa, S., et al., *A guide to gene regulatory network inference for obtaining predictive solutions: Underlying assumptions and fundamental biological and data constraints.* Biosystems, 2018. **174**: p. 37-48.
151. Delgado, F.M. and F. Gomez-Vela, *Computational methods for Gene Regulatory Networks reconstruction and analysis: A review.* Artif Intell Med, 2019. **95**: p. 133-145.
152. Wang, R.S., B.A. Maron, and J. Loscalzo, *Systems medicine: evolution of systems biology from bench to bedside.* Wiley Interdiscip Rev Syst Biol Med, 2015. **7**(4): p. 141-61.
153. Hu, Y.S., et al., *Analyzing the genes related to Alzheimer's disease via a network and pathway-based approach.* Alzheimers Res Ther, 2017. **9**(1): p. 29.
154. Ghiassian, S.D., J. Menche, and A.L. Barabasi, *A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome.* PLoS Comput Biol, 2015. **11**(4): p. e1004120.
155. Ferrari, R., et al., *Stratification of candidate genes for Parkinson's disease using weighted protein-protein interaction network analysis.* BMC Genomics, 2018. **19**(1): p. 452.
156. Dervishi, I., et al., *Protein-protein interactions reveal key canonical pathways, upstream regulators, interactome domains, and novel targets in ALS.* Sci Rep, 2018. **8**(1): p. 14732.
157. Beltran, S., et al., *Network approach identifies Pacer as an autophagy protein involved in ALS pathogenesis.* Mol Neurodegener, 2019. **14**(1): p. 14.
158. Zhang, S., et al., *Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis.* Neuron, 2022. **110**(6): p. 992-1008 e11.
159. Kahle, J.J., et al., *Comparison of an expanded ataxia interactome with patient medical records reveals a relationship between macular degeneration and ataxia.* Hum Mol Genet, 2011. **20**(3): p. 510-27.
160. Ferrari, R., et al., *Frontotemporal dementia: insights into the biological underpinnings of disease through gene co-expression network analysis.* Mol Neurodegener, 2016. **11**: p. 21.
161. Kong, P., et al., *Integrated microarray analysis provided a new insight of the pathogenesis of Parkinson's disease.* Neurosci Lett, 2018. **662**: p. 51-58.
162. Ouzounoglou, E., et al., *In silico modeling of the effects of alpha-synuclein oligomerization on dopaminergic neuronal homeostasis.* BMC Syst Biol, 2014. **8**: p. 54.
163. Kuznetsov, I.A. and A.V. Kuznetsov, *What can trigger the onset of Parkinson's disease - A modeling study based on a compartmental model of alpha-synuclein transport and aggregation in neurons.* Math Biosci, 2016. **278**: p. 22-9.
164. Mizuno, S., et al., *AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease.* BMC Syst Biol, 2012. **6**: p. 52.
165. Ogishima, S., et al., *AlzPathway, an Updated Map of Curated Signaling Pathways: Towards Deciphering Alzheimer's Disease Pathogenesis.* Methods Mol Biol, 2016. **1303**: p. 423-32.

166. Harding, A.E., *Classification of the hereditary ataxias and paraplegias.* Lancet, 1983. **1**(8334): p. 1151-5.
167. Deluca, G.C., G.C. Ebers, and M.M. Esiri, *The extent of axonal loss in the long tracts in hereditary spastic paraplegia.* Neuropathol Appl Neurobiol, 2004. **30**(6): p. 576-84.
168. Fink, J.K., *Hereditary spastic paraplegia: clinico-pathologic features and emerging molecular mechanisms.* Acta Neuropathol, 2013. **126**(3): p. 307-28.
169. Faber, I., et al., *Hereditary spastic paraplegia from 1880 to 2017: an historical review.* Arq Neuropsiquiatr, 2017. **75**(11): p. 813-818.
170. Boutry, M., S. Morais, and G. Stevanin, *Update on the Genetics of Spastic Paraplegias.* Curr Neurol Neurosci Rep, 2019. **19**(4): p. 18.
171. Stevanin, G., et al., *Mutations in SPG11 are frequent in autosomal recessive spastic paraplegia with thin corpus callosum, cognitive decline and lower motor neuron degeneration.* Brain, 2008. **131**(Pt 3): p. 772-84.
172. Ruano, L., et al., *The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies.* Neuroepidemiology, 2014. **42**(3): p. 174-83.
173. Koh, K., et al., *JASPAC: Japan Spastic Paraplegia Research Consortium.* Brain Sci, 2018. **8**(8).
174. Parkinson, J., *An Essay on the Shaking Palsy*. 1817: Whittingham and Rowland.
175. Li, S. and W. Le, *Milestones of Parkinson's Disease Research: 200 Years of History and Beyond.* Neurosci Bull, 2017. **33**(5): p. 598-602.
176. Zhang, Z.X., Z.H. Dong, and G.C. Roman, *Early descriptions of Parkinson disease in ancient China.* Arch Neurol, 2006. **63**(5): p. 782-4.
177. Bereczki, D., *The description of all four cardinal signs of Parkinson's disease in a Hungarian medical text published in 1690.* Parkinsonism Relat Disord, 2010. **16**(4): p. 290-3.
178. Brissaud, É. and H. Meige, *Leçons sur les maladies nerveuses*. 1899: Masson.
179. Trétiakoff, C., *Contribution a l'etude de l'Anatomie pathologique du Locus Niger de Soemmering avec quelques deduction relatives a la pathogenie des troubles du tonus musculaire et de la maladie de Parkinson.* Theses de Paris, 1919.
180. Bumke, O. and O. Foerster, *HANDBUCH DER NEUROLOGIE. Dreizehn-ter BAND. Spezielle Neurologie V. Erkrankungen Des Rücken-Marks Und Gehirns Iii. Infektionen. Intoxikationen IL*. 1937, LWW.
181. Ehringer, H. and O. Hornykiewicz, *[Distribution of noradrenaline and dopamine (3-hydroxytyramine) in the human brain and their behavior in diseases of the extrapyramidal system].* Klin Wochenschr, 1960. **38**: p. 1236-9.
182. Barbeau, A., G. Jasmin, and Y. Duchastel, *Biochemistry of Parkinson's disease.* Neurology, 1963. **13**: p. 56-8.
183. Burns, R.S., et al., *A primate model of parkinsonism: selective destruction of dopaminergic neurons in the pars compacta of the substantia nigra by N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine.* Proc Natl Acad Sci U S A, 1983. **80**(14): p. 4546-50.
184. Langston, J.W., et al., *Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis.* Science, 1983. **219**(4587): p. 979-80.
185. Calne, D.B., et al., *Bromocriptine in Parkinsonism.* Br Med J, 1974. **4**(5942): p. 442-4.
186. Cotzias, G.C., et al., *Similarities between neurologic effects of L-dopa and of apomorphine.* N Engl J Med, 1970. **282**(1): p. 31-3.
187. Polymeropoulos, M.H., et al., *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.* Science, 1997. **276**(5321): p. 2045-7.
188. Freed, C.R., et al., *Transplantation of embryonic dopamine neurons for severe Parkinson's disease.* N Engl J Med, 2001. **344**(10): p. 710-9.
189. Golbe, L.I., *Young-onset Parkinson's disease: a clinical review.* Neurology, 1991. **41**(2 (Pt 1)): p. 168-73.
190. Hughes, A.J., et al., *Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.* J Neurol Neurosurg Psychiatry, 1992. **55**(3): p. 181-4.

191. Lees, A., *The bare essentials: Parkinson's disease.* Pract Neurol, 2010. **10**(4): p. 240-6.
192. Schapira, A.H.V., K.R. Chaudhuri, and P. Jenner, *Non-motor features of Parkinson disease.* Nat Rev Neurosci, 2017. **18**(8): p. 509.
193. Pfeiffer, R.F., *Autonomic dysfunction in Parkinson's disease.* Expert Rev Neurother, 2012. **12**(6): p. 697-706.
194. St Louis, E.K., A.R. Boeve, and B.F. Boeve, *REM Sleep Behavior Disorder in Parkinson's Disease and Other Synucleinopathies.* Mov Disord, 2017. **32**(5): p. 645-658.
195. Ha, A.D. and J. Jankovic, *Pain in Parkinson's disease.* Mov Disord, 2012. **27**(4): p. 485-91.
196. Haehner, A., T. Hummel, and H. Reichmann, *A clinical approach towards smell loss in Parkinson's disease.* J Parkinsons Dis, 2014. **4**(2): p. 189-95.
197. Lewy, F., *Zur pathologischen Anatomie der Paralysis agitans.* Dtsch Z Nervenheilk, 1913. **50**: p. 50-55.
198. Forno, L.S., *Neuropathology of Parkinson's disease.* J Neuropathol Exp Neurol, 1996. **55**(3): p. 259-72.
199. Watkins, K.E. and N. Jenkinson, *Chapter 8 - The Anatomy of the Basal Ganglia*, in *Neurobiology of Language*, G. Hickok and S.L. Small, Editors. 2016, Academic Press: San Diego. p. 85-94.
200. Yager, L.M., et al., *The ins and outs of the striatum: role in drug addiction.* Neuroscience, 2015. **301**: p. 529-41.
201. Packard, M.G., *Basal Ganglia*, in *International Encyclopedia of the Social & Behavioral Sciences*, N.J. Smelser and P.B. Baltes, Editors. 2001, Pergamon: Oxford. p. 1044-1048.
202. Carlsson, A., M. Lindqvist, and T. Magnusson, *3,4-Dihydroxyphenylalanine and 5-hydroxytryptophan as reserpine antagonists.* Nature, 1957. **180**(4596): p. 1200.
203. Carlsson, A., et al., *On the presence of 3-hydroxytyramine in brain.* Science, 1958. **127**(3296): p. 471.
204. Sonne, J., V. Reddy, and M.R. Beato, *Neuroanatomy, Substantia Nigra*, in *StatPearls*. 2022, StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC.: Treasure Island (FL).
205. Goedert, M., et al., *100 years of Lewy pathology.* Nat Rev Neurol, 2013. **9**(1): p. 13-24.
206. Lashuel, H.A., et al., *The many faces of alpha-synuclein: from structure and toxicity to therapeutic target.* Nat Rev Neurosci, 2013. **14**(1): p. 38-48.
207. Arrasate, M., et al., *Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death.* Nature, 2004. **431**(7010): p. 805-10.
208. Li, J.Y., et al., *Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation.* Nat Med, 2008. **14**(5): p. 501-3.
209. Kordower, J.H., et al., *Lewy body-like pathology in long-term embryonic nigral transplants in Parkinson's disease.* Nat Med, 2008. **14**(5): p. 504-6.
210. Isacson, O., *The production and use of cells as therapeutic agents in neurodegenerative diseases.* Lancet Neurol, 2003. **2**(7): p. 417-24.
211. Braak, H., et al., *Staging of brain pathology related to sporadic Parkinson's disease.* Neurobiol Aging, 2003. **24**(2): p. 197-211.
212. Walsh, D.M. and D.J. Selkoe, *A critical appraisal of the pathogenic protein spread hypothesis of neurodegeneration.* Nat Rev Neurosci, 2016. **17**(4): p. 251-60.
213. Burke, R.E., W.T. Dauer, and J.P. Vonsattel, *A critical evaluation of the Braak staging scheme for Parkinson's disease.* Ann Neurol, 2008. **64**(5): p. 485-91.
214. Kitada, T., et al., *Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism.* Nature, 1998. **392**(6676): p. 605-8.
215. Paisan-Ruiz, C., et al., *Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease.* Neuron, 2004. **44**(4): p. 595-600.
216. Zimprich, A., et al., *Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology.* Neuron, 2004. **44**(4): p. 601-7.

217. Benamer, H.T. and R. de Silva, *LRRK2 G2019S in the North African population: a review.* Eur Neurol, 2010. **63**(6): p. 321-5.
218. Nalls, M.A., et al., *Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease.* Nat Genet, 2014. **46**(9): p. 989-93.
219. Vilarino-Guell, C., et al., *VPS35 mutations in Parkinson disease.* Am J Hum Genet, 2011. **89**(1): p. 162-7.
220. Mazzulli, J.R., et al., *Gaucher disease glucocerebrosidase and alpha-synuclein form a bidirectional pathogenic loop in synucleinopathies.* Cell, 2011. **146**(1): p. 37-52.
221. Park, J., et al., *Mitochondrial dysfunction in Drosophila PINK1 mutants is complemented by parkin.* Nature, 2006. **441**(7097): p. 1157-61.
222. Clark, I.E., et al., *Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin.* Nature, 2006. **441**(7097): p. 1162-6.
223. Ransohoff, R.M., *How neuroinflammation contributes to neurodegeneration.* Science, 2016. **353**(6301): p. 777-83.
224. Langston, J.W., *The MPTP Story.* J Parkinsons Dis, 2017. **7**(s1): p. S11-S19.
225. Brooks, A.I., et al., *Paraquat elicited neurobehavioral syndrome caused by dopaminergic neuron loss.* Brain Res, 1999. **823**(1-2): p. 1-10.
226. Betarbet, R., et al., *Chronic systemic pesticide exposure reproduces features of Parkinson's disease.* Nat Neurosci, 2000. **3**(12): p. 1301-6.
227. Casals, J., T.S. Elizan, and M.D. Yahr, *Postencephalitic parkinsonism--a review.* J Neural Transm (Vienna), 1998. **105**(6-7): p. 645-76.
228. Lees, A.J., J. Hardy, and T. Revesz, *Parkinson's disease.* Lancet, 2009. **373**(9680): p. 2055-66.
229. Foundation, P.s. *Parkinson's Foundation/Statistics*. 2021 [cited 2022 20-01-2022]; Available from: https://www.parkinson.org/Understanding-Parkinsons/Statistics.
230. Dorsey, E.R. and B.R. Bloem, *The Parkinson Pandemic-A Call to Action.* JAMA Neurol, 2018. **75**(1): p. 9-10.
231. Robakis, D. and S. Fahn, *Defining the Role of the Monoamine Oxidase-B Inhibitors for Parkinson's Disease.* CNS Drugs, 2015. **29**(6): p. 433-41.
232. Muller, T., *Catechol-O-methyltransferase inhibitors in Parkinson's disease.* Drugs, 2015. **75**(2): p. 157-74.
233. Wolters, E., Y.D. van der Werf, and O.A. van den Heuvel, *Parkinson's disease-related disorders in the impulsive-compulsive spectrum.* J Neurol, 2008. **255 Suppl 5**: p. 48-56.
234. Okun, M.S., *Deep-brain stimulation for Parkinson's disease.* N Engl J Med, 2012. **367**(16): p. 1529-38.
235. Health, N.I.o. *Clinical Trials*. 2022 [cited 2022 24/01]; Available from: https://clinicaltrials.gov/.
236. NHS. *Clinical trials*. 2022 [cited 2022 24/01]; Available from: https://bepartofresearch.nihr.ac.uk/.
237. Brady, R.O. and R. Schiffmann, *Enzyme-replacement therapy for metabolic storage disorders.* Lancet Neurol, 2004. **3**(12): p. 752-6.
238. Galatsis, P., *Leucine-rich repeat kinase 2 inhibitors: a patent review (2014-2016).* Expert Opin Ther Pat, 2017. **27**(6): p. 667-676.
239. Masliah, E., et al., *Effects of alpha-synuclein immunization in a mouse model of Parkinson's disease.* Neuron, 2005. **46**(6): p. 857-68.
240. Athauda, D. and T. Foltynie, *The ongoing pursuit of neuroprotective therapies in Parkinson disease.* Nat Rev Neurol, 2015. **11**(1): p. 25-40.
241. Barker, R.A., J. Drouin-Ouellet, and M. Parmar, *Cell-based therapies for Parkinson disease-past insights and future potential.* Nat Rev Neurol, 2015. **11**(9): p. 492-503.
242. Schilder, B.M., J. Humphrey, and T. Raj, *echolocatoR: an automated end-to-end statistical and functional genomic fine-mapping pipeline.* Bioinformatics, 2021. **38**(2): p. 536-539.

243. Corces, M.R., et al., *Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases.* Nat Genet, 2020. **52**(11): p. 1158-1168.
244. International Parkinson Disease Genomics, C., *Ten Years of the International Parkinson Disease Genomics Consortium: Progress and Next Steps.* J Parkinsons Dis, 2020. **10**(1): p. 19-30.
245. Choi, A.M., S.W. Ryter, and B. Levine, *Autophagy in human health and disease.* N Engl J Med, 2013. **368**(19): p. 1845-6.
246. Martinez-Vicente, M. and A.M. Cuervo, *Autophagy and neurodegeneration: when the cleaning crew goes on strike.* Lancet Neurol, 2007. **6**(4): p. 352-61.
247. De Duve, C., *The lysosome.* Sci Am, 1963. **208**: p. 64-72.
248. Levine, B. and G. Kroemer, *Autophagy in the pathogenesis of disease.* Cell, 2008. **132**(1): p. 27-42.
249. Levine, B. and G. Kroemer, *Autophagy in aging, disease and death: the true identity of a cell death impostor.* Cell Death Differ, 2009. **16**(1): p. 1-2.
250. Hara, T., et al., *Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice.* Nature, 2006. **441**(7095): p. 885-9.
251. Komatsu, M., et al., *Loss of autophagy in the central nervous system causes neurodegeneration in mice.* Nature, 2006. **441**(7095): p. 880-4.
252. Pankiv, S., et al., *p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy.* J Biol Chem, 2007. **282**(33): p. 24131-45.
253. Kirkin, V., et al., *A role for NBR1 in autophagosomal degradation of ubiquitinated substrates.* Mol Cell, 2009. **33**(4): p. 505-16.
254. Okamoto, K., N. Kondo-Okamoto, and Y. Ohsumi, *Mitochondria-anchored receptor Atg32 mediates degradation of mitochondria via selective autophagy.* Dev Cell, 2009. **17**(1): p. 87-97.
255. Richter, B., et al., *Phosphorylation of OPTN by TBK1 enhances its binding to Ub chains and promotes selective autophagy of damaged mitochondria.* Proc Natl Acad Sci U S A, 2016. **113**(15): p. 4039-44.
256. Teng, Y.H., et al., *Autophagy Protects from Raddeanin A-Induced Apoptosis in SGC-7901 Human Gastric Cancer Cells.* Evid Based Complement Alternat Med, 2016. **2016**: p. 9406758.
257. Kelekar, A., *Introduction to the Review Series Autophagy in Higher Eukaryotes- A matter of survival or death.* Autophagy, 2008. **4**(5): p. 555-556.
258. Jacob, J.A., et al., *Autophagy: An overview and its roles in cancer and obesity.* Clin Chim Acta, 2017. **468**: p. 85-89.
259. Li, W.W., J. Li, and J.K. Bao, *Microautophagy: lesser-known self-eating.* Cell Mol Life Sci, 2012. **69**(7): p. 1125-36.
260. Dice, J.F., *Peptide sequences that target cytosolic proteins for lysosomal proteolysis.* Trends Biochem Sci, 1990. **15**(8): p. 305-9.
261. Chiang, H.L., et al., *A role for a 70-kilodalton heat shock protein in lysosomal degradation of intracellular proteins.* Science, 1989. **246**(4928): p. 382-5.
262. Cuervo, A.M. and J.F. Dice, *A receptor for the selective uptake and degradation of proteins by lysosomes.* Science, 1996. **273**(5274): p. 501-3.
263. White, E., J.M. Mehnert, and C.S. Chan, *Autophagy, Metabolism, and Cancer.* Clin Cancer Res, 2015. **21**(22): p. 5037-46.
264. Deretic, V., *Autophagosome and phagosome.* Methods Mol Biol, 2008. **445**: p. 1-10.
265. Jing, K. and K. Lim, *Why is autophagy important in human diseases?* Exp Mol Med, 2012. **44**(2): p. 69-72.
266. Mizushima, N., T. Yoshimori, and Y. Ohsumi, *The role of Atg proteins in autophagosome formation.* Annu Rev Cell Dev Biol, 2011. **27**: p. 107-32.

267. Rubinsztein, D.C., P. Codogno, and B. Levine, *Autophagy modulation as a potential therapeutic target for diverse diseases.* Nat Rev Drug Discov, 2012. **11**(9): p. 709-30.
268. Itakura, E. and N. Mizushima, *Characterization of autophagosome formation site by a hierarchical analysis of mammalian Atg proteins.* Autophagy, 2010. **6**(6): p. 764-76.
269. Axe, E.L., et al., *Autophagosome formation from membrane compartments enriched in phosphatidylinositol 3-phosphate and dynamically connected to the endoplasmic reticulum.* J Cell Biol, 2008. **182**(4): p. 685-701.
270. Alers, S., et al., *Role of AMPK-mTOR-Ulk1/2 in the regulation of autophagy: cross talk, shortcuts, and feedbacks.* Mol Cell Biol, 2012. **32**(1): p. 2-11.
271. Matsunaga, K., et al., *Two Beclin 1-binding proteins, Atg14L and Rubicon, reciprocally regulate autophagy at different stages.* Nat Cell Biol, 2009. **11**(4): p. 385-96.
272. Zhong, Y., et al., *Distinct regulation of autophagic activity by Atg14L and Rubicon associated with Beclin 1-phosphatidylinositol-3-kinase complex.* Nat Cell Biol, 2009. **11**(4): p. 468-76.
273. Chan, E.Y., et al., *Kinase-inactivated ULK proteins inhibit autophagy via their conserved C-terminal domains using an Atg13-independent mechanism.* Mol Cell Biol, 2009. **29**(1): p. 157-71.
274. Hosokawa, N., et al., *Nutrient-dependent mTORC1 association with the ULK1-Atg13-FIP200 complex required for autophagy.* Mol Biol Cell, 2009. **20**(7): p. 1981-91.
275. Jung, C.H., et al., *ULK-Atg13-FIP200 complexes mediate mTOR signaling to the autophagy machinery.* Mol Biol Cell, 2009. **20**(7): p. 1992-2003.
276. Levy, J.M.M., C.G. Towers, and A. Thorburn, *Targeting autophagy in cancer.* Nat Rev Cancer, 2017. **17**(9): p. 528-542.
277. Carlsson, S.R. and A. Simonsen, *Membrane dynamics in autophagosome biogenesis.* J Cell Sci, 2015. **128**(2): p. 193-205.
278. Devereaux, K., et al., *Regulation of mammalian autophagy by class II and III PI 3-kinases through PI3P synthesis.* PLoS One, 2013. **8**(10): p. e76405.
279. Lamb, C.A., T. Yoshimori, and S.A. Tooze, *The autophagosome: origins unknown, biogenesis complex.* Nat Rev Mol Cell Biol, 2013. **14**(12): p. 759-74.
280. Polson, H.E., et al., *Mammalian Atg18 (WIPI2) localizes to omegasome-anchored phagophores and positively regulates LC3 lipidation.* Autophagy, 2010. **6**(4): p. 506-22.
281. Dooley, H.C., et al., *WIPI2 links LC3 conjugation with PI3P, autophagosome formation, and pathogen clearance by recruiting Atg12-5-16L1.* Mol Cell, 2014. **55**(2): p. 238-52.
282. Weidberg, H., et al., *LC3 and GATE-16/GABARAP subfamilies are both essential yet act differently in autophagosome biogenesis.* EMBO J, 2010. **29**(11): p. 1792-802.
283. Hanada, T., et al., *The Atg12-Atg5 conjugate has a novel E3-like activity for protein lipidation in autophagy.* J Biol Chem, 2007. **282**(52): p. 37298-302.
284. Fujita, N., et al., *The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy.* Mol Biol Cell, 2008. **19**(5): p. 2092-100.
285. Ricci, V., *Relationship between VacA Toxin and Host Cell Autophagy in Helicobacter pylori Infection of the Human Stomach: A Few Answers, Many Questions.* Toxins (Basel), 2016. **8**(7).
286. Hirano, S., et al., *Differing susceptibility to autophagic degradation of two LC3-binding proteins: SQSTM1/p62 and TBC1D25/OATL1.* Autophagy, 2016. **12**(2): p. 312-26.
287. Nakatogawa, H., Y. Ichimura, and Y. Ohsumi, *Atg8, a ubiquitin-like protein required for autophagosome formation, mediates membrane tethering and hemifusion.* Cell, 2007. **130**(1): p. 165-78.
288. Tanida, I., *Autophagy basics.* Microbiol Immunol, 2011. **55**(1): p. 1-11.
289. Geng, J. and D.J. Klionsky, *The Atg8 and Atg12 ubiquitin-like conjugation systems in macroautophagy. 'Protein modifications: beyond the usual suspects' review series.* EMBO Rep, 2008. **9**(9): p. 859-64.

290. Hamasaki, M., et al., *Autophagosomes form at ER-mitochondria contact sites.* Nature, 2013. **495**(7441): p. 389-93.
291. Itakura, E., et al., *Beclin 1 forms two distinct phosphatidylinositol 3-kinase complexes with mammalian Atg14 and UVRAG.* Mol Biol Cell, 2008. **19**(12): p. 5360-72.
292. Liang, C., et al., *Beclin1-binding UVRAG targets the class C Vps complex to coordinate autophagosome maturation and endocytic trafficking.* Nat Cell Biol, 2008. **10**(7): p. 776-87.
293. Kochl, R., et al., *Microtubules facilitate autophagosome formation and fusion of autophagosomes with endosomes.* Traffic, 2006. **7**(2): p. 129-45.
294. Kimura, S., T. Noda, and T. Yoshimori, *Dynein-dependent movement of autophagosomes mediates efficient encounters with lysosomes.* Cell Struct Funct, 2008. **33**(1): p. 109-22.
295. Metcalf, D.J., et al., *Autophagy and misfolded proteins in neurodegeneration.* Exp Neurol, 2012. **238**(1): p. 22-8.
296. Gutierrez, M.G., et al., *Rab7 is required for the normal progression of the autophagic pathway in mammalian cells.* J Cell Sci, 2004. **117**(Pt 13): p. 2687-97.
297. Jager, S., et al., *Role for Rab7 in maturation of late autophagic vacuoles.* J Cell Sci, 2004. **117**(Pt 20): p. 4837-48.
298. Filimonenko, M., et al., *Functional multivesicular bodies are required for autophagic clearance of protein aggregates associated with neurodegenerative disease.* J Cell Biol, 2007. **179**(3): p. 485-500.
299. Lee, J.A., et al., *ESCRT-III dysfunction causes autophagosome accumulation and neurodegeneration.* Curr Biol, 2007. **17**(18): p. 1561-7.
300. Urwin, H., et al., *Disruption of endocytic trafficking in frontotemporal dementia with CHMP2B mutations.* Hum Mol Genet, 2010. **19**(11): p. 2228-38.
301. Rusten, T.E., et al., *ESCRTing autophagic clearance of aggregating proteins.* Autophagy, 2008. **4**(2): p. 233-236.
302. Rusten, T.E., et al., *ESCRTs and Fab1 regulate distinct steps of autophagy.* Curr Biol, 2007. **17**(20): p. 1817-25.
303. Atlashkin, V., et al., *Deletion of the SNARE vti1b in mice results in the loss of a single SNARE partner, syntaxin 8.* Mol Cell Biol, 2003. **23**(15): p. 5198-207.
304. Furuta, N., et al., *Combinational soluble N-ethylmaleimide-sensitive factor attachment protein receptor proteins VAMP8 and Vti1b mediate fusion of antimicrobial and canonical autophagosomes with lysosomes.* Mol Biol Cell, 2010. **21**(6): p. 1001-10.
305. Laplante, M. and D.M. Sabatini, *mTOR signaling in growth control and disease.* Cell, 2012. **149**(2): p. 274-93.
306. Koren, I., E. Reem, and A. Kimchi, *DAP1, a novel substrate of mTOR, negatively regulates autophagy.* Curr Biol, 2010. **20**(12): p. 1093-8.
307. Sancak, Y., et al., *Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids.* Cell, 2010. **141**(2): p. 290-303.
308. Sancak, Y., et al., *The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1.* Science, 2008. **320**(5882): p. 1496-501.
309. Meijer, A.J. and P. Codogno, *Signalling and autophagy regulation in health, aging and disease.* Mol Aspects Med, 2006. **27**(5-6): p. 411-25.
310. Gwinn, D.M., et al., *AMPK phosphorylation of raptor mediates a metabolic checkpoint.* Mol Cell, 2008. **30**(2): p. 214-26.
311. Zheng, Q., et al., *Inhibition of AMPK accentuates prolonged caloric restriction-induced change in cardiac contractile function through disruption of compensatory autophagy.* Biochim Biophys Acta, 2015. **1852**(2): p. 332-42.
312. Fimia, G.M., et al., *Ambra1 regulates autophagy and development of the nervous system.* Nature, 2007. **447**(7148): p. 1121-5.
313. Liang, C., et al., *Autophagic and tumour suppressor activity of a novel Beclin1-binding protein UVRAG.* Nat Cell Biol, 2006. **8**(7): p. 688-99.

314. Takahashi, Y., et al., *Bif-1 interacts with Beclin 1 through UVRAG and regulates autophagy and tumorigenesis.* Nat Cell Biol, 2007. **9**(10): p. 1142-51.
315. Pattingre, S., et al., *Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy.* Cell, 2005. **122**(6): p. 927-39.
316. Ciechomska, I.A., et al., *Bcl-2 complexed with Beclin-1 maintains full anti-apoptotic function.* Oncogene, 2009. **28**(21): p. 2128-41.
317. Vicencio, J.M., et al., *The inositol 1,4,5-trisphosphate receptor regulates autophagy through its interaction with Beclin 1.* Cell Death Differ, 2009. **16**(7): p. 1006-17.
318. Wei, Y., et al., *JNK1-mediated phosphorylation of Bcl-2 regulates starvation-induced autophagy.* Mol Cell, 2008. **30**(6): p. 678-88.
319. Di Bartolomeo, S., et al., *The dynamic interaction of AMBRA1 with the dynein motor complex regulates mammalian autophagy.* J Cell Biol, 2010. **191**(1): p. 155-68.
320. Kim, J., et al., *AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1.* Nat Cell Biol, 2011. **13**(2): p. 132-41.
321. White, E., *Autophagy and p53.* Cold Spring Harb Perspect Med, 2016. **6**(4): p. a026120.
322. Kenzelmann Broz, D., et al., *Global genomic profiling reveals an extensive p53-regulated autophagy program contributing to key p53 responses.* Genes Dev, 2013. **27**(9): p. 1016-31.
323. Crighton, D., et al., *DRAM, a p53-induced modulator of autophagy, is critical for apoptosis.* Cell, 2006. **126**(1): p. 121-34.
324. Eby, K.G., et al., *ISG20L1 is a p53 family target gene that modulates genotoxic stress-induced autophagy.* Mol Cancer, 2010. **9**: p. 95.
325. Tasdemir, E., et al., *Regulation of autophagy by cytoplasmic p53.* Nat Cell Biol, 2008. **10**(6): p. 676-87.
326. Young, A.R., et al., *Starvation and ULK1-dependent cycling of mammalian Atg9 between the TGN and endosomes.* J Cell Sci, 2006. **119**(Pt 18): p. 3888-900.
327. Proikas-Cezanne, T. and P. Codogno, *Beclin 1 or not Beclin 1.* Autophagy, 2011. **7**(7): p. 671-2.
328. Tian, S., et al., *Beclin 1-independent autophagy induced by a Bcl-XL/Bcl-2 targeting compound, Z18.* Autophagy, 2010. **6**(8): p. 1032-41.
329. Grishchuk, Y., et al., *Beclin 1-independent autophagy contributes to apoptosis in cortical neurons.* Autophagy, 2011. **7**(10): p. 1115-31.
330. Cheong, H., et al., *Ammonia-induced autophagy is independent of ULK1/ULK2 kinases.* Proc Natl Acad Sci U S A, 2011. **108**(27): p. 11121-6.
331. Nishida, Y., et al., *Discovery of Atg5/Atg7-independent alternative macroautophagy.* Nature, 2009. **461**(7264): p. 654-8.
332. Blackstone, C., *Cellular pathways of hereditary spastic paraplegia.* Annu Rev Neurosci, 2012. **35**: p. 25-47.
333. Blackstone, C., *Converging cellular themes for the hereditary spastic paraplegias.* Curr Opin Neurobiol, 2018. **51**: p. 139-146.
334. Blackstone, C., C.J. O'Kane, and E. Reid, *Hereditary spastic paraplegias: membrane traffic and the motor pathway.* Nat Rev Neurosci, 2011. **12**(1): p. 31-42.
335. Bis-Brewer, D.M., et al., *A network biology approach to unraveling inherited axonopathies.* Sci Rep, 2019. **9**(1): p. 1692.
336. Synofzik, M. and R. Schule, *Overcoming the divide between ataxias and spastic paraplegias: Shared phenotypes, genes, and pathways.* Mov Disord, 2017. **32**(3): p. 332-345.
337. Novarino, G., et al., *Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders.* Science, 2014. **343**(6170): p. 506-511.
338. de Souza, P.V.S., et al., *Hereditary Spastic Paraplegia: Clinical and Genetic Hallmarks.* Cerebellum, 2017. **16**(2): p. 525-551.
339. Bouwkamp, C.G., et al., *ACO2 homozygous missense mutation associated with complicated hereditary spastic paraplegia.* Neurol Genet, 2018. **4**(2): p. e223.

340. Spiegel, R., et al., *Infantile cerebellar-retinal degeneration associated with a mutation in mitochondrial aconitase, ACO2.* Am J Hum Genet, 2012. **90**(3): p. 518-23.
341. Metodiev, M.D., et al., *Mutations in the tricarboxylic acid cycle enzyme, aconitase 2, cause either isolated or syndromic optic neuropathy with encephalopathy and cerebellar atrophy.* J Med Genet, 2014. **51**(12): p. 834-8.
342. Simone, M., et al., *KIF5A and ALS2 Variants in a Family With Hereditary Spastic Paraplegia and Amyotrophic Lateral Sclerosis.* Front Neurol, 2018. **9**: p. 1078.
343. Eymard-Pierre, E., et al., *Infantile-onset ascending hereditary spastic paralysis is associated with mutations in the alsin gene.* Am J Hum Genet, 2002. **71**(3): p. 518-27.
344. Yang, Y., et al., *The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis.* Nat Genet, 2001. **29**(2): p. 160-5.
345. Hadano, S., et al., *A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2.* Nat Genet, 2001. **29**(2): p. 166-73.
346. Kropatsch, R., et al., *BICD2 mutational analysis in hereditary spastic paraplegia and hereditary motor and sensory neuropathy.* Muscle Nerve, 2019. **59**(4): p. 484-486.
347. Neveling, K., et al., *Mutations in BICD2, which encodes a golgin and important motor adaptor, cause congenital autosomal-dominant spinal muscular atrophy.* Am J Hum Genet, 2013. **92**(6): p. 946-54.
348. Oates, E.C., et al., *Mutations in BICD2 cause dominant congenital spinal muscular atrophy and hereditary spastic paraplegia.* Am J Hum Genet, 2013. **92**(6): p. 965-73.
349. Peeters, K., et al., *Molecular defects in the motor adaptor BICD2 cause proximal spinal muscular atrophy with autosomal-dominant inheritance.* Am J Hum Genet, 2013. **92**(6): p. 955-64.
350. Storbeck, M., et al., *Phenotypic extremes of BICD2-opathies: from lethal, congenital muscular atrophy with arthrogryposis to asymptomatic with subclinical features.* Eur J Hum Genet, 2017. **25**(9): p. 1040-1048.
351. Alan Pestronk, M., *Neuromuscular Disease Center*. 2019, Washington University.
352. Bouhouche, A., et al., *Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5) gene causes autosomal recessive mutilating sensory neuropathy with spastic paraplegia.* J Med Genet, 2006. **43**(5): p. 441-3.
353. Halevy, A., et al., *Novel EXOSC3 mutation causes complicated hereditary spastic paraplegia.* J Neurol, 2014. **261**(11): p. 2165-9.
354. Wan, J., et al., *Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration.* Nat Genet, 2012. **44**(6): p. 704-8.
355. Lo Giudice, T., et al., *Hereditary spastic paraplegia: clinical-genetic characteristics and evolving molecular mechanisms.* Exp Neurol, 2014. **261**: p. 518-39.
356. Lynex, C.N., et al., *Homozygosity for a missense mutation in the 67 kDa isoform of glutamate decarboxylase in a family with autosomal recessive spastic cerebral palsy: parallels with Stiff-Person Syndrome and other movement disorders.* BMC Neurol, 2004. **4**(1): p. 20.
357. Hollstein, R., et al., *HACE1 deficiency causes an autosomal recessive neurodevelopmental syndrome.* J Med Genet, 2015. **52**(12): p. 797-803.
358. Akawi, N., et al., *Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families.* Nat Genet, 2015. **47**(11): p. 1363-9.
359. Liu, N., et al., *Hereditary spastic paraplegia associated with a rare IFIH1 mutation: a case report and literature review.* Hereditas, 2019. **156**: p. 28.

360. Smyth, D.J., et al., *A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region.* Nat Genet, 2006. **38**(6): p. 617-9.
361. Rice, G.I., et al., *Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling.* Nat Genet, 2014. **46**(5): p. 503-509.
362. Oda, H., et al., *Aicardi-Goutieres syndrome is caused by IFIH1 mutations.* Am J Hum Genet, 2014. **95**(1): p. 121-5.
363. Helbig, K.L., et al., *A recurrent mutation in KCNA2 as a novel cause of hereditary spastic paraplegia and ataxia.* Ann Neurol, 2016. **80**(4).
364. Syrbe, S., et al., *De novo loss- or gain-of-function mutations in KCNA2 cause epileptic encephalopathy.* Nat Genet, 2015. **47**(4): p. 393-399.
365. Pena, S.D. and R.L. Coimbra, *Ataxia and myoclonic epilepsy due to a heterozygous new mutation in KCNA2: proposal for a new channelopathy.* Clin Genet, 2015. **87**(2): p. e1-3.
366. Zhao, M., et al., *Genetic and Clinical Profile of Chinese Patients with Autosomal Dominant Spastic Paraplegia.* Mol Diagn Ther, 2019. **23**(6): p. 781-789.
367. Josifova, D.J., et al., *Heterozygous KIDINS220/ARMS nonsense variants cause spastic paraplegia, intellectual disability, nystagmus, and obesity.* Hum Mol Genet, 2016. **25**(11): p. 2158-2167.
368. Song, Y., et al., *Identification of a compound heterozygote in LYST gene: a case report on Chediak-Higashi syndrome.* BMC Med Genet, 2020. **21**(1): p. 4.
369. Shimazaki, H., et al., *Autosomal-recessive complicated spastic paraplegia with a novel lysosomal trafficking regulator gene mutation.* J Neurol Neurosurg Psychiatry, 2014. **85**(9): p. 1024-8.
370. Morrone, K., et al., *Two novel mutations identified in an african-american child with chediak-higashi syndrome.* Case Rep Med, 2010. **2010**: p. 967535.
371. Karim, M.A., et al., *Apparent genotype-phenotype correlation in childhood, adolescent, and adult Chediak-Higashi syndrome.* Am J Med Genet, 2002. **108**(1): p. 16-22.
372. Verny, C., et al., *Hereditary spastic paraplegia-like disorder due to a mitochondrial ATP6 gene point mutation.* Mitochondrion, 2011. **11**(1): p. 70-5.
373. Craig, K., et al., *Episodic ataxia and hemiplegia caused by the 8993T->C mitochondrial DNA mutation.* J Med Genet, 2007. **44**(12): p. 797-9.
374. Rantamaki, M.T., et al., *Adult-onset ataxia and polyneuropathy caused by mitochondrial 8993T-->C mutation.* Ann Neurol, 2005. **58**(2): p. 337-40.
375. Dionisi-Vici, C., et al., *Fulminant Leigh syndrome and sudden unexpected death in a family with the T9176C mutation of the mitochondrial ATPase 6 gene.* J Inherit Metab Dis, 1998. **21**(1): p. 2-8.
376. Takahashi, S., et al., *De novo mtDNA nt 8993 (T-->G) mutation resulting in Leigh syndrome.* Am J Hum Genet, 1998. **62**(3): p. 717-9.
377. Campos, Y., et al., *Leigh syndrome associated with the T9176C mutation in the ATPase 6 gene of mitochondrial DNA.* Neurology, 1997. **49**(2): p. 595-7.
378. de Vries, D.D., et al., *A second missense mutation in the mitochondrial ATPase 6 gene in Leigh's syndrome.* Ann Neurol, 1993. **34**(3): p. 410-2.
379. Castagna, A.E., et al., *Late onset Leigh syndrome and ataxia due to a T to C mutation at bp 9,185 of mitochondrial DNA.* Am J Med Genet A, 2007. **143A**(8): p. 808-16.
380. Lamminen, T., et al., *A mitochondrial mutation at nt 9101 in the ATP synthase 6 gene associated with deficient oxidative phosphorylation in a family with Leber hereditary optic neuroretinopathy.* Am J Hum Genet, 1995. **56**(5): p. 1238-40.
381. Holt, I.J., et al., *A new mitochondrial disease associated with mitochondrial DNA heteroplasmy.* Am J Hum Genet, 1990. **46**(3): p. 428-33.

382. Tiranti, V., et al., *A novel frameshift mutation of the mtDNA COIII gene leads to impaired assembly of cytochrome c oxidase in a patient affected by Leigh-like syndrome.* Hum Mol Genet, 2000. **9**(18): p. 2733-42.
383. Johns, D.R. and M.J. Neufeld, *Cytochrome c oxidase mutations in Leber hereditary optic neuropathy.* Biochem Biophys Res Commun, 1993. **196**(2): p. 810-5.
384. Keightley, J.A., et al., *A microdeletion in cytochrome c oxidase (COX) subunit III associated with COX deficiency and recurrent myoglobinuria.* Nat Genet, 1996. **12**(4): p. 410-6.
385. Clarencon, F., et al., *Spastic paraparesis as a manifestation of Leber's disease.* J Neurol, 2006. **253**(4): p. 525-6.
386. Sudoyo, H., et al., *Leber's hereditary optic neuropathy in Indonesia: two families with the mtDNA 11778G>A and 14484T>C mutations.* Hum Mutat, 1998. **Suppl 1**: p. S271-4.
387. Kormann, B.A., et al., *Detection of the G to A mitochondrial DNA mutation at position 11778 in German families with Leber's hereditary optic neuropathy.* Hum Genet, 1991. **88**(1): p. 98-100.
388. Majander, A., et al., *Electron transfer properties of NADH:ubiquinone reductase in the ND1/3460 and the ND4/11778 mutations of the Leber hereditary optic neuroretinopathy (LHON).* FEBS Lett, 1991. **292**(1-2): p. 289-92.
389. Wallace, D.C., et al., *Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy.* Science, 1988. **242**(4884): p. 1427-30.
390. De Vries, D.D., et al., *Genetic and biochemical impairment of mitochondrial complex I activity in a family with Leber hereditary optic neuropathy and hereditary spastic dystonia.* Am J Hum Genet, 1996. **58**(4): p. 703-11.
391. Ilgaz Aydinlar, E., et al., *Mutation in FAM134B causing hereditary sensory neuropathy with spasticity in a Turkish family.* Muscle Nerve, 2014. **49**(5): p. 774-5.
392. Kurth, I., et al., *Mutations in FAM134B, encoding a newly identified Golgi protein, cause severe sensory and autonomic neuropathy.* Nat Genet, 2009. **41**(11): p. 1179-81.
393. Horibata, Y., et al., *EPT1 (selenoprotein I) is critical for the neural development and maintenance of plasmalogen in humans.* J Lipid Res, 2018. **59**(6): p. 1015-1026.
394. Ahmed, M.Y., et al., *A mutation of EPT1 (SELENOI) underlies a new disorder of Kennedy pathway phospholipid biosynthesis.* Brain, 2017. **140**(3): p. 547-554.
395. Tomkins, J.E., et al., *PINOT: an intuitive resource for integrating protein-protein interactions.* Cell Commun Signal, 2020. **18**(1): p. 92.
396. The Gene Ontology, C., *The Gene Ontology Resource: 20 years and still GOing strong.* Nucleic Acids Res, 2019. **47**(D1): p. D330-D338.
397. Jassal, B., et al., *The reactome pathway knowledgebase.* Nucleic Acids Res, 2020. **48**(D1): p. D498-D503.
398. Wagner, M., et al., *Bi-allelic variants in RNF170 are associated with hereditary spastic paraplegia.* Nat Commun, 2019. **10**(1): p. 4790.
399. Bakula, D., et al., *WIPI3 and WIPI4 beta-propellers are scaffolds for LKB1-AMPK-TSC signalling circuits in the control of autophagy.* Nat Commun, 2017. **8**: p. 15637.
400. Saitsu, H., et al., *De novo mutations in the autophagy gene WDR45 cause static encephalopathy of childhood with neurodegeneration in adulthood.* Nat Genet, 2013. **45**(4): p. 445-9, 449e1.
401. Bonham, L.W., et al., *Protein network analysis reveals selectively vulnerable regions and biological processes in FTD.* Neurol Genet, 2018. **4**(5): p. e266.
402. Tomkins, J.E., et al., *Comparative Protein Interaction Network Analysis Identifies Shared and Distinct Functions for the Human ROCO Proteins.* Proteomics, 2018. **18**(10): p. e1700444.
403. Vavouraki, N., et al., *Integrating protein networks and machine learning for disease stratification in the Hereditary Spastic Paraplegias.* iScience, 2021. **24**(5): p. 102484.
404. Leveille, E., et al., *SPTAN1 variants as a potential cause for autosomal recessive hereditary spastic paraplegia.* J Hum Genet, 2019. **64**(11): p. 1145-1151.

405. Jones, H.F., et al., *Autosomal dominant ADAR c.3019G>A (p.(G1007R)) variant is an important mimic of hereditary spastic paraplegia and cerebral palsy.* Brain Dev, 2022. **44**(2): p. 153-160.
406. Morejon-Garcia, P., et al., *Dysfunctional Homozygous VRK1-D263G Variant Impairs the Assembly of Cajal Bodies and DNA Damage Response in Hereditary Spastic Paraplegia.* Neurol Genet, 2021. **7**(5): p. e624.
407. Sahin, I. and H. Saat, *Hereditary spastic paraplegia: new insights into clinical variability and spasticity-ataxia phenotype, and novel mutations.* Acta Neurol Belg, 2021.
408. Sager, G., et al., *HACE1, GLRX5, and ELP2 gene variant cause spastic paraplegies.* Acta Neurol Belg, 2021.
409. Parodi, L., et al., *Hereditary ataxias and paraparesias: clinical and genetic update.* Curr Opin Neurol, 2018. **31**(4): p. 462-471.
410. Oughtred, R., et al., *The BioGRID interaction database: 2019 update.* Nucleic Acids Res, 2019. **47**(D1): p. D529-D541.
411. Singapore, N.U.o. *MBInfo-Mechanobiology Institute*. 18/06/2019; Available from: https://www.mechanobio.info/.
412. Licata, L., et al., *MINT, the molecular interaction database: 2012 update.* Nucleic Acids Res, 2012. **40**(Database issue): p. D857-61.
413. UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.
414. Matsuto, M., F. Kano, and M. Murata, *Reconstitution of the targeting of Rab6A to the Golgi apparatus in semi-intact HeLa cells: A role of BICD2 in stabilizing Rab6A on Golgi membranes and a concerted role of Rab6A/BICD2 interactions in Golgi-to-ER retrograde transport.* Biochim Biophys Acta, 2015. **1853**(10 Pt A): p. 2592-609.
415. Crow, Y.J., J. Shetty, and J.H. Livingston, *Treatments in Aicardi-Goutieres syndrome.* Dev Med Child Neurol, 2020. **62**(1): p. 42-47.
416. Faigle, W., et al., *Deficient peptide loading and MHC class II endosomal sorting in a human genetic immunodeficiency disease: the Chediak-Higashi syndrome.* J Cell Biol, 1998. **141**(5): p. 1121-34.
417. Mirel, D.B., et al., *Characterization of the human mitochondrial aconitase gene (ACO2).* Gene, 1998. **213**(1-2): p. 205-18.
418. Liao, Y.H., S.M. Hsu, and P.H. Huang, *ARMS depletion facilitates UV irradiation induced apoptotic cell death in melanoma.* Cancer Res, 2007. **67**(24): p. 11547-56.
419. Leong, W.F. and V.T. Chow, *Transcriptomic and proteomic analyses of rhabdomyosarcoma cells reveal differential cellular gene expression in response to enterovirus 71 infection.* Cell Microbiol, 2006. **8**(4): p. 565-80.
420. Freund, A., et al., *Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1.* Cell, 2014. **159**(6): p. 1389-403.
421. Sleat, D.E., et al., *Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis.* Science, 1997. **277**(5333): p. 1802-5.
422. Vazza, G., et al., *A new locus for autosomal recessive spastic paraplegia associated with mental retardation and distal motor neuropathy, SPG14, maps to chromosome 3q27-q28.* Am J Hum Genet, 2000. **67**(2): p. 504-9.
423. Ebbing, B., et al., *Effect of spastic paraplegia mutations in KIF5A kinesin on transport activity.* Hum Mol Genet, 2008. **17**(9): p. 1245-52.
424. Caballero Oteyza, A., et al., *Motor protein mutations cause a new form of hereditary spastic paraplegia.* Neurology, 2014. **82**(22): p. 2007-16.
425. Tarrade, A., et al., *A mutation of spastin is responsible for swellings and impairment of transport in a region of axon characterized by changes in microtubule composition.* Hum Mol Genet, 2006. **15**(24): p. 3544-58.

426. White, S.R., et al., *Recognition of C-terminal amino acids in tubulin by pore loops in Spastin is important for microtubule severing.* J Cell Biol, 2007. **176**(7): p. 995-1005.
427. Zhu, P.P., et al., *SPG3A protein atlastin-1 is enriched in growth cones and promotes axon elongation during neuronal development.* Hum Mol Genet, 2006. **15**(8): p. 1343-53.
428. Namekawa, M., et al., *Mutations in the SPG3A gene encoding the GTPase atlastin interfere with vesicle trafficking in the ER/Golgi interface and Golgi morphogenesis.* Mol Cell Neurosci, 2007. **35**(1): p. 1-13.
429. Bakowska, J.C., et al., *Troyer syndrome protein spartin is mono-ubiquitinated and functions in EGF receptor trafficking.* Mol Biol Cell, 2007. **18**(5): p. 1683-92.
430. Goytain, A., et al., *NIPA1(SPG6), the basis for autosomal dominant form of hereditary spastic paraplegia, encodes a functional Mg2+ transporter.* J Biol Chem, 2007. **282**(11): p. 8060-8.
431. Cheon, C.K., et al., *Autosomal dominant transmission of complicated hereditary spastic paraplegia due to a dominant negative mutation of KIF1A, SPG30 gene.* Sci Rep, 2017. **7**(1): p. 12527.
432. Hirst, J., et al., *Loss of AP-5 results in accumulation of aberrant endolysosomes: defining a new type of lysosomal storage disease.* Hum Mol Genet, 2015. **24**(17): p. 4984-96.
433. Chang, J., S. Lee, and C. Blackstone, *Spastic paraplegia proteins spastizin and spatacsin mediate autophagic lysosome reformation.* J Clin Invest, 2014. **124**(12): p. 5249-62.
434. Tesson, C., J. Koht, and G. Stevanin, *Delving into the complexity of hereditary spastic paraplegias: how unexpected phenotypes and inheritance modes are revolutionizing their nosology.* Hum Genet, 2015. **134**(6): p. 511-38.
435. Orso, G., et al., *Homotypic fusion of ER membranes requires the dynamin-like GTPase atlastin.* Nature, 2009. **460**(7258): p. 978-83.
436. Rismanchi, N., et al., *Atlastin GTPases are required for Golgi apparatus and ER morphogenesis.* Hum Mol Genet, 2008. **17**(11): p. 1591-604.
437. Hu, J., et al., *Membrane proteins of the endoplasmic reticulum induce high-curvature tubules.* Science, 2008. **319**(5867): p. 1247-50.
438. Evans, K., et al., *Interaction of two hereditary spastic paraplegia gene products, spastin and atlastin, suggests a common pathway for axonal maintenance.* Proc Natl Acad Sci U S A, 2006. **103**(28): p. 10666-71.
439. Park, S.H., et al., *Hereditary spastic paraplegia proteins REEP1, spastin, and atlastin-1 coordinate microtubule interactions with the tubular ER network.* J Clin Invest, 2010. **120**(4): p. 1097-110.
440. Falk, J., et al., *Functional mutation analysis provides evidence for a role of REEP1 in lipid droplet biology.* Hum Mutat, 2014. **35**(4): p. 497-504.
441. Klemm, R.W., et al., *A conserved role for atlastin GTPases in regulating lipid droplet size.* Cell Rep, 2013. **3**(5): p. 1465-75.
442. Renvoise, B., et al., *Reep1 null mice reveal a converging role for hereditary spastic paraplegia proteins in lipid droplet regulation.* Hum Mol Genet, 2016. **25**(23): p. 5111-5125.
443. Boukhris, A., et al., *Alteration of ganglioside biosynthesis responsible for complex hereditary spastic paraplegia.* Am J Hum Genet, 2013. **93**(1): p. 118-23.
444. Harlalka, G.V., et al., *Mutations in B4GALNT1 (GM2 synthase) underlie a new disorder of ganglioside biosynthesis.* Brain, 2013. **136**(Pt 12): p. 3618-24.
445. Hammer, M.B., et al., *Mutations in GBA2 cause autosomal-recessive cerebellar ataxia with spasticity.* Am J Hum Genet, 2013. **92**(2): p. 245-51.
446. Martin, E., et al., *Loss of function of glucocerebrosidase GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia.* Am J Hum Genet, 2013. **92**(2): p. 238-44.
447. Raju, D., et al., *Accumulation of glucosylceramide in the absence of the beta-glucosidase GBA2 alters cytoskeletal dynamics.* PLoS Genet, 2015. **11**(3): p. e1005063.

448. Schule, R., et al., *Marked accumulation of 27-hydroxycholesterol in SPG5 patients with hereditary spastic paresis.* J Lipid Res, 2010. **51**(4): p. 819-23.
449. Bross, P., et al., *The Hsp60-(p.V98I) mutation associated with hereditary spastic paraplegia SPG13 compromises chaperonin function both in vitro and in vivo.* J Biol Chem, 2008. **283**(23): p. 15694-700.
450. Ferreirinha, F., et al., *Axonal degeneration in paraplegin-deficient mice is associated with abnormal mitochondria and impairment of axonal transport.* J Clin Invest, 2004. **113**(2): p. 231-42.
451. Lu, J., F. Rashid, and P.C. Byrne, *The hereditary spastic paraplegia protein spartin localises to mitochondria.* J Neurochem, 2006. **98**(6): p. 1908-19.
452. Yang, Y., et al., *A Newly Identified Missense Mutation in FARS2 Causes Autosomal-Recessive Spastic Paraplegia.* Hum Mutat, 2016. **37**(2): p. 165-9.
453. Lavie, J., et al., *Mitochondrial morphology and cellular distribution are altered in SPG31 patients and are linked to DRP1 hyperphosphorylation.* Hum Mol Genet, 2017. **26**(4): p. 674-685.
454. Gabrych, D.R., et al., *Going Too Far Is the Same as Falling Short(dagger): Kinesin-3 Family Members in Hereditary Spastic Paraplegia.* Front Cell Neurosci, 2019. **13**: p. 419.
455. Yang, Y., et al., *The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis.* Nature Genetics, 2001. **29**(2): p. 160-165.
456. Lesca, G., et al., *Infantile ascending hereditary spastic paralysis (IAHSP): clinical features in 11 families.* Neurology, 2003. **60**(4): p. 674-82.
457. Wiethoff, S., et al., *Pure Cerebellar Ataxia with Homozygous Mutations in the PNPLA6 Gene.* Cerebellum, 2017. **16**(1): p. 262-267.
458. Synofzik, M., et al., *PNPLA6 mutations cause Boucher-Neuhauser and Gordon Holmes syndromes as part of a broad neurodegenerative spectrum.* Brain, 2014. **137**(Pt 1): p. 69-77.
459. Pfeffer, G., et al., *SPG7 mutations are a common cause of undiagnosed ataxia.* Neurology, 2015. **84**(11): p. 1174-6.
460. Dor, T., et al., *KIF1C mutations in two families with hereditary spastic paraparesis and cerebellar dysfunction.* J Med Genet, 2014. **51**(2): p. 137-42.
461. Soehn, A.S., et al., *Uniparental disomy of chromosome 16 unmasks recessive mutations of FA2H/SPG35 in 4 families.* Neurology, 2016. **87**(2): p. 186-91.
462. Pierson, T.M., et al., *Exome sequencing and SNP analysis detect novel compound heterozygosity in fatty acid hydroxylase-associated neurodegeneration.* Eur J Hum Genet, 2012. **20**(4): p. 476-9.
463. Dick, K.J., et al., *Mutation of FA2H underlies a complicated form of hereditary spastic paraplegia (SPG35).* Hum Mutat, 2010. **31**(4): p. E1251-60.
464. Mademan, I., et al., *Multisystemic SYNE1 ataxia: confirming the high frequency and extending the mutational and phenotypic spectrum.* Brain, 2016. **139**(Pt 8): p. e46.
465. Synofzik, M., et al., *SYNE1 ataxia is a common recessive ataxia with major non-cerebellar features: a large multi-centre study.* Brain, 2016. **139**(Pt 5): p. 1378-93.
466. Dupre, N., et al., *Clinical and genetic study of autosomal recessive cerebellar ataxia type 1.* Ann Neurol, 2007. **62**(1): p. 93-8.
467. Noreau, A., et al., *SYNE1 mutations in autosomal recessive cerebellar ataxia.* JAMA Neurol, 2013. **70**(10): p. 1296-31.
468. Khateeb, S., et al., *PLA2G6 mutation underlies infantile neuroaxonal dystrophy.* Am J Hum Genet, 2006. **79**(5): p. 942-8.
469. Durr, A., *Movement disorders: Are umbrella terms for rare genetic diseases still useful?* Nat Rev Neurol, 2016. **12**(6): p. 321-2.
470. Fridman, V. and M.A. Saporta, *Mechanisms and Treatments in Demyelinating CMT.* Neurotherapeutics, 2021. **18**(4): p. 2236-2268.

471. Sambuughin, N., et al., *Adult-onset autosomal dominant spastic paraplegia linked to a GTPase-effector domain mutation of dynamin 2.* BMC Neurol, 2015. **15**: p. 223.
472. Strickland, A.V., et al., *Mutation screen reveals novel variants and expands the phenotypes associated with DYNC1H1.* J Neurol, 2015. **262**(9): p. 2124-34.
473. Kett, L.R. and W.T. Dauer, *Endolysosomal dysfunction in Parkinson's disease: Recent developments and future challenges.* Mov Disord, 2016. **31**(10): p. 1433-1443.
474. Klein, A.D. and J.R. Mazzulli, *Is Parkinson's disease a lysosomal disorder?* Brain, 2018. **141**(8): p. 2255-2262.
475. Robak, L.A., et al., *Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease.* Brain, 2017. **140**(12): p. 3191-3203.
476. Manzoni, C., *The LRRK2-macroautophagy axis and its relevance to Parkinson's disease.* Biochem Soc Trans, 2017. **45**(1): p. 155-162.
477. Plotegher, N., et al., *Ceramides in Parkinson's Disease: From Recent Evidence to New Hypotheses.* Front Neurosci, 2019. **13**: p. 330.
478. Suzuki, R. and H. Shimodaira, *Pvclust: an R package for assessing the uncertainty in hierarchical clustering.* Bioinformatics, 2006. **22**(12): p. 1540-2.
479. Lobo, A., et al., *Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group.* Neurology, 2000. **54**(11 Suppl 5): p. S4-9.
480. Masters, C.L., et al., *Amyloid plaque core protein in Alzheimer disease and Down syndrome.* Proc Natl Acad Sci U S A, 1985. **82**(12): p. 4245-9.
481. Ihara, Y., et al., *Phosphorylated tau protein is integrated into paired helical filaments in Alzheimer's disease.* J Biochem, 1986. **99**(6): p. 1807-10.
482. Pickford, F., et al., *The autophagy-related protein beclin 1 shows reduced expression in early Alzheimer disease and regulates amyloid beta accumulation in mice.* J Clin Invest, 2008. **118**(6): p. 2190-9.
483. Cataldo, A.M., et al., *Presenilin mutations in familial Alzheimer disease and transgenic mouse models accelerate neuronal lysosomal pathology.* J Neuropathol Exp Neurol, 2004. **63**(8): p. 821-30.
484. Lee, J.H., et al., *Lysosomal proteolysis and autophagy require presenilin 1 and are disrupted by Alzheimer-related PS1 mutations.* Cell, 2010. **141**(7): p. 1146-58.
485. Yamamoto, A., et al., *Bafilomycin A1 prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes in rat hepatoma cell line, H-4-II-E cells.* Cell Struct Funct, 1998. **23**(1): p. 33-42.
486. Ross, O.A., et al., *Genomic investigation of alpha-synuclein multiplication and parkinsonism.* Ann Neurol, 2008. **63**(6): p. 743-50.
487. Winslow, A.R., et al., *alpha-Synuclein impairs macroautophagy: implications for Parkinson's disease.* J Cell Biol, 2010. **190**(6): p. 1023-37.
488. Cuervo, A.M., et al., *Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy.* Science, 2004. **305**(5688): p. 1292-5.
489. Oh, S.E. and M.M. Mouradian, *Regulation of Signal Transduction by DJ-1.* Adv Exp Med Biol, 2017. **1037**: p. 97-131.
490. Berger, Z., et al., *Rapamycin alleviates toxicity of different aggregate-prone proteins.* Hum Mol Genet, 2006. **15**(3): p. 433-42.
491. Menzies, F.M., B. Ravikumar, and D.C. Rubinsztein, *Protective roles for induction of autophagy in multiple proteinopathies.* Autophagy, 2006. **2**(3): p. 224-5.
492. Williams, A., et al., *Aggregate-prone proteins are cleared from the cytosol by autophagy: therapeutic implications.* Curr Top Dev Biol, 2006. **76**: p. 89-101.
493. Alvarez-Erviti, L., et al., *Influence of microRNA deregulation on chaperone-mediated autophagy and alpha-synuclein pathology in Parkinson's disease.* Cell Death Dis, 2013. **4**: p. e545.

494. Sivasathiaseelan, H., et al., *Frontotemporal Dementia: A Clinical Review.* Semin Neurol, 2019. **39**(2): p. 251-263.
495. Bang, J., S. Spina, and B.L. Miller, *Frontotemporal dementia.* Lancet, 2015. **386**(10004): p. 1672-82.
496. Rohrer, J.D., et al., *The heritability and genetics of frontotemporal lobar degeneration.* Neurology, 2009. **73**(18): p. 1451-6.
497. Le Ber, I., *Genetics of frontotemporal lobar degeneration: an up-date and diagnosis algorithm.* Rev Neurol (Paris), 2013. **169**(10): p. 811-9.
498. Gijselinck, I., et al., *Loss of TBK1 is a frequent cause of frontotemporal dementia in a Belgian cohort.* Neurology, 2015. **85**(24): p. 2116-25.
499. Pottier, C., et al., *Whole-genome sequencing reveals important role for TBK1 and OPTN mutations in frontotemporal lobar degeneration without motor neuron disease.* Acta Neuropathol, 2015. **130**(1): p. 77-92.
500. Freischmidt, A., et al., *Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia.* Nat Neurosci, 2015. **18**(5): p. 631-6.
501. Mackenzie, I.R., P. Frick, and M. Neumann, *The neuropathology associated with repeat expansions in the C9ORF72 gene.* Acta Neuropathol, 2014. **127**(3): p. 347-57.
502. Urwin, H., et al., *FUS pathology defines the majority of tau- and TDP-43-negative frontotemporal lobar degeneration.* Acta Neuropathol, 2010. **120**(1): p. 33-41.
503. Kwong, L.K., et al., *TDP-43 proteinopathy: the neuropathology underlying major forms of sporadic and familial frontotemporal lobar degeneration and motor neuron disease.* Acta Neuropathol, 2007. **114**(1): p. 63-70.
504. Lee, V.M., M. Goedert, and J.Q. Trojanowski, *Neurodegenerative tauopathies.* Annu Rev Neurosci, 2001. **24**: p. 1121-59.
505. Majid, T., et al., *In vivo axonal transport deficits in a mouse model of fronto-temporal dementia.* Neuroimage Clin, 2014. **4**: p. 711-7.
506. Hochgrafe, K., et al., *Preventive methylene blue treatment preserves cognition in mice expressing full-length pro-aggregant human Tau.* Acta Neuropathol Commun, 2015. **3**: p. 25.
507. Melis, V., et al., *Effects of oxidized and reduced forms of methylthioninium in two transgenic mouse tauopathy models.* Behav Pharmacol, 2015. **26**(4): p. 353-68.
508. Congdon, E.E., et al., *Methylthioninium chloride (methylene blue) induces autophagy and attenuates tauopathy in vitro and in vivo.* Autophagy, 2012. **8**(4): p. 609-22.
509. Shin, S.Y., et al., *SIRT1 activation by methylene blue, a repurposed drug, leads to AMPK-mediated inhibition of steatosis and steatohepatitis.* Eur J Pharmacol, 2014. **727**: p. 115-24.
510. Holler, C.J., et al., *Trehalose upregulates progranulin expression in human and mouse models of GRN haploinsufficiency: a novel therapeutic lead to treat frontotemporal dementia.* Mol Neurodegener, 2016. **11**(1): p. 46.
511. Wang, I.F., K.J. Tsai, and C.K. Shen, *Autophagy activation ameliorates neuronal pathogenesis of FTLD-U mice: a new light for treatment of TARDBP/TDP-43 proteinopathies.* Autophagy, 2013. **9**(2): p. 239-40.
512. Stolz, A., A. Ernst, and I. Dikic, *Cargo recognition and trafficking in selective autophagy.* Nat Cell Biol, 2014. **16**(6): p. 495-501.
513. Zhang, K.Y., et al., *Ubiquilin 2: a component of the ubiquitin-proteasome system with an emerging role in neurodegeneration.* Int J Biochem Cell Biol, 2014. **50**: p. 123-6.
514. Pilli, M., et al., *TBK-1 promotes autophagy-mediated antimicrobial defense by controlling autophagosome maturation.* Immunity, 2012. **37**(2): p. 223-34.
515. Ince, P.G., et al., *Amyotrophic lateral sclerosis associated with genetic abnormalities in the gene encoding Cu/Zn superoxide dismutase: molecular pathology of five new cases, and comparison with previous reports and 73 sporadic cases of ALS.* J Neuropathol Exp Neurol, 1998. **57**(10): p. 895-904.

516. Sacks, O., *The origin of "Awakenings".* Br Med J (Clin Res Ed), 1983. **287**(6409): p. 1968-9.
517. Hay, J., et al., *Chronic Traumatic Encephalopathy: The Neuropathological Legacy of Traumatic Brain Injury.* Annu Rev Pathol, 2016. **11**: p. 21-45.
518. Parkinson, N., et al., *ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B).* Neurology, 2006. **67**(6): p. 1074-7.
519. Skibinski, G., et al., *Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia.* Nat Genet, 2005. **37**(8): p. 806-8.
520. Kohler, K., et al., *A combined proteomic and genetic analysis identifies a role for the lipid desaturase Desat1 in starvation-induced autophagy in Drosophila.* Autophagy, 2009. **5**(7): p. 980-90.
521. Zalckvar, E., et al., *A systems level strategy for analyzing the cell death network: implication in exploring the apoptosis/autophagy connection.* Cell Death Differ, 2010. **17**(8): p. 1244-53.
522. Ojha, R., M. Ishaq, and S.K. Singh, *Caspase-mediated crosstalk between autophagy and apoptosis: Mutual adjustment or matter of dominance.* J Cancer Res Ther, 2015. **11**(3): p. 514-24.
523. Lipinski, M.M., et al., *A genome-wide siRNA screen reveals multiple mTORC1 independent signaling pathways regulating autophagy under normal nutritional conditions.* Dev Cell, 2010. **18**(6): p. 1041-52.
524. Caron, E., et al., *A comprehensive map of the mTOR signaling network.* Mol Syst Biol, 2010. **6**: p. 453.
525. Philipp, O., et al., *The autophagy interaction network of the aging model Podospora anserina.* BMC Bioinformatics, 2017. **18**(1): p. 196.
526. Huett, A., et al., *A novel hybrid yeast-human network analysis reveals an essential role for FNBP1L in antibacterial autophagy.* J Immunol, 2009. **182**(8): p. 4917-30.
527. Wang, J.Y., et al., *Network analysis reveals crosstalk between autophagy genes and disease genes.* Sci Rep, 2017. **7**: p. 44391.
528. Jegga, A.G., et al., *Systems biology of the autophagy-lysosomal pathway.* Autophagy, 2011. **7**(5): p. 477-89.
529. Shu, J., et al., *Detection of molecular signatures and pathways shared by Alzheimer's disease and type 2 diabetes.* Gene, 2022. **810**: p. 146070.
530. Zhu, Y., et al., *Exploring Shared Pathogenesis of Alzheimer's Disease and Type 2 Diabetes Mellitus via Co-expression Networks Analysis.* Curr Alzheimer Res, 2020. **17**(6): p. 566-575.
531. Rahman, M.H., et al., *A Network-Based Bioinformatics Approach to Identify Molecular Biomarkers for Type 2 Diabetes that Are Linked to the Progression of Neurological Diseases.* Int J Environ Res Public Health, 2020. **17**(3).
532. Battaglia, C., et al., *Candidate Genes and MiRNAs Linked to the Inverse Relationship Between Cancer and Alzheimer's Disease: Insights From Data Mining and Enrichment Analysis.* Front Genet, 2019. **10**: p. 846.
533. Caberlotto, L., et al., *The central role of AMP-kinase and energy homeostasis impairment in Alzheimer's disease: a multifactor network analysis.* PLoS One, 2013. **8**(11): p. e78919.
534. Li, Q.S. and L. De Muynck, *Differentially expressed genes in Alzheimer's disease highlighting the roles of microglia genes including OLR1 and astrocyte gene CDK2AP1.* Brain Behav Immun Health, 2021. **13**: p. 100227.
535. Caberlotto, L. and T.P. Nguyen, *A systems biology investigation of neurodegenerative dementia reveals a pivotal role of autophagy.* BMC Syst Biol, 2014. **8**: p. 65.
536. Estrada, E., *Cascading from SARS-CoV-2 to Parkinson's Disease through Protein-Protein Interactions.* Viruses, 2021. **13**(5).
537. Keane, H., et al., *Protein-protein interaction networks identify targets which rescue the MPP+ cellular model of Parkinson's disease.* Sci Rep, 2015. **5**: p. 17004.

538. He, P.K., et al., *Idebenone-Activating Autophagic Degradation of alpha-Synuclein via Inhibition of AKT-mTOR Pathway in a SH-SY5Y-A53T Model of Parkinson's Disease: A Network Pharmacological Approach.* Evid Based Complement Alternat Med, 2021. **2021**: p. 8548380.
539. Monti, C., et al., *Systems biology analysis of the proteomic alterations induced by MPP(+), a Parkinson's disease-related mitochondrial toxin.* Front Cell Neurosci, 2015. **9**: p. 14.
540. Torres-Odio, S., et al., *Progression of pathology in PINK1-deficient mouse brain from splicing via ubiquitination, ER stress, and mitophagy changes to neuroinflammation.* J Neuroinflammation, 2017. **14**(1): p. 154.
541. Kalvari, I., et al., *iLIR: A web resource for prediction of Atg8-family interacting proteins.* Autophagy, 2014. **10**(5): p. 913-25.
542. Xie, Q., et al., *hfAIM: A reliable bioinformatics approach for in silico genome-wide identification of autophagy-associated Atg8-interacting motifs in various organisms.* Autophagy, 2016. **12**(5): p. 876-87.
543. Gouw, M., et al., *Exploring Short Linear Motifs Using the ELM Database and Tools.* Curr Protoc Bioinformatics, 2017. **58**: p. 8 22 1-8 22 35.
544. Krystkowiak, I. and N.E. Davey, *SLiMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions.* Nucleic Acids Res, 2017. **45**(W1): p. W464-W469.
545. Jacomin, A.C., et al., *iLIR database: A web resource for LIR motif-containing proteins in eukaryotes.* Autophagy, 2016. **12**(10): p. 1945-1953.
546. Deng, W., et al., *THANATOS: an integrative data resource of proteins and post-translational modifications in the regulation of autophagy.* Autophagy, 2018. **14**(2): p. 296-310.
547. Homma, K., K. Suzuki, and H. Sugawara, *The Autophagy Database: an all-inclusive information resource on autophagy that provides nourishment for research.* Nucleic Acids Res, 2011. **39**(Database issue): p. D986-90.
548. Wu, D., et al., *ncRDeathDB: A comprehensive bioinformatics resource for deciphering network organization of the ncRNA-mediated cell death system.* Autophagy, 2015. **11**(10): p. 1917-26.
549. Chen, Y., J. Huang, and B. Liu, *AutomiRDB: a web resource connecting microRNAs and autophagy in cancer.* Apoptosis, 2015. **20**(7): p. 1016-7.
550. Zhang, L., et al., *GAMDB: a web resource to connect microRNAs with autophagy in gerontology.* Cell Prolif, 2016. **49**(2): p. 246-51.
551. Turei, D., et al., *Autophagy Regulatory Network - a systems-level bioinformatics resource for studying the mechanism and regulation of autophagy.* Autophagy, 2015. **11**(1): p. 155-65.
552. Yuan, W.C., et al., *A Cullin3-KLHL20 Ubiquitin ligase-dependent pathway targets PML to potentiate HIF-1 signaling and prostate cancer progression.* Cancer Cell, 2011. **20**(2): p. 214-28.
553. Lee, Y.R., et al., *The Cullin 3 substrate adaptor KLHL20 mediates DAPK ubiquitination to control interferon responses.* EMBO J, 2010. **29**(10): p. 1748-61.
554. Scott, D.C., et al., *Two Distinct Types of E3 Ligases Work in Unison to Regulate Substrate Ubiquitylation.* Cell, 2016. **166**(5): p. 1198-1214 e24.
555. Luo, W., et al., *Jab1, a novel protease-activated receptor-2 (PAR-2)-interacting protein, is involved in PAR-2-induced activation of activator protein-1.* J Biol Chem, 2006. **281**(12): p. 7927-36.
556. Wang, H., et al., *Fank1 interacts with Jab1 and regulates cell apoptosis via the AP-1 pathway.* Cell Mol Life Sci, 2011. **68**(12): p. 2129-39.
557. Zhou, J., et al., *Jab1 interacts with brain-specific kinase 2 (BRSK2) and promotes its degradation in the ubiquitin-proteasome pathway.* Biochem Biophys Res Commun, 2012. **422**(4): p. 647-52.

558. Cooper, E.M., et al., *K63-specific deubiquitination by two JAMM/MPN+ complexes: BRISC-associated Brcc36 and proteasomal Poh1.* EMBO J, 2009. **28**(6): p. 621-31.
559. Uhle, S., et al., *Protein kinase CK2 and protein kinase D are associated with the COP9 signalosome.* EMBO J, 2003. **22**(6): p. 1302-12.
560. Groisman, R., et al., *The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage.* Cell, 2003. **113**(3): p. 357-67.
561. Lyapina, S., et al., *Promotion of NEDD-CUL1 conjugate cleavage by COP9 signalosome.* Science, 2001. **292**(5520): p. 1382-5.
562. Bech-Otschir, D., et al., *COP9 signalosome-specific phosphorylation targets p53 to degradation by the ubiquitin system.* EMBO J, 2001. **20**(7): p. 1630-9.
563. Seeger, M., et al., *A novel protein complex involved in signal transduction possessing similarities to 26S proteasome subunits.* FASEB J, 1998. **12**(6): p. 469-78.
564. Zheng, J., et al., *CAND1 binds to unneddylated CUL1 and regulates the formation of SCF ubiquitin E3 ligase complex.* Mol Cell, 2002. **10**(6): p. 1519-26.
565. Fica, S.M., et al., *A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation.* Science, 2019. **363**(6428): p. 710-714.
566. Zhang, X., et al., *Structures of the human spliceosomes before and after release of the ligated exon.* Cell Res, 2019. **29**(4): p. 274-285.
567. Zhan, X., et al., *Structure of a human catalytic step I spliceosome.* Science, 2018. **359**(6375): p. 537-545.
568. Zhang, X., et al., *Structure of the human activated spliceosome in three conformational states.* Cell Res, 2018. **28**(3): p. 307-322.
569. Haselbach, D., et al., *Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex.* Cell, 2018. **172**(3): p. 454-464 e11.
570. Bertram, K., et al., *Cryo-EM structure of a human spliceosome activated for step 2 of splicing.* Nature, 2017. **542**(7641): p. 318-323.
571. Zhang, X., et al., *An Atomic Structure of the Human Spliceosome.* Cell, 2017. **169**(5): p. 918-929 e14.
572. Grote, M., et al., *Molecular architecture of the human Prp19/CDC5L complex.* Mol Cell Biol, 2010. **30**(9): p. 2105-19.
573. Jurica, M.S., et al., *Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis.* RNA, 2002. **8**(4): p. 426-39.
574. Groenen, P.M., et al., *Rearrangement of the human CDC5L gene by a t(6;19)(p21;q13.1) in a patient with multicystic renal dysplasia.* Genomics, 1998. **49**(2): p. 218-29.
575. Woolf, A.S. and P.J. Winyard, *Gene expression and cell turnover in human renal dysplasia.* Histol Histopathol, 2000. **15**(1): p. 159-66.
576. Strappazzon, F., et al., *Mitochondrial BCL-2 inhibits AMBRA1-induced autophagy.* EMBO J, 2011. **30**(7): p. 1195-208.
577. Chen, D., et al., *Parkin mono-ubiquitinates Bcl-2 and regulates autophagy.* J Biol Chem, 2010. **285**(49): p. 38214-23.
578. Lariviere, R.C. and J.P. Julien, *Functions of intermediate filaments in neuronal development and disease.* J Neurobiol, 2004. **58**(1): p. 131-48.
579. Barclay, M., et al., *Neuronal expression of peripherin, a type III intermediate filament protein, in the mouse hindbrain.* Histochem Cell Biol, 2007. **128**(6): p. 541-50.
580. McKeon, J.E., et al., *Parkin-mediated K63-polyubiquitination targets ubiquitin C-terminal hydrolase L1 for degradation by the autophagy-lysosome system.* Cell Mol Life Sci, 2015. **72**(9): p. 1811-24.
581. Hunn, B.H.M., et al., *Impairment of Macroautophagy in Dopamine Neurons Has Opposing Effects on Parkinsonian Pathology and Behavior.* Cell Rep, 2019. **29**(4): p. 920-931 e7.

582. Martin, K.R., et al., *Computational model for autophagic vesicle dynamics in single cells.* Autophagy, 2013. **9**(1): p. 74-92.
583. Sakai, Y., et al., *Modeling Membrane Morphological Change during Autophagosome Formation.* iScience, 2020. **23**(9): p. 101466.
584. Ortega, F., et al., *Interplay between alpha-, beta-, and gamma-secretases determines biphasic amyloid-beta protein level in the presence of a gamma-secretase inhibitor.* J Biol Chem, 2013. **288**(2): p. 785-92.
585. Proctor, C.J., et al., *Aggregation, impaired degradation and immunization targeting of amyloid-beta dimers in Alzheimer's disease: a stochastic modelling approach.* Mol Neurodegener, 2012. **7**: p. 32.
586. Manzoni, C., et al., *mTOR independent regulation of macroautophagy by Leucine Rich Repeat Kinase 2 via Beclin-1.* Sci Rep, 2016. **6**: p. 35106.
587. Manzoni, C., et al., *mTOR independent alteration in ULK1 Ser758 phosphorylation following chronic LRRK2 kinase inhibition.* Biosci Rep, 2018. **38**(2).
588. Manzoni, C., et al., *Inhibition of LRRK2 kinase activity stimulates macroautophagy.* Biochim Biophys Acta, 2013. **1833**(12): p. 2900-2910.
589. Fernandes, H.J., et al., *ER Stress and Autophagic Perturbations Lead to Elevated Extracellular alpha-Synuclein in GBA-N370S Parkinson's iPSC-Derived Dopamine Neurons.* Stem Cell Reports, 2016. **6**(3): p. 342-56.
590. Zavodszky, E., et al., *Mutation in VPS35 associated with Parkinson's disease impairs WASH complex association and inhibits autophagy.* Nat Commun, 2014. **5**: p. 3828.
591. Deng, Z., et al., *Autophagy deficiency in neurodevelopmental disorders.* Cell Biosci, 2021. **11**(1): p. 214.
592. Shalev, I., J. Somekh, and A. Eran, *Multimodal bioinformatic analyses of the neurodegenerative disease-associated TECPR2 gene reveal its diverse roles.* J Med Genet, 2021.
593. Pircs, K., et al., *Distinct subcellular autophagy impairments in induced neurons from Huntington's disease patients.* Brain, 2021.
594. Ramirez-Jarquin, U.N., et al., *Deletion of SUMO1 attenuates behavioral and anatomical deficits by regulating autophagic activities in Huntington disease.* Proc Natl Acad Sci U S A, 2022. **119**(5).
595. Kuo, S.H., et al., *Mutant glucocerebrosidase impairs alpha-synuclein degradation by blockade of chaperone-mediated autophagy.* Sci Adv, 2022. **8**(6): p. eabm6393.
596. Liu, P., et al., *Rab21 Protein Is Degraded by Both the Ubiquitin-Proteasome Pathway and the Autophagy-Lysosome Pathway.* Int J Mol Sci, 2022. **23**(3).
597. Manzoni, C., *LRRK2 and autophagy: a common pathway for disease.* Biochem Soc Trans, 2012. **40**(5): p. 1147-51.
598. Manzoni, C. and P.A. Lewis, *LRRK2 and Autophagy.* Adv Neurobiol, 2017. **14**: p. 89-105.
599. Luzon-Toro, B., et al., *Mechanistic insight into the dominant mode of the Parkinson's disease-associated G2019S LRRK2 mutation.* Hum Mol Genet, 2007. **16**(17): p. 2031-9.
600. Di Fonzo, A., et al., *A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease.* Lancet, 2005. **365**(9457): p. 412-5.
601. Kachergus, J., et al., *Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations.* Am J Hum Genet, 2005. **76**(4): p. 672-80.
602. Roosen, D.A. and M.R. Cookson, *LRRK2 at the interface of autophagosomes, endosomes and lysosomes.* Mol Neurodegener, 2016. **11**(1): p. 73.
603. Paisan-Ruiz, C., P.A. Lewis, and A.B. Singleton, *LRRK2: cause, risk, and mechanism.* J Parkinsons Dis, 2013. **3**(2): p. 85-103.
604. Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans.* Nature, 2020. **581**(7809): p. 434-443.

605. Trinh, J., I. Guella, and M.J. Farrer, *Disease penetrance of late-onset parkinsonism: a meta-analysis.* JAMA Neurol, 2014. **71**(12): p. 1535-9.
606. Reed, X., et al., *The role of monogenic genes in idiopathic Parkinson's disease.* Neurobiol Dis, 2019. **124**: p. 230-239.
607. Civiero, L., et al., *The role of LRRK2 in cytoskeletal dynamics.* Biochem Soc Trans, 2018. **46**(6): p. 1653-1663.
608. Blauwendraat, C., et al., *Frequency of Loss of Function Variants in LRRK2 in Parkinson Disease.* JAMA Neurol, 2018. **75**(11): p. 1416-1422.
609. Kluss, J.H., A. Mamais, and M.R. Cookson, *LRRK2 links genetic and sporadic Parkinson's disease.* Biochem Soc Trans, 2019. **47**(2): p. 651-661.
610. Heckman, M.G., et al., *Protective effect of LRRK2 p.R1398H on risk of Parkinson's disease is independent of MAPT and SNCA variants.* Neurobiol Aging, 2014. **35**(1): p. 266 e5-14.
611. Chen, L., et al., *LRRK2 R1398H polymorphism is associated with decreased risk of Parkinson's disease in a Han Chinese population.* Parkinsonism Relat Disord, 2011. **17**(4): p. 291-2.
612. Guerreiro, P.S., et al., *LRRK2 Promotes Tau Accumulation, Aggregation and Release.* Mol Neurobiol, 2016. **53**(5): p. 3124-3135.
613. Takanashi, M., et al., *Isolated nigral degeneration without pathological protein aggregation in autopsied brains with LRRK2 p.R1441H homozygous and heterozygous mutations.* Acta Neuropathol Commun, 2018. **6**(1): p. 105.
614. Henderson, M.X., et al., *Alzheimer's disease tau is a prominent pathology in LRRK2 Parkinson's disease.* Acta Neuropathol Commun, 2019. **7**(1): p. 183.
615. Ling, H., et al., *TDP-43 pathology in a patient carrying G2019S LRRK2 mutation and a novel p.Q124E MAPT.* Neurobiol Aging, 2013. **34**(12): p. 2889 e5-9.
616. Neumann, M., et al., *Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.* Science, 2006. **314**(5796): p. 130-3.
617. Arai, T., et al., *TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.* Biochem Biophys Res Commun, 2006. **351**(3): p. 602-11.
618. Zhao, Y.G. and H. Zhang, *Autophagosome maturation: An epic journey from the ER to lysosomes.* J Cell Biol, 2019. **218**(3): p. 757-770.
619. Khandia, R., et al., *A Comprehensive Review of Autophagy and Its Various Roles in Infectious, Non-Infectious, and Lifestyle Diseases: Current Knowledge and Prospects for Disease Prevention, Novel Drug Design, and Therapy.* Cells, 2019. **8**(7).
620. Lorincz, P. and G. Juhasz, *Autophagosome-Lysosome Fusion.* J Mol Biol, 2020. **432**(8): p. 2462-2482.
621. Dyshlovoy, S.A., *Blue-Print Autophagy in 2020: A Critical Review.* Mar Drugs, 2020. **18**(9).
622. Cheung, Y.W.S., S.E. Nam, and C.K. Yip, *Recent Advances in Single-Particle Electron Microscopic Analysis of Autophagy Degradation Machinery.* Int J Mol Sci, 2020. **21**(21).
623. Helikar, T., B. Kowal, and J.A. Rogers, *A cell simulator platform: the cell collective.* Clin Pharmacol Ther, 2013. **93**(5): p. 393-5.
624. Galluzzi, L., et al., *Molecular definitions of autophagy and related processes.* EMBO J, 2017. **36**(13): p. 1811-1836.
625. Malik-Sheriff, R.S., et al., *BioModels-15 years of sharing computational models in life science.* Nucleic Acids Res, 2020. **48**(D1): p. D407-D415.
626. Le Novere, N., et al., *BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems.* Nucleic Acids Res, 2006. **34**(Database issue): p. D689-91.
627. Nishimura, T. and S.A. Tooze, *Emerging roles of ATG proteins and membrane lipids in autophagosome formation.* Cell Discov, 2020. **6**(1): p. 32.

628. Mercer, T.J., A. Gubas, and S.A. Tooze, *A molecular perspective of mammalian autophagosome biogenesis.* J Biol Chem, 2018. **293**(15): p. 5386-5395.
629. Gomez-Sanchez, R., S.A. Tooze, and F. Reggiori, *Membrane supply and remodeling during autophagosome biogenesis.* Curr Opin Cell Biol, 2021. **71**: p. 112-119.
630. De Tito, S., et al., *The Golgi as an Assembly Line to the Autophagosome.* Trends Biochem Sci, 2020. **45**(6): p. 484-496.
631. Zavodszky, E., M. Vicinanza, and D.C. Rubinsztein, *Biology and trafficking of ATG9 and ATG16L1, two proteins that regulate autophagosome formation.* FEBS Lett, 2013. **587**(13): p. 1988-96.
632. Noda, T., *Autophagy in the context of the cellular membrane-trafficking system: the enigma of Atg9 vesicles.* Biochem Soc Trans, 2017. **45**(6): p. 1323-1331.
633. He, C. and D.J. Klionsky, *Atg9 trafficking in autophagy-related pathways.* Autophagy, 2007. **3**(3): p. 271-4.
634. Gomez-Suaga, P., et al., *Leucine-rich repeat kinase 2 regulates autophagy through a calcium-dependent pathway involving NAADP.* Hum Mol Genet, 2012. **21**(3): p. 511-25.
635. Noda, T. and Y. Ohsumi, *Tor, a phosphatidylinositol kinase homologue, controls autophagy in yeast.* J Biol Chem, 1998. **273**(7): p. 3963-6.
636. Moloney, P.B., G.L. Cavalleri, and N. Delanty, *Epilepsy in the mTORopathies: opportunities for precision medicine.* Brain Commun, 2021. **3**(4): p. fcab222.
637. Zhang, S., et al., *Regulation of mTORC1 by amino acids in mammalian cells: A general picture of recent advances.* Anim Nutr, 2021. **7**(4): p. 1009-1023.
638. Rehbein, U., et al., *The TSC Complex-mTORC1 Axis: From Lysosomes to Stress Granules and Back.* Front Cell Dev Biol, 2021. **9**: p. 751892.
639. Lu, X., et al., *Meta-analysis of the association between mTORC1-related genes polymorphisms and cancer risk.* Pathol Res Pract, 2022. **229**: p. 153696.
640. Mehrpour, M.B., J; Codogno, P, *Mechanisms and regulation of autophagy in mammalian cells.* Atlas Genet Cytogenet Oncol Haematol, 2011. **16**(2): p. 18.
641. Auger, C., et al., *Autophagy and Extracellular Vesicles in Colorectal Cancer: Interactions and Common Actors?* Cancers (Basel), 2021. **13**(5).
642. Gaudet, P., et al., *Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium.* Brief Bioinform, 2011. **12**(5): p. 449-62.
643. Cashikar, A.G., et al., *Structure of cellular ESCRT-III spirals and their relationship to HIV budding.* Elife, 2014. **3**.
644. Hurley, J.H. and P.I. Hanson, *Membrane budding and scission by the ESCRT machinery: it's all in the neck.* Nat Rev Mol Cell Biol, 2010. **11**(8): p. 556-66.
645. Morita, E., et al., *Human ESCRT-III and VPS4 proteins are required for centrosome and spindle maintenance.* Proc Natl Acad Sci U S A, 2010. **107**(29): p. 12889-94.
646. Bache, K.G., et al., *The ESCRT-III subunit hVps24 is required for degradation but not silencing of the epidermal growth factor receptor.* Mol Biol Cell, 2006. **17**(6): p. 2513-23.
647. Nazarko, T.Y., et al., *Trs85 is required for macroautophagy, pexophagy and cytoplasm to vacuole targeting in Yarrowia lipolytica and Saccharomyces cerevisiae.* Autophagy, 2005. **1**(1): p. 37-45.
648. Pahari, S., et al., *Induction of autophagy through CLEC4E in combination with TLR4: an innovative strategy to restrict the survival of Mycobacterium tuberculosis.* Autophagy, 2020. **16**(6): p. 1021-1043.
649. Lynch-Day, M.A., et al., *Trs85 directs a Ypt1 GEF, TRAPPIII, to the phagophore to promote autophagy.* Proc Natl Acad Sci U S A, 2010. **107**(17): p. 7811-6.
650. Lipatova, Z., et al., *Regulation of selective autophagy onset by a Ypt/Rab GTPase module.* Proc Natl Acad Sci U S A, 2012. **109**(18): p. 6981-6.
651. Pang, S.Y., et al., *LRRK2, GBA and their interaction in the regulation of autophagy: implications on therapeutics in Parkinson's disease.* Transl Neurodegener, 2022. **11**(1): p. 5.

652. MacLeod, D., et al., *The familial Parkinsonism gene LRRK2 regulates neurite process morphology.* Neuron, 2006. **52**(4): p. 587-93.
653. Plowey, E.D., et al., *Role of autophagy in G2019S-LRRK2-associated neurite shortening in differentiated SH-SY5Y cells.* J Neurochem, 2008. **105**(3): p. 1048-56.
654. Hernandez-Diaz, S., et al., *Endophilin-B regulates autophagy during synapse development and neurodegeneration.* Neurobiol Dis, 2022. **163**: p. 105595.
655. Bravo-San Pedro, J.M., et al., *The LRRK2 G2019S mutant exacerbates basal autophagy through activation of the MEK/ERK pathway.* Cell Mol Life Sci, 2013. **70**(1): p. 121-36.
656. Zhao, Y., et al., *Reduced LRRK2 in association with retromer dysfunction in post-mortem brain tissue from LRRK2 mutation carriers.* Brain, 2018. **141**(2): p. 486-495.
657. Murphy, K.E., et al., *Lysosomal-associated membrane protein 2 isoforms are differentially affected in early Parkinson's disease.* Mov Disord, 2015. **30**(12): p. 1639-47.
658. Henry, A.G., et al., *Pathogenic LRRK2 mutations, through increased kinase activity, produce enlarged lysosomes with reduced degradative capacity and increase ATP13A2 expression.* Hum Mol Genet, 2015. **24**(21): p. 6013-28.
659. Ramirez, A., et al., *Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase.* Nat Genet, 2006. **38**(10): p. 1184-91.
660. Purves, D., et al., *Neuroscience*. International sixth edition ed. 1 volume (various pagings).
661. Ivanisevic, J., et al., *Brain region mapping using global metabolomics.* Chem Biol, 2014. **21**(11): p. 1575-84.
662. Roosen, D.A., et al., *DNAJC proteins and pathways to parkinsonism.* FEBS J, 2019. **286**(16): p. 3080-3094.
663. Nascimbeni, A.C., P. Codogno, and E. Morel, *Phosphatidylinositol-3-phosphate in the regulation of autophagy membrane dynamics.* FEBS J, 2017. **284**(9): p. 1267-1278.
664. Proctor, C.J., M. Tsirigotis, and D.A. Gray, *An in silico model of the ubiquitin-proteasome system that incorporates normal homeostasis and age-related decline.* BMC Syst Biol, 2007. **1**: p. 17.
665. Hui, W., et al., *Oxidative changes and signalling pathways are pivotal in initiating age-related changes in articular cartilage.* Ann Rheum Dis, 2016. **75**(2): p. 449-58.
666. Dalle Pezze, P., et al., *A systems study reveals concurrent activation of AMPK and mTOR by amino acids.* Nat Commun, 2016. **7**: p. 13254.
667. Thoreen, C.C., et al., *An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1.* J Biol Chem, 2009. **284**(12): p. 8023-32.
668. Ping, L., et al., *Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease.* Sci Data, 2018. **5**: p. 180036.
669. Ramasamy, A., et al., *Genetic variability in the regulation of gene expression in ten regions of the human brain.* Nat Neurosci, 2014. **17**(10): p. 1418-1428.
670. Nakatogawa, H., et al., *Dynamics and diversity in autophagy mechanisms: lessons from yeast.* Nat Rev Mol Cell Biol, 2009. **10**(7): p. 458-67.
671. Franchin, C., et al., *Exploring the CK2 Paradox: Restless, Dangerous, Dispensable.* Pharmaceuticals (Basel), 2017. **10**(1).
672. Buontempo, F., et al., *Therapeutic targeting of CK2 in acute and chronic leukemias.* Leukemia, 2018. **32**(1): p. 1-10.
673. Lee, Y.S., et al., *Multitarget-directed cotreatment with cilostazol and aripiprazole for augmented neuroprotection against oxidative stress-induced toxicity in HT22 mouse hippocampal cells.* Eur J Pharmacol, 2019. **857**: p. 172454.
674. Kim, G.S., et al., *CK2 is a novel negative regulator of NADPH oxidase and a neuroprotectant in mice after cerebral ischemia.* J Neurosci, 2009. **29**(47): p. 14779-89.
675. Ampofo, E., et al., *Protein Kinase CK2-A Putative Target for the Therapy of Diabetes Mellitus?* Int J Mol Sci, 2019. **20**(18).

676. Husain, K., et al., *Protein kinase 2 (CK2): a potential regulator of immune cell development and function in cancer.* Immunol Med, 2021. **44**(3): p. 159-174.
677. Castello, J., et al., *CK2-An Emerging Target for Neurological and Psychiatric Disorders.* Pharmaceuticals (Basel), 2017. **10**(1).
678. Rosenberger, A.F., et al., *Increased occurrence of protein kinase CK2 in astrocytes in Alzheimer's disease pathology.* J Neuroinflammation, 2016. **13**: p. 4.
679. Waxman, E.A. and B.I. Giasson, *Specificity and regulation of casein kinase-mediated phosphorylation of alpha-synuclein.* J Neuropathol Exp Neurol, 2008. **67**(5): p. 402-16.
680. Tenreiro, S., K. Eckermann, and T.F. Outeiro, *Protein phosphorylation in neurodegeneration: friend or foe?* Front Mol Neurosci, 2014. **7**: p. 42.
681. Jang, D.E., et al., *Protein kinase CK2 activates Nrf2 via autophagic degradation of Keap1 and activation of AMPK in human cancer cells.* BMB Rep, 2020. **53**(5): p. 272-277.
682. Kim, J.E. and T.C. Kang, *CDDO-Me Attenuates Astroglial Autophagy via Nrf2-, ERK1/2-SP1- and Src-CK2-PTEN-PI3K/AKT-Mediated Signaling Pathways in the Hippocampus of Chronic Epilepsy Rats.* Antioxidants (Basel), 2021. **10**(5).
683. Park, J.W., J. Jeong, and Y.S. Bae, *Protein Kinase CK2 Is Upregulated by Calorie Restriction and Induces Autophagy.* Mol Cells, 2022. **45**(3): p. 112-121.
684. Richardson, C.D., et al., *Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.* Nat Biotechnol, 2016. **34**(3): p. 339-44.
685. Gijselinck, I., et al., *A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study.* Lancet Neurol, 2012. **11**(1): p. 54-65.
686. DeJesus-Hernandez, M., et al., *Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS.* Neuron, 2011. **72**(2): p. 245-56.
687. Zhang, Y., et al., *The C9orf72-interacting protein Smcr8 is a negative regulator of autoimmunity and lysosomal exocytosis.* Genes Dev, 2018. **32**(13-14): p. 929-943.
688. O'Rourke, J.G., et al., *C9orf72 is required for proper macrophage and microglial function in mice.* Science, 2016. **351**(6279): p. 1324-9.
689. McAlpine, W., et al., *Excessive endosomal TLR signaling causes inflammatory disease in mice with defective SMCR8-WDR41-C9ORF72 complex function.* Proc Natl Acad Sci U S A, 2018. **115**(49): p. E11523-E11531.
690. Corrionero, A. and H.R. Horvitz, *A C9orf72 ALS/FTD Ortholog Acts in Endolysosomal Degradation and Lysosomal Homeostasis.* Curr Biol, 2018. **28**(10): p. 1522-1535 e5.
691. Amick, J., A. Roczniak-Ferguson, and S.M. Ferguson, *C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling.* Mol Biol Cell, 2016. **27**(20): p. 3040-3051.
692. Sullivan, P.M., et al., *The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway.* Acta Neuropathol Commun, 2016. **4**(1): p. 51.
693. Xiao, S., et al., *C9orf72 isoforms in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration.* Brain Res, 2016. **1647**: p. 43-49.
694. Ugolino, J., et al., *Loss of C9orf72 Enhances Autophagic Activity via Deregulated mTOR and TFEB Signaling.* PLoS Genet, 2016. **12**(11): p. e1006443.
695. Sellier, C., et al., *Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death.* EMBO J, 2016. **35**(12): p. 1276-97.
696. Pang, W. and F. Hu, *Cellular and physiological functions of C9ORF72 and implications for ALS/FTD.* J Neurochem, 2021. **157**(3): p. 334-350.
697. Iyer, S., V. Subramanian, and K.R. Acharya, *C9orf72, a protein associated with amyotrophic lateral sclerosis (ALS) is a guanine nucleotide exchange factor.* PeerJ, 2018. **6**: p. e5815.

698. Yang, M., et al., *A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy.* Sci Adv, 2016. **2**(9): p. e1601167.
699. Amick, J., et al., *WDR41 supports lysosomal response to changes in amino acid availability.* Mol Biol Cell, 2018. **29**(18): p. 2213-2227.
700. Talaia, G., J. Amick, and S.M. Ferguson, *Receptor-like role for PQLC2 amino acid transporter in the lysosomal sensing of cationic amino acids.* Proc Natl Acad Sci U S A, 2021. **118**(8).
701. Amick, J., et al., *PQLC2 recruits the C9orf72 complex to lysosomes in response to cationic amino acid starvation.* J Cell Biol, 2020. **219**(1).
702. Ibanez, K., et al., *Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: a retrospective diagnostic accuracy and prospective clinical validation study.* Lancet Neurol, 2022. **21**(3): p. 234-245.
703. Koks, S., et al., *Longitudinal intronic RNA-Seq analysis of Parkinson's disease patients reveals disease-specific nascent transcription.* Exp Biol Med (Maywood), 2022. **247**(11): p. 945-957.
704. Shi, M.M., C.H. Shi, and Y.M. Xu, *Rab GTPases: The Key Players in the Molecular Pathway of Parkinson's Disease.* Front Cell Neurosci, 2017. **11**: p. 81.
705. Cogo, S., et al., *PKA-mediated phosphorylation of SPG11/spatacsin regulates binding with a subset of 14-3-3 proteins.* bioRxiv, 2020: p. 2020.09.09.289009.
706. Chavez, J.D., et al., *Multiplexed Cross-Linking with Isobaric Quantitative Protein Interaction Reporter Technology.* Anal Chem, 2021. **93**(50): p. 16759-16768.
707. Claussnitzer, M., et al., *FTO Obesity Variant Circuitry and Adipocyte Browning in Humans.* N Engl J Med, 2015. **373**(10): p. 895-907.
708. Liu, Y., A. Beyer, and R. Aebersold, *On the Dependency of Cellular Protein Levels on mRNA Abundance.* Cell, 2016. **165**(3): p. 535-50.
709. Papachristou, E.K., et al., *A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes.* Nat Commun, 2018. **9**(1): p. 2311.
710. Arul, A.B. and R.A.S. Robinson, *Sample Multiplexing Strategies in Quantitative Proteomics.* Anal Chem, 2019. **91**(1): p. 178-189.
711. Vistain, L.F. and S. Tay, *Single-Cell Proteomics.* Trends Biochem Sci, 2021. **46**(8): p. 661-672.
712. Christopher, J.A., et al., *Subcellular proteomics.* Nat Rev Methods Primers, 2021. **1**.
713. Kim, C., *iPSC technology--Powerful hand for disease modeling and therapeutic screen.* BMB Rep, 2015. **48**(5): p. 256-65.
714. Filippi-Chiela, E.C., et al., *Single-cell analysis challenges the connection between autophagy and senescence induced by DNA damage.* Autophagy, 2015. **11**(7): p. 1099-113.
715. Xu, Y., J. Yuan, and M.M. Lipinski, *Live imaging and single-cell analysis reveal differential dynamics of autophagy and apoptosis.* Autophagy, 2013. **9**(9): p. 1418-30.
716. Hu, Y., et al., *Single-cell RNA sequencing highlights transcription activity of autophagy-related genes during hematopoietic stem cell formation in mouse embryos.* Autophagy, 2017. **13**(4): p. 770-771.
717. Tindall, M.J. and A. Clerk, *Modelling negative feedback networks for activating transcription factor 3 predicts a dominant role for miRNAs in immediate early gene regulation.* PLoS Comput Biol, 2014. **10**(5): p. e1003597.
718. Dunster, J.L., et al., *Regulation of Early Steps of GPVI Signal Transduction by Phosphatases: A Systems Biology Approach.* PLoS Comput Biol, 2015. **11**(11): p. e1004589.

Appendix A

A1.

In order to show how a variable (e.g., protein concentration, Y) changes over time (t), two graphs can be made: (i) $Y(t)$, in which the y-axis is visualising the concentration of Y, and the x-axis the time, or (ii) $\frac{dY}{dt}(t)$, in which the y-axis shows the rate of change of the concentration of Y, and the x-axis the time. For example, if X is made into Y with a constant rate (named α) until the timepoint t_1, and then from t_1 and on Y is made into Z, also with a constant rate (named β), the corresponding graphs would look like those of Fig 1-4A and 1-4B, respectively. In order to graph the system correctly, the initial values of the variables of the model, which are called initial conditions, are also needed. In this example, the initial concentration of Y and Z (at $t = 0s$, $Y(t) = Y(0) = Y_0$, $Z(t) = Z(0) = Z_0$) are assumed to be equal to 0.

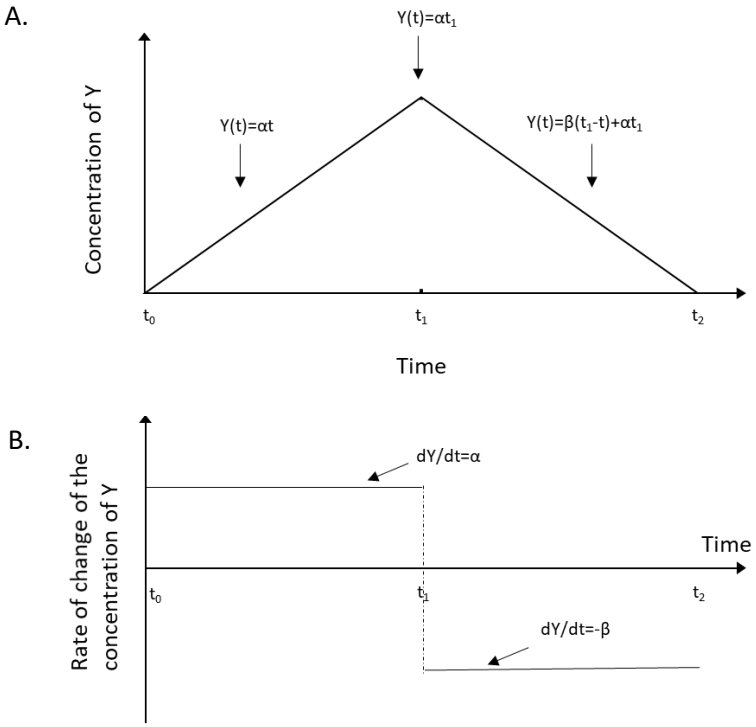


Figure S1-1. Graphs describing the change of the concentration of Y

*An initial amount of X is transformed into Y until t_1, and then into Z until t_2. (**A**) A graph of the concentration of Y over time. (**B**) A graph of the rate of change of the concentration of Y over time.*

In each model, it is important to identify the components. The model of Fig 1-4 is composed of 3 dependent variables (X, Y & Z, depend on time), an independent variable (time) and 2 parameters

(the rates α and β) and the initial conditions (at t=0s: $X=X_0$, Y=Z=0). The dynamics of Y can be described by 2 ODEs ($\frac{dY}{dt} = at$, for $0 \leq t < t_1$; and $\frac{dY}{dt} = -\beta t$, for $t_1 \leq t < t_2$).

A2.

Reactions with 2 substrates are described with a slightly different form of the previously mentioned differential expression. For instance, for the reaction: $X_1 + X_2 \rightarrow X_3$, the ODE is:

$$\frac{dX_3}{dt} = k_3 X_1 X_2,$$

with k_3 the rate of the creation of X_3.

Note that the equation is not expressed as:

$$\frac{dX_3}{dt} = k_3 X_1 + k_3 X_2,$$

because the two molecules need to come into physical contact with each other to then form X_3, which adds the effect of probability of this event into the system that needs to be accounted for.

Similarly to the previous example, the equation describing the concentration of the substrates over time, are:

$$\frac{dX_1}{dt} = \frac{dX_2}{dt} = -k_3 X_1 X_2,$$

There is also the case of bimolecular reactions, in which a product is made out of two molecules of the same substrate. Then:

Reaction: $2X_1 \rightarrow X_3$

ODEs: $\frac{dX_3}{dt} = k_3 X_1 X_1 = k_3 X_1^2$, $and\ \frac{dX_1}{dt} = -2k_3 X_1^2$.

As expected X_1 is used up in double the speed as X_3 is being produced, as two molecules of X_1 are required for the generation of one molecule of X_3.

A3.

The necessity of computational solutions can be demonstrated through the identification of the value of x for:

$$e^x - 4x = 0$$

Even though when visualised through a graph it is evident that there are two solutions and they can be estimated based on the values of the axes (Fig S1-2), there is a lack of an algebraic method for computing the solutions in an exact manner. So, the combination of mathematical analysis and computer methods can be insightful.

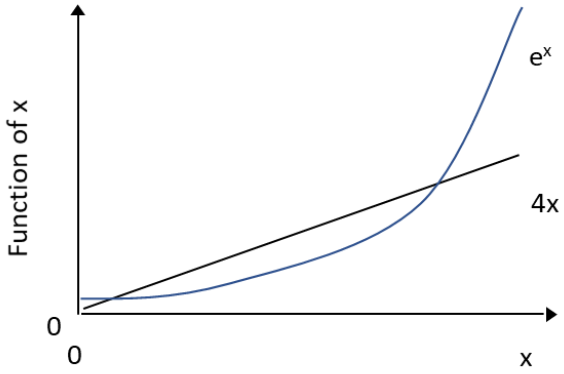


Figure S1-2. ***Solving the equation e^x-4x=0 graphically***

Appendix B

Table S2-1. The gene (or locus) and protein name responsible for each HSP type			
Gene name (or locus)	**Protein name/ previous name**	**UniProt identifier (SwissProt)**	**HSP type**
L1CAM	neural cell adhesion molecule L1	P32004	SPG1
PLP1	myelin proteolipid protein	P60201	SPG2
ATL1	atlastin-1	Q8WXF7	SPG3A
SPAST	spastin	Q9UBP0	SPG4
CYP7B1	25-hydroxycholesterol 7-alpha-hydroxylase	O75881	SPG5A
NIPA1	magnesium transporter NIPA1	Q7RTP0	SPG6
SPG7	paraplegin	Q9UQ90	SPG7
WASHC5	WASH complex subunit 5 (strumpellin)	Q12768	SPG8
ALDH18A1	delta-1-pyrroline-5-carboxylate synthase	P54886	SPG9A/SPG9B
KIF5A	kinesin heavy chain isoform 5A	Q12840	SPG10
SPG11	spatacsin	Q96JI7	SPG11
RTN2	reticulon-2	O75298	SPG12
HSPD1	60 kDa heat shock protein, mitochondrial	P10809	SPG13
(3q27-q28)			SPG14
ZFYVE26	zinc finger FYVE domain-containing protein 26 (spastizin)	Q68DK2	SPG15
(Xq11.2)			SPG16
BSCL2	Seipin	Q96G97	SPG17
ERLIN2	erlin-2	O94905	SPG18/SPG37
(9q33-q34)			SPG19
SPART	Spartin	Q8N0X7	SPG20
SPG21	maspardin	Q9NZD8	SPG21
SLC16A2	monocarboxylate transporter 8	P36021	SPG22
DSTYK	dual serine/threonine and tyrosine protein kinase	Q6XUX3	SPG23

Table S2-1. (continued) The gene (or locus) and protein name responsible for each HSP type			
(13q14)			SPG24
(6q23-24.1)			SPG25
B4GALNT1	beta-1,4 N-acetylgalactosaminyltransferase 1	Q00973	SPG26
(10q22.1-q24.1)			SPG27
DDHD1	phospholipase DDHD1	Q8NEL9	SPG28
(1p31.1-p21.1)			SPG29
KIF1A	kinesin-like protein KIF1A	Q12756	SPG30
REEP1	receptor expression-enhancing protein 1	Q9H902	SPG31
(14q12-q21)			SPG32
ZFYVE27	protrudin	Q5T4F4	SPG33
(Xq24-q25)			SPG34
FA2H	fatty acid 2-hydroxylase	Q7L5A8	SPG35
(12q23-q24)			SPG36
(4p16-p15)			SPG38
PNPLA6	neuropathy target esterase	Q8IY17	SPG39
(11p14.1-p11.2)			SPG41
SLC33A1	acetyl-coenzyme A transporter 1	O00400	SPG42
C19orf12	protein C19orf12	Q9NSK7	SPG43
GJC2	gap junction gamma-2 protein	Q5T442	SPG44
NT5C2	cytosolic purine 5'-nucleotidase	P49902	SPG45/SPG65
GBA2	non-lysosomal glucosylceramidase	Q9HCG7	SPG46
AP4B1	AP-4 complex subunit beta-1	Q9Y6B7	SPG47
AP5Z1	AP-5 complex subunit zeta-1	O43299	SPG48
TECPR2	tectonin beta-propeller repeat-containing protein 2	O15040	SPG49
AP4M1	AP-4 complex subunit mu-1	O00189	SPG50

Table S2-1. (continued) The gene (or locus) and protein name responsible for each HSP type			
AP4E1	AP-4 complex subunit epsilon-1	Q9UPM8	SPG51
AP4S1	AP-4 complex subunit sigma-1	Q9Y587	SPG52
VPS37A	vacuolar protein sorting-associated protein 37A	Q8NEZ2	SPG53
DDHD2	phospholipase DDHD2	O94830	SPG54
C12orf65	probable peptide chain release factor C12orf65, mitochondrial	Q9H3J6	SPG55
CYP2U1	cytochrome P450 2U1	Q7Z449	SPG56
TFG	protein TFG	Q92734	SPG57
KIF1C	kinesin-like protein KIF1C	O43896	SPG58
USP8	ubiquitin carboxyl-terminal hydrolase 8	P40818	SPG59
WDR48	WD repeat-containing protein 48	Q8TAF3	SPG60
ARL6IP1	ADP-ribosylation factor-like protein 6-interacting protein 1	Q15041	SPG61
ERLIN1	erlin-1	O75477	SPG62
AMPD2	AMP deaminase 2	Q01433	SPG63
ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1	P49961	SPG64
ARSI	arylsulfatase I	Q5FYB1	SPG66
PGAP1	GPI inositol-deacylase	Q75T13	SPG67
KLC2	kinesin light chain 2	Q9H0B6	SPG68
RAB3GAP2	rab3 GTPase-activating protein non-catalytic subunit	Q9H2M9	SPG69
MARS	methionine--tRNA ligase, cytoplasmic	P56192	SPG70
ZFR	zinc finger RNA-binding protein	Q96KR1	SPG71
REEP2	receptor expression-enhancing protein 2	Q9BRK0	SPG72
CPT1C	carnitine O-palmitoyltransferase 1, brain isoform	Q8TCG5	SPG73
IBA57	putative transferase CAF17, mitochondrial	Q5T440	SPG74
MAG	myelin-associated glycoprotein	P20916	SPG75

Table S2-1. (continued) The gene (or locus) and protein name responsible for each HSP type			
CAPN1	calpain-1 catalytic subunit	P07384	SPG76
FARS2	phenylalanine--tRNA ligase, mitochondrial	O95363	SPG77
ATP13A2	cation-transporting ATPase 13A2	Q9NQ11	SPG78
UCHL1	ubiquitin carboxyl-terminal hydrolase isozyme L1	P09936	SPG79
UBAP1	ubiquitin-associated protein 1	Q9NZ09	SPG80
TPP1	tripeptidyl-peptidase 1	O14773	-
Note: *Adapted from [403].*			

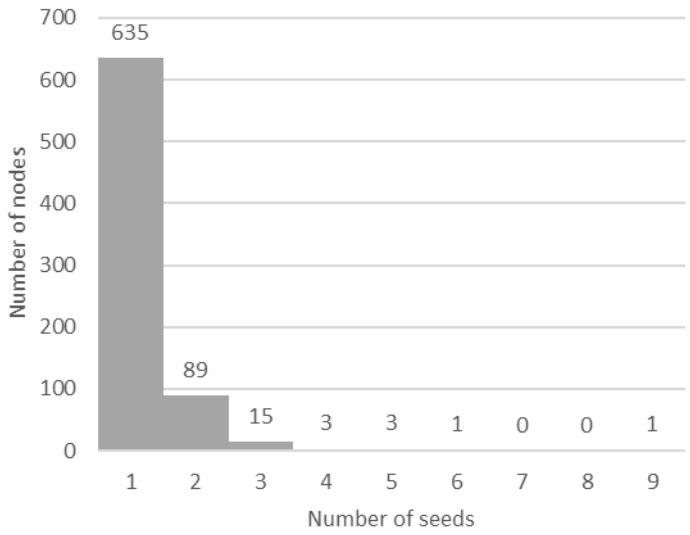


Figure S2-1. Degree distribution of the proteins of the global HSP-PPIN based on their connectivity with seeds

The nodes of the global HSP-PPIN were analysed to calculate the number of seeds to which they connect. The nodes connected to one seed (635/746, 85.0%) were excluded from further analyses. Adapted from [403].

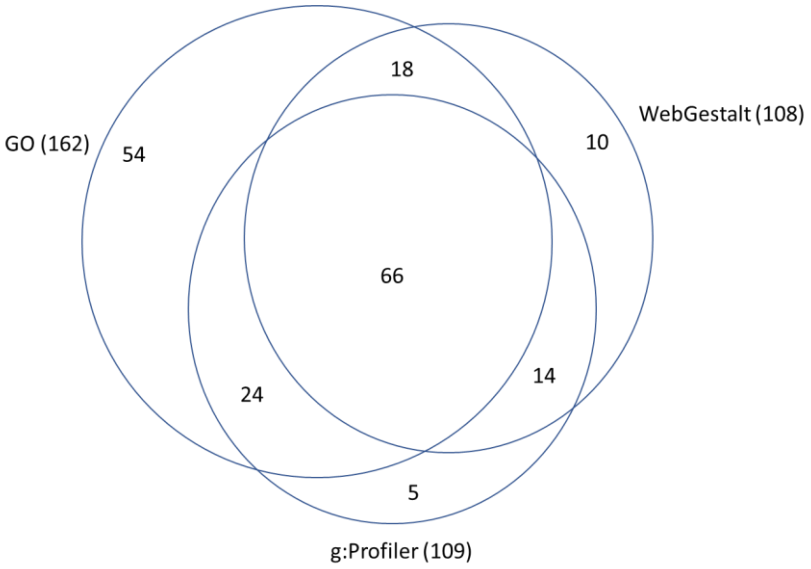


Figure S2-2. Overlap of the three functional enrichment tools for the analysis of the core HSP-PPIN in the level of GO-BP terms

The single GO terms resulted from the analysis of the core HSP-PPIN were compared across all the used functional enrichment tools. Most were common in at least two tools (122, 63.8%).

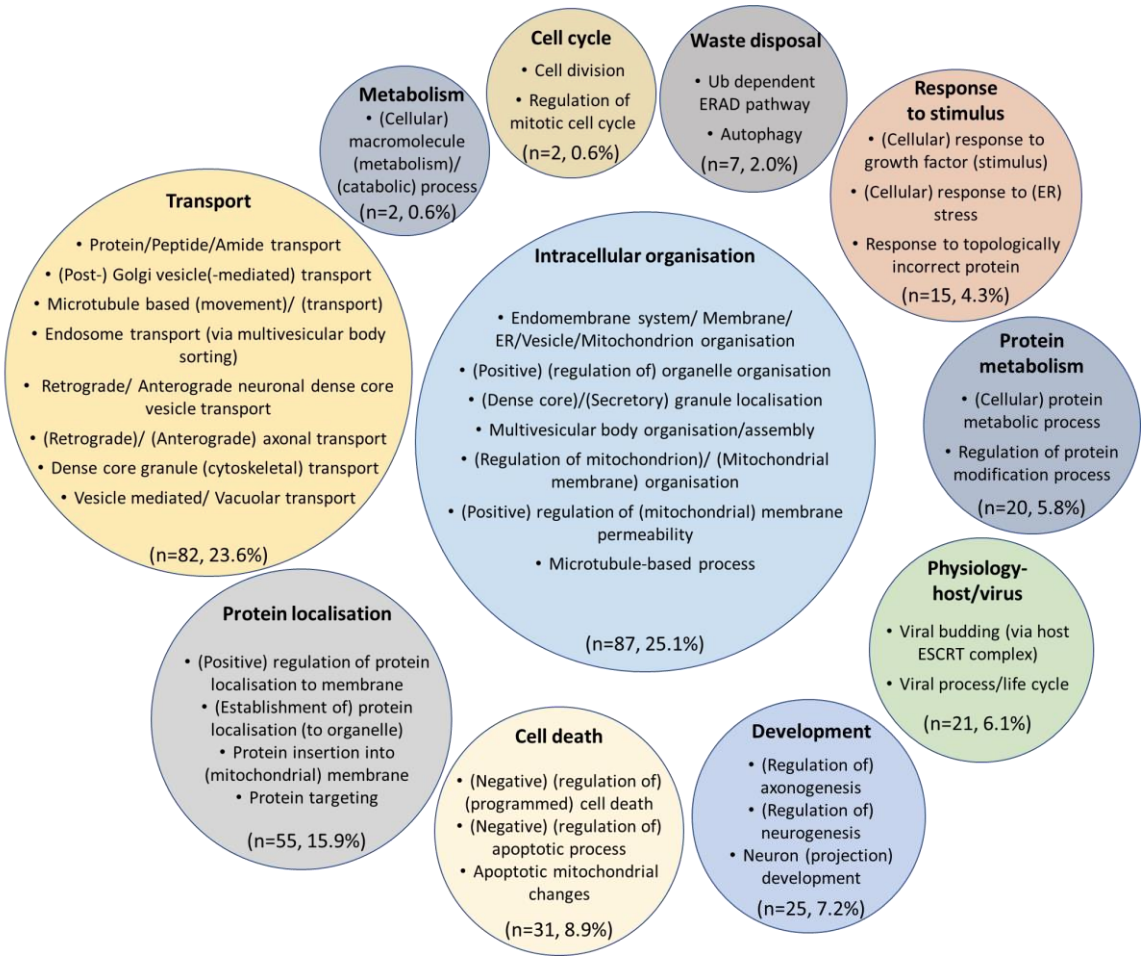
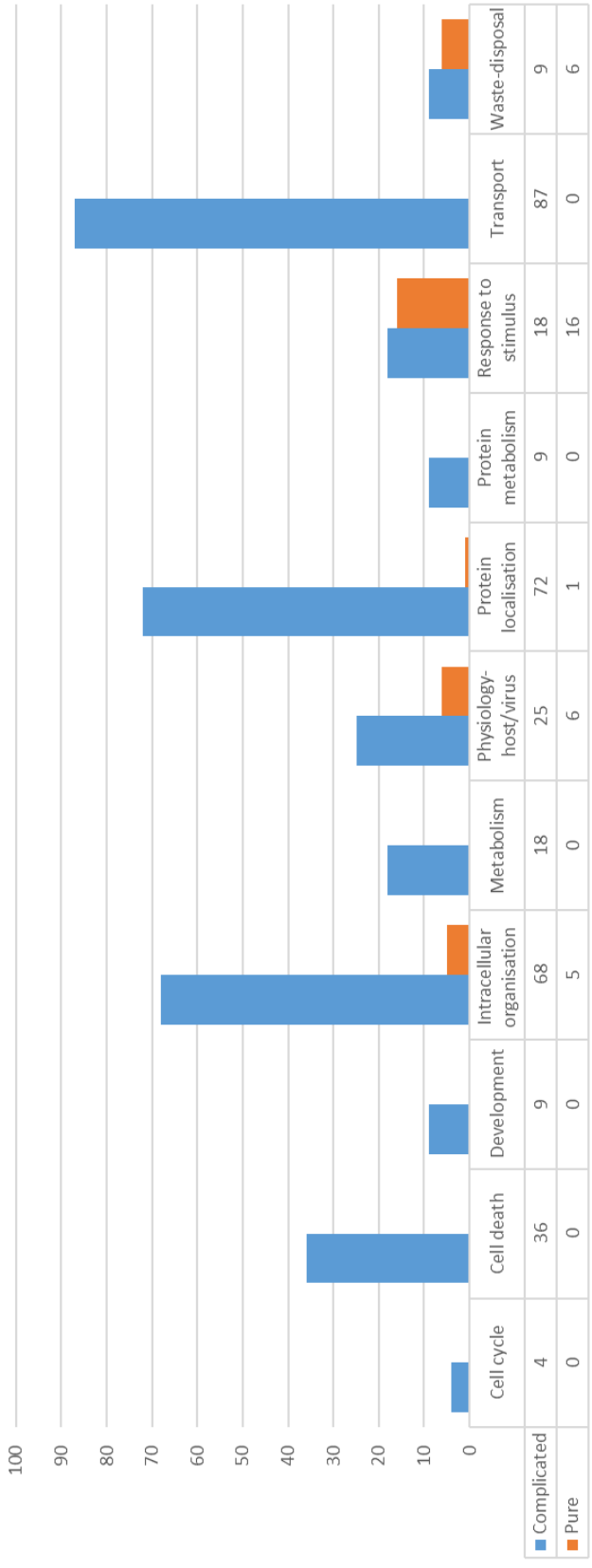


Figure S2-3. Detailed graphical representation of the functional enrichment of the core HSP-PPIN

Functional enrichment was performed for the components of the core HSP-PPIN. The resulted GO terms (n=379) were grouped into level 1 groups (n=54; excluding "metabolism") using in-house R script and then into level 2 groups (n=11; excluding "general"). The number of GO terms of each level 2 was calculated from g:Profiler, WebGestalt, and GO (through -PANTHER) using the GO terms of semantic classes that were present in at least 2 tools, and their distribution is shown here. Examples of GO terms are included for each functional block.

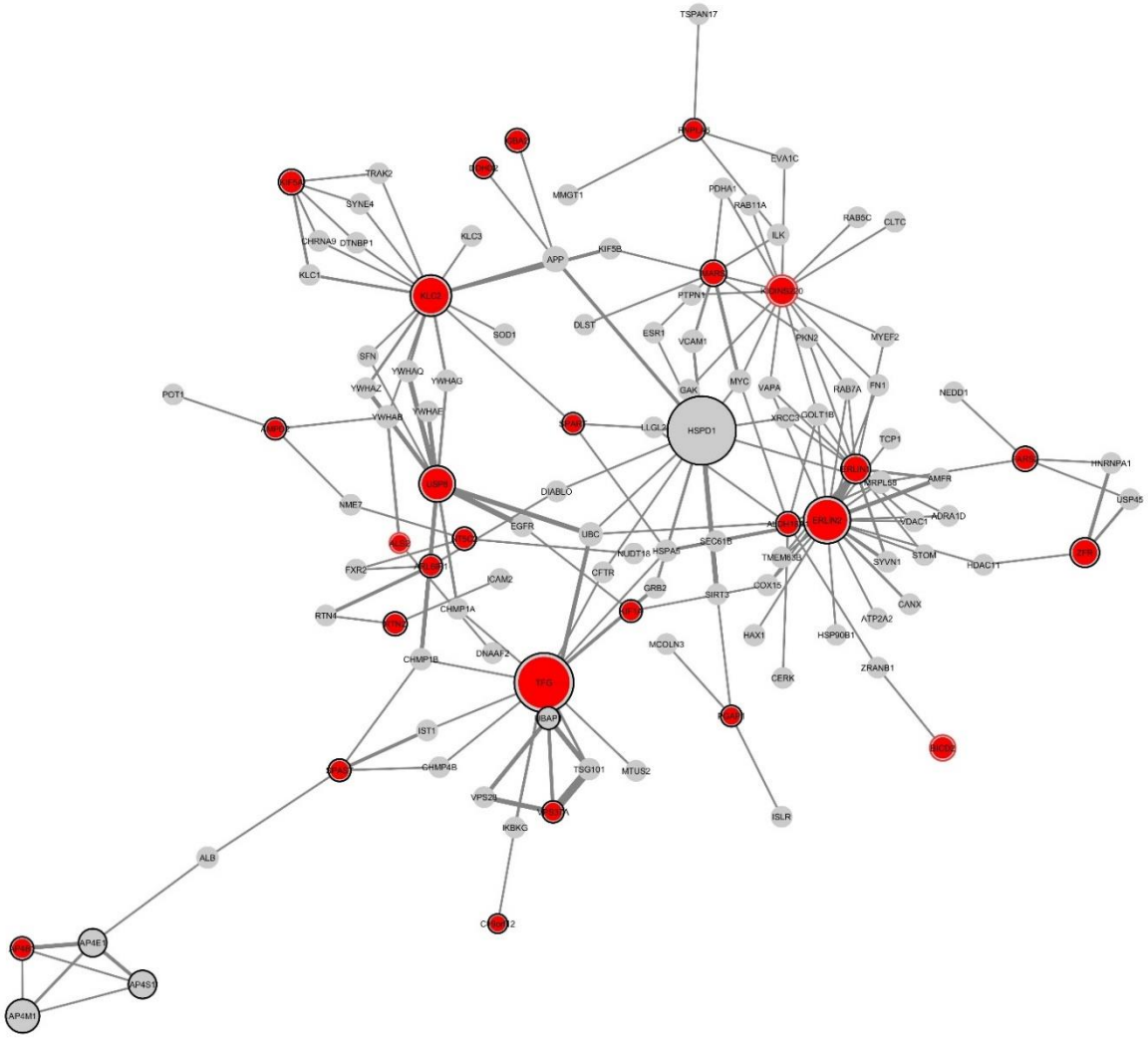
Appendix C



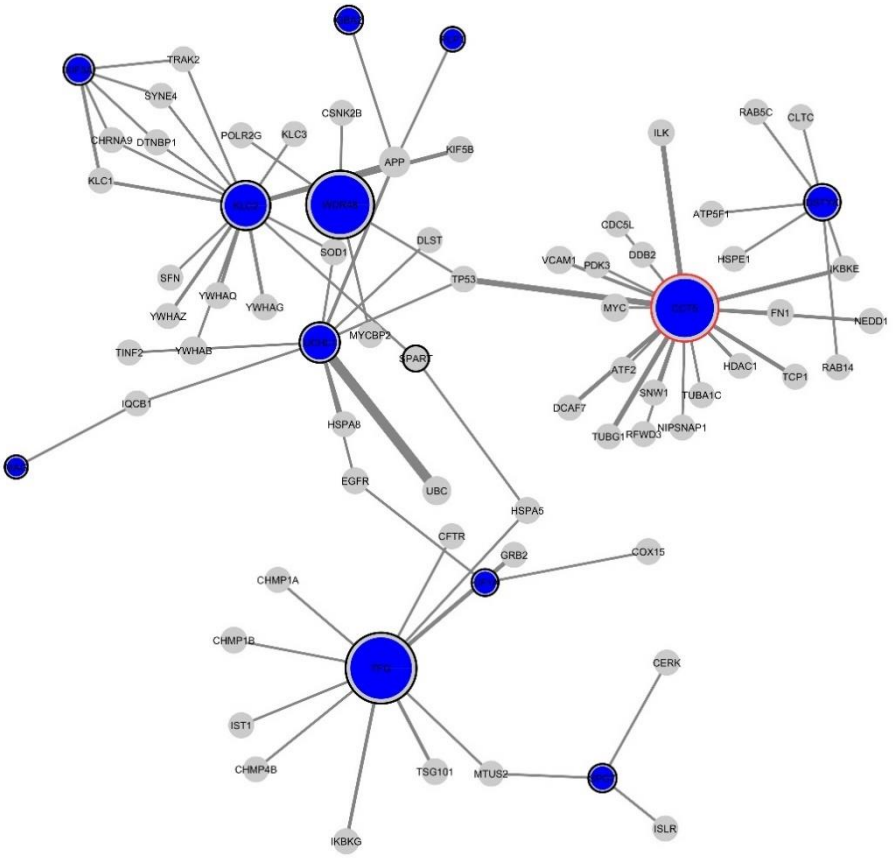
	Cell cycle	Cell death	Development	Intracellular organisation	Metabolism	Physiology-host/virus	Protein localisation	Protein metabolism	Response to stimulus	Transport	Waste-disposal
Complicated	4	36	9	68	18	25	72	9	18	87	9
Pure	0	0	0	5	0	6	1	0	16	0	6

Figure S3-1. Comparison of GO-BP terms between complicated and pure form of HSPs based on their number

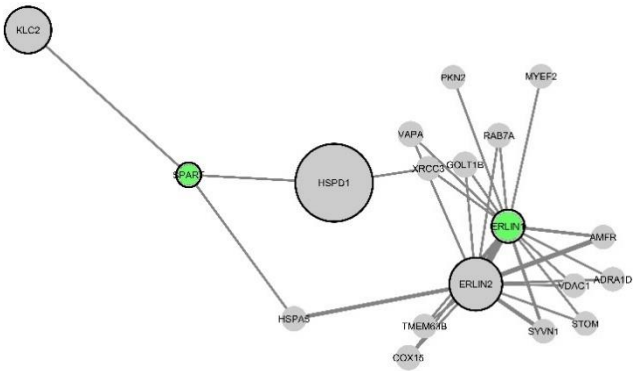
A.


TSPAN17
EVA1C
MMGT1
PDHA1
RAB11A
RAB5C
CLTC
ILK
TRAK2
SYNE4
CHRNA9
DTNBP1
KLC3
KLC1
APP
KIF5B
PTPN1
SOD1
DLST
ESR1
VCAM1
PKN2
MYEF2
SFN
YWHAQ
YWHAZ
YWHAG
YWHAB
YWHAE
POT1
GAK
MYC
VAPA
RAB7A
FN1
NEDD1
XRCC5
GOLT1B
HSPD1
LLGL2
TCP1
HNRNPA1
AMFR
VRPL55
USP45
DIABLO
NME7
EGFR
UBC
VDAC1
ADRA1D
SEC61B
HSPA5
MUDT18
TMEM135
SYVN1
STOM
HDAC11
FXR2
CFTR
ICAM2
GRB2
SIRT3
COX15
CANX
RTN4
CHMP1A
ATP2A2
HAX1
HSP90B1
MCOLN3
DNAAF2
CHMP1B
CERK
ZRANB1
IST1
CHMP4B
TSG101
MTUS2
VPS28
IKBKG
ISLR
ALB
AP4E1
AP4S1
AP4M1

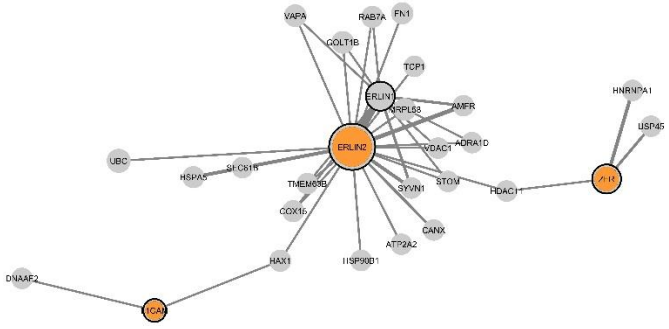
B.


TRAK2
SYNE4
CSNK2B
CHRNA9
DTNBP1
POLR2G
KLC3
APP
KIF5B
KLC1
SOD1
DLST
SFN
YWHAQ
YWHAZ
YWHAG
TP53
MYC
FN1
NEDD1
TINF2
YWHAB
MYCBP2
SPART
IQCB1
HSPA8
EGFR
UBC
HSPA5
CFTR
GRB2
COX15
CHMP1A
CHMP1B
CERK
IST1
CHMP4B
TSG101
MTUS2
ISLR
IKBKG
RAB5C
CLTC
ILK
ATP5F1
CDC5L
VCAM1
PDK3
DDB2
HSPE1
IKBKE
ATF2
HDAC1
TCP1
RAB14
SNW1
TUBA1C
DCAF7
TUBG1
RFWD3
NIPSNAP1

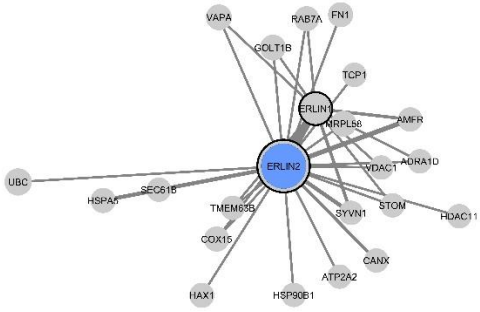
C.


KLC2
SPART
HSPD1
PKN2
MYEF2
VAPA
RAB7A
XRCC3
GOLT1B
ERLIN1
AMFR
ERLIN2
VDAC1
ADRA1D
HSPA5
TMEM63B
SYVN1
STOM
COX15

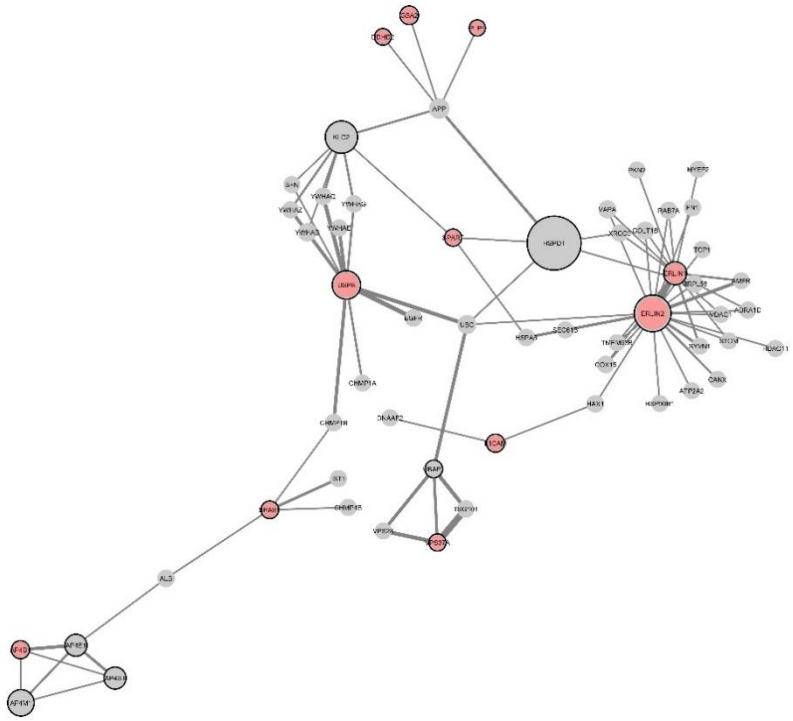
D.


VAPA
RAB7A
FN1
GOLT1B
TCP1
ERLIN1
AMFR
NRPL58
ADRA1D
VDAC1
UBC
ERLIN2
HSPA5
SEC61B
TMEM63B
SYVN1
STOM
HDAC11
HNRNPA1
USP45
COX15
CANX
ATP2A2
HAX1
HSP90B1
DNAAF2
MCAM

E.


VAPA
RAB7A
FN1
GOLT1B
TCP1
ERLIN1
AMFR
NRPL58
ADRA1D
VDAC1
UBC
ERLIN2
HSPA5
SEC61B
TMEM63B
SYVN1
STOM
HDAC11
COX15
CANX
ATP2A2
HAX1
HSP90B1

F.



G.

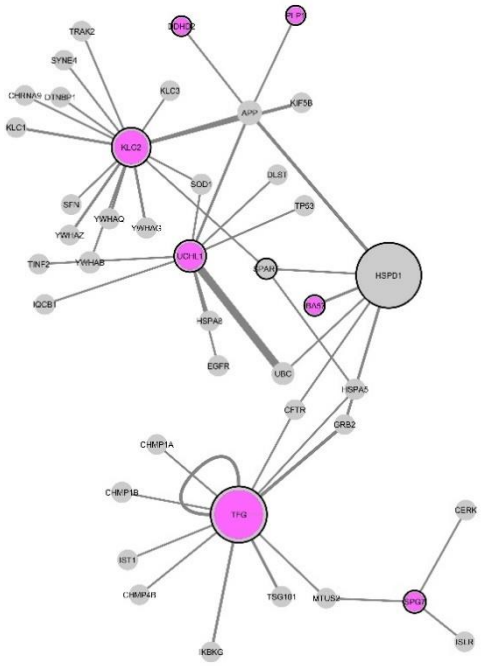
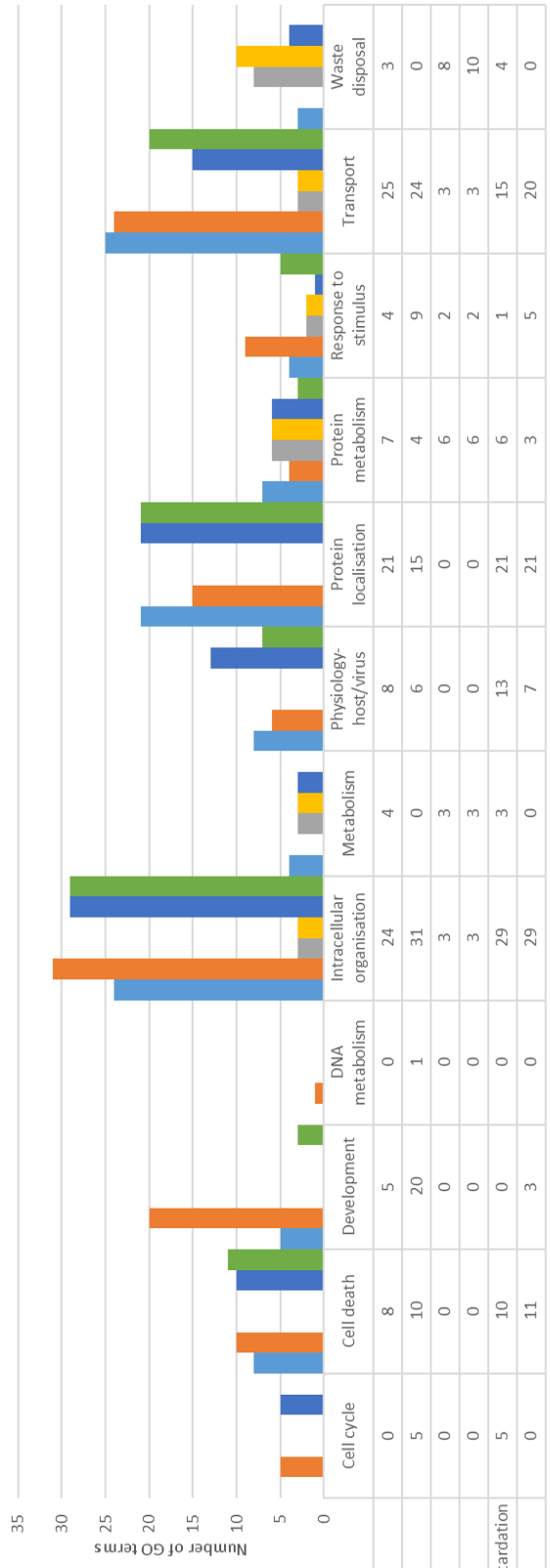


Figure S3-2. Comparison of the core network of seeds with the analysed clinical features

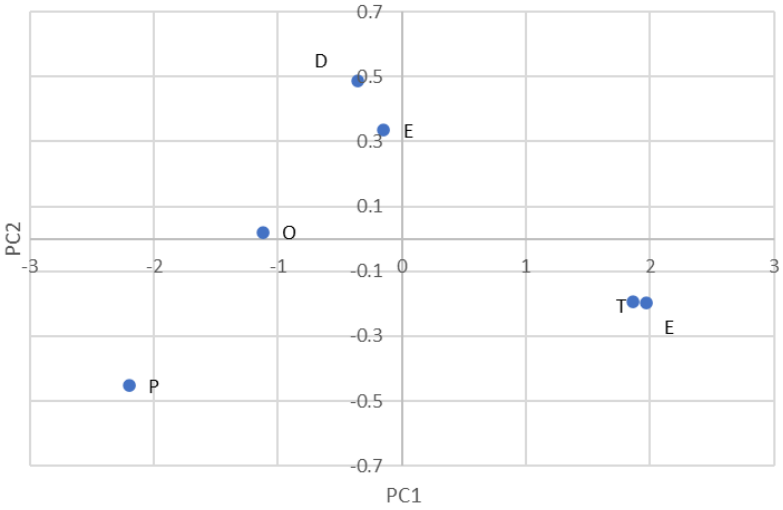
*The presence of clinical characteristics in HSPs is visualised in the core networks by the colour of each node for early onset (**A**), peripheral neuropathy (**B**), motor neuropathy (**C**), thin corpus callosum (**D**), seizures (**E**), dementia or mental retardation (**F**) and optic atrophy (**G**). The nodes corresponding to the HSP-seeds have a black border, while the test-seeds have a red border. The size of each node correlates with its degree and all pictures are of the same scale. The thickness of each edge correlates with its final score as calculated by PINOT. The network was visualised using Cytoscape.*



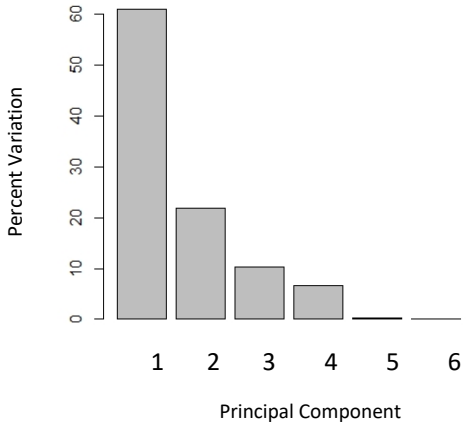
	Cell cycle	Cell death	Development	DNA metabolism	Intracellular organisation	Metabolism	Physiology-host/virus	Protein localisation	Protein metabolism	Response to stimulus	Transport	Waste disposal
Early onset	0	8	5	0	24	4	8	21	7	4	25	3
Peripheral neuropathy	5	10	20	1	31	0	6	15	4	9	24	0
Thin corpus callosum	0	0	0	0	3	3	0	0	6	2	3	8
Seizures	0	0	0	0	3	3	0	0	6	2	3	10
Dementia or mental retardation	5	10	0	0	29	3	13	21	6	1	15	4
Optical atrophy	0	11	3	0	29	0	7	21	3	5	20	0

Figure S3-3. Comparison of the functional enrichment results of the clinical subnetworks using the number of the GO terms of each functional group

A.



B.



C.

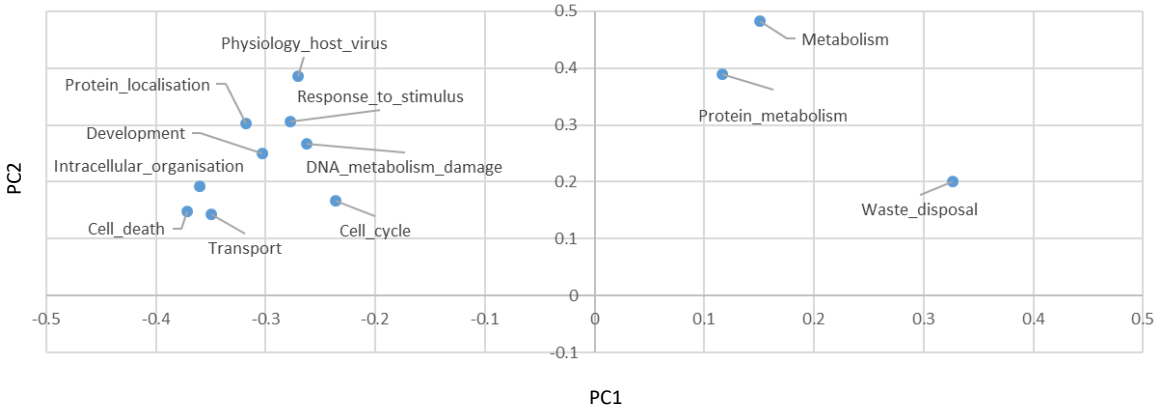


Figure S3-4. Comparison of the functional profiles of the 6 subdivisions of the HSP-PPIN based on clinical features using Principal Component Analysis

*The number of GO-BP terms for each functional group were analysed with PCA through R. (**A**) The PCA graph showing the distribution of the gene groups in the PC1 and PC2 axes. (**B**) The scree plot is showing the explained variation of the data for PC1 to PC6. (**C**) The loading scores of each variable (here functional groups) are plotted against PC1 and PC2, indicating which functions drive the localisation of the gene groups in the PCA graph. In (A) and (C) the PC axes were transformed to equate their significance.* ***Adapted from [403].***

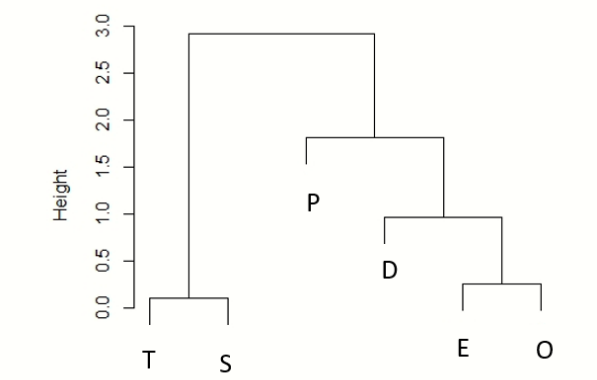


Figure S3-5. Cluster dendrogram for the number of GO-terms in the enrichment of the clinical subnetworks following PCA

Cluster dendrogram produced based on hierarchical clustering of the gene groups of the clinical subnetworks using R. Adapted from [403].

A. B.

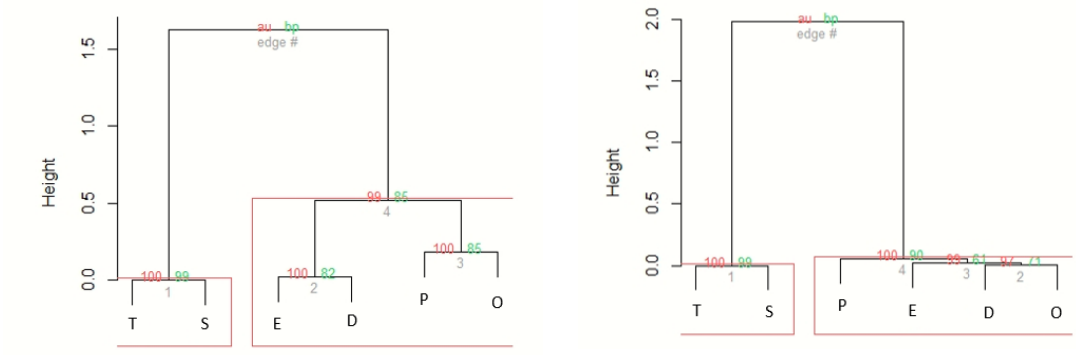


Figure S3-6. Evaluating the optimal number of clusters using the multiscale bootstrap resampling

*The analysis was based on the number (**A**) and percentage (**B**) of the GO terms in functional block. The graphs based on the multiscale bootstrap resampling the recommended clusters are framed in red boxes. The analysis was performed through R using the function pvclust()..Adapted from [403].*

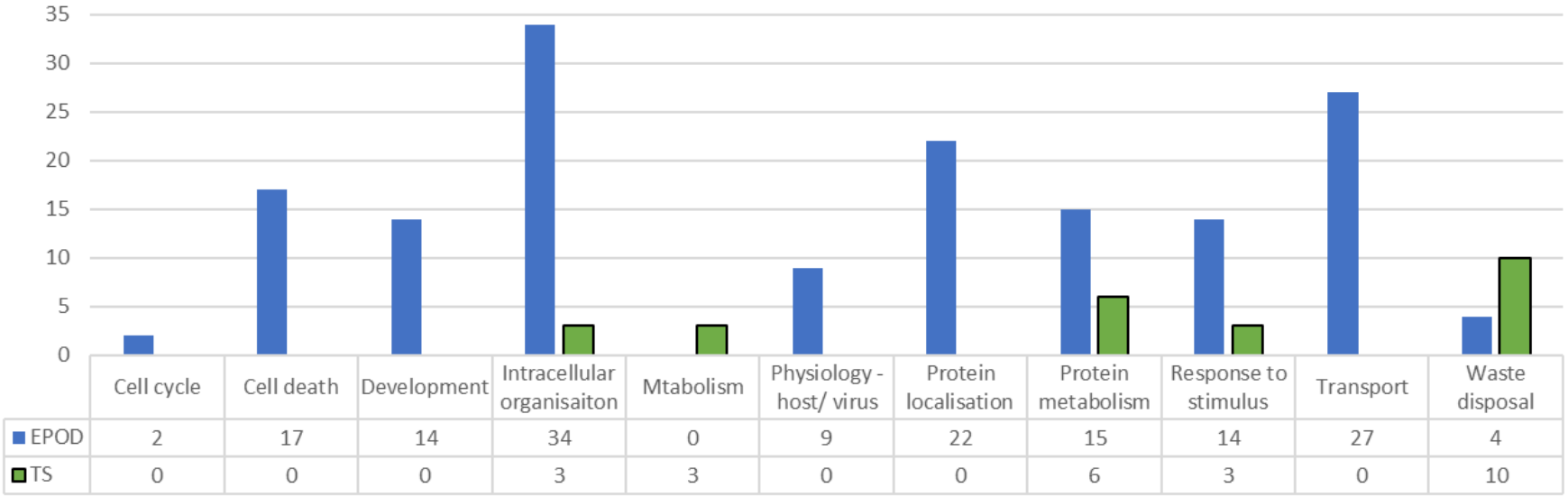
Table S3-1. The distance index resulted from the comparison of the co-occurrence of clinical features in HSP patients

	E	P	T	E	D	O
E						
P	0.84					
T	1.76	2.59				
E	2.06	3.00	0.30			
D	0.62	1.55	1.00	1.23		
O	0.99	0.45	2.49	2.87	1.70	

Note: The smaller the value of the distance index of a pair of clinical features, the more they tend to co-occur in HSP patients

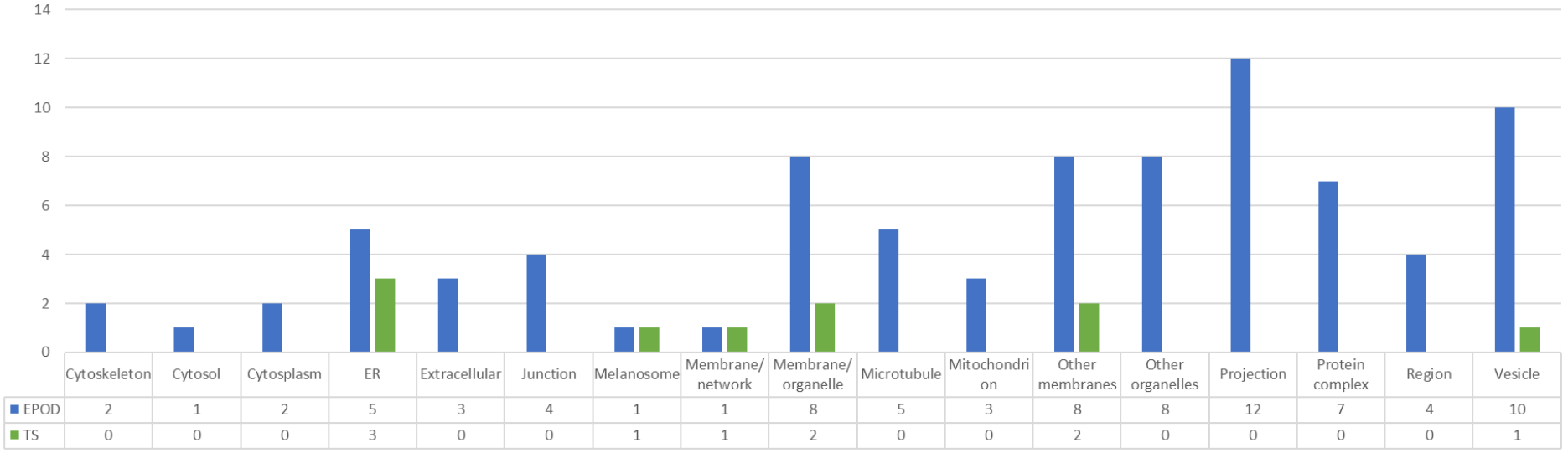
Table S3-2. p values from the Pearson correlation analysis for the co-occurrence of clinical features in HSP patients.

	E	P	T	E	D	O
E						
P	0.347473786					
T	0.877808201	0.875683027				
E	0.866587401	0.712382365	0.000271037			
D	0.866597665	0.282766009	0.172178712	0.231163467		
O	0.1602981	0.000977093	0.725008956	0.56306008	0.402797979	



	Cell cycle	Cell death	Development	Intracellular organisaiton	Mtabolism	Physiology - host/ virus	Protein localisation	Protein metabolism	Response to stimulus	Transport	Waste disposal
EPOD	2	17	14	34	0	9	22	15	14	27	4
TS	0	0	0	3	3	0	0	6	3	0	10

Figure S3-7. Comparison of the functional enrichment profiles of EPOD and TS using the number of the GO-BP terms in each functional group

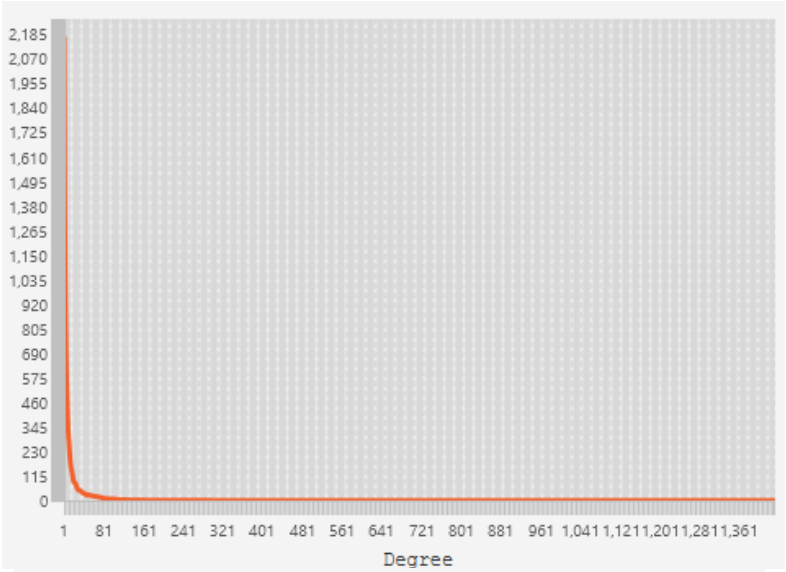


	Cytoskeleton	Cytosol	Cytosplasm	ER	Extracellular	Junction	Melanosome	Membrane/ network	Membrane/ organelle	Microtubule	Mitochondrion	Other membranes	Other organelles	Projection	Protein complex	Region	Vesicle
EPOD	2	1	2	5	3	4	1	1	8	5	3	8	8	12	7	4	10
TS	0	0	0	3	0	0	1	1	2	0	0	2	0	0	0	0	1

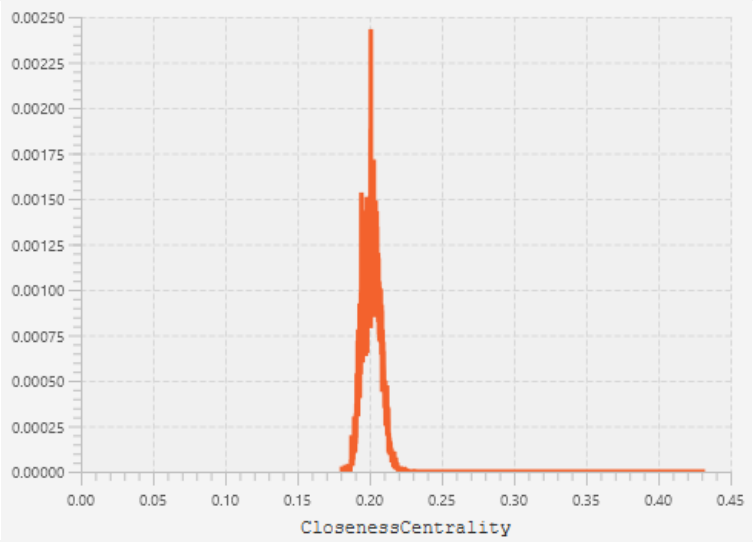
Figure S3-8. Comparison of the location enrichment profiles of EPOD and TS using the number of the GO-CC terms in each location group

Appendix D

A.



B.



C.

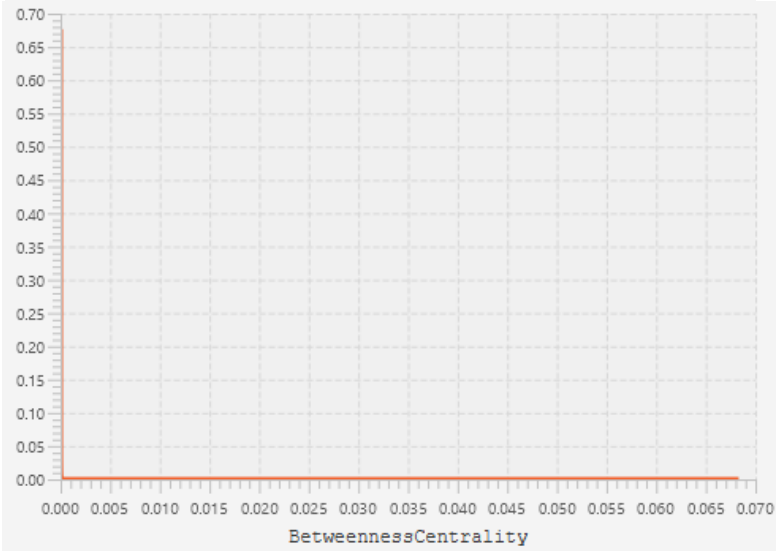
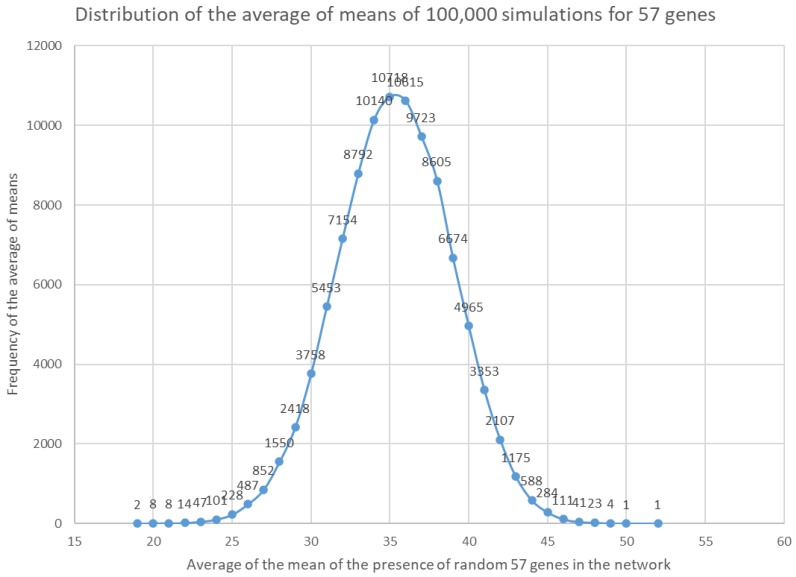


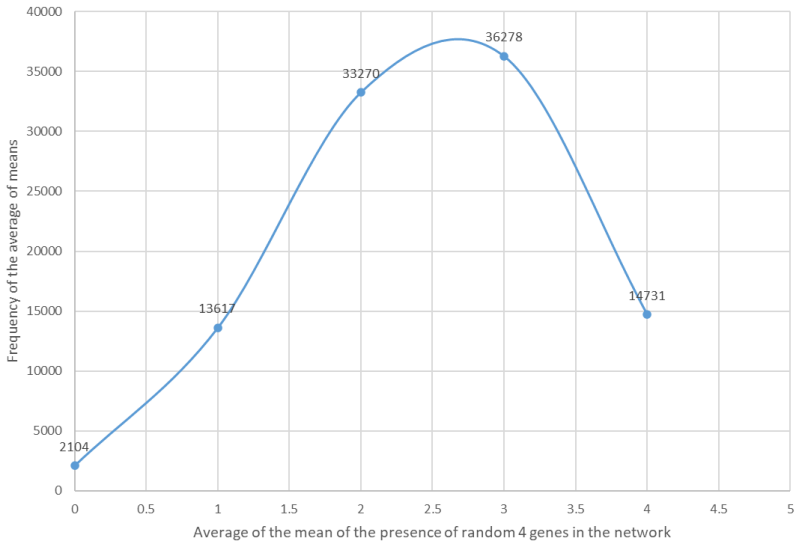
Figure S2-1. Centralities of the macroautophagy network

The node degree (A), closeness centrality (B), and betweenness centrality (C) were calculated for the 2nd layer of the macroautophagy network. The calculation was made through Cytoscape. Image produced using Cytoscape.

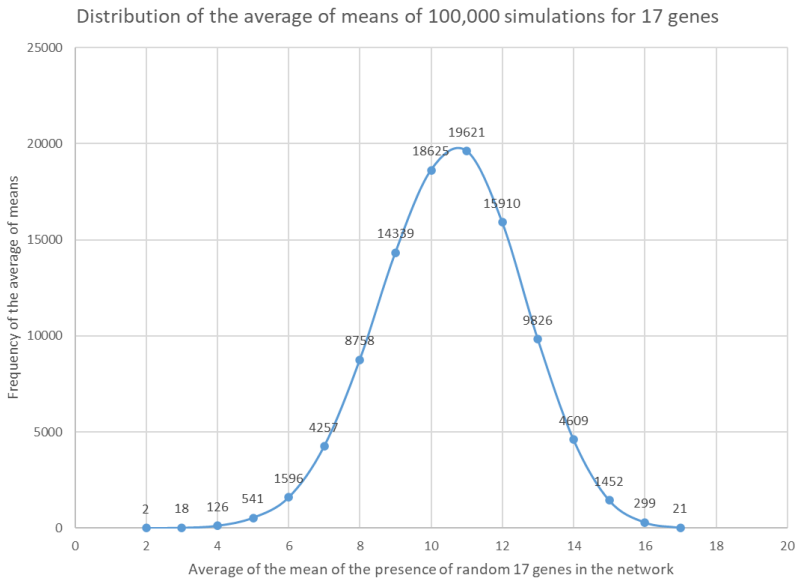
A.



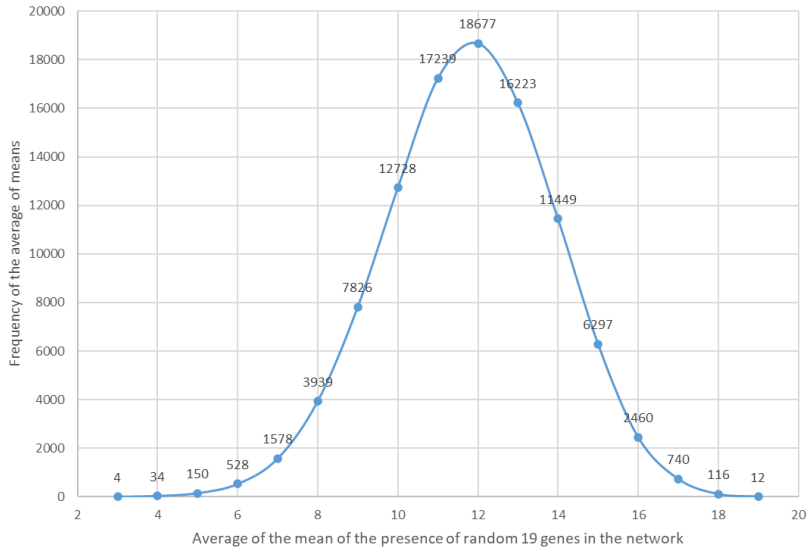
B.



C.


Distribution of the average of means of 100,000 simulations for 17 genes
Frequency of the average of means
Average of the mean of the presence of random 17 genes in the network
2
18
126
541
1596
4257
8758
14339
19621
18625
15910
9826
4609
1452
299
21

D.


Distribution of the average of means of 100,000 simulations for 19 genes
Frequency of the average of means
Average of the mean of the presence of random 19 genes in the network
4
34
150
528
1578
3939
7826
12728
18677
17239
16223
11449
6297
2460
740
116
12

E.

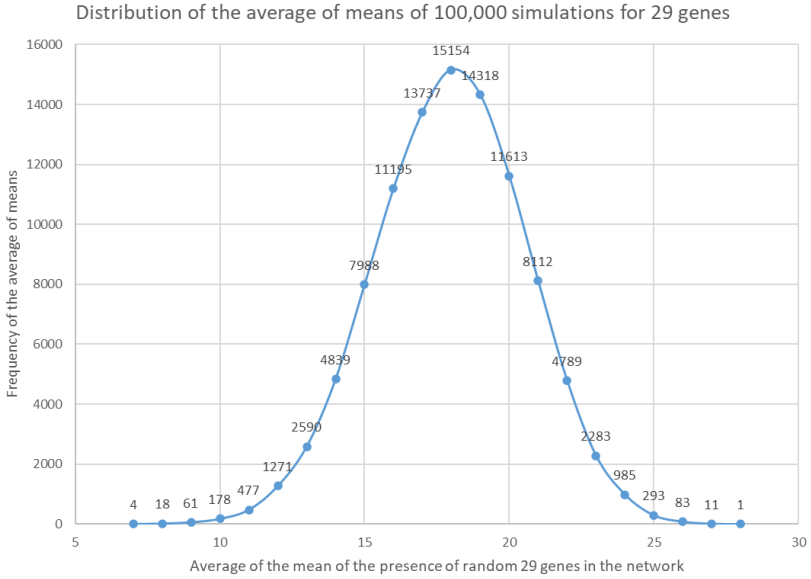

Distribution of the average of means of 100,000 simulations for 29 genes
Frequency of the average of means
Average of the mean of the presence of random 29 genes in the network
4
18
61
178
477
1271
2590
4839
7988
11195
15154
14318
13737
11613
8112
4789
2283
985
293
83
11
1

Figure S4-2. Analysis of the overlap of the neurodegenerative seeds with the macroautophagy network

The distributions were produced by randomly picking a list of genes -of equal length to each neurodegenerative disease seed list-, and evaluating their presence in the macroautophagy network. This process was repeated 100,000 times and produced the distribution of the overlap for the whole ND list (A), AD (B), FTD(C), PD (D), and ALS (E). ND: neurodegenerative disease, AD: Alzheimer's disease, FTD: Frontotemporal Dementia, ALS: Amyotrophic Lateral Sclerosis

Table S4-1. Connectivity of interactors of ND seeds with the MA network

ND seed	Interactors	N MA seeds connected through two proteins	Gene name of connected MA seeds
ANG (n=6/7, 85.7%)	ACTN2	**9**	DYNC1H1, DYNC1I1, HDAC6, HMGB1, NBR1, OPTN, RAB8B, SQSTM1, TOMM70
	ANXA2	**55**	ATG3, ATG9A, BCL2, CALCOCO2, CAMKK2, CAPN1, CAPNS1, CHMP2A, CHMP4B, CLTC, CSNK2A1, CSNK2B, DAPK3, DYNC1H1, DYNLL1, DYNLL2, FYCO1, HDAC6, HMGB1, HSPA8, HTT, ITPR1, KEAP1, MAP1B, MAP1LC3A, MAPK10, MAPK8, MAPK9, NEDD4, PARK7, PGAM5, PIK3CB, PIK3R2, PRKAB1, PRKACA, PRKN, RAB1A, RAB7A, RAB8B, SH3GLB1, SNAP29, SQSTM1, STK11, TBC1D17, TMEM173, TOMM20, TOMM22, TOMM40, TRIM21, UBQLN1, VAMP8, VCP, VDAC1, WDFY3, WIPI2
	FST	**0**	-
	PCNA	**70**	AMBRA1, ATG101, ATG16L1, BCL2, BCL2L1, BECN1, CALCOCO2, CLTC, CSNK2A1, CSNK2A2, CSNK2B, DAPK1, DAPK3, DYNC1H1, DYNC1I1, DYNC1I2, DYNC1LI1, DYNC1LI2, DYNLL1, FYCO1, GABARAP, GABARAPL1, GABARAPL2, HDAC6, HMGB1, HSPA8, HTT, KEAP1, LAMP2, MAP1B, MAPK10, MAPK15, MAPK8, MAPK9, MFN2, MTMR3, NBR1, NEDD4, NRBF2, OPTN, PARK7, PGAM5, PIK3R2, PINK1, PLAA, PRKAA1, PRKAB1, PRKAB2, PRKACA, PRKAG1, PRKAG2, PRKN, RAB7A, RB1CC1, RUBCN, SH3GLB1, SNAP29, SQSTM1, TBC1D17, TOMM22, TOMM40, TP53INP1, TRIM21, UBQLN1, USP10, VCP, VDAC1, VPS18, WIPI1, YOD1
	PLAUR	**13**	HDAC10, HDAC6, HSPA8, KEAP1, LAMP2, PIK3R4, PRKAB1, SQSTM1, TOMM22, TOMM5, UBQLN1, UBQLN2, WDFY3

Table S4-1. (continued) Connectivity of interactors of ND seeds with the MA network			
	RNH1	**24**	ATG4C, BCL2, CALCOCO2, CLTC, CSNK2A1, CSNK2B, DYNC1H1, DYNLL1, FYCO1, HDAC6, HMGB1, HSPA8, MAP1B, MAPK8, PGAM5, PRKN, RAB1A, RAB7A, SQSTM1, TRIM21, UBQLN4, UBXN6, ULK1, USP10
	S100A10	**22**	ATG9A, BCL2, BCL2L1, CAMKK2, CHMP4B, CLTC, CSNK2A1, DAPK3, HDAC6, HSPA8, MAP1LC3B2, MAPK10, MAPK8, MAPK9, NEDD4, PIK3CB, PIK3R2, PRKACA, RAB7A, RAB8B, SQSTM1, VDAC1
FIG4 (n=4/4, 100%)	ANK1	7	NBR1, OPTN, PIK3CB, SQSTM1, TOMM70, VPS18, ZFYVE1
	PIKFYVE	2	KEAP1, PRKAG2
	SNX27	12	ATG3, CALCOCO2, CAPN1, HSPA8, MAP1B, MAPK15, MAPK8, MAPK9 NEDD4, SQSTM1, TBC1D5, TRIM21
	VAC14	9	ATG9A, CALCOCO2, CHMP6, DYNLL1, HSPA8, MFN1, OPTN, PRKACA, TMEM173
PRPH2	0	-	-
RAB38 (n=1/1, 100%)	RAB32	22	CISD2, CLTC, DAPK1, DYNC1H1, DYNLL1, HSPA8, LAMP2, MAP1B, MAP1LC3A, MFN2, PRKACA, PRKAG3, PRKN, RAB1A, RAB7A, RAB7B, RAB8B, SNAP29, SQSTM1, TOMM22, and TOMM40

Table S4-2 Distribution of the ND network in the MA networks (absolute values)				
		MA		
		Seeds	1st layer	2nd layer
AD	Seeds	0	3	4
	1st layer	10	155	465
	2nd layer	86	2089	6274
ALS	Seeds	4	14	26
	1st layer	43	603	1092
	2nd layer	93	2222	8833
FTD	Seeds	4	10	16
	1st layer	35	476	691
	2nd layer	96	2272	8405
PD	Seeds	5	13	19
	1st layer	33	499	787
	2nd layer	96	2219	8406

Appendix E

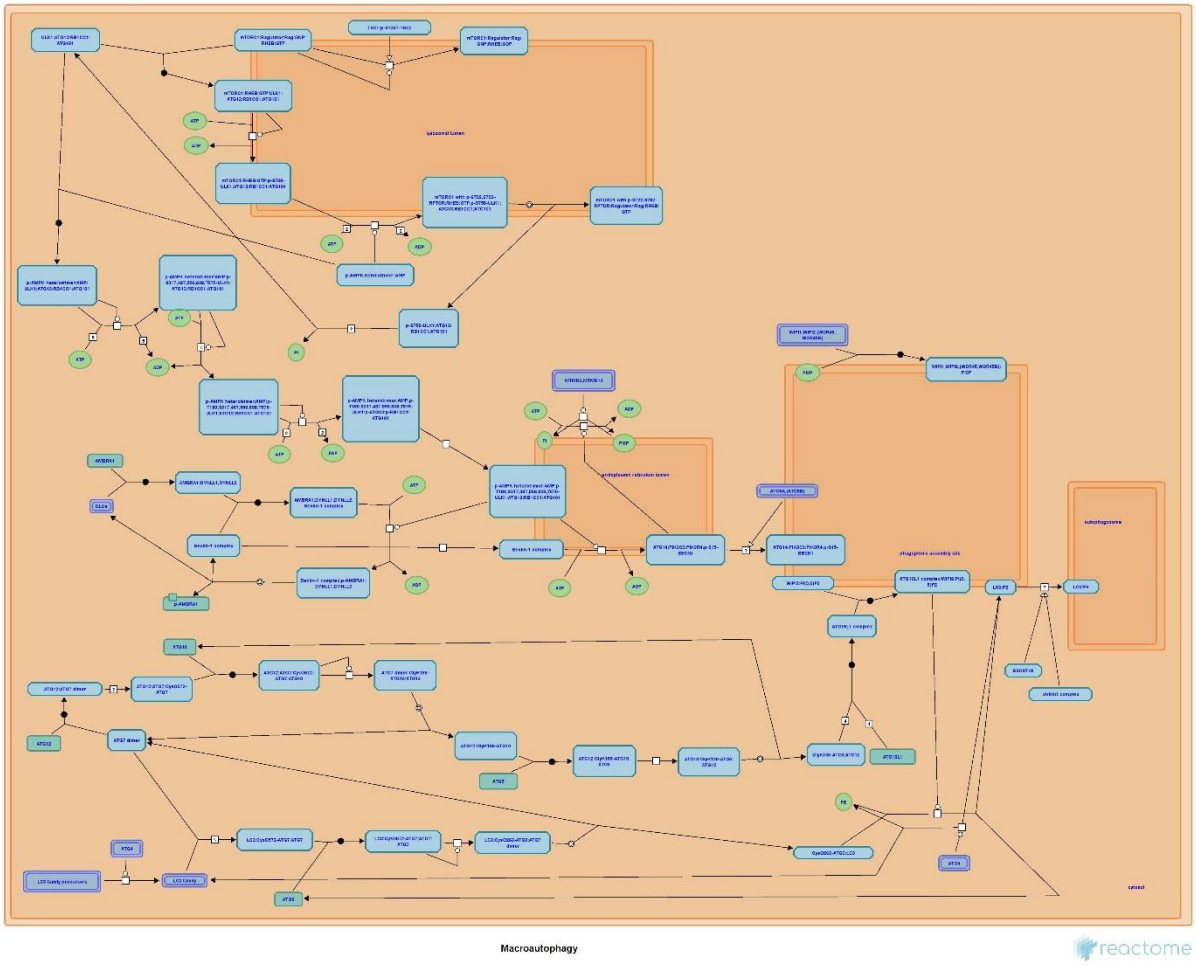


Figure S5-1. Macroautophagy diagram of Reactome

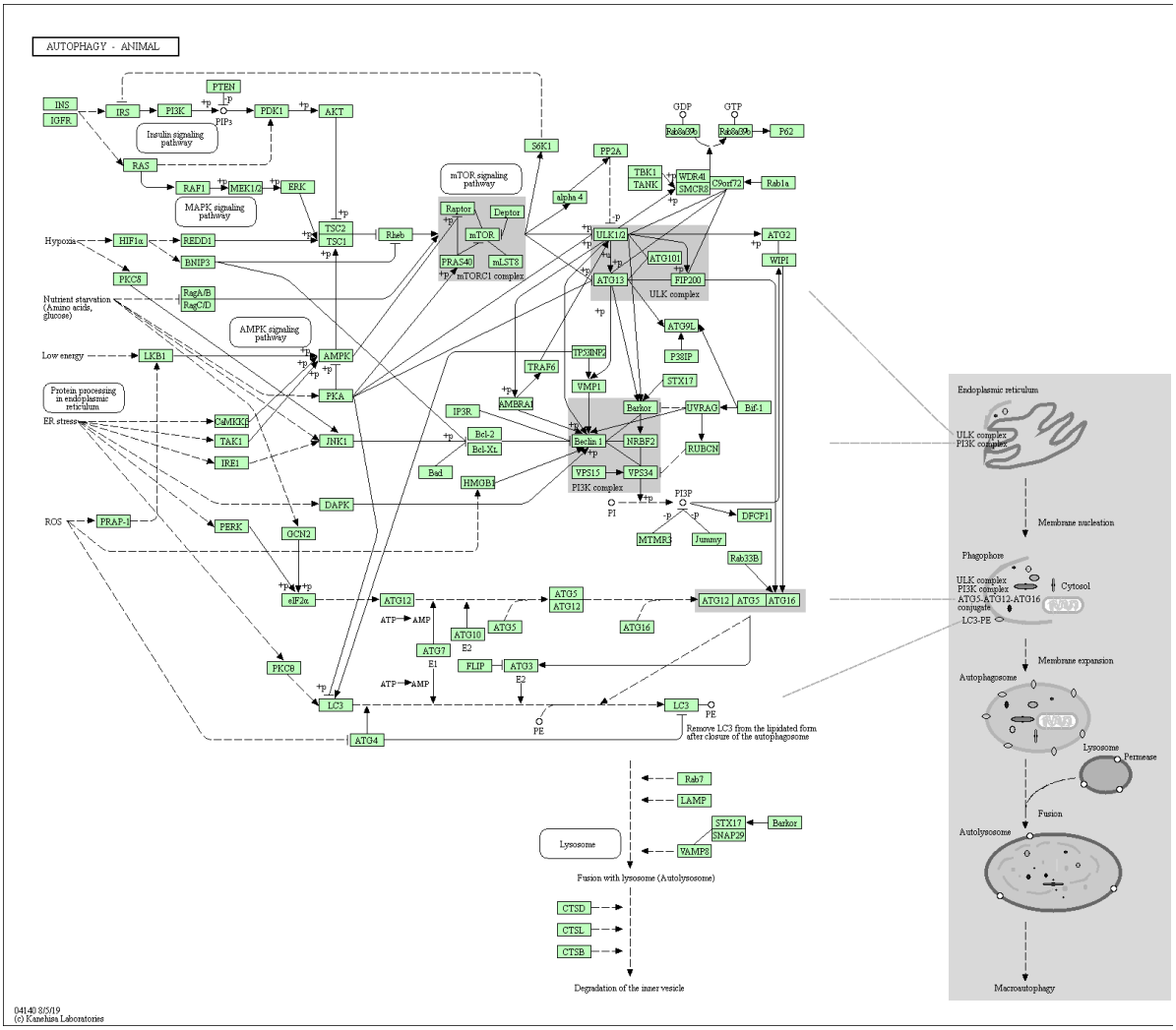


Figure S5-2. Macroautophagy diagram of KEGG

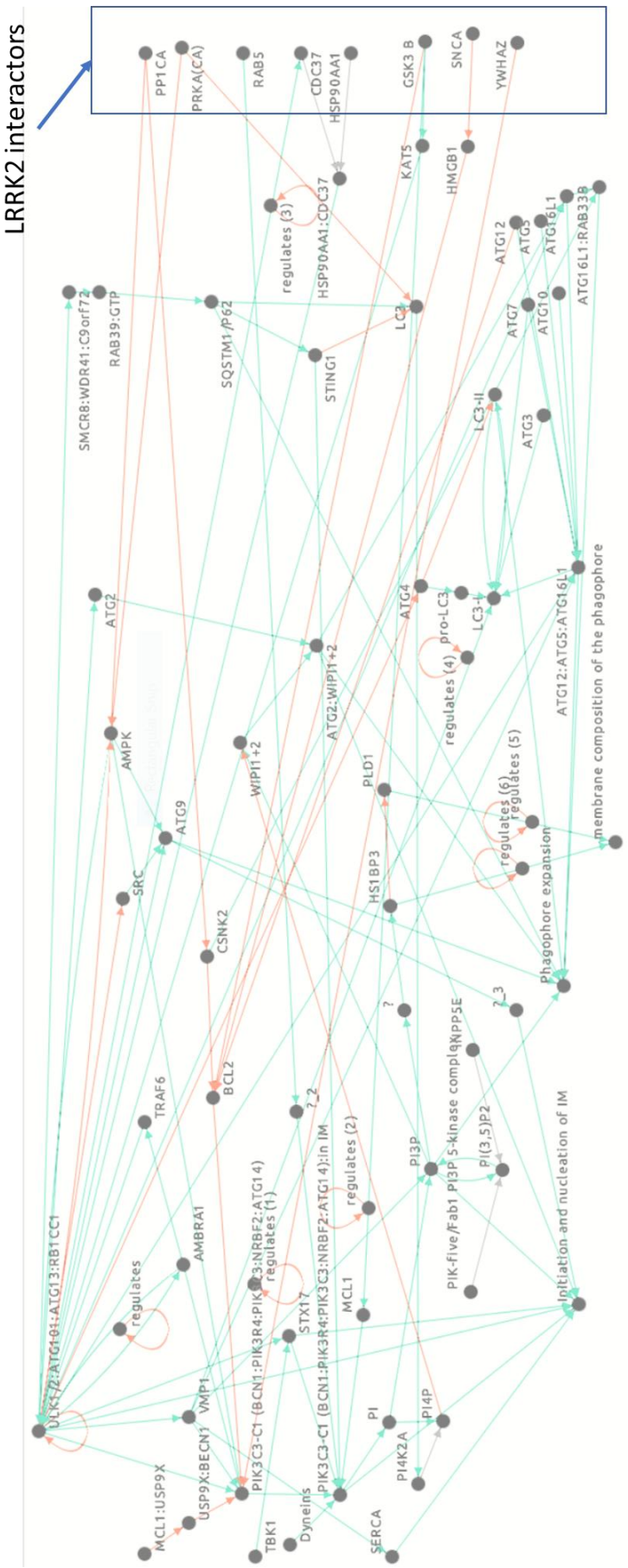


Figure S5-3. Enriched diagram D

The connections with the LRRK2 interactors have been filtered to include only those interactors for which there are data showing that the interactions are affected by LRRK2 mutations (SQSTM1, CDC37, GSK3B, PPP1CA, PRKACA, RAB5B, SNCA, and YWHAZ) and other necessary proteins (HSP90AA1). On the right-hand side of the figure, the LRRK2 interactors are included. This image was produced with Cell Collective.

Table S5-1. Protein interactions of LRRK2 affected by mutations				
#Feature AC	Feature short label	Feature type & annotation	PubMedID	Interactor
EBI-9104709	Q5S007:p.Arg1441Gly	mutation disrupting (MI:0573)	24351927	1433Z
EBI-9104715	Q5S007:p.Ser1444Ala	mutation disrupting (MI:0573)	24351927	1433Z
EBI-6508728	Q5S007:p.Gly2385Arg	mutation increasing (MI:0382)	22612223	CDC37
EBI-8842326	Q5S007:p.Gly2019Ser	mutation increasing (MI:0382)	24165324	GSK3B
EBI-22228908	Q5S007:p.Gly2385Arg	mutation decreasing rate (MI:1130)	29519959	SQSTM1
EBI-22229066	Q5S007:p.Asp1994Ala	mutation disrupting rate (MI:1129)	29519959	SQSTM1
EBI-22228788	Q5S007:p.Asp1994Ala	mutation disrupting rate (MI:1129)	29519959	SQSTM1
EBI-22228906	Q5S007:p.Gly2019Ser	mutation increasing rate (MI:1131)	29519959	SQSTM1
EBI-22229046	Q5S007:p.Gly2019Ser	mutation increasing rate (MI:1131)	29519959	SQSTM1
EBI-9691043	Q5S007:p.Asn1437His	mutation increasing strength (MI:1132)	23937259	PPP1CA
EBI-9691045	Q5S007:p.Arg1441Gly	mutation increasing strength (MI:1132)	23937259	PPP1CA
EBI-9691047	Q5S007:p.Tyr1699Cys	mutation increasing strength (MI:1132)	23937259	PPP1CA
EBI-9691049	Q5S007:p.Ser910Ala	mutation increasing strength (MI:1132)	23937259	PPP1CA
EBI-9691049	Q5S007:p.Ser935Ala	mutation increasing strength (MI:1132)	23937259	PPP1CA
EBI-9691049	Q5S007:p.Ser955Ala	mutation increasing strength (MI:1132)	23937259	PPP1CA
EBI-9691049	Q5S007:p.Ser973Ala	mutation increasing strength (MI:1132)	23937259	PPP1CA
EBI-12509246	Q5S007:p.Gly2385Arg	mutation decreasing rate (MI:1130)	27314038	RAB5B
EBI-12509232	Q5S007:p.Arg1441Cys	mutation decreasing rate (MI:1130)	27314038	RAB5B
EBI-12509228	Q5S007:p.Asp1994Ala	mutation disrupting rate (MI:1129)	27314038	RAB5B
EBI-10688243	Q5S007:p.Asp1994Ala	mutation disrupting rate (MI:1129)	25605758	RAB5B
EBI-12509242	Q5S007:p.Ile2020Thr	mutation increasing rate (MI:1131)	27314038	RAB5B
EBI-10688241	Q5S007:p.Gly2019Ser	mutation increasing rate (MI:1131)	25605758	RAB5B
EBI-12509230	Q5S007:p.Gly2019Ser	mutation increasing rate (MI:1131)	27314038	RAB5B
EBI-6309768	Q5S007:p.Gly2019Ser	mutation decreasing (MI:0119)	21454543	RAC1
EBI-6309786	Q5S007:p.Lys1906Met	mutation decreasing (MI:0119)	21454543	RAC1
EBI-6309770	Q5S007:p.Arg1441Cys	mutation disrupting (MI:0573)	21454543	RAC1
EBI-6309857	Q5S007:p.Arg1441Cys	mutation disrupting (MI:0573)	21454543	RAC1

Table S5-1. (continued) Protein interactions of LRRK2 affected by mutations				
EBI-6309784	Q5S007:p.Ile2020Thr	mutation increasing (MI:0382)	21454543	RAC1
EBI-6309776	Q5S007:p.Tyr1699Cys	mutation increasing (MI:0382)	21454543	RAC1
EBI-8844589	Q5S007:p.Asp1994Ala	mutation disrupting rate (MI:1129)	24165324	GSK3B
EBI-9104120	Q5S007:p.Ser1444Ala	mutation decreasing rate (MI:1130)	24351927	PRKACA
EBI-9104139	Q5S007:p.Arg1441Cys	mutation disrupting rate (MI:1129)	24351927	PRKACA
EBI-9104141	Q5S007:p.Arg1441Gly	mutation disrupting rate (MI:1129)	24351927	PRKACA
EBI-9104122	Q5S007:p.Ser1443_Ser1444delinsAlaAla	mutation disrupting rate (MI:1129)	24351927	PRKACA
EBI-9104143	Q5S007:p.Arg1441His	mutation disrupting rate (MI:1129)	24351927	PRKACA
EBI-6507337	Q5S007:p.Gly2019Ser	mutation increasing (MI:0382)	23183827	SNCA

Note: For each interaction the affected protein AC was uniprotkb:Q5S007, the affected protein symbol was LRRK2 and the affected protein organism was 9606-*Homo sapiens.*

Table S5-2. Decisions based on feedback from Dr Sharon Tooze	
Question	**Decision**
Are AMBRA1 and NRBF2 part of the PIK3C3 complex I?	Can be considered as true for this project
Should P62 bodies be considered as scaffolds for macroautophagy?	This is true only for selective autophagy. Therefore, p62 bodies will not be taken into consideration in this project.
Which should be considered the membrane sources in this project? (e.g., ERGIC-derived COPII vesicles and/or recycling endosomes and/or ATG9 vesicles)	Combining the different membrane sources might be tricky, so either different models should be made or one membrane source to be chosen. Here, ATG9 was considered the main membrane source and the rest were excluded.
Should the ATG9 trafficking system be included in the diagram?	Due to its complexity and partial knowledge of this system it will be excluded.
Lipid composition and phosphatidic acid levels of the autophagosome precursor membranes pay a role in the formation of the phagophore. Should they be included?	Too complex to be included in a model of the decided scale. Can be added in a future, more complicated version of the model.
Exclusion of LRRK2 interactors due to lack of specific information regarding their link with macroautophagy or a very complex relationship with macroautophagy?	Both are acceptable
Are there any other key players that are missing?	Added TMEM41B in the detailed diagram but was later removed (see Section 5.4.5.2).

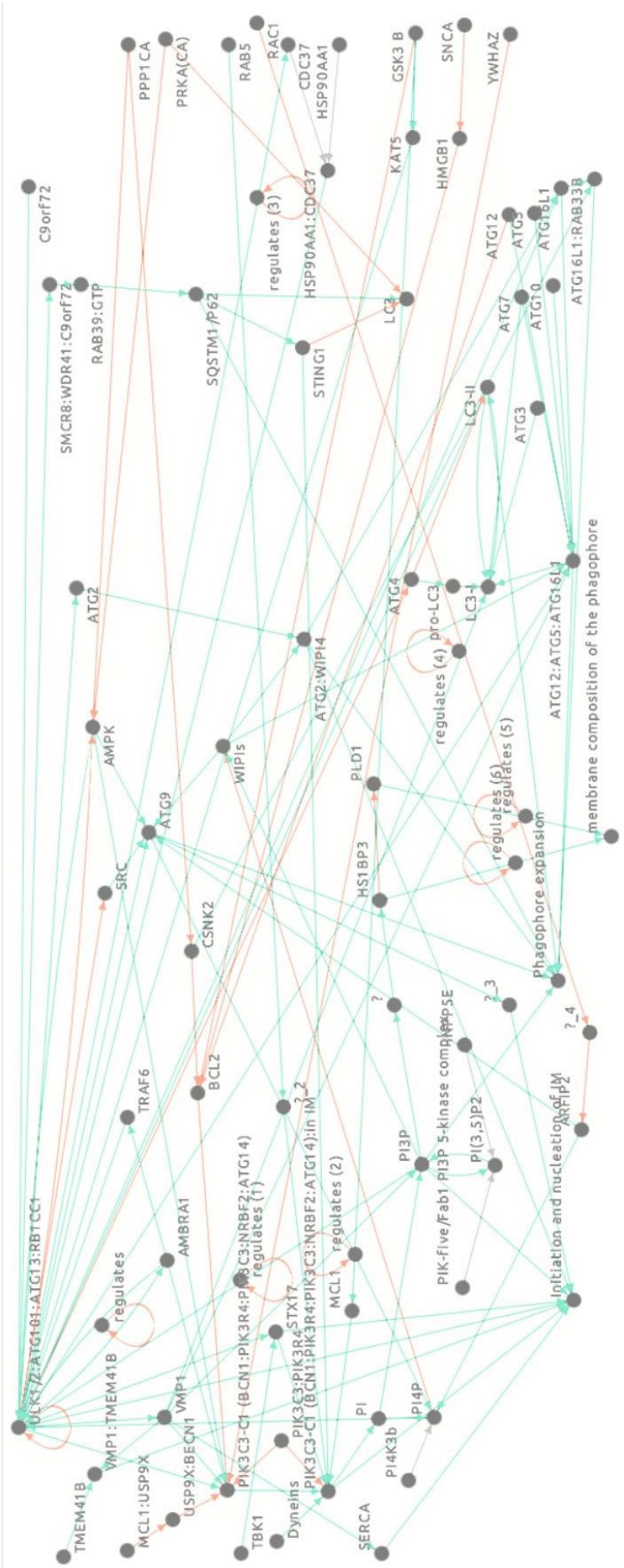


Figure S5-4. Enriched diagram E

Updated detailed diagram of macroautophagy based on the feedback from an autophagy expert, Dr Sharon Tooze, and information from some of her publications. One key difference is the incorporation of RAC1, an additional LRRK2 interactor.

Appendix F

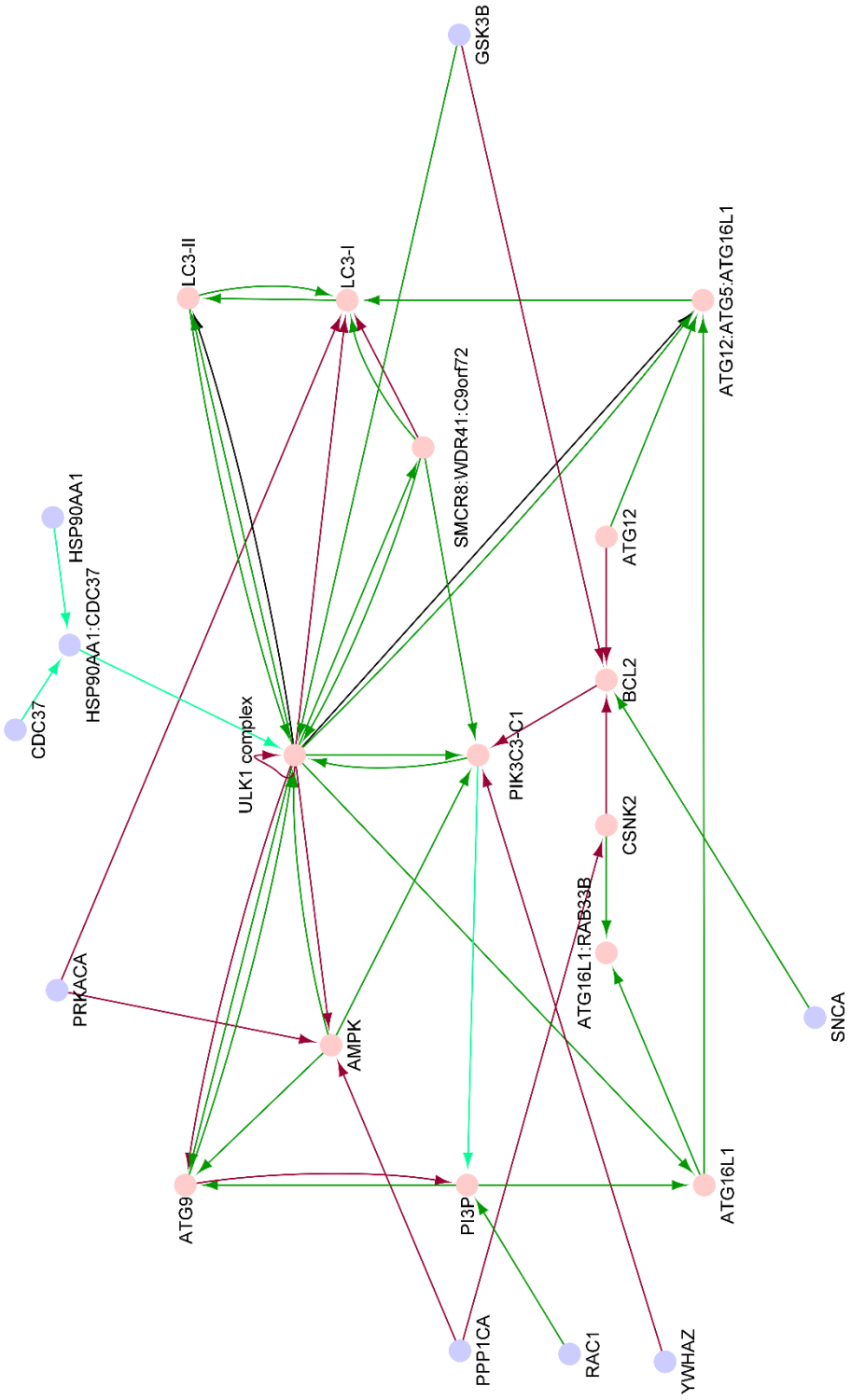


Figure S6-1. Macroautophagy diagram for modelling including the connections of LRRK2 interactors
In green are the activations, in red the inhibitions, and in cyan the productions. This picture was created using Cytoscape (v3.7.1).

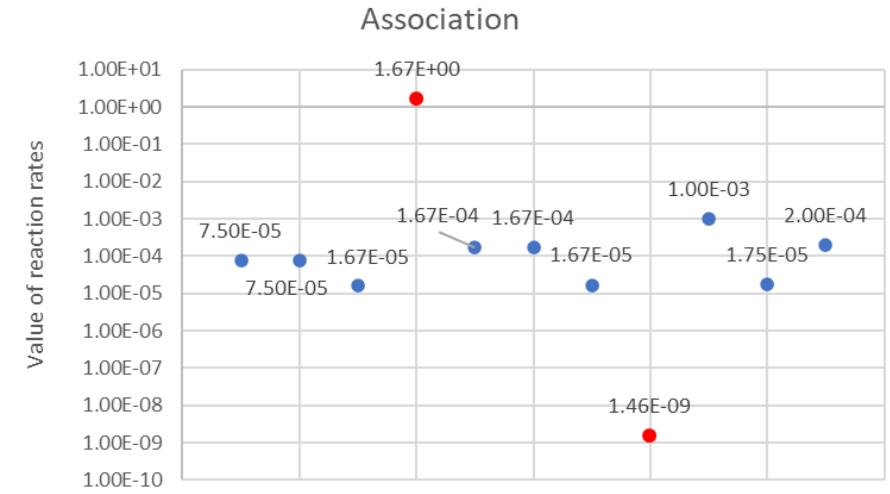
Table S6-1. Rate reaction equations describing the macroautophagy model and the calculation of the reaction rate

Reaction number	Rate reaction equation	Reaction Rate
0	PIK3C3_C1_act + ULK1complex -> ULK1complex_act + PIK3C3_C1_act	kf*PIK3C3_C1_act*ULK1complex
1	ULK1complex_act + PIK3C3_C1 -> ULK1complex_act + PIK3C3_C1_act	kf_1*ULK1complex_act*PIK3C3_C1
2	ULK1complex_act + AMPK_act -> AMPK + ULK1complex_act	kf_2*ULK1complex_act*AMPK_act
4	AMPK_act + ULK1complex -> AMPK_act + ULK1complex_act	kf_4*AMPK_act*ULK1complex
5	AMPK_act + ATG9 -> AMPK_act + ATG9_act	kf_5*AMPK_act*ATG9
6	AMPK_act + PIK3C3_C1 -> AMPK_act + PIK3C3_C1_act	kf_6*AMPK_act*PIK3C3_C1
7	ATG9 + ULK1complex_act -> ATG9_act + ULK1complex_act	kf_7*ATG9*ULK1complex_act
9	PIK3C3_C1_act + PI -> PIK3C3_C1_act + PI3P	kf_9*PIK3C3_C1_act*PI
8	ATG9_act + PI -> ATG9_act + PI4P	kf_8*ATG9_act*PI
11	ATG9 + PI3P -> ATG9_act + PI3P	kf_11*ATG9*PI3P
13	PI3P + ATG16L1 -> PI3P + ATG16L1_act	kf_13*PI3P*ATG16L1
14	ATG16L1 + ULK1complex_act -> ATG16L1_act + ULK1complex_act	kf_14*ATG16L1*ULK1complex_act
15	ATG16L1_act + ATG12_act + ATG5_act -> ATG12ATG5ATG16L1_act	kf_15*ATG16L1_act*ATG12_act*ATG5_act
18	CSNK2_act + BCL2_act -> BCL2 + CSNK2_act	kf_18*CSNK2_act*BCL2_act
19	ATG12_act + BCL2_act -> BCL2 + ATG12_act	kf_19*ATG12_act*BCL2_act
20	BCL2_act + PIK3C3_C1_act -> BCL2_act + PIK3C3_C1	kf_20*BCL2_act*PIK3C3_C1_act
22	LC3I_act + ATG12ATG5ATG16L1_act -> LC3II_act + ATG12ATG5ATG16L1_act	kf_22*LC3I_act*ATG12ATG5ATG16L1_act

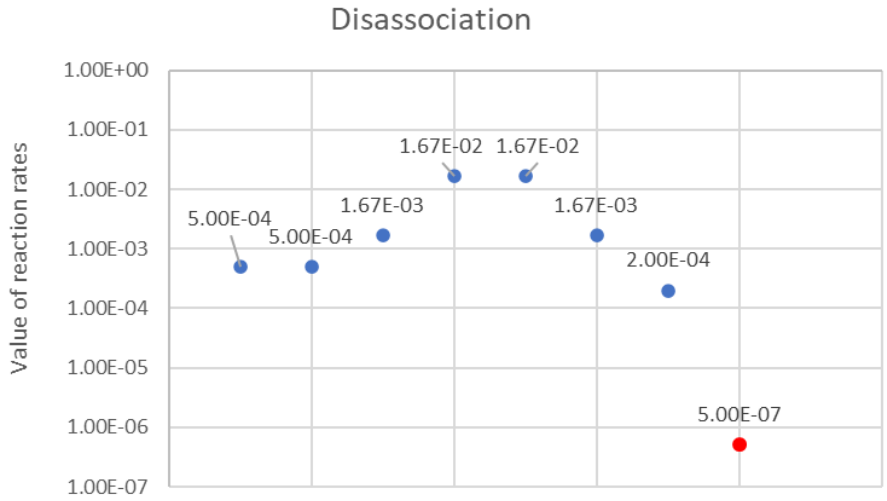
Table S6-1. (continued) Rate reaction equations describing the macroautophagy model and the calculation of the reaction rate		
23	ULK1complex_act + LC3I_act -> ULK1complex_act + LC3I	kf_23*ULK1complex_act*LC3I_act
24	ULK1complex_act + LC3I_act -> ULK1complex_act + LC3II_act	kf_24*ULK1complex_act*LC3I_act
25	ULK1complex + LC3II_act -> ULK1complex_act + LC3II_act	kf_25*ULK1complex*LC3II_act
34	SMCR8WDR41C9ORF72_act + LC3I_act -> LC3II_act + SMCR8WDR41C9ORF72_act	kf_34*SMCR8WDR41C9ORF72_act*LC3I_act
28	SMCR8_act + WDR41_act + C9ORF72_act + ULK1complex_act -> ULK1complex_act + SMCR8WDR41C9ORF72_act	kf_28*SMCR8_act*WDR41_act*C9ORF72_act* ULK1complex_act
29	ULK1complex + C9ORF72_act -> ULK1complex_act + C9ORF72_act	kf_29*ULK1complex*C9ORF72_act
32	PIK3C3_C1 + SMCR8WDR41C9ORF72_act -> PIK3C3_C1_act + SMCR8WDR41C9ORF72_act	kf_32*PIK3C3_C1*SMCR8WDR41C9ORF72_act
35	ULK1complex_act -> ULK1complex	kf_35*ULK1complex_act
36	PIK3C3_C1_act -> PIK3C3_C1	kf_36*PIK3C3_C1_act
37	AMPK_act -> AMPK	kf_37*AMPK_act
38	ATG9_act -> ATG9	kf_38*ATG9_act
39	ATG16L1_act -> ATG16L1	kf_39*ATG16L1_act
41	ATG12ATG5ATG16L1_act -> ATG12_act + ATG5_act + ATG16L1_act	kf_41*ATG12ATG5ATG16L1_act
16	PI3P -> PI	kf_16*PI3P
44	BCL2_act -> BCL2	kf_44*BCL2_act
46	SMCR8WDR41C9ORF72_act -> SMCR8_act + WDR41_act + C9ORF72_act	kf_46*SMCR8WDR41C9ORF72_act
47	LC3I_act -> LC3I	kf_47*LC3I_act
48	LC3II_act -> LC3I_act	kf_48*LC3II_act

Table S6-1. (continued) Rate reaction equations describing the macroautophagy model and the calculation of the reaction rate		
49	PI4P + ULK1complex -> ULK1complex_act + PI4P	kf_49*PI4P*ULK1complex
50	ATG16L1 + CSNK2_act -> ATG16L1_act + CSNK2_act	kf_50*ATG16L1*CSNK2_act
52	ULK1complex + PI3P -> PI3P + ULK1complex_act	kf_52*ULK1complex*PI3P
53	ULK1complex_act + BCL2_act -> BCL2 + ULK1complex_act	kf_53*ULK1complex_act*BCL2_act
54	ULK1complex_act + LC3I_act -> LC3I_act + ULK1complex	kf_54*ULK1complex_act*LC3I_act
12	PI4P -> PI	kf_12*PI4P
3	LC3I -> LC3I_act	kf_3*LC3I
26	CSNK2_act + LC3I_act -> CSNK2_act + LC3II_act	kf_26*CSNK2_act*LC3I_act
27	CSNK2_act + PIK3C3_C1 -> CSNK2_act + PIK3C3_C1_act	kf_27*CSNK2_act*PIK3C3_C1
10	SMCR8WDR41C9ORF72_act + LC3II_act -> SMCR8WDR41C9ORF72_act + LC3I_act	kf_10*SMCR8WDR41C9ORF72_act*LC3II_act

A.



B.



C.

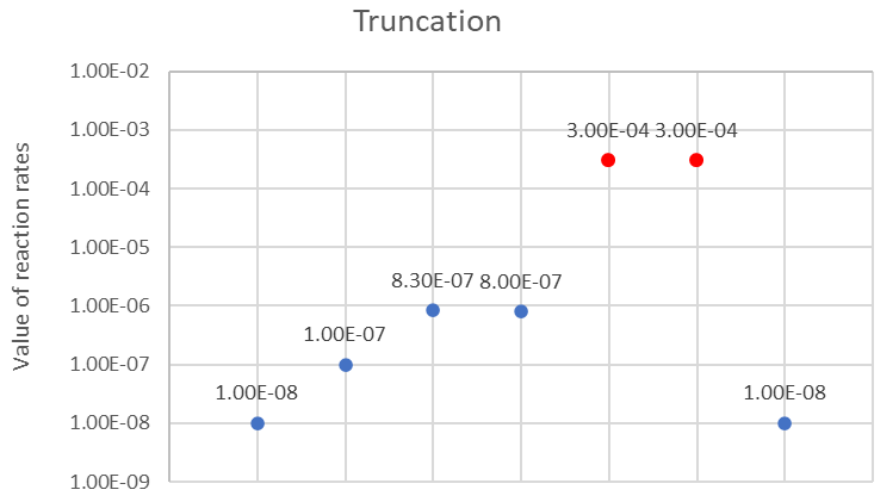


Figure S6-2. Removing outliers of reaction rates during parametrisation

*Parameter values with more than 2 magnitudes of difference compared to the rest of the values were considered outliers and removed from further analysis for association (**A**), disassociation (**B**) and truncation (**C**).*

Table S6-2. Initial conditions- Amounts of the components of the mathematical model of macroautophagy

Name of entity	Amount in molecules
ULK1complex / ULK1complex_act	0 / 8.88×10^{-5}
PIK3C3_C1 / PIK3C3_C1_act	0.000149 / 0
AMPK / AMPK_act	0 / 0.000103
ATG9 / ATG9_act	0.00107 / 0
PI / PI3P / PI4P	0.0165 / 0 / 0
ATG16L1 / ATG16L1_act	0.00159 / 0
ATG12_act	0.000111
ATG5_act	0.00114
ATG12ATG5ATG16L1_act	0
CSNK2_act	0.00123
BCL2 / BCL2_act	0 / 0.00077
LC3I / LC3I_act / LC3II_act	0.00143 / 0 / 0
C9ORF72_act	0.000477
SMCR8_act	0.00119
WDR41_act	0.000669
SMCR8WDR41C9ORF72_act	0

Table S6-3. Parameter values for the mathematical model of macroautophagy

Name	Value (units)	Name	Value (units)
kf	9.5 $(\text{molecule sec})^{-1}$	**kf_35**	0.0000025 $(\text{sec})^{-1}$
kf_1	5.216865 $(\text{molecule sec})^{-1}$	**kf_36**	0.0000025 $(\text{sec})^{-1}$
kf_2	5 $(\text{molecule sec})^{-1}$	**kf_37**	0.0000025 $(\text{sec})^{-1}$
kf_4	5.216865 $(\text{molecule sec})^{-1}$	**kf_38**	0.0000025 $(\text{sec})^{-1}$
kf_5	5 $(\text{molecule sec})^{-1}$	**kf_39**	0.0000025 $(\text{sec})^{-1}$
kf_6	5 $(\text{molecule sec})^{-1}$	**kf_41**	0.0000025 $(\text{sec})^{-1}$
kf_7	9.716865 $(\text{molecule sec})^{-1}$	**kf_44**	0.0000025 $(\text{sec})^{-1}$

Table S6-3. (continued) Parameter values for the mathematical model of macroautophagy			
kf_9	5 $(molecule\ sec)^{-1}$	**kf_46**	0.0000025 $(sec)^{-1}$
kf_11	0.3903525 $(molecule\ sec)^{-1}$	**kf_47**	0.0000025 $(sec)^{-1}$
kf_13	0.3903525 $(molecule\ sec)^{-1}$	**kf_48**	0.0000025 $(sec)^{-1}$
kf_14	0.433725 $(molecule\ sec)^{-1}$	**kf_49**	0.4335 $(molecule\ sec)^{-1}$
kf_15	63363769.15 $(molecule^2\ sec)^{-1}$	**kf_50**	5 $(molecule\ sec)^{-1}$
kf_18	1.084315 $(molecule\ sec)^{-1}$	**kf_52**	0.2168625 $(molecule\ sec)^{-1}$
kf_19	0.86745 $(molecule\ sec)^{-1}$	**kf_53**	0.86745 $(molecule\ sec)^{-1}$
kf_20	0.86745 $(molecule\ sec)^{-1}$	**kf_54**	0.2168625 $(molecule\ sec)^{-1}$
kf_22	4 $(molecule\ sec)^{-1}$	**kf_3**	0.000614645 $(sec)^{-1}$
kf_23	0.4298215 $(molecule\ sec)^{-1}$	**kf_8**	5 $(molecule\ sec)^{-1}$
kf_24	4.5 $(molecule\ sec)^{-1}$	**kf_12**	0.0000025 $(sec)^{-1}$
kf_25	0.86745 $(molecule\ sec)^{-1}$	**kf_16**	0.0000025 $(sec)^{-1}$
kf_28	21975221 $(molecule^3\ sec)^{-1}$	**kf_26**	4.5 $(molecule\ sec)^{-1}$
kf_29	0.3903525 $(molecule\ sec)^{-1}$	**kf_27**	4 $(molecule\ sec)^{-1}$
kf_32	0.607215 $(molecule\ sec)^{-1}$	**kf_10**	0.607215 $(molecule\ sec)^{-1}$
kf_34	0.607215 $(molecule\ sec)^{-1}$		

Table S6-4. ODEs describing the mathematical model of macroautophagy

Equations

$$d(U)/dt = -(kf*PCa*U) - (kf\_4*AKa*U) - (kf\_25*U*L2a) - (kf\_29*U*C9a) + (kf\_35*Ua) - (kf\_49*P4*U) - (kf\_52*U*P3) + (kf\_54*Ua*L1a) \quad (\mathbf{1})$$

$$d(Ua)/dt = (kf*PCa*U) + (kf\_4*AKa*U) + (kf\_25*U*L2a) + (kf\_29*U*C9a) - (kf\_35*Ua) + (kf\_49*P4*U) + (kf\_52*U*P3) - (kf\_54*Ua*L1a) \quad (\mathbf{2})$$

$$d(PC)/dt = -(kf\_1*Ua*PC) - (kf\_6*AKa*PC) + (kf\_20*Ba*PCa) - (kf\_32*PC*SCMa) + (kf\_36*PCa) - (kf\_27*Ca*PC) \quad (\mathbf{3})$$

$$d(PCa)/dt = (kf\_1*Ua*PC) + (kf\_6*AKa*PC) - (kf\_20*Ba*PCa) + (kf\_32*PC*SCMa) - (kf\_36*PCa) + (kf\_27*Ca*PC) \quad (\mathbf{4})$$

$$d(AK)/dt = (kf\_2*Ua*AKa) + (kf\_37*AKa) \quad (\mathbf{6})$$

$$d(AKa)/dt = -(kf\_2*Ua*AKa) - (kf\_37*AKa) \quad (\mathbf{5})$$

$$d(A9)/dt = -(kf\_5*AKa*A9) - (kf\_7*A9*Ua) - (kf\_11*A9*P3) - (kf\_30*A9*Ua) + (kf\_38*A9a) \quad (\mathbf{7})$$

d(A9a)/dt = (kf_5*AKa*A9) + (kf_7*A9*Ua) + (kf_11*A9*P3) + (kf_30*A9*Ua) - (kf_38*A9a) (**8**)

d(P3)/dt = (kf_9*PCa*P) - (kf_16*P3) (**9**)

d(P4)/dt = (kf_8*A9a*P) - (kf_12*P4) (**24**)

d(P)/dt = -(kf_9*PCa*P) - (kf_8*A9a*P) + (kf_16*P3) + (kf_12*P4) (**25**)

d(A16)/dt = -(kf_13*P3*A16) - (kf_14*A16*Ua) + (kf_39*A16a) - (kf_50*A16*Ca) (**10**)

d(A16a)/dt = (kf_13*P3*A16) + (kf_14*A16*Ua) - (kf_15*A16a*A12a*A5a) - (kf_39*A16a) + (kf_41*ACa) + (kf_50*A16*Ca) (**11**)

d(A12a)/dt = -(kf_15*A16a*A12a*A5a) + (kf_41*ACa) (**12**)

d(A5a)/dt = -(kf_15*A16a*A12a*A5a) + (kf_41*ACa) (**13**)

d(ACa)/dt = (kf_15*A16a*A12a*A5a) - (kf_41*ACa) (**14**)

d(B)/dt = (kf_18*Ca*Ba) + (kf_19*A12a*Ba) + (kf_44*Ba) + (kf_53*Ua*Ba) (**16**)

d(Ba)/dt = -(kf_18*Ca*Ba) - (kf_19*A12a*Ba) - (kf_44*Ba) - (kf_53*Ua*Ba) (**15**)

d(L1)/dt = (kf_23*Ua*L1a) + (kf_47*L1a) - (kf_3*L1) (**19**)

d(L1a)/dt = -(kf_22*L1a*ACa) - (kf_23*Ua*L1a) - (kf_24*Ua*L1a) - (kf_34*SCMa*L1a) - (kf_47*L1a) + (kf_48*L2a) + (kf_3*L1) - (kf_26*Ca*L1a) (**17**)

d(L2a)/dt = (kf_22*L1a*ACa) + (kf_24*Ua*L1a) + (kf_34*SCMa*L1a) - (kf_48*L2a) + (kf_26*Ca*L1a) (**18**)

d(SCMa)/dt = (kf_28*Sa*Wa*C9a*U) + (kf_31*Sa*Wa*C9a) - (kf_46*SCMa) (**20**)

d(C9a)/dt = -(kf_28*Sa*Wa*C9a*U) - (kf_31*Sa*Wa*C9a) + (kf_46*SCMa) (**21**)

d(Sa)/dt = -(kf_28*Sa*Wa*C9a*U) - (kf_31*Sa*Wa*C9a) + (kf_46*SCMa) (**22**)

d(Wa)/dt = -(kf_28*Sa*Wa*C9a*U) - (kf_31*Sa*Wa*C9a) + (kf_46*SCMa) (**23**)

Name of entity	**Symbol of entity**
ULK1complex / ULK1complex_act	U / Ua
PIK3C3_C1 / PIK3C3_C1_act	PC / PCa
AMPK / AMPK_act	AK / AKa
ATG9 / ATG9_ac	A9 / A9a
PI / PI3P / PI4P	P / P3 / P4
ATG16L1 / ATG16L1_act	A16 / A16a
ATG12_act /	A12
ATG5_act	A12a
ATG12ATG5ATG16L1_act	ACa
BCL2 / BCL2_act	B / Ba
LC3I / LC3I_act / LC3II_act	L1 / L1a / L2a
SMCR8WDR41C9ORF72_act	SCMa
C9ORF72_act	C9a
SMCR8_act	Sa
WDR41_act	Wa
CSNK2_act	Ca

Note: The ODEs exactly as produced by SimBiology are presented in Table S6-5.

Table S6-5 ODEs describing the mathematical model of macroautophagy as produced by SimBiology, a MATLAB toolbox

d(ULK1complex)/dt = 1/[L.A.F.]*(-(kf*PIK3C3_C1_act*ULK1complex) - kf_4*AMPK_act*ULK1complex) - (kf_25*ULK1complex*LC3II_act) - (kf_29*ULK1complex*C9ORF72_act) + (kf_35*ULK1complex_act) - (kf_49*PI4P*ULK1complex) - (kf_52*ULK1complex*PI3P) + (kf_54*ULK1complex_act*LC3I_act))

d(ULK1complex_act)/dt = 1/[L.A.F.]*((kf*PIK3C3_C1_act*ULK1complex) + (kf_4*AMPK_act*ULK1complex) + (kf_25*ULK1complex*LC3II_act) + (kf_29*ULK1complex*C9ORF72_act) - (kf_35*ULK1complex_act) + (kf_49*PI4P*ULK1complex) + (kf_52*ULK1complex*PI3P) - (kf_54*ULK1complex_act*LC3I_act))

d(PIK3C3_C1)/dt = 1/[L.A.F.]*(-(kf_1*ULK1complex_act*PIK3C3_C1) - (kf_6*AMPK_act* PIK3C3_C1) + (kf_20*BCL2_act*PIK3C3_C1_act) - (kf_32*PIK3C3_C1*SMCR8WDR41C9ORF72_act) + (kf_36*PIK3C3_C1_act) - (kf_27*CSNK2_act*PIK3C3_C1))

d(PIK3C3_C1_act)/dt = 1/[L.A.F.]*((kf_1*ULK1complex_act*PIK3C3_C1) + (kf_6*AMPK_act* PIK3C3_C1) - (kf_20*BCL2_act*PIK3C3_C1_act) + (kf_32*PIK3C3_C1*SMCR8WDR41C9ORF72_act) - (kf_36*PIK3C3_C1_act) + (kf_27*CSNK2_act*PIK3C3_C1))

d(AMPK_act)/dt = 1/[L.A.F.]*(-(kf_2*ULK1complex_act*AMPK_act) - (kf_37*AMPK_act))

d(AMPK)/dt = 1/[L.A.F.]*((kf_2*ULK1complex_act*AMPK_act) + (kf_37*AMPK_act))

d(ATG9)/dt = 1/[L.A.F.]*(-(kf_5*AMPK_act*ATG9) - (kf_7*ATG9*ULK1complex_act) - (kf_11*ATG9*PI3P) + (kf_38*ATG9_act))

d(ATG9_act)/dt = 1/[L.A.F.]*((kf_5*AMPK_act*ATG9) + (kf_7*ATG9*ULK1complex_act) + (kf_11*ATG9*PI3P) - (kf_38*ATG9_act))

d(PI3P)/dt = 1/[L.A.F.]*((kf_9*PIK3C3_C1_act*PI) - (kf_16*PI3P))

d(ATG16L1)/dt = 1/[L.A.F.]*(-(kf_13*PI3P*ATG16L1) - (kf_14*ATG16L1*ULK1complex_act) + (kf_39*ATG16L1_act) - (kf_50*ATG16L1*CSNK2_act))

d(ATG16L1_act)/dt = 1/[L.A.F.]*((kf_13*PI3P*ATG16L1) + (kf_14*ATG16L1*ULK1complex_act) - (kf_15*ATG16L1_act*ATG12_act*ATG5_act) - (kf_39*ATG16L1_act) + (kf_41*ATG12ATG5ATG16L1_act) + (kf_50*ATG16L1*CSNK2_act))

d(ATG12_act)/dt = 1/[L.A.F.]*(-(kf_15*ATG16L1_act*ATG12_act*ATG5_act) + (kf_41*ATG12ATG5ATG16L1_act))

d(ATG5_act)/dt = 1/[L.A.F.]*(-(kf_15*ATG16L1_act*ATG12_act*ATG5_act) + (kf_41*ATG12ATG5ATG16L1_act))

d(ATG12ATG5ATG16L1_act)/dt = 1/[L.A.F.]*((kf_15*ATG16L1_act*ATG12_act*ATG5_act) - (kf_41*ATG12ATG5ATG16L1_act))

d(BCL2_act)/dt = 1/[L.A.F.]*(-(kf_18*CSNK2_act*BCL2_act) - (kf_19*ATG12_act*BCL2_act) - (kf_44*BCL2_act) - (kf_53*ULK1complex_act*BCL2_act))

d(BCL2)/dt = 1/[L.A.F.]*((kf_18*CSNK2_act*BCL2_act) + (kf_19*ATG12_act*BCL2_act) + (kf_44*BCL2_act) + (kf_53*ULK1complex_act*BCL2_act))

d(LC3I_act)/dt = 1/[L.A.F.]*(-(kf_22*LC3I_act*ATG12ATG5ATG16L1_act) - (kf_23*ULK1complex_act*LC3I_act) - (kf_24*ULK1complex_act*LC3I_act) - (kf_34*SMCR8WDR41C9ORF72_act*LC3I_act) - (kf_47*LC3I_act) + (kf_48*LC3II_act) + (kf_3*LC3I) - (kf_26*CSNK2_act*LC3I_act) + (kf_10*SMCR8WDR41C9ORF72_act*LC3II_act))

d(LC3II_act)/dt = 1/[L.A.F.]*((kf_22*LC3I_act*ATG12ATG5ATG16L1_act) + (kf_24*ULK1complex_act*LC3I_act) + (kf_34*SMCR8WDR41C9ORF72_act*LC3I_act) - (kf_48*LC3II_act) + (kf_26*CSNK2_act*LC3I_act) - (kf_10*SMCR8WDR41C9ORF72_act*LC3II_act))

d(LC3I)/dt = 1/[L.A.F.]*((kf_23*ULK1complex_act*LC3I_act) + (kf_47*LC3I_act) - (kf_3*LC3I))

d(SMCR8WDR41C9ORF72_act)/dt = 1/[L.A.F.]*((kf_28*SMCR8_act*WDR41_act*C9ORF72_act*ULK1complex_act) - (kf_46*SMCR8WDR41C9ORF72_act))

d(C9ORF72_act)/dt = 1/[L.A.F.]*(-(kf_28*SMCR8_act*WDR41_act*C9ORF72_act*ULK1complex_act) + (kf_46*SMCR8WDR41C9ORF72_act))

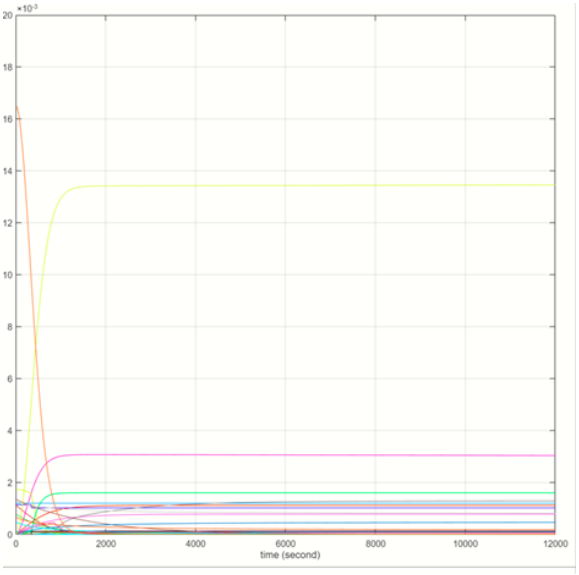
d(SMCR8_act)/dt = 1/[L.A.F.]*(-(kf_28*SMCR8_act*WDR41_act*C9ORF72_act*ULK1complex_act) + (kf_46*SMCR8WDR41C9ORF72_act))

d(WDR41_act)/dt = 1/[L.A.F.]*(-(kf_28*SMCR8_act*WDR41_act*C9ORF72_act*ULK1complex_act) + (kf_46*SMCR8WDR41C9ORF72_act))

d(PI4P)/dt = 1/[L.A.F.]*((kf_8*ATG9_act*PI) - (kf_12*PI4P))

d(PI)/dt = 1/[L.A.F.]*(-(kf_9*PIK3C3_C1_act*PI) - (kf_8*ATG9_act*PI) + (kf_16*PI3P) + (kf_12*PI4P))

A.



B.

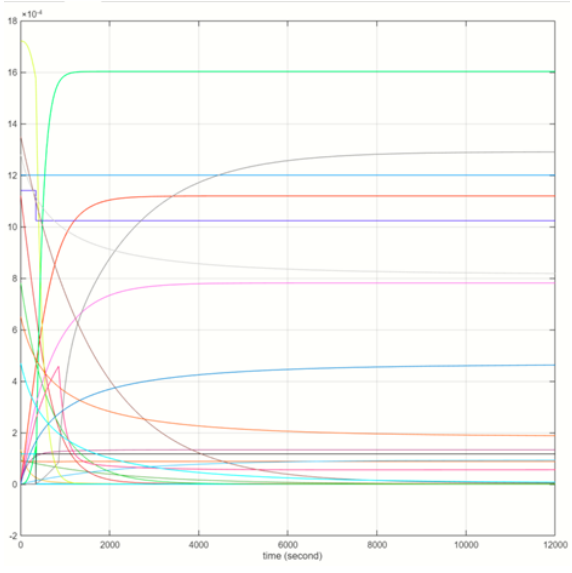
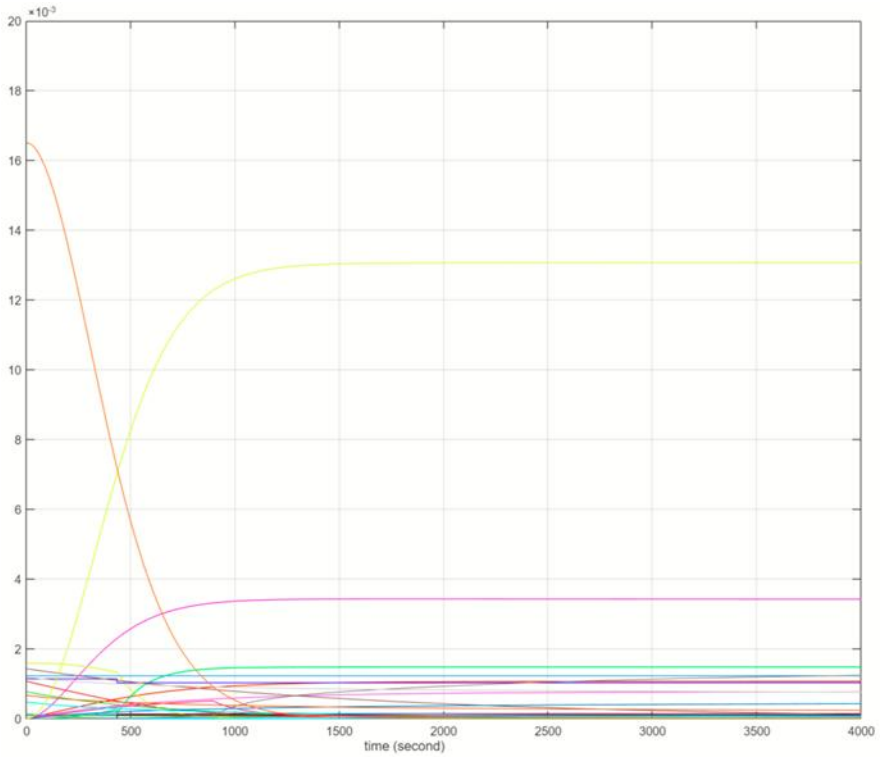


Figure S6-3. Steady state of the macroautophagy model is reached at around 10,000s.

*(**A**) Includes all components of the model. (**B**) PI3P, PI4P, and PI were removed to allow better visualisation of the rest of the components.*

A.



B.

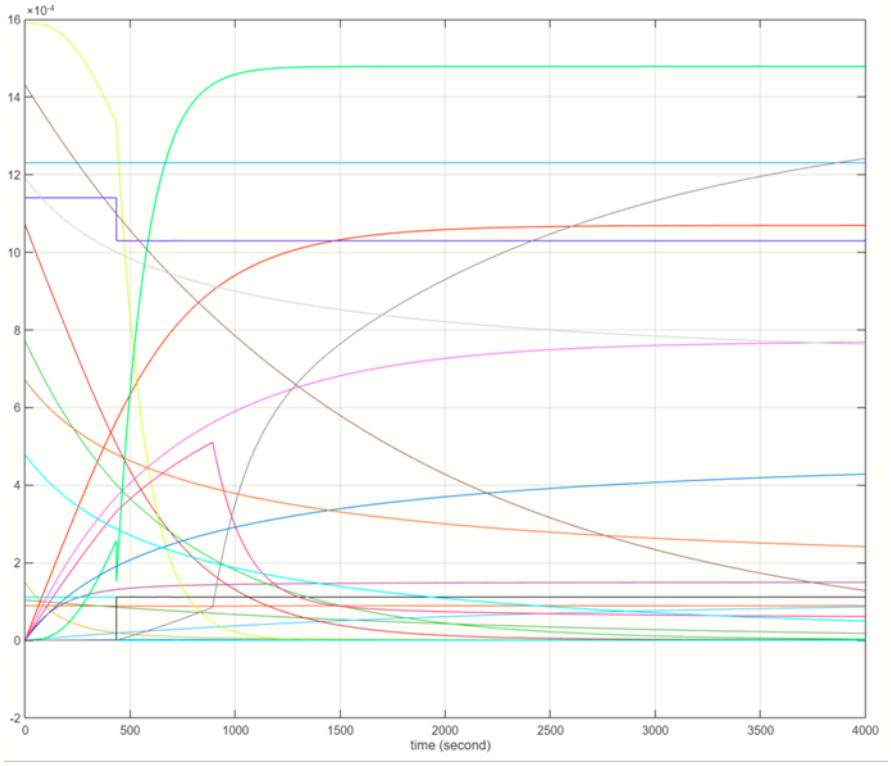
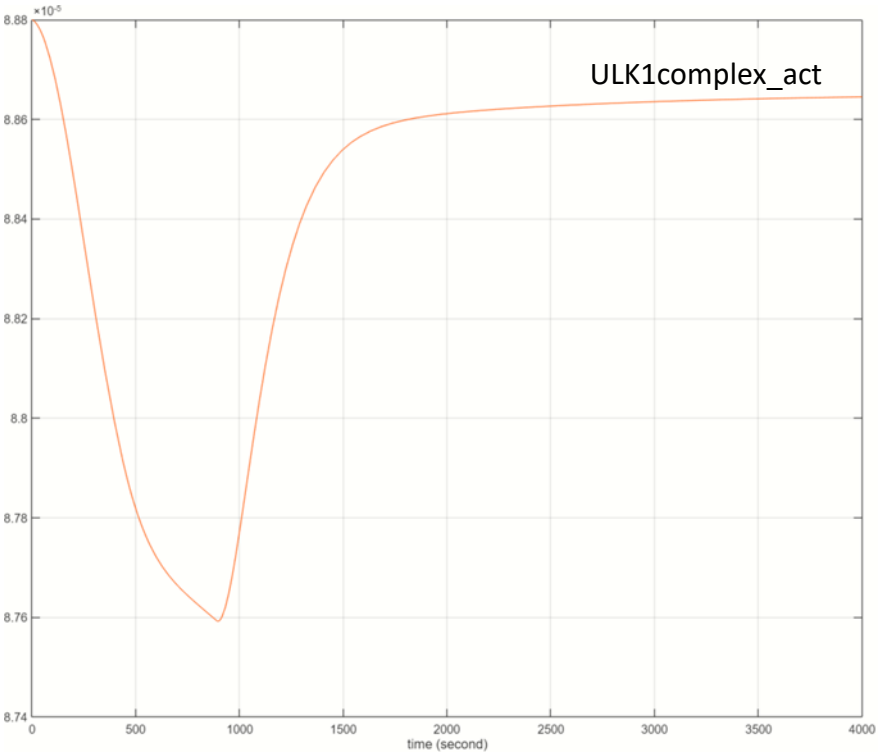


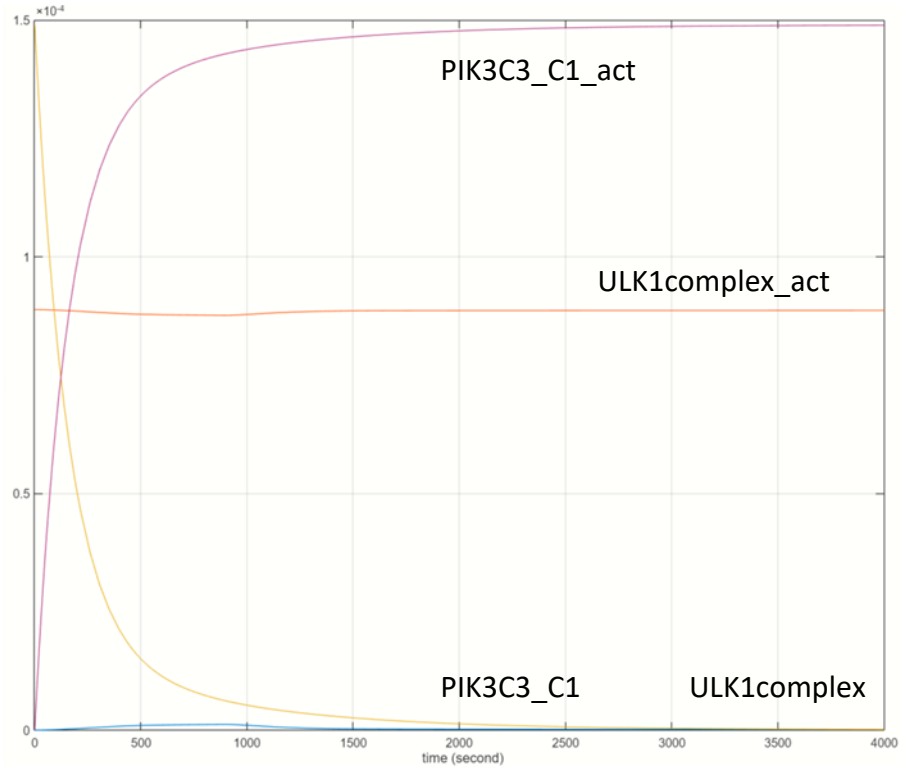
Figure S6-4. Simulation of macroautophagy as produced by the model.

(A) All the components of the model are included. (B) PI3P (pink), PI (orange) and PI4P (yellow) were excluded to allow better visualisation of components with lower amounts. See Fig S6-5 for the individual graphs.

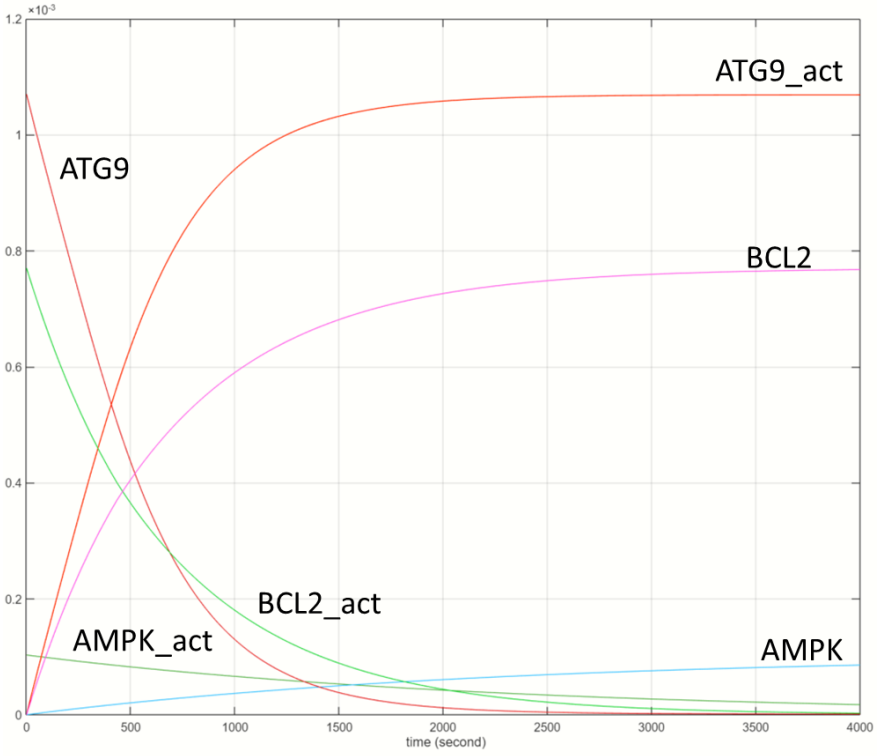
A.


×10⁻⁵
8.88
8.86
8.84
8.82
8.8
8.78
8.76
8.74
0
500
1000
1500
2000
2500
3000
3500
4000
time (second)
ULK1complex_act

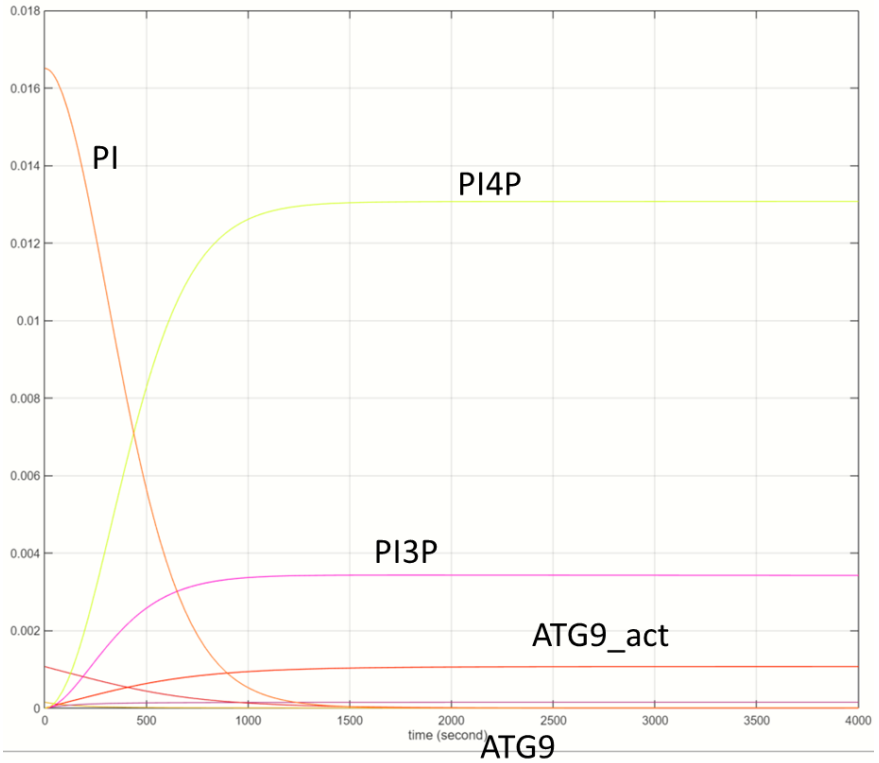
B.


×10⁻⁴
1.5
1
0.5
0
0
500
1000
1500
2000
2500
3000
3500
4000
time (second)
PIK3C3_C1_act
ULK1complex_act
PIK3C3_C1
ULK1complex

C.

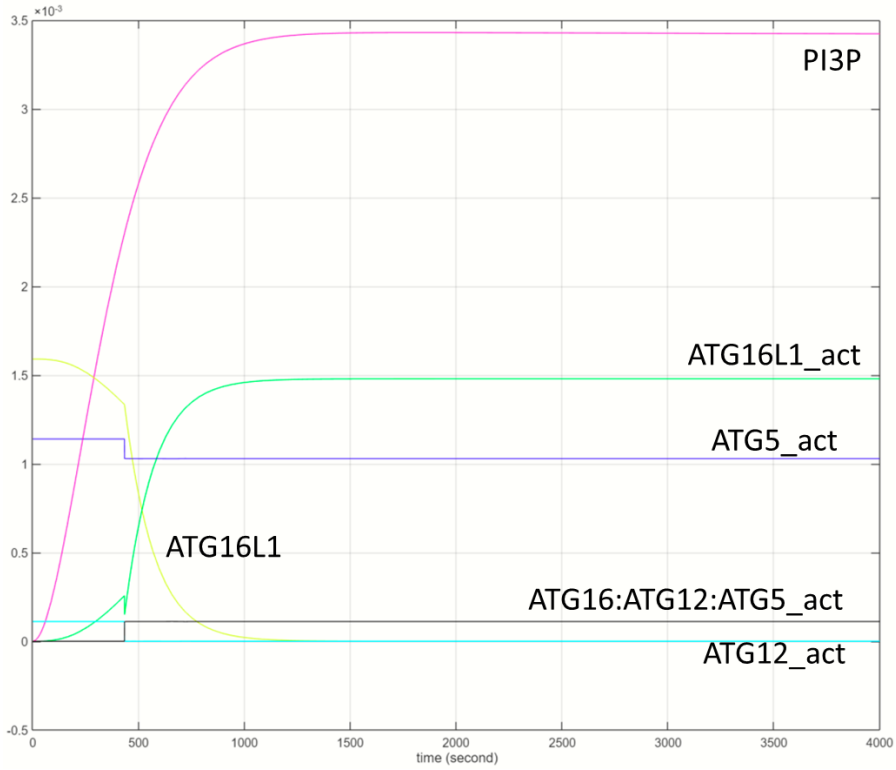


D.



PIK3C3_C1_act PIK3C3_C1

E.



F.

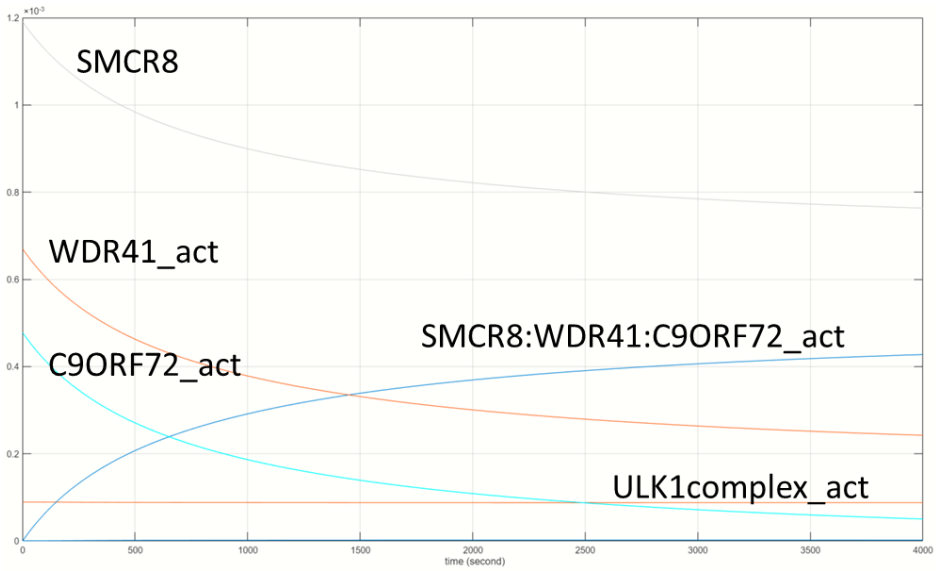


Figure S6-5. Changes in the amounts of groups of entities of the model of macroautophagy over time.

*(**A**) ULK1 complex in its activated form. (**B**) PIK3C3-C1 in activated and inactivated form, including the activated form of ULK1 complex for reference. (**C**) ATG9, BCL2, and AMPK in their activated and inactivated forms. (**D**) PI, PI3P, and PI4P, together with ATG9 and PIK3C3 in both their activated and inactivated forms for reference. (**E**) ATG16L1:ATG12:ATG5 and its components, together with PI3P for reference. (**F**) SMCR8:WDR41:C9ORF72 and its components, together with the activated form of ULK1complex for reference. The timescale for the simulation was: 0≤t≤4,000s.*

Table S6-6. Amounts of the components of the mathematical model of macroautophagy for people with PD	
Name	**Number in molecules**
ULK1complex / ULK1complex_act	0 / 8.81×10^{-5}
PIK3C3_C1 / PIK3C3_C1_act	1.33×10^{-4} / 0
AMPK / AMPK_act	0 / 9.30×10^{-5}
ATG9 / ATG9_act	1.12×10^{-3} / 0
PI / PI3P / PI4P	0.0165 / 0 / 0
ATG16L1 / ATG16L1_act	1.72×10^{-3} / 0
ATG12_act	1.17×10^{-4}
ATG5_act	1.14×10^{-3}
ATG12ATG5ATG16L1_act	0
CSNK2_act	1.20×10^{-3}
BCL2 / BCL2_act	0 / 7.81×10^{-4}
LC3I / LC3I_act / LC3II_act	1.35×10^{-3} / 0 / 0
SMCR8WDR41C9ORF72_act	0
C9ORF72_act	4.69×10^{-4}
SMCR8_act	1.28×10^{-3}
WDR41_act	6.50×10^{-4}

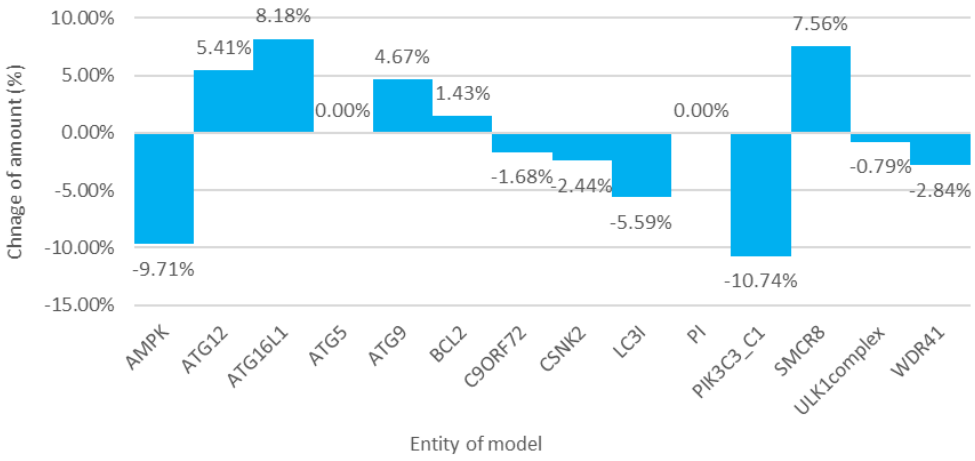
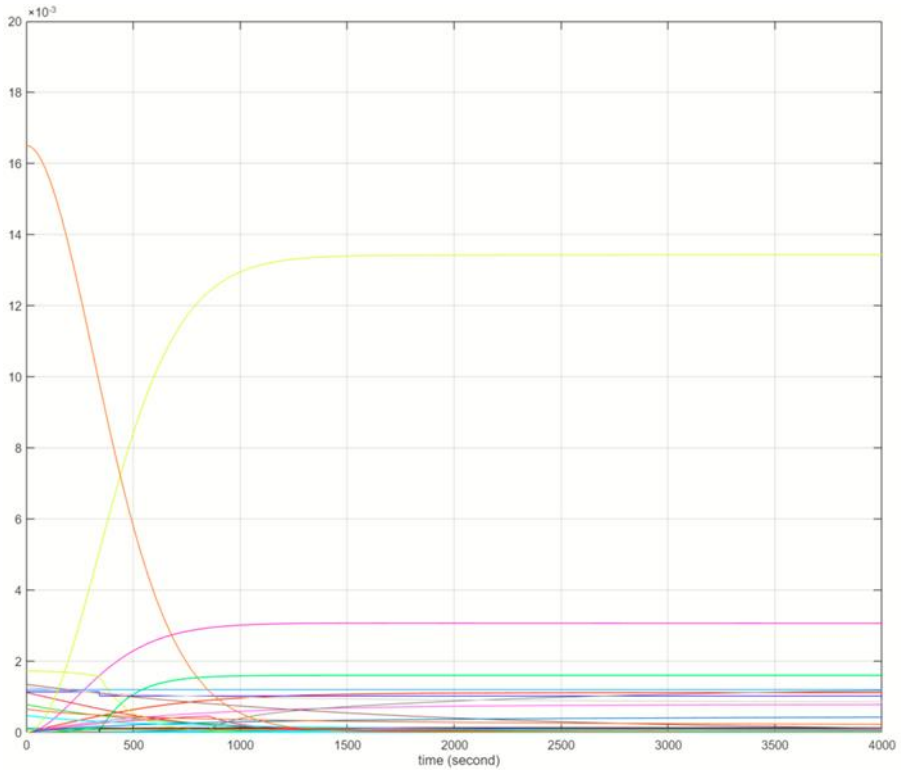
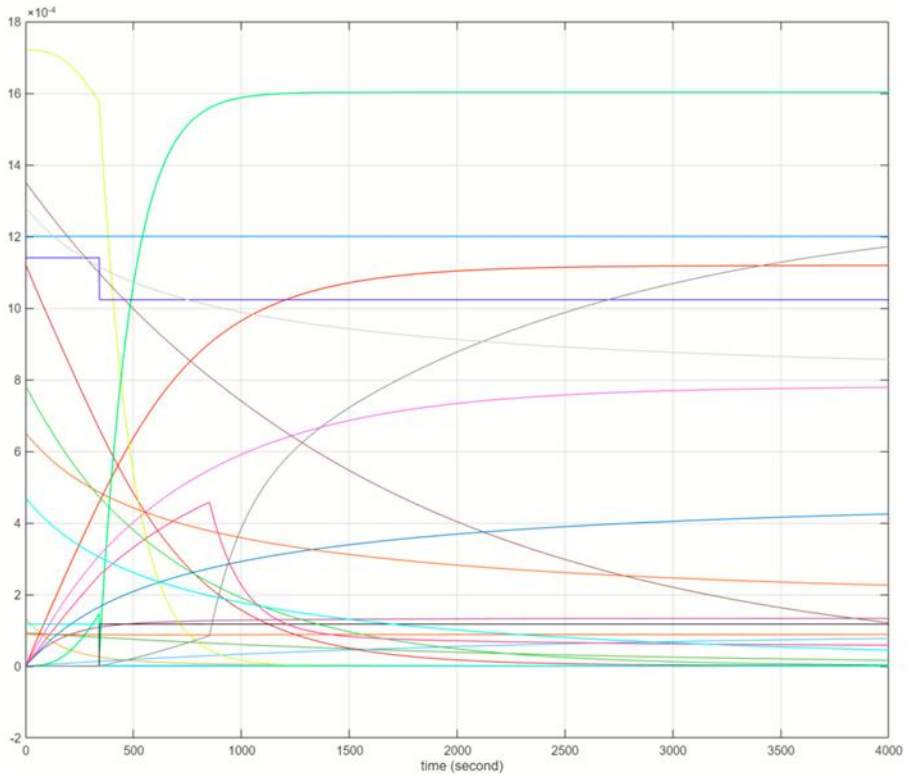


Figure S6-6. Comparison of the relative total amount of the entities of the macroautophagy model in the PD state compared to the healthy state.

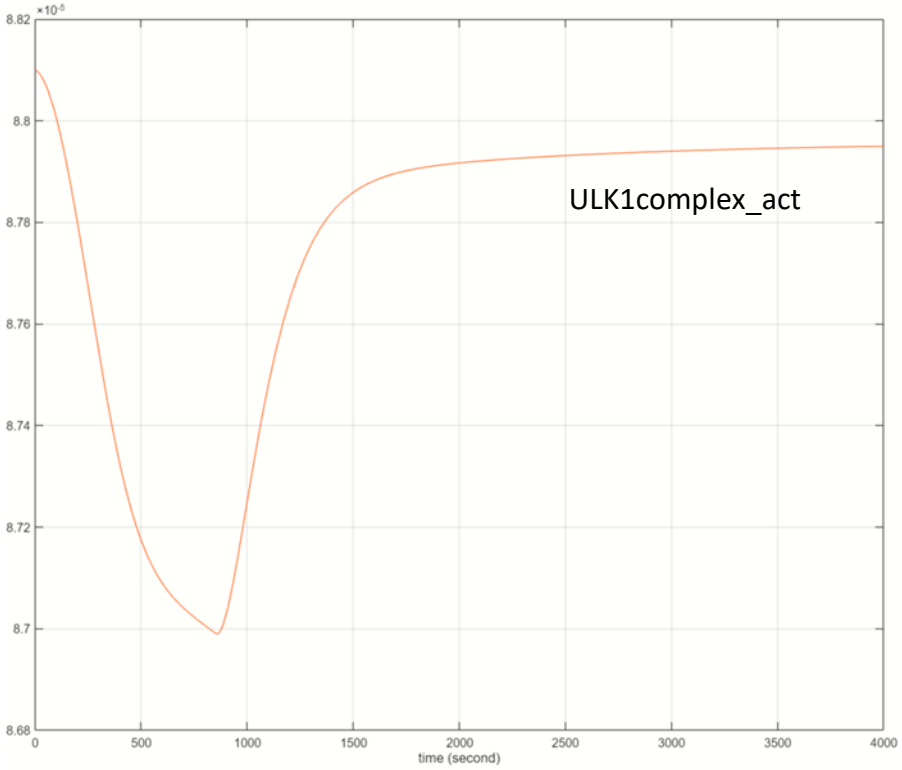
A.


×10^-3
20
18
16
14
12
10
8
6
4
2
0
0
500
1000
1500
2000
2500
3000
3500
4000
time (second)

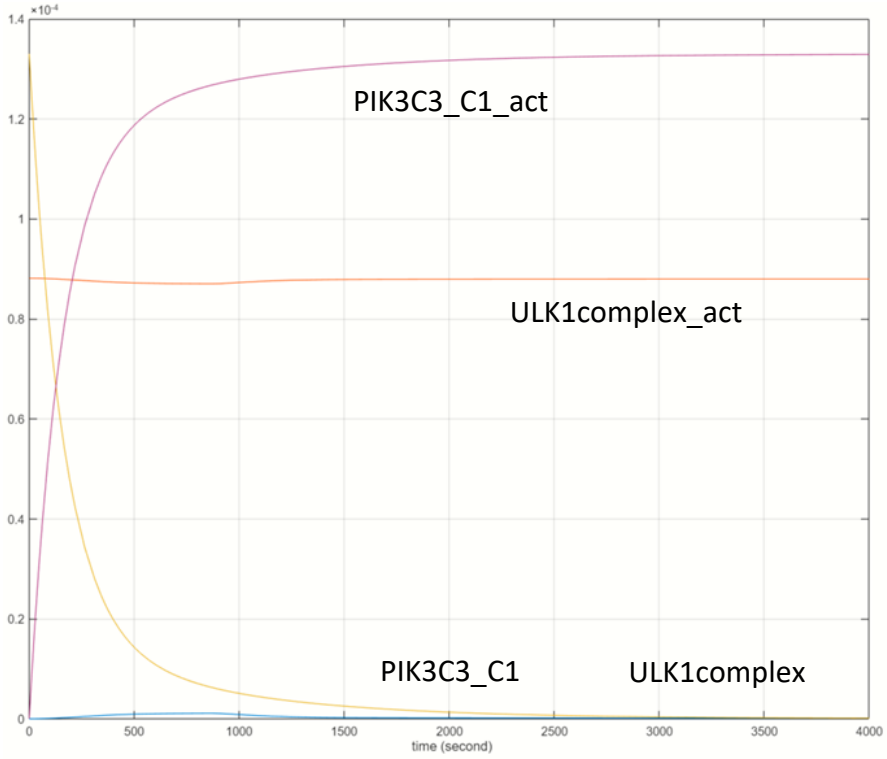
B.


×10^-4
18
16
14
12
10
8
6
4
2
0
-2
0
500
1000
1500
2000
2500
3000
3500
4000
time (second)

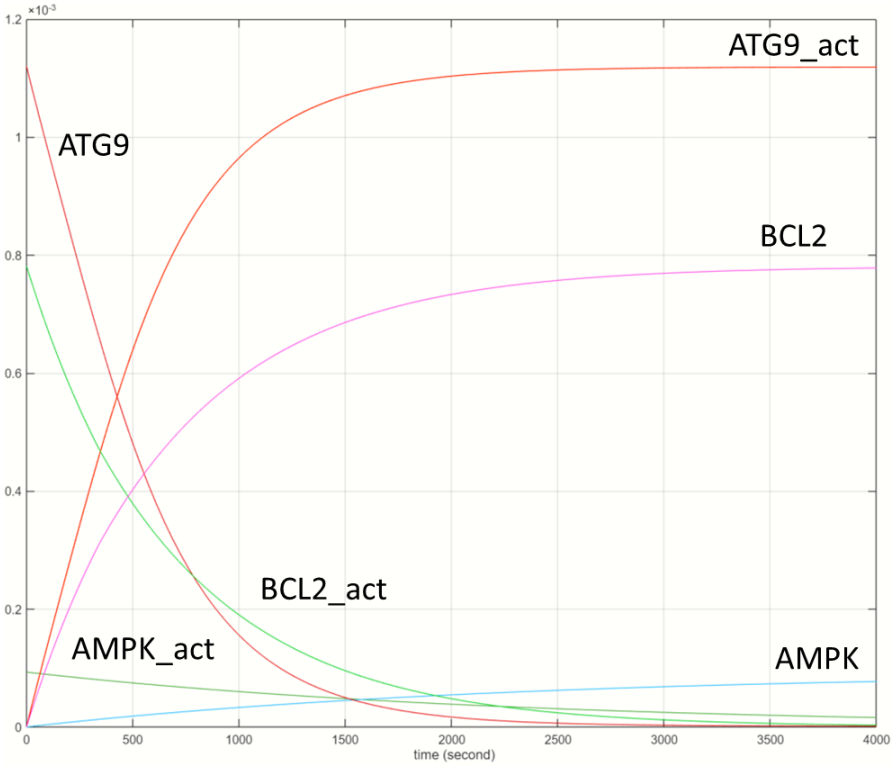
C.


×10^-5
8.82
8.8
8.78
8.76
8.74
8.72
8.7
8.68
0
500
1000
1500
2000
2500
3000
3500
4000
time (second)
ULK1complex_act

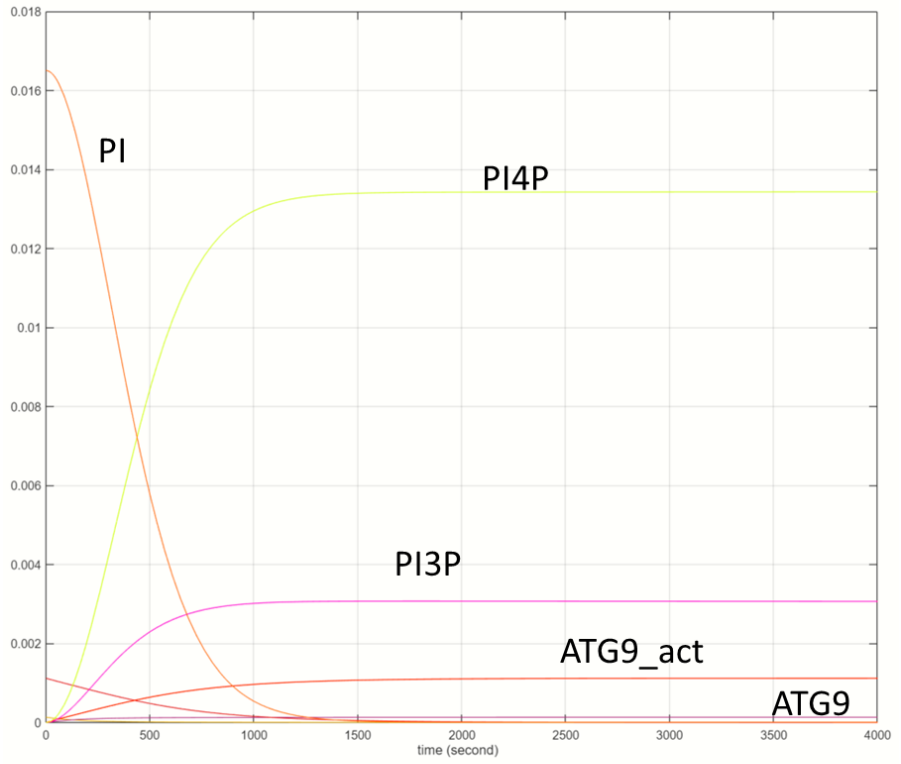
D.


×10^-4
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
500
1000
1500
2000
2500
3000
3500
4000
time (second)
PIK3C3_C1_act
ULK1complex_act
PIK3C3_C1
ULK1complex

E.

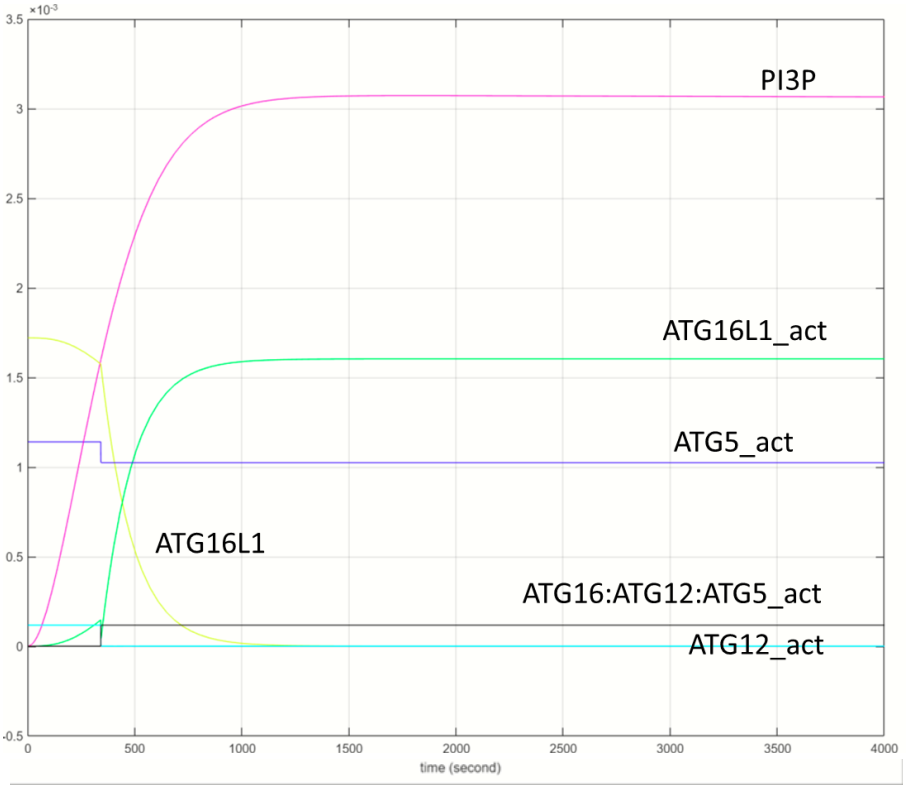


F.

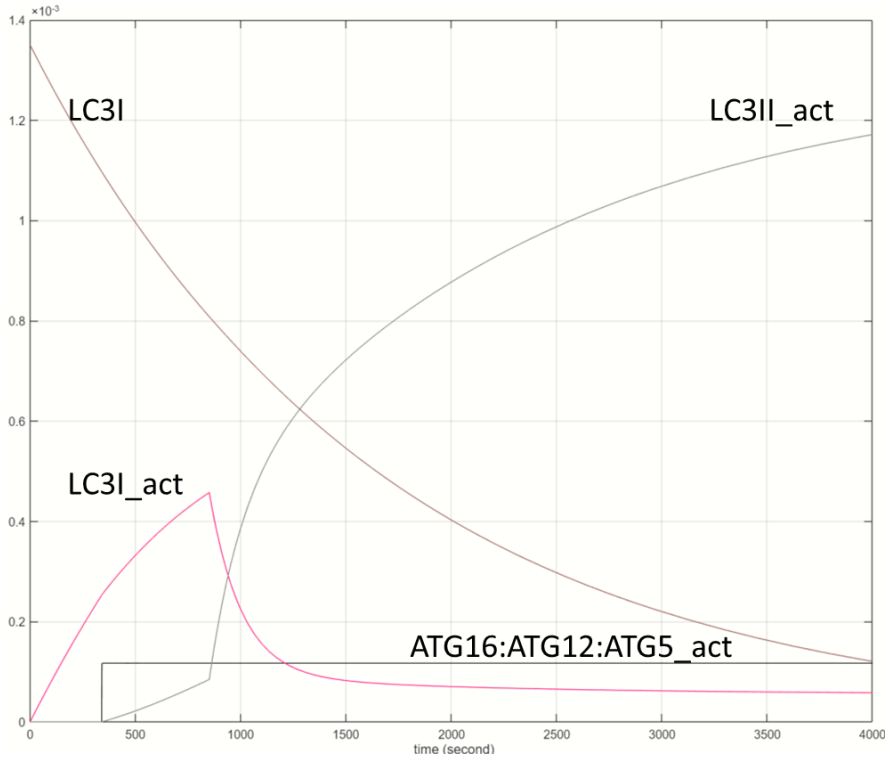


PIK3C3_C1_act PIK3C3_C1

G.


×10⁻³
PI3P
ATG16L1_act
ATG5_act
ATG16L1
ATG16:ATG12:ATG5_act
ATG12_act
time (second)

H.


×10⁻³
LC3I
LC3II_act
LC3I_act
ATG16:ATG12:ATG5_act
time (second)

I.

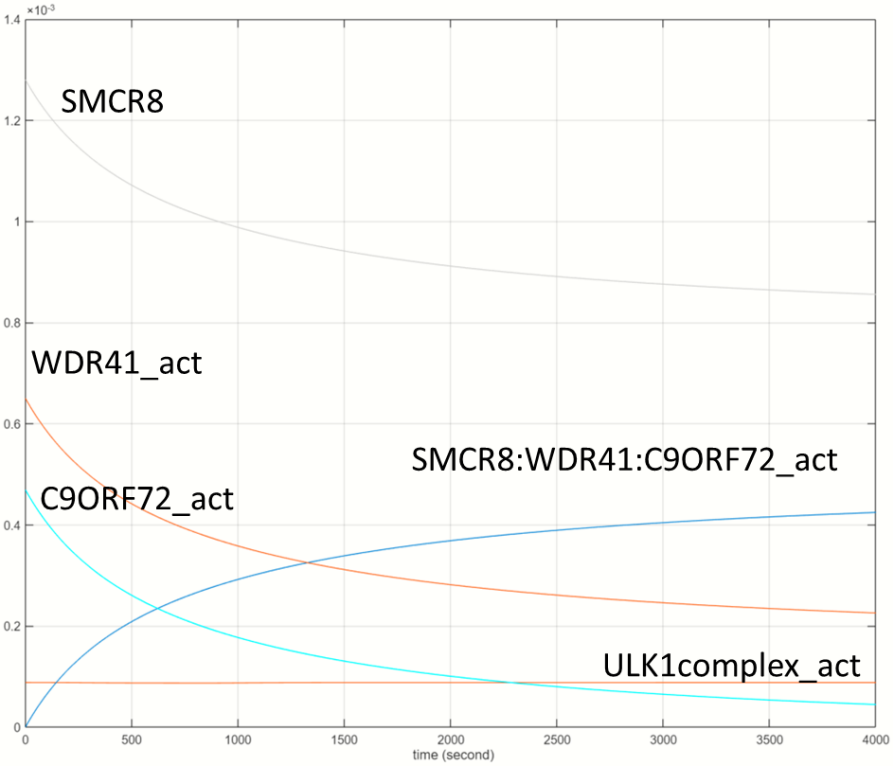


Figure S6-7. The totality of and individual components of macroautophagy simulated with the model for better visualisation- Parkinson's disease.

*(**A**) All the components of the model are included. (**B**) PI3P (pink), PI (orange) and PI4P (yellow) were excluded to allow better visualisation of components with lower amounts. (**C**) ULK1 complex in its activated form. (**D**) PIK3C3-C1 in activated and inactivated form, including the activated form of ULK1 complex for reference. (**E**) ATG9, BCL2, and AMPK in their activated and inactivated forms. (**F**) PI, PI3P, and PI4P, together with ATG9 and PIK3C3 in both their activated and inactivated forms for reference. (**G**) ATG16L1:ATG12:ATG5 and its components, together with PI3P for reference. (**H**) LC3s together with ATG16L1:ATG12:ATG5 for reference. (**I**) SMCR8:WDR41:C9ORF72 and its components, together with the activated form of ULK1complex for reference.*

Analysis for simulating the effect of LRRK2 mutations

Data for the identity and manner in which mutations of LRRK2 affect its interactions were collected from IntAct (on 12/07/2021). Each mutation affecting an interaction had a label of a feature type (e.g., mutation increasing strength, and mutation disrupting rate). The effect of each mutation on each interaction is unknown in most cases. Therefore, it will be assumed that the effect of a decreased rate of interaction with LRRK2, leads to an increased amount of the free form of the LRRK2 interactor, which enhances its concentration/amount. The opposite will be assumed for an increased rate of interaction. It will be assumed that the concentration/amount of the affected component of the model will be changed by 5% in either direction. The resulted effects of each

LRRK2 mutation on components of the model are shown in Table S6-7. For each LRRK2 mutation a version of the model can be created, and the effect can be simulated by changing the amounts of the affected components of the model. The ratio of LC3II and LC3I can be used as a readout to compare macroautophagy among the different mutations.

Table S6-7. Rates of reaction representing the effect of LRRK2 mutations to the model

LRRK2 mutation	Affected LRRK2 interactor (direction of change of interaction by the mutation)	Affected component of model (relationship with LRRK2 interactor)	Affected component of model (direction of change of amount)
R1441C	PRKACA (↓) RAC1 (↓)	LC3I (negatively regulated) PI3P (positively regulated) AMPK (negatively regulated)	LC3I (↓) PI3P (↑) AMPK (↓)
R1441G	PRKACA (↓) PPP1CA (↑) YWHAZ (↓)	LC3I (negatively regulated) AMPK (negatively regulated) CSNK2 (negatively regulated) PIK3C3-C1 (negatively regulated)	LC3I (↓) AMPK (↑) CSNK2 (↑) PIK3C3-C1 (↓)
D1994A	GSK3B (↓) SQSTM1(↓)	ULK1 complex (positively regulated) BCL2 (negatively regulated) LC3I (positively regulated) LC3I (negatively regulated) PIK3C3-C1 (positively regulated)	ULK1 complex (↑) BCL2 (↓) LC3I (↑) LC3I (↓) PIK3C3-C1 (↑)
G2019S	SNCA (↑) GSK3B (↑) SQSTM1 (↑) RAC1 (↓)	BCL2 (positively regulated) ULK1 complex (positively regulated) BCL2 (negatively regulated) LC3I (positively regulated) LC3I (negatively regulated) PIK3C3-C1 (positively regulated) PI3P (positively regulated)	BCL2 (↓) ULK1 complex (↓) BCL2 (↑) LC3I (↓) LC3I (↑) PIK3C3-C1 (↓) PI3P (↑)
G2385R	SQSTM1(↓) CDC37 (↑)	LC3I (positively regulated) LC3I (negatively regulated) PIK3C3-C1 (positively regulated) ULK1 complex (positively regulated)	LC3I (↑) LC3I (↓) PIK3C3-C1 (↑) ULK1 complex (↓)

Table S6-8. Number of molecules of the components of the mathematical model of macroautophagy (in molecules)-Different brain regions											
Gene name	**aveALL**	**CRBL**	**FCTX**	**HIPP**	**MEDU**	**OCTX**	**PUTM**	**SNIG**	**TCTX**	**THAL**	**WHMT**
ULK1	1.80E-03	1.76E-03	1.83E-03	1.83E-03	1.74E-03	1.85E-03	1.96E-03	1.74E-03	1.83E-03	1.79E-03	1.69E-03
ATG101	1.48E-03	1.42E-03	1.54E-03	1.48E-03	1.45E-03	1.53E-03	1.51E-03	1.46E-03	1.52E-03	1.47E-03	1.45E-03
ATG13	1.98E-03	1.89E-03	2.04E-03	1.99E-03	1.95E-03	2.02E-03	2.02E-03	1.94E-03	2.02E-03	1.97E-03	1.97E-03
RB1CC1	1.64E-03	1.62E-03	1.77E-03	1.67E-03	1.51E-03	1.73E-03	1.59E-03	1.55E-03	1.78E-03	1.63E-03	1.54E-03
BECN1	1.29E-03	1.26E-03	1.32E-03	1.28E-03	1.28E-03	1.32E-03	1.28E-03	1.29E-03	1.32E-03	1.29E-03	1.27E-03
PIK3R4	1.83E-03	1.83E-03	1.87E-03	1.81E-03	1.82E-03	1.85E-03	1.85E-03	1.81E-03	1.87E-03	1.82E-03	1.78E-03
PIK3C3	1.82E-03	1.93E-03	1.82E-03	1.80E-03	1.79E-03	1.82E-03	1.84E-03	1.77E-03	1.84E-03	1.77E-03	1.84E-03
NRBF2	1.46E-03	1.41E-03	1.48E-03	1.45E-03	1.47E-03	1.47E-03	1.45E-03	1.45E-03	1.47E-03	1.48E-03	1.45E-03
ATG14	1.45E-03	1.44E-03	1.45E-03	1.45E-03	1.47E-03	1.45E-03	1.43E-03	1.44E-03	1.46E-03	1.42E-03	1.54E-03
AMBRA1	1.64E-03	1.66E-03	1.66E-03	1.64E-03	1.60E-03	1.66E-03	1.67E-03	1.61E-03	1.66E-03	1.63E-03	1.62E-03
ATG9A	1.51E-03	1.52E-03	1.53E-03	1.52E-03	1.49E-03	1.51E-03	1.52E-03	1.51E-03	1.53E-03	1.52E-03	1.45E-03
ATG9B	1.37E-03	1.34E-03	1.38E-03	1.37E-03	1.39E-03	1.38E-03	1.38E-03	1.38E-03	1.37E-03	1.40E-03	1.37E-03
PRKAA2	1.79E-03	1.94E-03	1.96E-03	1.80E-03	1.72E-03	1.86E-03	1.78E-03	1.70E-03	1.94E-03	1.67E-03	1.55E-03
PRKAA1	1.57E-03	1.66E-03	1.54E-03	1.55E-03	1.63E-03	1.54E-03	1.56E-03	1.62E-03	1.55E-03	1.59E-03	1.55E-03
PRKAB1	1.72E-03	1.70E-03	1.74E-03	1.69E-03	1.72E-03	1.76E-03	1.70E-03	1.69E-03	1.72E-03	1.71E-03	1.72E-03
PRKAB2	1.76E-03	1.88E-03	1.78E-03	1.71E-03	1.77E-03	1.76E-03	1.74E-03	1.72E-03	1.77E-03	1.73E-03	1.76E-03
PRKAG1	1.96E-03	1.89E-03	1.98E-03	1.95E-03	1.99E-03	1.96E-03	1.93E-03	2.00E-03	1.99E-03	2.03E-03	1.91E-03
PRKAG2	1.65E-03	1.64E-03	1.71E-03	1.72E-03	1.59E-03	1.65E-03	1.75E-03	1.61E-03	1.70E-03	1.68E-03	1.47E-03
PRKAG3	1.17E-03	1.14E-03	1.16E-03	1.15E-03	1.17E-03	1.17E-03	1.21E-03	1.16E-03	1.16E-03	1.17E-03	1.17E-03
ATG16L1	1.45E-03	1.49E-03	1.50E-03	1.43E-03	1.39E-03	1.53E-03	1.48E-03	1.39E-03	1.50E-03	1.41E-03	1.39E-03
RAB33B	1.64E-03	1.76E-03	1.61E-03	1.59E-03	1.67E-03	1.59E-03	1.56E-03	1.62E-03	1.60E-03	1.66E-03	1.70E-03
CSNK2A1	1.82E-03	1.84E-03	1.81E-03	1.82E-03	1.83E-03	1.80E-03	1.81E-03	1.83E-03	1.80E-03	1.83E-03	1.82E-03
CSNK2A2	1.85E-03	1.77E-03	1.92E-03	1.87E-03	1.81E-03	1.89E-03	1.87E-03	1.82E-03	1.92E-03	1.85E-03	1.81E-03
CSNK2B	1.91E-03	1.93E-03	1.90E-03	1.90E-03	1.90E-03	1.91E-03	1.91E-03	1.92E-03	1.90E-03	1.89E-03	1.91E-03

Table S6-8. (continued) Number of molecules of the components of the mathematical model of macroautophagy (in molecules)-Different brain regions											
BCL2	1.50E-03	1.29E-03	1.49E-03	1.51E-03	1.55E-03	1.48E-03	1.52E-03	1.60E-03	1.50E-03	1.59E-03	1.51E-03
ATG12	1.48E-03	1.54E-03	1.48E-03	1.47E-03	1.48E-03	1.46E-03	1.44E-03	1.47E-03	1.49E-03	1.46E-03	1.49E-03
ATG5	1.54E-03	1.52E-03	1.64E-03	1.56E-03	1.48E-03	1.59E-03	1.51E-03	1.49E-03	1.63E-03	1.51E-03	1.49E-03
MAP1LC3B	1.75E-03	1.87E-03	1.74E-03	1.71E-03	1.76E-03	1.76E-03	1.70E-03	1.77E-03	1.75E-03	1.73E-03	1.73E-03
MAP1LC3A	1.65E-03	1.66E-03	1.67E-03	1.66E-03	1.63E-03	1.65E-03	1.59E-03	1.68E-03	1.67E-03	1.66E-03	1.60E-03
SMRC8	1.58E-03	1.64E-03	1.64E-03	1.59E-03	1.52E-03	1.64E-03	1.59E-03	1.54E-03	1.63E-03	1.55E-03	1.50E-03
WDR41	1.84E-03	1.80E-03	1.88E-03	1.83E-03	1.84E-03	1.85E-03	1.84E-03	1.83E-03	1.87E-03	1.85E-03	1.83E-03
C9ORF72	1.50E-03	1.83E-03	1.51E-03	1.41E-03	1.50E-03	1.51E-03	1.34E-03	1.48E-03	1.52E-03	1.47E-03	1.43E-03
CDC37	1.89E-03	1.83E-03	1.95E-03	1.89E-03	1.88E-03	1.93E-03	1.86E-03	1.88E-03	1.94E-03	1.90E-03	1.85E-03
HSP90AA1	1.75E-03	1.68E-03	1.72E-03	1.74E-03	1.80E-03	1.72E-03	1.72E-03	1.76E-03	1.72E-03	1.76E-03	1.91E-03
GSK3B	1.94E-03	2.00E-03	2.01E-03	1.95E-03	1.92E-03	1.99E-03	1.90E-03	1.94E-03	2.02E-03	1.96E-03	1.79E-03
SNCA	2.18E-03	2.14E-03	2.32E-03	2.21E-03	2.19E-03	2.23E-03	2.10E-03	2.22E-03	2.33E-03	2.01E-03	2.13E-03
PPP1CA	1.80E-03	1.50E-03	1.68E-03	1.81E-03	2.02E-03	1.70E-03	1.74E-03	1.90E-03	1.70E-03	1.79E-03	2.15E-03
PRKACA	1.71E-03	1.72E-03	1.77E-03	1.73E-03	1.70E-03	1.75E-03	1.65E-03	1.73E-03	1.76E-03	1.77E-03	1.58E-03
YWHAZ	1.64E-03	1.64E-03	1.67E-03	1.67E-03	1.61E-03	1.65E-03	1.65E-03	1.62E-03	1.68E-03	1.65E-03	1.59E-03

Note: The abbreviations CRBL, FCTX, HIPP, MEDU, OCTX, PUTM, SNIG, TCTX, THAL, and WHMT refer to the following brain areas: cerebellar cortex, frontal cortex, hippocampus, medulla (inf olivary nucleus), occipital cortex, putamen, substantia nigra, temporal cortex, thalamus, and intralobular white matter, respectively. The first category, "averALL", is simply the average of all the brain areas, as calculated by Braineac.