LRRK2 Protein Interactome in Health and Parkinson's Disease

Yibo Zhao

A thesis submitted for the degree of Doctor of Philosophy

December 2023

Declaration of original authorship

I, Yibo Zhao confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Background: Mutations in LRRK2 are the most common genetic cause of familial Parkinson's disease (PD). However, the physiological/pathological roles of LRRK2 are still unclear. This project aims at collecting the vast data of LRRK2 research and integrating them in a homogeneous, computational model to describe LRRK2's activity in physiological and pathological pathways. **Method:** The protein interactome of LRRK2 was built in consecutive steps: 1) peer-reviewed protein-protein interactions (PPIs) were derived to define the general LRRK2 interactome ($LRRK2_{int}$); 2) interactions among LRRK2 interactors were derived to construct the LRRK2-centred protein-protein interactions network ($LRRK2_{net}$); 3) topological analysis was performed on the $LRRK2_{net}$ to identify clusters of LRRK2 interactors with high-density PPIs; 4) each topological cluster was functionally annotated; 5) the $LRRK2_{net}$ were then merged with transcriptomic data of healthy human tissues to define the tissue specificity LRRK2 interactors; 6) the $LRRK2_{net}$ was merged with transcriptomic and genomic data of patients with sporadic PD, LRRK2 genetic PD and healthy controls to investigate the molecular mechanisms of the 2 types of PD. **Results:** A total of 418 proteins were included in the $LRRK2_{int}$, involving a range of different protein families. PPIs among these interactors formed a scale-free network, in which 7 topological clusters with biological significance were identified, associated with ribosomal biosynthesis, cytoskeleton dynamics, synaptic transport, mitophagy and protein metabolism. LRRK2 interactors presented distinct expression signatures and functional patterns in the brain and the periphery. Of note, a striatal unit of putamen, caudate and nucleus accumbens was identified where LRRK2 interactors presented the highest co-expression with LRRK2 and similar expression profiles. At last, PD-associated expression analysis identified 100 LRRK2 interactors with significant differential expression in PD cases vs. control. **Conclusion:** This study defined a comprehensive protein interactome of LRRK2 with high tissue specificity and substantial association with sporadic PD and LRRK2-related PD.

Impact Statement

This study offers a comprehensive exploration of the protein interactome associated with LRRK2, a protein closely implicated in Parkinson's disease (PD) and other inflammatory conditions like Crohn’s disease (CD). The generated LRRK2 interactome was meticulously annotated to unveil a "functional core," comprising a conserved list of LRRK2 interactors participating across diverse biological processes enriched in the LRRK2 interactome (LRRK2int). This functional core represents the foundational elements of the LRRK2 functional network and highlights critical players in LRRK2-associated signalling pathways.

Moreover, PPIs among LRRK2 interactors were integrated into a network model, unveiling a "topological core" comprised of highly-centred proteins forming the structural backbone of the LRRK2 interaction network. The combination of the "functional core" and "topological core," offers a concise list of potential hubs in LRRK2 signalling pathways, presenting promising avenues for therapeutic intervention in both sporadic PD (sPD) and LRRK2-PD. Identifying these hubs can guide targeted drug development strategies, focusing on key players in the intricate network of LRRK2-mediated processes.

Furthermore, the study detected 7 topological clusters within the LRRK2 network, each enriched for distinct biological functions. These clusters serve as valuable starting points for pathway modelling, facilitating a deeper understanding of the player implicated in specific LRRK2-mediated signalling cascades associated with functions such as ribosomal biosynthesis, cytoskeleton dynamics, synaptic transport, mitophagy, and protein metabolism. Exploring these clusters provides insights into the multifaceted roles of LRRK2 in cellular processes and helps unravel the intricacies of its functional contributions.

The investigation also emphasizes the tissue-specificity of the LRRK2 interactome, highlighting the need to differentiate LRRK2's role in the brain and peripheral tissues. This insight is crucial for LRRK2-focused research and targeted drug development, acknowledging the distinct functions and implications of LRRK2 in different tissues. The study suggests the development of a "trimmed-LRRK2int" by filtering the general LRRK2 interactome based on expression and co-expression levels in each tissue. This refined dataset enhances the tissue-

specificity of functional research, reducing complexity and focusing on key participants in LRRK2 signalling pathways within specific tissues.

Additionally, the study provides novel insights into the early-stages of PD pathology by identifying alterations in LRRK2 interactors in both the sPD and LRRK2-PD conditions. While further research is warranted to validate these changes, they offer a foundation for the understanding of the molecular differences of sporadic vs familial PD (which are currently considered as one single disease) and aid in the development of potential therapeutic strategies tailored to sporadic and LRRK2-associated Parkinson's disease. Overall, this research significantly contributes to our understanding of LRRK2's role in health and disease, providing a roadmap for future investigations and therapeutic developments in the realm of neurodegenerative disorders.

Research Paper Declaration Form

The research presented in this thesis contributed to the following manuscripts:

1. **Published: Zhao Y**, Vavouraki N, Lovering RC, Escott-Price V, Harvey K, Lewis PA, Manzoni C. ***Tissue specific LRRK2 interactomes reveal a distinct striatal functional unit.*** PLoS Comput Biol. 2023 Jan 30;19(1):e1010847. doi: 10.1371/journal.pcbi.1010847. PMID: 36716346; PMCID: PMC9910798.

2. **Preprint: Zhao, Y.**, Harvey, K., Escott-Price, V., Lewis, P. A., & Manzoni, C. (2023). ***Transcriptomics analyses of the LRRK2 protein interactome reveal distinct molecular signatures for sporadic and LRRK2 Parkinson's Disease.*** BioRxiv, 2023.09.12.557373. https://doi.org/10.1101/2023.09.12.557373

3. **Published:** Shao L, **Zhao Y**, Heinrich M, Prieto JM, Manzoni C. ***Active natural compounds perturb the melanoma risk-gene network.*** G3 (Bethesda). 2023 Nov 30:jkad274. doi: 10.1093/g3journal/jkad274. Epub ahead of print. PMID: 38035793.

4. **Preprint:** Tombesi, G., Kompella, S., Favetta, G., Chen, C., **Zhao, Y.**, Sevegnani, M., Marte, A., Battisti, I., Morosin, E., Ornaghi, M., Iannotta, L., Plotegher, N., Civiero, L., Onofri, F., Eickholt, B. J., Piccoli, G., Arrigoni, G., Beccano-Kelly, D., Manzoni, C., … Greggio, E. (2023). ***LRRK2 regulates synaptic function through BDNF signaling and actin cytoskeleton***. *BioRxiv*, 2022.10.31.514622. https://doi.org/10.1101/2022.10.31.514622

5. **Preprint:** Lucia Iannotta, Rachel Fasiczka, Giulia Favetta, **Yibo Zhao**, Elena Giusto, Elena Dall'Ara, Jianning Wei, Franz Ho, Claudia Ciriani, Susanna Cogo, Isabella Tessari, Ciro Iaccarino, Maxime Liberelle, Luigi Bubacco, Jean-Marc Taymans, Claudia Manzoni, Arjan Kortholt, Laura Civiero, Sabine Hilfiker, Michael Lu, Elisa Greggio. ***PAK6 rescues pathogenic LRRK2-mediated ciliogenesis and centrosomal cohesion defects in a mutation-specific manner.*** BioRxiv 2024.04.11.589075; doi: https://doi.org/10.1101/2024.04.11.589075

Acknowledgement

I would like to express my deepest gratitude to my supervisor team, Dr Claudia Manzoni, Professor Kirsten Harvey and Professor Patrick A. Lewis for their unwavering guidance, insightful feedback, and continuous support throughout the entire journey of my doctoral research. Their mentorship has been invaluable, and I am truly grateful for the time and expertise they generously shared.

I extend my sincere appreciation to Professor Valentina Escott-Price and Professor Ruth Lovering for their valuable guidance and support for my publications. I also extend my thanks to my colleagues and fellow researchers who shared their ideas, provided assistance, and created a supportive academic community. Their camaraderie made the research process more enjoyable and fulfilling.

I am thankful to the patients and their families who participated in the PPMI study. Part of the data used in the preparation of this thesis were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/access-data-specimens/download-data), RRID:SCR_006431. For up-to-date information on the study, visit www.ppmi-info.org PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, Abbvie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Janssen Neuroscience, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation and Yumanity Therapeutics

I would like to acknowledge the studentship provided by UCL School of Pharmacy, which made this research possible. Their investment in my work has been instrumental in the successful completion of this thesis.

Heartfelt thanks go to my parents for their unwavering encouragement, understanding, and love. Their support sustained me during challenging times, and I am grateful for the sacrifices they made to see me succeed.

Table of Contents

Table of figures

Table of tables

Introduction

1.1 LRRK2 (Leucine-rich Repeat Kinase 2)

LRRK2 is a is a large, multifunctional protein possessing both kinase and GTPase enzymatic domains. In 2004, mutations in the LRRK2 gene were firstly associated with familial Parkinson's disease (PD) (Paisán-Ruíz et al., 2004; Zimprich et al., 2004). Since then, considerable efforts have been made to understand the exact functions of LRRK2 in the physiological and pathological scenarios. It has been associated with a wide range of cellular processes, such as autophagy, cytoskeleton organisation and immune response. However, due to both its complex structure and its multifaceted involvement in the regulation of cellular homeostasis, the exact role of LRRK2 is still unclear.

1.1.1 Structure of LRRK2

LRRK2 is a large, complex protein consisting of 2527 amino acids with a molecular weight of 285 kDa. It belongs to the ROCO protein superfamily. The term "ROCO" is derived from the unique combination of two functional domains: ROC (Ras of Complex) and COR (C-terminal of ROC), in which the ROC domain exhibits GTPase activity while the COR domain is the hub of protein-protein interactions. In addition, LRRK2 contains a serine-threonine kinase domain, which is linked to the GTPase domain by the COR domain (**Figure 1**). With these 2 distinct enzymatic domains, LRRK2 is engaged in a variety of complex biological process, either as a direct participant or as a regulator. Of note, previous studies have found that the kinase domain of LRRK2 is regulated by its GTPase domain (A. P. T. Nguyen & Moore, 2017). Possibly, the reverse is also true as LRRK2:LRRK2 self-interaction has been reported in multiple studies with assorted detection methods (Manzoni et al., 2015) and the kinase domain of LRRK2 is able to phosphorylate the LRRK2 binding partner at different residues (those mainly studied are Ser935 and Ser910) (Marchand et al., 2020). Consequently, the interplay between the 2 enzymatic domains adds another layer of complexity to the study of LRRK2 also considering that these self-interactions are affected by LRRK2 pathogenic mutations in the ROC-COR and kinase domains, thereby leading to enzymatic dysregulation (Sen et al., 2009). LRRK2 contains 4 protein-protein interaction domains, namely an Armadillo (ARM) region, an Ankyrin (ANK) region, a leucine-rich repeat (LRR) domain, and a WD40 domain. These domains enable another crucial role of LRRK2: a scaffold protein, serving as a platform for the assembly of

protein complexes in cellular signalling events. For example, previous studies suggest that LRRK2 may function as a scaffold in Wnt sigalling pathway, ASK1-mediated neuronal apoptosis pathway (Berwick & Harvey, 2012; Yoon et al., 2017). Many single nucleotide polymorphisms (SNPs) have been found in the *LRRK2* gene, all coding and associated with aminoacid changes along the entire LRRK2 molecule. However, the pathogenicity of these variants is not fully understood. The recognized Parkinson's Disease (PD) -related pathogenic LRRK2 mutations concentrate in the LRRK2 catalytic core: the ROC domain contains the R1441C/G/H and the N1437D/H mutations; the COR domain contains the Y1699C and the R1628P mutations; and the kinase domain contains the G2019S and I2020T mutations (**Figure 1**) (Cookson, 2010; Mills et al., 2012). Only 1 PD risk variant occurs outside the catalytic core (G2385R) (Tezuka et al., 2022). This implies that the enzymatic activities of LRRK2, rather than other functions of LRRK2, might be associated with PD development. However, how these mutations and the corresponding LRRK2 structural/functional alterations lead to PD is still unclear.

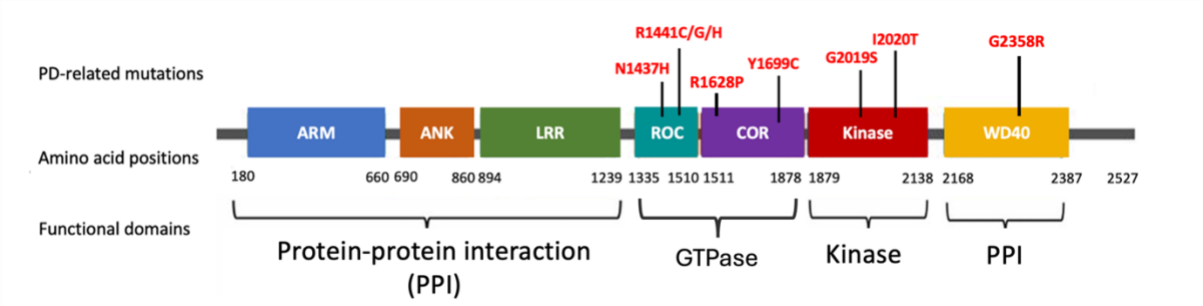


Figure 1. LRRK2 domain structure and PD-related pathogenic mutations

LRRK2 is a multidomain proteins comprised of 1 GTPase domain (ROC domain), 1 kinase domain and 5 protein-protein interaction domains (ARM, ANK, LRR, COR and WD40 domain). A total of 7 pathogenic mutations of LRRK2 have been identified in the previous studies, located in different domains of the LRRK2 protein. Abbreviations: ARM: Armadillo; ANK: Ankyrin; LRR: leucine-rich repeat; ROC: Ras of Complex; COR: C-terminal of ROC. Originally designed with Power Point.

1.1.2 Localisation of LRRK2

LRRK2 is prominently expressed in the periphery, especially in lung, liver and kidney (**Figure 2**) (Madureira et al., 2020; Paisán-Ruíz et al., 2004). It also presents a high expression level in the whole blood, as compared to other biofluids such as urine and cerebrospinal fluid (CSF). On the contrary, LRRK2 exhibits relatively lower expression levels in the brain regions. It has

been shown that LRRK2 was mainly detected in dopamine-innervated regions such as frontal cortex, striatum and cerebellum, while at a lower level in dopaminergic area like *substantial nigra* (Biskup et al., 2006; Gaiter et al., 2006; Higashi et al., 2007). However, reports regarding LRRK2's distribution in the brain seems to vary according to the different detection techniques used. Hence, further confirmation is required for the expression profile of LRRK2 in the Central Nervous System (CNS). From a cell-specific point of view, LRRK2 presents higher expression levels in peripheral innate and adaptive immune cells (B cells, monocytes, and neutrophils) (Hakimi et al., 2011a; Westerlund et al., 2008). In the brain, higher LRRK2 expression is observed in glia cells (especially in microglia) in comparison with neurons (Moehle et al., 2012). In cells, LRRK2 is mainly found associated with the intracellular membranes and vesicular structures throughout the cytoplasm, such as endosomes, lysosomes, and synaptic vesicles, Golgi complex and outer mitochondrial membrane (Alegre-Abarrategui et al., 2009). It has been suggested that alterations of LRRK2 expression level are probably associated with PD. For example, significantly increased LRRK2 expression was found in immune cells (B cells, and T cells) of PD patients as compared to healthy controls (Cook et al., 2017a). Another study observed decreased LRRK2 level in sigmoid colon biopsy specimens from PD patients (De Guilhem De Lataillade et al., 2021).

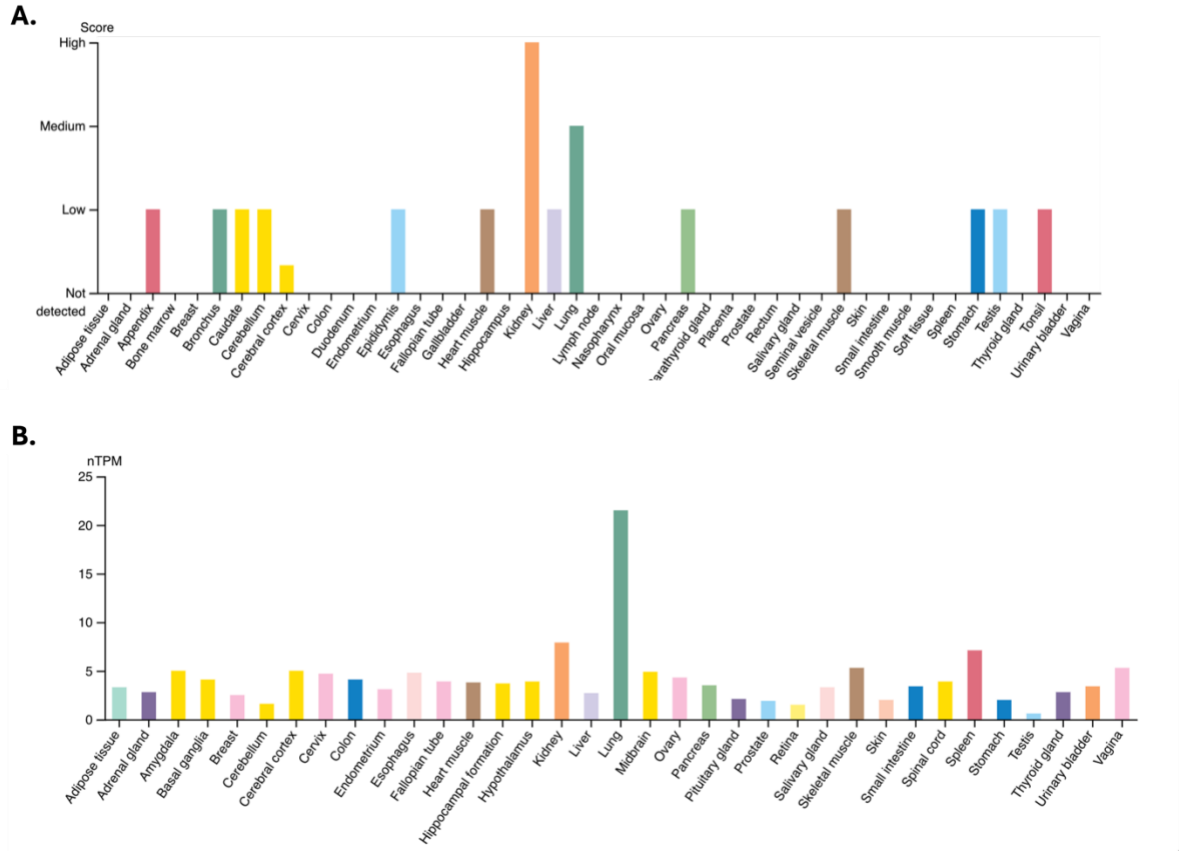


Figure 2. LRRK2 expression profiles in different tissues

(Adapted from The Human Protein Atlas, https://www.proteinatlas.org/) (Sjöstedt et al., 2020)

1.1.3 Functions of LRRK2

Since 2004, when mutations in *LRRK2* were firstly associated with PD, a great amount of research has been conducted to understand the role of LRRK2 in health and disease, associating LRRK2 with a wide range of cellular functions.

1.1.3.1 Autophagy

Autophagy is a highly orchestrated lysosomal degradative process designed to degrade and recycle obsolete cellular components, such as aged or dysfunctional organelles and proteins. Dysregulation of autophagy can lead to the accumulation of misfolded proteins and other cellular debris, contributing to the pathogenesis of various neurodegenerative disorders; in PD reduced autophagy has been linked with the accumulation of a-synuclein, the main constituent of Lewy bodies (LB) (Lu et al., 2020). There are 3 main types of autophagy: macroautophagy, microautophagy and chaperone-medicated autophagy (CMA). Previous studies have associated LRRK2 with macroautophagy and CMA, while no sufficient evidence support its linkage with microautophagy, which may be due to the fact that microautophagy is the least characterised form of autophagy.

1.1.3.1.1 LRRK2 and macroautophagy

Under normal circumstance, macroautophagy is primarily a cytoprotective mechanism, involving the formation and intracellular transport of autophagosomes (a type of intracellular double-membrane vesicles to sequester waste molecules), autophagosome-lysosome fusion and the degradation and recycling of cell waste in lysosome. However, dysregulated macroautophagy in disease scenario can lead to cell death. Previous studies have suggested a profound association between LRRK2 and macroautophagy. For example, increased number of autophagosomes was identified in rat neurons while overexpressing LRRK2-G2019S, a PD pathogenic variant that increases LRRK2's kinase function, in comparison with neurons overexpressing wild-type LRRK2 or kinase-dead LRRK2 mutant (D. MacLeod et al., 2006). Similar alterations have been observed in fibroblasts from LRRK2-G2019S PD patients (Bravo-San Pedro et al., 2013). Furthermore, pharmacological inhibition of LRRK2 function seems to

decrease macroautophagy in monocyte cell lines (Bravo-San Pedro et al., 2013). However, LRRK2 kinase inhibition shows opposite effect in in primary astrocytes, HEK293T cells, H4 neuroglioma cells and SH-SY5Y neuroblastoma cells, in which it stimulates macroautophagy, suggesting that the regulation of LRRK2 on the process of macroautophagy is primarily dependent on its kinase function and is potentially cell-type specific (Manzoni et al., 2013, 2016; Saez-Atienzar et al., 2014).

The mechanism by which LRRK2 modulates macroautophagy is still unclear. Previous studies have found that LRRK2 may regulate the early and late stage of macroautophagy by modulating autophagy signal induction and lysosomal function via its kinase activities. In the normal scenario, macroautophagy is induced by cellular response to stress, involving the inhibition of mammalian target of rapamycin (mTOR). However, multiple lines of evidence have suggested that LRRK2-mediated autophagy induction is independent from mTOR pathway, although the down-stream alterations are similar. For example, it was reported that inhibitor of mTOR-activated autophagy has no effect on the macroautophagy induced by LRRK2-G2019S overexpression in in HEK293T cells (Gómez-Suaga et al., 2012). Similar results were found in H4 neuroglioma cells (Manzoni et al., 2016). Rather, LRRK2-mediated autophagy was deactivated by the inhibition of the Ca^{2+}/CaMKK (calcium-dependent protein kinase kinase)/AMPK (adenosine monophosphate activated protein kinase) pathway. In addition, while under long-term stress, cells with mutant LRRK2 exhibit significantly higher death rate, which can be alleviated by enhancing the mTOR-mediated autophagy (Gómez-Suaga et al., 2012). These results indicate that LRRK2-mediated and mTOR-mediated macroautophagy are independent but interplay with each other. Furthermore, LRRK2 mutants have been found delaying autophagosome-lysosome fusion by interrupting autophagosomes transportation and increased lysosomal Ca^{2+} levels and reduced lysosomal Ca^{2+} release in both mouse and human neurons, which may further perturb the early stage of macroautophagy (Boecker et al., 2021; Dou et al., 2022). The last stage of macroautophagy involves waste degradation and recycling in lysosome. Previous studies have revealed that LRRK2-G2019S and LRRK2-R1441C lead to significant increase in lysosomal pH and thereby causing decreased lysosomal protein degradation in mouse and rat neurons (Schapansky et al., 2014; R. Wallings et al., 2019). In addition, increased lysosomal count and decreased lysosomal size were observed in these neurons. These changes lead to increased intracellular

accumulation of α-synuclein, which is the hallmark of PD, and neuronal α-synuclein release into the cell culture media, suggesting that LRRK2-related lysosomal function changes and sequential autophagy dysregulation are associated with PD progression.

1.1.3.1.2 LRRK2 and CMA

Compared to macroautophagy, CMA is a highly selective form of autophagy that targets individual proteins for degradation. These targeted proteins typically contain a specific amino acid motif, called the KFERQ-like motif, which can be recognised by the chaperone protein HSC70. Subsequently, the protein:chaperon complex is translocated to the lysosomal surface, where HSC70 binds to the lysosomal protein LAMP2A and facilitating the transport of the target protein into the lysosome for degradation by proteases. Aberrant CMA is highly linked to PD pathology since it is responsible for accumulation of α-synuclein, which is the hallmark of the disease pathology (Cuervo et al., 2004). Additionally, LRRK2 is also a substrate of CMA and a protein interactor of LAMP2A. Previous studies suggest that LRRK2-G2019S mutant presents more stable binding with LAMP2A and seem to prevent its multimerization and formation of the lysosomal membrane translocation complex, leading to aberrant degradation of other CMA substrates including α-synuclein. This hypothesis has been supported by observations in both human and mouse neurons. For example, impaired CMA has been observed in dopaminergic neurons from PD patients with LRRK2-G2019S variant (Sánchez-Danés et al., 2012). Similar alterations have been found in PD mouse model with LRRK2-R1441C (P. W. L. Ho et al., 2020). In addition, accumulation of α-synuclein was observed in these mouse neurons.

1.1.3.2 LRRK2 and synaptic function

Synaptic functionality is crucial for efficient communication between neurons. Synaptic communication encompasses the synthesis and release of neurotransmitter-loaded synaptic vesicles from the presynaptic terminal to the postsynaptic terminal, where vesicles are merged with cell membrane and the neurotransmitters are released to bind with receptors. At the end, the vesicles are either re-uptaken for recycling or enzymatically degraded. It is vulnerable to multifactorial neuronal stress, such as abnormal protein aggregation, mitochondrial dysfunction, oxidative stress and inflammation, etc. Collective loss or impairment of synaptic health leads to neurodegenerative diseases. For example, loss of

dopaminergic synapses in the *substantia nigra* contributes to typical motor symptoms in PD. Previous studies have associated LRRK2 with both presynaptic and post synaptic functions.

1.1.3.2.1 The role of LRRK2 in the presynaptic site

Multiple lines of evidence have linked *LRRK2* pathogenic mutation with aberrant presynaptic function. For example, decreased dopamine release in the striatum has been observed in a LRRK2-G2019S mouse model. Intraneuronal metabolism analysis showed that extracellular metabolites:dopamine ratios in the striatum were significantly higher in LRRK2-G2019S mutant mice, suggesting that in the presence of the *LRRK2* variant, newly synthesised dopamine is trapped in presynaptic terminal (Hammes et al., 2019a). This can either due to the impaired vesicle packing and/or disrupted exocytotic release process in the presynaptic site. Furthermore, another study found that overexpression of wild-type LRRK2 and LRRK2-G2019S increased Ca^{2+} current density in the presynaptic membrane, which dramatically influenced neurotransmitter vesicle release (Bedford et al., 2016). Collective upregulated Ca^{2+} signalling may affect other organelle that crucial for maintaining presynaptic homeostasis, such as mitochondria. Presynaptic mitochondria are important for providing ATP to continuously support synaptic activity and buffering Ca^{2+} signals. Therefore, LRRK2 mutant induced Ca^{2+} increase may overwhelm mitochondria's buffering capability and thereby further perturb vesicle release in the presynaptic terminal.

The molecular mechanism behind these alterations is not fully understood yet, though LRRK2 has been shown to interact and phosphorylate a panel of presynaptic proteins, among which Rab proteins are unambiguous substrates of LRRK2's kinase domain. Rabs are a group of small GTPase functioning as key regulators of membrane trafficking and selective vesicle targeting (Jeong et al., 2018). These proteins participate in each single step of synaptic vesicle transport, from the formation of endosome in endoplasmic reticulum (ER) and the Golgi (RAB1 and RAB43), intracellular vesicle transport (RAB3 and RAB8), vesicle-membrane fusion (RAB8 and RAB35) to endosome recycling and exocytosis (RAB10) (Steger et al., 2016a). LRRK2 phosphorylation blocks the effector binding sites in Rab proteins so that they get trapped on intracellular membranes and unable to bind with upstream and downstream protein interactors, thereby perturbing vesicle endocytosis. In addition, LRRK2 phosphorylates presynaptic N-ethylmaleimide sensitive fusion (NSF) protein and enhances its ATPase activity

(Belluzzi et al., 2016). NSF is a housekeeping protein which is broadly required for intercellular membrane fusion. It also plays an important role in sustaining the pool of synaptic vesicles that are readily available for release. Furthermore, LRRK2 has been found to directly interact with synaptic vesicle via the synapsin I protein via its kinase domain and WD40 protein-protein interaction domain (Marte et al., 2019). These findings suggest that the understanding of the protein-protein interaction network around LRRK2 is the key to understand its role in regulating presynaptic functions.

1.1.3.2.2 The role of LRRK2 in the postsynaptic site

When the synaptic vesicles diffuses across the cleft and bind with receptors at the postsynaptic site, the vesicles fuse with the postsynaptic neuron membrane and release the neurotransmitter, leading to the sequential neuron activation. Within this process, LRRK2 has been shown to affect the expression and activity of postsynaptic receptor, the endocytosis of synaptic vesicles and the postsynaptic morphology. Studies have suggested that the postsynaptic expression of glutamatergic AMPAR and dopamine D1 receptor (D1R) could be affected by overexpression of wild-type LRRK2 and its pathogenic variants. Of note, LRRK2 is primarily expressed in glutamatergic and dopaminergic projection regions such as striatum, where the postsynaptic terminals are located, while hardly expressed in substantia nigra where presynaptic dopamine neurons are located. Therefore, these findings suggest that LRRK2-mediated dopamine synaptic transmission impairment may mainly occur at the postsynaptic sites and then reflect at presynaptic sites. In fact, it has been suggested that postsynaptic overexpression of LRRK2 leads to retrograde upregulation of presynaptic neurotransmitter release (Foffani & Obeso, 2018; Kuhlmann & Milnerwood, 2020; Lamonaca & Volta, 2020). Further research is required to understand this regulation mechanism of LRRK2. In addition, aberrant synaptic vesicle endocytosis has been observed in ventral midbrain neurons in the presence of LRRK2-G2019S variant. This alteration can be rescued by LRRK2 kinase inhibition (Pan et al., 2017). However, no such change was observed in other brain regions, suggesting that the regulatory role of LRRK2 is probably brain-region specific. Furthermore, a critical role of LRRK2 in modulating dendritic spine morphology has been reported. Increased LRRK2 kinase activity induced by extracellular stimuli or pathogenic mutations (LRRK2-G2019S) engages LRRK2 with a macromolecular complex required to regulate actin dynamics during postsynaptic plasticity. Additionally, enhanced kinase function

significantly affects its protein-protein interactions with cytoskeleton proteins such as drebrin, ACTR2, ACTR3 and ARPC2, etc., which are important for spine maturation (Tombesi et al., 2022).

Taken together, by interacting with a panel of essential proteins at both presynaptic and postsynaptic terminals, LRRK2 plays a crucial and complicated role in synaptic signal transduction. Therefore, protein interaction analysis might bring valuable insight into the molecular mechanism of LRRK2-mediated synaptic regulation.

1.1.3.3 LRRK2 and immune system

The immune system is a sophisticated network crucial for maintaining homeostasis. It can be broadly categorized into 2 major components: the central nervous system (CNS) and the peripheral immune system. The CNS, composed of the brain and spinal cord, traditionally operates within an immune-privileged environment owing to the blood-brain barrier, which restricts the access of immune cells (Proulx & Engelhardt, 2022). Nevertheless, recent studies have revealed that the CNS is not entirely immune-isolated; instead, it maintains a delicate balance to ensure protection against pathogens while preventing excessive immune activation that may lead to neuroinflammation (Ampie & McGavern, 2022; Matejuk et al., 2021; Waltl & Kalinke, 2022). On the other hand, the peripheral immune system operates outside the CNS, encompassing various tissues and organs. Its primary function is to defend against invading pathogens and maintain overall systemic health. Recent studies have reported an intricate role of LRRK2 as the interplay between CNS and peripheral immune system and thereby further highlight its essential role in physiology and pathology.

1.1.3.3.1 The role of LRRK2 in CNS immunity

The immune system in the CNS is distinct from its peripheral counterpart in terms of the composition of immune cells. In the CNS, glia, especially microglia and astrocytes, play the main roles in innate and adaptive immune responses (Ousman & Kubes, 2012).Microglia are the tissue-resident macrophages in the brain, mainly skilled in removing dysfunctional or damaged protein aggregates and neurons; while astrocytes predominately function as maintaining a normal biochemical environment for neuronal signalling. These glial cells

constitute the first barrier of CNS immunity, and are associated with pathological neuroinflammation in numerous diseases (Stevenson et al., 2020). In 2012, a high level of LRRK2 expression was firstly observed in activated murine microglia, though no accompanied high level of mRNA was detected(Moehle et al., 2012). Since then, numerous studies have been conducted to understand the role of LRRK2 in microglia. Multiple studies suggest that the LRRK2 protein, especially its kinase function, is a key regulator of microglia-mediated neuroinflammation (reviewed in (Filippini et al., 2021)). For example, increased LRRK2 kinase function in microglia is paired with increased production of pro-inflammatory cytokines, reduced microglial migration and increased microglial phagocytosis compared with respective control cells (Marek et al., 2018; Panagiotakopoulou et al., 2020). In parallel, such elevated microglial activation can be attenuated by downregulating LRRK2 expression. On the other hand, it has been reported that LRRK2-KO mice exhibit reduced microglial activation, dopaminergic neuron degeneration and movement deficit after LPS exposure, suggesting normal-functioning LRRK2 protein is essential for maintaining neuronal functions under inflammation (Dwyer et al., 2020).So far, the mechanisms underlying these observations are still unclear. Some studies suggest LRRK2 might impact microglial functions through PKA-NFκB pathway, which is supported by the evidence that reduced NFκB transcription has been observed in microglia with LRRK2 KO(C. Kim et al., 2020; Russo et al., 2015). However, more work is required to understand the details of these signalling pathways. The role of LRRK2 in astrocytes has been prevalently linked with endosomal-lysosomal functions and regulation of α-synuclein degradation (Bonet-Ponce et al., 2020; Booth et al., 2019; di Domenico et al., 2019). For example, astrocytes derived from transgenic mice carrying the LRRK2-G2019S mutation, which increases LRRK2's kinase activity, exhibited enlarged lysosomes and deficits in protein aggregates degradation. These dysfunctions were attenuated by chemical LRRK2 kinase inhibition (Henry et al., 2015). Similarly, patient-derived induced pluripotent stem cells (iPSC) carrying the same LRRK2 mutation exhibited increased cytokine release, reduced ability to degrade α-synuclein compared to the control cells during the exposure of inflammatory stimuli (Sonninen et al., 2020). Inhibiting LRRK2 kinase activity downregulated immune response in these astrocytes.

1.1.3.3.2 The role of LRRK2 in peripheral immunity

The role of LRRK2 at the periphery has been underestimated and poorly investigated for a long time. However, a high expression of LRRK2 has been observed in immune-related organs (lung, kidney, spleen), myeloid cells and peripheral immune cells and the expression level increases under inflammatory conditions (Biskup et al., 2007; Melrose et al., 2007; Thévenet et al., 2011a). Furthermore, LRRK2 is upregulated in B cells, T cells, macrophages and non-classical monocytes of sporadic PD patients, accompanied by enhanced pro-inflammatory cytokines secretion (Cook et al., 2017b; Hakimi et al., 2011b; Miklossy et al., 2006; Thévenet et al., 2011a). Taken together, these findings suggest a crucial role of LRRK2 at the interface of periphery immune response and PD. In fact, the correlation between PD and peripheral inflammation has been observed in the past 2 decades. For example, increased levels of peripheral inflammatory cytokines, such as TNF, IL-1β and IL-6 have been found in the serum of patients with sporadic PD, LRRK2-related PD as well as asymptomatic LRRK2-G2019S carriers as compared to healthy controls (Bu et al., 2015; Dursun et al., 2015; Dzamko et al., 2016; Williams-Gray et al., 2016). In the meantime, PD patients who experience viral or bacterial infections often display a decline in both motor and cognitive functions, suggesting that peripheral inflammation may contribute to the PD progression (Brugger et al., 2015). Therefore, LRRK2 has been suggested as a target for modulating peripheral inflammation induced by PD progression.

How LRRK2 regulates immune response is still unclear. However, previous findings have associated this function with its kinase activity. For example, it has been reported that LRRK2 autophosphorylation increases in human peripheral blood mononuclear cells (PBMCs) and mouse in bone-marrow derived macrophages (BMDMs) under immune stimulation of cytokine, while the altered pattern was not observed in Lrrk2 knock-out mouse macrophages (Cook et al., 2017b; Dzamko et al., 2012). In contrast, it was observed that the phosphorylation of LRRK2 decreased in the peripheral blood mononuclear cells (PBMCs) of individuals carrying the LRRK2-G2019S mutation. This suggests that the increased kinase activity resulting from the pathogenic mutation might trigger a compensatory mechanism aimed at reducing LRRK2 autophosphorylation, thereby helping to maintain immune balance in the periphery (R. L. Wallings et al., 2020). In addition, previous studies have suggested that LRRK2:RAB10 interaction may play an essential role in mediating cellular response to

inflammation. RAB10 is a robust substrate of LRRK2 kinase. It is expressed in peripheral immune cells including B cells, monocytes and neutrophils. It has been shown that in the healthy status, phosphorylated-RAB10 (pRAB10) decreased in the above-mentioned cells by pharmacological LRRK2 inhibition, suggesting that reduced LRRK2 kinase activity may affect immune response conducted via these blood cells (Fan et al., 2018; Thirstrup et al., 2017a). In addition, the level of LRRK2 and pRAB10 in mouse and human primary macrophages were found associated with PD-implicated plasma cytokine levels TNF-α and IL-6 (Atashrazm et al., 2019; Z. Liu et al., 2020). Considering that the main function of RAB10 is to modulate membrane trafficking and its phosphorylation is pro-inflammatory stimulation, these findings suggest a potential mechanism by which LRRK2 regulates peripheral immunity via adjusting the phosphorylation of RAB10 and thereby affecting the release of PD-related cytokine (Z. Liu et al., 2020; Nazish et al., 2021; D. Wang et al., 2010). However, further research is required to fully understand this complicated pathway.

1.1.4 LRRK2's relevance to Disease

Considering the complicated functionality of the LRRK2 protein, its relevance to disease becomes increasingly apparent. Up to now, LRRK2 has been primarily associated with Parkinson's disease (PD) and Crohn's disease (CD).

1.1.4.1 PD

1.1.4.1.1 Prevalence, symptoms, pathology and genetics of PD

According to the PD technical brief launched in 2022 by the World Health Organisation (WHO) , the global prevalence of PD has doubled in the past 25 years, with an estimate of 8.5 million people living with PD in 2019 (https://www.who.int/news/item/14-06-2022-launch-of-who-s-parkinson-disease-technical-brief). Furthermore, disability and mortality attributed to PD are rising at a more rapid pace compared to any other neurological condition, as indicated by the 2019 estimates, which show PD causing 5.8 million disability-adjusted life years, marking a growth of over 100% since 2000. As the world's population continues to age, the prevalence of Parkinson's disease is expected to rise, making PD an increasingly significant global health concern.

PD was first identified by British physician James Parkinson in 1817, the disease was initially known as "shaking palsy", which refer to a set of progressive motor deficits including resting tremors, bradykinesia, muscle rigidity, and postural instability, which remain the defining features of the disease to this day (Parkinson, 1817). Over the years, apart from its characteristic motor symptoms, a wide array of non-motor symptoms such as cognitive impairment, depression, sleep disorder and hallucinations were described with PD (Poewe, 2008). In 1960s, researchers found that damage to dopaminergic neurons in the *substantia nigra* and deletion of dopamine in the brain is responsible for the movement symptoms of PD (Fénelon & Walusinski, 2021).

Treatment for Parkinson's disease aims to manage and alleviate the various motor and non-motor symptoms associated with the condition, as there is currently no cure for the neurodegeneration. In terms of medications, levodopa is one of the most effective and commonly used drugs, helping to replenish dopamine levels (Lewitt & Library, 2015). Other medications like dopamine agonists, MAO-B inhibitors, and anticholinergic drugs can be prescribed to address specific symptoms or enhance the effectiveness of levodopa. In some cases, surgical interventions, such as deep brain stimulation, may be considered to alleviate motor symptoms when medications become less effective. A holistic approach to treatment often includes addressing non-motor symptoms, such as depression, anxiety, and sleep disturbances, through counselling and appropriate medications. Ongoing research and clinical trials continue to explore new treatments and potential disease-modifying therapies, providing hope for improved management and outcomes for those living with Parkinson's disease.

The cause of movement disorders during PD progression results from the impairment of the dopaminergic nigrostriatal system. However, the pathology underlying the dopaminergic neurons remains unclear, but it is believed to involve a complex interplay of assorted cellular and molecular mechanisms in both CNS and periphery, among which the pathological hallmark is the aggregations of α-synuclein in the form of Lewy bodies (LB). LB pathology is found in multiple brain regions of PD patients: it starts from early affected areas such as amygdala and anterior cingulate cortex and the reaches late affected areas such as insula cortex, middle temporal cortex and anterior middle frontal gyrus (de Boni et al., 2022). LB is

also found in peripheral tissues such as heart, gastrointestinal tract, saliva gland, skin and lung (Devic et al., 2011; Ennemoser et al., 2022a, 2022b; Y. yan Li et al., 2022; Niemann et al., 2021). These findings suggest that PD is a multi-system disorder rather than a localised brain condition. Dysregulation of different cellular pathways has been found associated with PD, such as autophagy, synaptic vesicle transport, mitochondrial function, ribosomal function etc, indicating a complicated cellular and molecular mechanism for PD (Gonçalves et al., 2021; Lu et al., 2020; M. Nguyen et al., 2019; Picca et al., 2021).

Additionally, many lines of evidence suggest that PD involves a complex interplay of environmental and genetic factors. Environmental risk factors associated with PD include the exposure to toxins, heavy metals, pesticides, traumatic brain lesions, and bacterial or viral infections (Wirdefeldt et al., 2011). Long-term expose to these factors elicits intense neuroinflammation especially in the substantia nigra. In addition, some viral proteins (such as HSV-1 and EBV) exhibit molecular similarity with α-synuclein, leading to α-synuclein accumulation and the formation of LB (Maries et al., 2003). Moreover, previous studies have shown that viral infection in the enteric nervous system (ENS) induce accumulation of α-synuclein in the neurites of the upper gastrointestinal tract (Stolzenberg et al., 2017). Considering the fact that α-synuclein produced in the ENS can be trafficked from the gut to the brain; therefore the viral infection may eventually cause LB pathology in the CNS (Del Tredici & Braak, 2016). In terms of genetic factors, until now, rare variants in more than 20 genes have been reported to cause PD, in *which SNCA, PRKN, PARK7, LRRK2, PINK1, ATP13A2, GBA1, PLA2G6* and *VPS35* are found with the highest correlation (Wirdefeldt et al., 2011). It is important to note that these mutations are responsible for a minority of PD cases (familial cases) as the vast majority of PD patients have a sporadic origin with absence of any classical mutation. Familial mutations in these genes show different frequencies across populations and may have varying effects on the set of disease symptoms manifested by the patients, age of onset, and the rate of disease progression. For example, mutations in LRRK2 are rare in general populations but are considerably higher in Ashkenazi Jews (about 15%) and North African Imazighen (about 40%), though patients with LRRK2 mutations present similar symptoms as sporadic PD patients (Bar-Shira et al., 2009; El Haj et al., 2017). In comparison, PD patients with mutations in SNCA gene are more likely to develop dementia as compared to sporadic PD patients while *PRKN, PRAK7, PINK1* and *PLA2G6* mutations usually cause early-

onset parkinsonism (Wirdefeldt et al., 2011). The complexity of PD genetics implies a complicated etiopathogenesis of the disease.

1.1.4.1.2 LRRK2 and Familial PD (fPD)

- Pathogenic LRRK2 mutations

Among the above-mentioned PD-related genes, mutations in *LRRK2* are the most common genetic causes of PD, accounting for 2-40% of familial PD cases depending on the population under analysis (Bras et al., 2005; Drolet et al., 2011; Lesage et al., 2006). The first PD-related *LRRK2* mutation was discovered in 2004 within a large Japanese family exhibiting autosomal dominant parkinsonism (Funayama et al., 2002, 2005). To the date, there are > 100 coding and non-coding mutations identified within the *LRRK2* gene that have been confirmed to be associated with fPD, within which 7 coding mutations have been confirmed as disease-causing, namely G2019S, R1441C/G/H, Y1699C, N1437D/H and I2020T, among which the G2019S mutation is by far the most prevalent. Despite its worldwide distribution, the G2019S mutation exhibits a higher frequency in North African Imazighen (42%), Ashkenazi Jewish (28%) and Portuguese (16%) populations, while barely present in Asia (<0.1%) (Gaig et al., 2009; Lesage et al., 2006; Ozelius et al., 2006). With respect to the R1441C/G/H mutations, although these 3 variants occur at the same position on the LRRK2 protein, they are actually observed in different populations. After LRRK2-G2019S, LRRK2-R1441C represents the second most common LRRK2 mutation identified in Europe, especially in the Belgian population, where R1441C accounts for 10.7% of fPD cases (Nuytemans et al., 2008). In comparison, LRRK2-R1441G is present with a frequency of 46% in fPD within the Basque population, but it is barely presents in other European populations (Gorostidi et al., 2009). A wider distribution has been found for the R1441H mutation, including Asia (Taiwanese population), Europe (Portuguese and Greek populations) and North America (US) (Ferreira et al., 2007; Mata et al., 2005; Spanaki et al., 2006; Zabetian et al., 2005). In comparison, the I2020T has only been reported in 9 cases from one family in Japan (the Sagamihara kindred), and it was the first PD-related LRRK2 mutation was observed (Funayama et al., 2005). Similarly, the Y1699C mutation has only been reported in 3 families, 1 German-Canadian family and 2 British families (Nicholl et al., 2002; Wszolek et al., 1997), while the other mutation N1437H present primarily in European populations, especially in Norwegian and Polish (Aasly et al., 2010; Turski et al.,

2022). It is worth to highlight that the association between LRRK2 mutations and PD is affected by age-dependent penetrance. For example, penetrance of the G2019S mutation increases from 17% at 50 years old to 85% at 70 years old (Kachergus et al., 2005; Lee et al., 2017; Luciano et al., 2010).

- Clinical symptoms and pathology of LRRK2-related fPD (LRRK2-PD)

In general, patients with LRRK2-PD present typical clinical symptoms similar to those with sporadic PD (sPD), though subtle variance has been observed among patients with different LRRK2 variants. For example, patients with the G2019 mutation were reported with longer disease duration, slower progression and various age of onset (from 41 to 79 years). PD pathology (neuronal loss in the *substantia nigra* and LB pathology) is different when comparing LRRK2-PD vs sPD. Neuron loss is observed in the *substantia nigra* of patients with all types of LRRK2 variants while LB pathology was only reported to consistently occur in 90% of LRRK2-PD cases. However, though less α-synuclein aggregation was observed in the other 10% LRRK2-PD cases, tau inclusions and TDP-43 accumulation seem to increase, especially with the LRRK2-G2019S cases (Kalia et al., 2015). These findings might suggest that the same clinical motor symptoms consequent to the PD neurodegeneration may result from different pathologies when comparing sPD and LRRK2-PD.

- Effects of the pathogenic variants on LRRK2's enzymatic functions

All pathogenic mutations have been found to increase the kinase activity of LRRK2, although to different extents (Ito & Utsunomiya-Tate, 2023). *In vitro* kinase assays (performed using the isolated LRRK2 kinase in an artificial reaction setting, checking the LRRK2 rate of auto-phosphorylation) consistently showed that the G2019S mutation significantly increases LRRK2's kinase activity, while the impacts of other mutations vary among studies, which is likely due to differences in the assay systems (Marchand et al., 2020). On the contrary, in cell culture experiments, the non-G2019S mutations seem to increase the LRRK2 kinase activity to a larger extent as compared to the G2019S mutation. For example, overexpression of LRRK2-R1441G and Rab proteins led to a significant increase of RAB10 phosphorylation (~3-4 folds), which is larger than the impact of the LRRK2-G2019S mutation (~2 folds)(Ito et al., 2016a). Also *in vivo*, Rab10 phosphorylation was increased in peripheral blood neutrophils

obtained from R1441G carriers but not from G2019S carriers (Fan et al., 2021). These findings may suggest that the mechanism by which non-G2019S mutations upregulate LRRK2 substrate phosphorylation might be different from the G2019S mutation and that the LRRK2 substrate of phosphorylation (either LRRK2 itself or downstream substrates) could be differently affected by the type of mutation.

Multiple lines of evidence have shown that some mutations also alter (decrease) LRRK2's GTPase activity, though this effect it is not as well-studied as the impact of mutations on the kinase function and the results vary among studies. For example, the R1441C/G/H and N1437H mutations have been reported to decrease *in vitro* LRRK2 GTPase activity to different extents, while there was not enough evidence for the impact of the G2019S mutation (Guo et al., 2007; C. X. Wu et al., 2019; Xiong et al., 2010). However, these findings have not been confirmed *in vivo* yet. Considering the structural proximity between the kinase and the GTPase domains, and the fact that LRRK2 can phosphorylate itself probably as a self-regulatory process, it is possible that alterations in the kinase domain might affect the GTPase activity of LRRK2 and vice-versa. Therefore, further investigations are needed on the co-regulation of the 2 enzymatic domains by LRRK2 mutations to gain a more comprehensive view of the LRRK2-fPD pathogenesis. Finally, the R1441C/G/H, Y1699C, and I2020T mutations have been shown to downregulate autophosphorylation of LRRK2 at Ser910/935, while the G2019S mutation does not (Doggett et al., 2012; X. Li et al., 2011; Nichols et al., 2010a). Since the Ser910/935 phosphorylation is crucial for LRRK2's interaction with 14-3-3 proteins by impacting its subcellular localization, it is likely that this localization mechanism does contribute to the fPD pathology (Nichols et al., 2010a).

1.1.4.1.2 Involvement of LRRK2 in Sporadic PD (sPD)

The association between LRRK2 mutations/variants and sPD is complex. On the one hand, mutations that cause fPD, such as G2019S and R1441C, are present in sPD cases and not only in fPD cases (Bardien et al., 2011; Correia Guedes et al., 2010). However, it is worth to highlight that the presence of LRRK2 mutations does not always lead to sPD as there are asymptomatic LRRK2 mutations carriers in the general population (Trinh et al., 2014). Additionally, some other variants, even though have not been confirmed as causing mutations, have been found to increase the risk for sPD, such as the G2385R mutation in the

WD40 domain and the R1628P mutation in the ROC domain. These 2 variants are more common in PD patients with Asian ancestry as compared to the ethnicity-matching healthy individuals (Di Fonzo et al., 2005; O. A. Ross et al., 2008). Further studies have linked these 2 mutations with alteration of the LRRK2 structure and increased kinase activity, which may cause perturbed vesicle transport and excessive cell death (Carrion et al., 2017; Shu et al., 2016). On the other hand, some LRRK2 mutations (such as the R1398H mutation in the ROC domain) have been shown to have protective effect against PD by reducing the kinase activity and the strength of GTP binding to the LRRK2 protein (Nixon-Abell et al., 2016; Tan et al., 2010). Furthermore, non-coding LRRK2 variants have also been suggested as risk factors for sPD by altering the expression level of the LRRK2 protein (Ryan et al., 2017). Previous studies have shown that increased LRRK2 expression may impair inflammatory response and apoptotic signalling, thereby contribute to the PD pathology (Härtlova, Herbst, Peltier, Rodgers, Bilkei-Gorzo, et al., 2018; K. S. Kim et al., 2018). Taken together, the linkage between LRRK2 and sPD is far more complicated than just a simple cause-and-effect relationship, involving intricate genetic, environmental, and molecular factors that interact in multiple ways to modulate the development and progression of sPD.

1.1.4.1.3 LRRK2-targeted PD therapy: oppurtunities and challenges

- Current development of LRRK2-targeted drug development for PD

Despite our incomplete understanding of the exact functions of LRRK2 in health and PD conditions, there has been successful development of small-molecule LRRK2 kinase inhibitors as potential PD therapy. These inhibitors seem to be appropriate for clinical use considering their pharmacological and pharmacokinetics properties confirmed in preclinical assessment, such as kinase targeting selectivity, bioavailability and blood-brain barrier (BBB) permeability. Two LRRK2 kinase inhibitors, DNL201 and DNL151/BIIB122 were evaluated in Phase 1 clinical trials, in which no adverse events were observed after short-term administration, though DNL151/BIIB122 presented better pharmacokinetics features and thereby was selected for further clinical trials (Jennings et al., 2022). Multi-centre Phase 2 (NCT05348785) and Phase 3 (NCT05418673) clinical trials are on-going to evaluate the efficacy and safety of DNL151/BIIB122 in participants with early clinical stage PD (30–80 years old) and LRRK2-PD

patients with the G2019S mutation. Other LRRK2 kinase inhibitors under Phase 1 or preclinical development include S221237, NEU-723, PFE-360, MLi-2 and PF-06685360 (Kingwell, 2023). However, it is worth to mention that even if these trials are successful, further assessment would be required to validate the efficacy and safety of LRRK2 inhibitors as a viable option for patients with sPD, which is much more common than LRRK2-PD in the worldwide. Apart from kinase inhibitors, other approached have been considered to modify other domains of the LRRK2 protein, such as the GTPase domain or the PPI domains. However, little evidence suggests the efficacy of alterations of non-kinase domains in PD models, which may result from the difficulty of selective targeting. Moreover, 1 clinical trial (NCT03976349) at Phase II has been conducted to assess the use of antisense oligonucleotides (ASOs) to decrease the entire LRRK2 protein expression levels, thereby blocking both kinase and GTPase activities as well as removing the interaction domains functions.

- Challenges in LRRK2-target drug development
    - *Selectivity*

The above-mentioned LRRK2 kinase inhibitors under pre-clinical or clinical trials are non-selective LRRK2 kinase inhibitors, i.e., they inhibit both mutant and wild-type LRRK2. Considering the crucial role of LRRK2's kinase function in maintaining normal cellular processes, non-selective suppression of wide-type kinase activity may lead to undesired pathological alterations. For example, most preclinical studies suggested that pharmacologic inhibition of LRRK2 kinase induce morphological changes to the lung cell in rodents and non-human primates, though no further changes in respiratory function were observed (Lesniak et al., 2022). Furthermore, these alterations were found to be washed out after drug withdrawal. No adverse events were observed in pulmonary function during the monitoring time frame in clinical trials of DNLI-201 and DNL-151/BIIB122. It remains to be determined if long-term exposure of LRRK2 kinase inhibitors lead to pathological changes in lung in human.

    - *Lack of biomarker*

Another major challenge in PD clinical trials is the absence of dependable biomarkers and scalable assays for assessing LRRK2 activity in patients. Theoretically, some LRRK2 substrates can be potential biomarkers for the assessment of LRRK2 kinase inhibition, such as specific Rab GTPases. Pathogenic LRRK2 mutations have been shown to enhance Rab phosphorylation,

implying that measuring phosphorylated Rab levels may serve as a reflection of LRRK2 activity (Steger et al., 2016b, 2017a). Additionally, these mutations also modulate phosphorylation of the LRRK2 protein at Ser910 and Ser935 (Dhekne et al., 2018). Developing these biomarkers for clinical use is crucial for drug development and clinical trials, enabling patient stratification based on the efficacy of kinase inhibition, and supporting personalized dosing and adaptive trial designs.

- *Timing of trials*

Most reviews of PD have suggested that the diagnosable motor symptoms of PD first appear at relatively late stage in the disease progression, when approximately 50-70% of dopaminergic neurons have been lost in the *substantia nigra* (Noyce et al., 2016; G. W. Ross et al., 2004). By then, PD pathology may be too advanced to be affected by pharmacological treatments. Therefore, identifying individuals at the earliest stages of disease would facilitate clinical trials. Since identifying prodromal sPD patients is nearly an unreachable goal due to the mildness and diversity of pre-motor symptoms, the population of prodromal LRRK2 mutation carriers is the probably most appealing group for clinical trials. However, screening of the general population to identify this subgroup of people is still a challenge. Tracing and including relatives of LRRK2-PD patients will help alleviate this problem, but it requires well-organised collaboration of multiple clinical centres. There are a few international PD-focused research project that recruit pre-symptomatic cohort with PD-related gene mutation for pathological study and drug development, such as the PPMI (Parkinson's Progression Markers Initiative) study sponsored by the Michael J. Fox Foundation (https://www.ppmi-info.org/). These studies may provide promising subject cohort for further clinical trials.

1.1.4.2 CD

Crohn's disease (CD) is an inflammatory bowel disease (IBD) that causes chronic inflammation of the gastrointestinal tract. Common symptoms of Crohn's disease include abdominal pain, diarrheal, weight loss, fatigue, and sometimes rectal bleeding. The exact cause of Crohn's disease is still not fully understood, but it is believed to result from a combination of genetic, environmental, and immunological factors. For example, active and passive smoking, high-saturated-fat diet and bacterial or viral infection (Ananthakrishnan et al., 2014; Chapman-Kiddell et al., 2010; Gradel et al., 2009; Higuchi et al., 2012; Hou et al., 2011; Mahid et al.,

2006). In terms of genetic factors, multiple GWAS have identified LRRK2 as a major susceptibility gene for CD (Anderson et al., 2011; Barrett et al., 2008a; Franke et al., 2010). GWAS has linked the LRRK2 M2397T coding variant in the WD40 domain to sporadic CD cases in European population (Barrett et al., 2008b; Fava et al., 2016a; Ikezu et al., 2020; Z. Liu et al., 2011). In addition, a newly identified CD mutation, LRRK2-N2081D in the kinase domain, has been associated with a similar kinase function increase as the PD LRRK2-G2019S mutation which is highly linked to PD. Moreover, 2 other variants (LRRK2-N551K and R1398H in the ROC domain) seemed to increase the GTPase activity and reduce risk for both diseases (Gopalai et al., 2019; Hui et al., 2018). This suggests that the pathogeneses of PD and CD may overlap in altering specific LRRK2 enzymatic functions. On this note, several epidemiological studies have identified the connection between PD and CD. For example, increased risk of PD was found in CD patients in multiple populations in US, Taiwan, South Korea and Sweden (Lai et al., 2014; Lin et al., 2016; Peter et al., 2018; Weimers et al., 2019). Anti-TNF-α treatment of CD patients eliminated the increased risk for PD (Peter et al., 2018). On the other hand, gastrointestinal dysfunction is known as a non-motor symptom among PD patients that often appears prior to the typical motor symptoms (Suzuki et al., 2019). In addition, α-synuclein aggregates have been detected in the gastrointestinal tract years before the motor-symptom onset, which may gradually precede into the CNS and lead to sequential neuronal damage (J. S. Kim et al., 2017; Rota et al., 2019; Stokholm et al., 2016). These findings suggest a potential pathological association between the 2 diseases. Multiple studies have tried to determine the role of LRRK2 at the interface between PD and CD. For example, upregulated expression level of LRRK2 and increased LRRK2 kinase function have been observed in gastrointestinal tract cells as well as peripheral immune cells such as the CD14+ monocytes in both LRRK2-CD and LRRK2-PD patients (Cook et al., 2017c; Gardet et al., 2010a). Also, an enhanced pro-inflammatory response to multiple types of cytokines was observed in these cases as well as LRRK2 mutation carriers (Gardet et al., 2010a; Ikezu et al., 2020). These finding suggest that LRRK2 functions either as an upstream regulator or a downstream responder of increased immune response in the disease scenario. In addition, since the 2 diseases seem to share a LRRK2-mediated inflammatory pathway, it is worth to keep caution in drug development since it is still unknown whether these alterations are beneficial or harmful for the affected area. Further research is required to understand the function of LRRK2 at the interplay of these 2 diseases.

Apart from Parkinson's disease (PD) and Crohn's disease (CD), LRRK2 variants have been implicated in the susceptibility to other immune-related diseases. For example, the LRRK2-M2397T variant was reported to aggravate the type-1 reaction (T1R) in leprosy, accompanied with excessive inflammatory response (Fava et al., 2016b, 2019). In addition, LRRK2 has been linked to the regulation of the innate immune response against Mycobacterium tuberculosis (Mtb) (CG et al., 2020; Weindel et al., 2019). This alteration has been associated with the protein's kinase function, which inhibits the formation and maturation of Mtb phagosome (Härtlova et al., 2018). The intricate association between LRRK2 and various disease indicates its crucial role in maintaining homeostasis and regulating CNS and peripheral pathology.

- Protein interactions

Protein-protein interactions (PPIs) are the elementary molecular interactions responsible for sustaining cellular functions. In homeostasis, PPIs coordinate biological processes such as subcellular molecular biosynthesis and trafficking, signal transduction and the regulation of organelle dynamics. Alterations of PPIs can lead to functional impairment and thereby cause disease. Therefore, PPI analysis is essential for understanding the physiological and pathological role of less-recognised proteins like LRRK2. Up to now, a large number of studies have been focused on identifying LRRK2 protein interactors.

1.1.5.1 LRRK2:Rab interaction

Out of the many LRRk2 protein interactors reported in literature, only 3 are widely accepted by the scientific community: LRRK2 itself and Rab proteins (as both binding partner and kinase substrate); 14-3-3 proteins (just as binding partners).

Rab proteins constitute the largest family of small GTPases. They regulate almost every stage of intracellular membrane traffic, including the formation, motility and fusion of vesicles with target membranes (Mizuno-Yamasaki et al., 2012; Pfeffer, 2017; Zerial & McBride, 2001). Phosphoproteomics study showed that a subset of Rab proteins are endogenous substrates of LRRK2, including a total of 10 Rab GTPases: RAB3A/B/C/D, RAB 8A/B, RAB 10, RAB 12, RAB 35, and RAB43 (Ito et al., 2016b; Steger et al., 2016b, 2017b; Thirstrup et al., 2017b). These proteins are key regulators of neurotransmitter release (RAB3 isoforms), ciliogenesis (RAB8 and RAB10), endocytosis (RAB10, 12 and 35) and ER-Golgi trafficking (RAB43) (Blacque et al.,

2018; C. Li et al., 2017; Schlüter et al., 2004; Thirstrup et al., 2017b). Previous studies suggested that LRRK2 phosphorylation may block effector binding sites of some Rab proteins (e.g. RAB8A) and perturb intracellular localisation of other Rab proteins (e.g. RAB10). However, how LRRK2 phosphorylation affect structures and functions of Rab proteins have not been fully understood yet. Activities of Rab proteins are usually downregulated after LRRK2 phosphorylation, which have been observed especially in the presence of pathogenic *LRRK2* variants which increase its kinase function. For example, the *LRRK2-G2019S* and *LRRK2-R1441C* variants have been associated with primary cilia formation inhibition as well as perturbated endosomal trafficking due to dysregulated LRRK2:Rab interaction (Dhekne et al., 2018; Madero-Pérez et al., 2018a; Rivero-Ríos et al., 2019; Steger et al., 2017a). Further research is required to understand the mechanisms of these alterations. Apart from phosphorylation, LRRK2 interacts with other Rab proteins through other functional domains. For example, LRRK2 interacts with RAB29 via its ANK domain (Beilina et al., 2014; D. A. MacLeod et al., 2013). Unlike the above-mentioned Rab proteins, RAB29 functions upstream of LRRK2. It recruits LRRK2 to specific cellular compartments, such as the trans-Golgi network, where RAB29 is normally located, and activate its kinase activity(Beilina et al., 2014; Helip-Wooley & Thoene, 2004; D. A. MacLeod et al., 2013; Madero-Pérez et al., 2018b; Purlyte et al., 2018a). Increased LRRK2 kinase function further affects the downstream recruitment and phosphorylation of other Rab proteins (such as RAB8 and RAB10), thereby regulating their cellular functions. These processes together form the "RAB29-LRRK2-RAB8/10 cascade" (Kuwahara & Iwatsubo, 2020). This signalling pathway exhibits heightened activation in the presence of pathogenic LRRK2 mutations, as it appears that the mutated LRRK2 is significantly more activated by RAB29 compared to the wild-type LRRK2 (Purlyte et al., 2018b). Nonetheless, the precise mechanism underlying this cascade remains to be fully investigated.

1.1.5.2 LRRK2:14-3-3 interaction

The 14-3-3 protein family constitute a group universally conserved and highly abundant regulatory molecules. The family contains a total of 7 known mammalian isoforms, namely 14-3-3 β (*YWHAB*), 14-3-3 γ (*YWHAG*), 14-3-3 δ (*YWHAQ*), 14-3-3 η (*YWHAH*), 14-3-3 ζ (*YWHAZ*), 14-3-3 ε (*YWHAE*), and 14-3-3 θ (*SFN*). These interact with a wide range of intracellular proteins and regulate various cellular processes such as apoptosis, transcription, protein transport and cytoskeleton organisation. Furthermore, 14-3-3 proteins have been

associated with preventing the formation of neurotoxic aggregates by directly interact with related proteins, such as α-synuclein, tau and neurofilament (NFL). These findings have identified 14-3-3 proteins as potential drug targets for a variety of neurodegenerative diseases such as PD, Alzheimer's disease (AD) and amyotrophic lateral sclerosis (ALS). Structurally, 14-3-3 proteins share a comparable domain organisation and sequence, resulting in an overlap in their sets of protein interactors, which includes LRRK2. The interactions between LRRK2 and 14-3-3 proteins involve multiple binding sites spreading across the entire LRRK2 structure, including the ANK domain at the N-terminal, ROC domain in the middle and the WD40 domain at the C-terminal (Manschwetus et al., 2020a; Nichols et al., 2010b). These binding sites are highly influenced by the kinase activity of the LRRK2 protein. Therefore, LRRK2:14-3-3 interaction can be perturbed by LRRK2 kinase inhibitors or pathogenic mutations that alters its kinase function (Lavalley et al., 2016; Manschwetus et al., 2020b; Nichols et al., 2010c; Stevers et al., 2017). Also, impaired or weakened LRRK2:14-3-3 interaction has been related with dysregulated exosome secretion of LRRK2 and sequential cytoplasmic accumulation of LRRK2, which was observed in CSF and urine samples of PD cases (Fraser et al., 2013; Giusto et al., 2021; Gu et al., 2020; Kilisch et al., 2016). These findings suggest an important role of LRRK2:14-3-3 protein in PD pathology.

1.1.5.3 Previous in-silico protein interactome analysis of LRRK2

In-silico Protein interactome analysis refers to the systematic study of the complete set of PPIs within a biological system based on the vast biodata resources. This approach offers a comprehensive exploration of potential interaction partners of an unknown protein, enabling researchers to infer its functional associations, predict its involvement in specific cellular pathways and thereby generate hypotheses for experimental validation. Up to now, 2 protein interactome studies have been performed on the LRRK2 protein, which were independently published in 2015.

In the first study (Porras et al., 2015a), both human and mouse LRRK2 (Q5S007 and Q5S006, respectively) were queried in primary PPI database IntAct (Orchard et al., 2014) returning 1075 binary interactions for 612 interacting pairs and 598 interactors of LRRK2. These interactions were reported from literature as detected with various of methods, which

reduces the impact of false positives and negatives resulting from single detection method. However, due to the variety in the accuracy and precision among these detection methods, the author suggested that a quality control process is needed to keep only high-quality interactions in the network. In the study, the authors defined three selection criteria: 1) detection of the protein interaction in native tissue, 2) the use of the MI score scoring system, 3) evidence of the protein to be phosphorylated by LRRK2 *in vitro*. Native detecting conditions were defined as the interactions that were obtained using untagged, full-length LRRK2 *in vivo* methods. However, due to the complex structure and the large size of the LRRK2 protein, most of interactions were unable to fit this criterion. In fact, out of 598 human or mouse LRRK2 interactions, only 25, including LRRK2 and Lrrk2, were detected in native condition. The MI score scoring system was defined based on the number of literature reports of a given interaction. Theoretically, higher MI score associates to interactions that were repeated in a consistent number of studies and thereby may be more reliable. However, the author pointed that some nonspecific kinase substrates get comparatively higher MI score because their wide usage in protein assays. In parallel, an interaction with low MI score may result from limited scientific interest in repeating the analysis. Due to the important role of LRRK2's kinase activity in biological processes, which is supported by multiple lines of evidence from functional studies of LRRK2, the author added more confidence on those interactors of LRRK2 that were detected with kinase assays. By combining these three criteria, the author generated a high-confidence LRRK2 interactome comprising 25 proteins including LRRK2 itself and performed pathway annotation on it. The result showed that the LRRK2 interactome is enriched in functions related to axon development, cell cycle, intrinsic apoptosis, membrane trafficking, EGFR signalling and response to stress.

In the second study (Manzoni et al., 2015), only human LRRK2 (Q5S007) was queried in 2 primary PPI databases, IntAct and BioGrid, returning 542 and 260 interactions. The author defined two filters to keep human-human LRRK2 interactions only and to remove replicated entries. After filtering and merging, the two databases gave a total of 422 LRRK2 protein interactors. Similar to the first study, a scoring process was also included to control the quality of interactions, though the process here was simpler considering for each interactor only the number of publications and interaction detection methods. Following this strategy, 62 interactors, which were reported in more than one publication or with multiple detection

methods, were selected for the final LRRK2 interactome. Functional analysis and family domain check were then performed on the interactome. The result indicates that LRRK2 interactors are more likely associated with intracellular vesicle transport and cell projections. Additionally, the author compared the complete LRRK2 interactome (comprising 269 interactors after removing the scoring filter) and the result of PD genome-wide association study (GWAS), identifying four genes coding for proteins in the LRRK2 interactome (*SNCA*, *RAB29*, *GAK* and *MAPT*) as present within a short physical distance from PD GWAS risk signals. In conclusion, these two independent comprehensive LRRK2 PPI study returned two vast interactomes consisting of more than 200 proteins found in literature as able to interact with LRRK2. In parallel, these studies also revealed the importance of quality control during the construction of interactomes, considering the massive data sources and the great variety of detection methods with different accuracy and specificity. Functional enrichment analysis in both studies presented similar results to the observations from wet-lab research, suggesting LRRK2 interactors are enriched in intracellular transport, apoptosis and cell development. However, it is worth noticing that the information regarding LRRK2 PPIs within these studies is non-tissue specific. Multiple lines of evidence have shown that the expression levels of LRRK2 differs in different tissues and cell types. Therefore, it is possible to hypothesise that the LRRK2 functions can also change based on tissue and cell specificity. So far, nobody has attempted an investigation of the LRRK2 interactome with a focus on its variability across different tissues and cell types. This approach will be part of the project within this thesis and the obtained information would be instrumental in furthering understanding the role of LRRK2 in health and disease, as well as in drug development for targeting the LRRK2 protein.

1.2. Systems biology

1.2.1 Development of systems biology

Systems biology is a dynamic and interdisciplinary field that has revolutionized our understanding of complex biological systems. It emerged as a response to the limitations of traditional reductionist approaches, which involve dissecting complex systems into their constituent parts, such as individual genes, proteins, or specific cellular pathways. Instead,

systems biology aims to elucidate the intricate interplay of biological components in a complex system and the emergent properties that arise from these interactions.

While systems biology started to gain its prominence in the late 20th century, its origins can be traced back to the mid-20th century. General systems theory, introduced by Ludwig von Bertalanffy in the 1930s, laid the conceptual groundwork for systems thinking (Fries, 1936). Bertalanffy's theory focused on the relationships and interactions between biological components rather than their isolated properties, emphasizing the importance of a holistic approach. In 1968, Mihajlo Mesarovic introduced the term "systems biology" as he applied general systems theory to study the relationships between biological entities (Mesarović, 1968; Mesarovic et al., 2004). This marked an essential step in the convergence of biology and systems thinking, though the term "systems biology" would not become widely adopted for several decades. In the late 20th century, as molecular biology advanced and high-throughput techniques were developed, the vast volume of molecular data was generated, highlighting the complexity of biological systems. This made it challenging to study molecular processes in a reductional manner. Since then, it became evident that understanding molecular networks was essential. This realization led to the convergence of systems molecular biology and systems mathematical biology, with mathematic and computational theories applied to analyse gene regulatory networks, protein interactions, metabolic pathways, and other biological systems, which formed the basis of contemporary systems biology.

Nowadays, systems biology stands as a crucial field, seeking to enhance the understanding of complex biological systems by integrating data produced from a wide range of disciplines, including molecular biology, multi-omics study, structural biology and so on, via computational and mathematical approaches. At its core, systems biology relies on the generation and analysis of big datasets generated by recent high-throughput techniques such as next-generation sequencing and microarray, and encompass diverse molecular information about a biological system.

1.2.2 Network analysis

1.2.2.1 Types of networks

Network analysis is a fundamental tool in systems biology. It involves the construction and examination of biological networks, where individual entities, such as genes, proteins, metabolites or other biological elements, are represented as nodes while the connections or relationships between these entities are represented as edges (i.e. connections among nodes). According to the categories of included entitles and types of their interactions, molecular networks can be classified into protein-protein interaction networks (PPIN) that describe physical interactions between proteins, Gene Regulatory Networks (GRNs) that illustrate the regulatory relationships between genes, Metabolic Networks, which present metabolic reactions in organisms, and Gene Co-expression Networks (GCN) which help identify functionally related genes based on their expression patterns, etc (Lü & Wang, 2020). These networks represent a combination of topological features and biological features and provide a more comprehensive view of a biological system.

Moreover, based on how connections are established between nodes, biological networks can be allocated into “direct network”, in which the linkage between nodes have a specific direction or orientation, indicating that signal flows from one node to anther but not in the reverse direction (Xiao et al., 2022). A typical example of direct network is metabolic networks, where metabolic reactions involve direct chemical transformation of metabolites and thereby make the edges in these networks direct (Junker & Schreiber, 2007). On the contrary, in an “indirect network”, connections between nodes do not have a specified direction. For example, in a gene co-expression network, edges represent co-expression behaviour between genes which are based on the similarity in their expression profiles and are thereby indirect. Additionally, nodes and edges can be weighted based on assorted features such as expression levels, confident scores, reaction rates, etc., thereby generating “weighted networks” (Gillis & Pavlidis, 2011).

Taken together, various networks provide a holistic representation of how biological molecules collaborate and influence one another. Network analysis enables the identification of crucial nodes (hubs) and their connectivity, unveiling the most influential elements within a system. It also offers insights into the flow of information, energy, or materials through the

network, helping researchers understand the underlying principles governing a biological process. By applying topological analysis, statistical analysis, and machine learning algorithms, network analysis aids in understanding the dynamics, robustness, and vulnerabilities of complex biological systems.

1.2.2.2 Graph theory

Graph theory is a branch of mathematics dating back to the 18th century. It provides a framework for understanding and analysing relationships and connections between individual nodes or node groups in a network. In systems biology, graph theory plays a pivotal role in the representation, analysis, evaluation and exploration of various types of biological networks, enabling the identification of critical nodes, pathways, and patterns in biological processes. Key concepts in graph theory includes degree of nodes, shortest path, closeness centrality and betweenness (Ma'ayan, 2011).

1.2.2.2.1 Degree of nodes

The degree of a node in a graph is the number of edges connected to it. In a directed graph, nodes have both in-degree (incoming edges) and out-degree (outgoing edges). Nodes with high degrees are often considered hubs and are crucial for maintaining network connectivity and information flow.

1.2.2.2.2 Shortest path

Shortest path represents the minimum number of edges or the lowest total edge weight (in weighted graphs) required to traverse from one node to another. It provides a way to determine the most efficient route between 2 nodes within a network.

1.2.2.2.3 Closeness centrality (CC)

Closeness centrality is used to quantify the centrality or importance of a node within a network based on its ability to quickly reach other nodes (Bavelas, 1948). It measures how efficiently a node can access all other nodes in the network, considering the shortest path lengths between them. Nodes with high closeness centrality are those that can be reached more rapidly from the rest of the network, making them central in terms of communication

efficiency. Mathematically, CC is calculated as $CC(i) = \frac{N-1}{\sum_j d(i,j)}$, where $i \neq j$, d_{ij} is the length of the shortest path between nodes *i* and *j* in the network, N is the number of nodes.

1.2.2.2.4 Edge betweenness (EB)

Edge betweenness is defined by the number of shortest paths going through an edge (Freeman, 1977). It quantifies the importance of an edge within a network by measuring its role in maintaining efficient communication between nodes and helps reveal potential bottlenecks or weak points in a network, as the removal of edges with high betweenness can disrupt communication and influence the network's overall robustness.

1.2.2.3 Topological analysis

In the context of network analysis and graph theory, topological analysis involves the examination of the structural properties and connectivity patterns of a network graph and provide insights into the behaviour, functionality, and properties of the represented biological system.

1.2.2.3.1 Degree distribution analysis

Degree distribution in a network refers to the statistical pattern of node degrees. It evaluates whether a network follows a specific structural pattern, such as a scale-free network where a few nodes have a significantly higher degree (hubs), or a random network where node degrees are more evenly distributed. Of note, most of known molecular networks are scale-free (Arita, 2005). Therefore, topological features such as degree of nodes are usually with biological significance. For example, in a cellular signalling network, key proteins like kinases may possess higher degrees, serving as essential hubs that relay information to various downstream targets (Buljan et al., 2020). Hence, degree distribution analysis aids in identifying hub nodes, assessing network robustness, and gaining information regarding the network's functional properties and organizational structure.

1.2.2.3.2 Module detection and Fast Greedy Clustering

Module detection in network analysis involves the identification of densely connected subgroups or communities (often referred as “modules” or “clusters”) within a network. These modules usually reveal functional or interactional units that connect more frequently

with each other than with the rest of the network (i.e. higher local density). For example, in a protein-protein interaction network, module detection helps uncover protein complexes or pathways with related functions, whereas in a gene regulation network, a detected module may represent a unit of co-regulated or co-expressed genes that contribute to a specific biological process. Various clustering algorithms have been developed to identify modules (clusters), such as the Markov Cluster Algorithm (MCL) (Van Dongen, 2008) which is based on random walk, the Lancichinetti-Fortunato-Radicchi (LFR) benchmark (Lancichinetti & Fortunato, 2009) which is based on modularity (a measure that quantifies the strength of division of a network into distinct communities or groups of nodes), the K-Means Clustering (Krishna & Murty, 1999) which is based on unsupervised machine learning algorithms, etc .

The Fast Greedy Clustering is particularly known for its speed and efficiency in module detection for large networks, which has wildly used in molecular network analysis (Clauset et al., 2004; He et al., 2012; Rahiminejad et al., 2019; R. Wang et al., 2020). The algorithm's key idea is to iteratively merge or split communities to maximize modularity, identifying cohesive groups of nodes within the network. It begins by considering each node as a separate community and then combines them to form larger communities, pursuing an increase in modularity at each step. This process continues until no further modularity improvement can be achieved, resulting in the identification of communities with high internal connectivity and relatively fewer connections between them. Such an algorithm excels in optimizing modularity and helps identify communities that maximize the network's division strength. It also supports hierarchical clustering based on different topological parameters such as edge betweenness and degree of node. However, this is also its limitations encompassing the need to carefully select suitable parameters and accounting for potential sensitivity to initialization conditions during the community detection process. In addition, Fast Greedy Clustering may have limited power in identifying overlapping modules (Mahé et al., 2014).

1.2.2.4 Examples of network analysis

1.2.2.4.1 Protein-protein interaction network (PPIN) analysis

Protein-protein interactions (PPIs) are fundamentals of all cellular pathways, such as DNA replication and repair, signal transduction, molecular trafficking, regulation of apoptosis,

response to stimuli, etc. PPIN analysis collects the enormous amount of experimental data produced by a variety of PPI detection methods, as well as predictive data generated by machine learning algorithms, providing a systematic framework to explore the organisation and dynamics within a list of proteins (Koh et al., 2012). The process of PPIN analysis typically includes PPI detection, data collection, network construction, analysis and interpretation.

- *PPI detection*

PPI detection methods can be classified into 3 categories: *in vitro*, *in vivo* and *in silico* approaches. *In vitro* detection techniques are performed outside a living cell in a controlled environment, such as under a particular pH or temperature. Examples of *in vitro* methods include Tandem Affinity Purification-Mass Spectroscopy (TAP-MS), Affinity Chromatography, Coimmunoprecipitation, Protein microarrays, and X-ray crystallography. These methods normally have high specificity since they directly assess interactions between purified proteins. Also, methods like X-ray crystallography provides structural insights about protein complexes. However, these methods are not able to fully present the interactions occurring within living cells, which are normally more complex due to cellular dynamics. This may result in false negatives since certain PPIs may only occur under specific cellular conditions. In comparison*, in vivo* techniques involve detecting PPIs within the living organism, such as the Yeast 2-Hybrid (Y2H) system. These methods provide a detection environment with higher cellular relevance but are normally lower throughput than *in vitro* approaches. Distinct from the other 2 types, *in silico* methods predict PPIs based on a range of properties of queried proteins such as amino acid sequence, 3D structure, chromosome proximity, expression profiles, etc. Taking advantage of machine learning algorithms, these methods are always high-throughput and time-efficient. Also, after years of development, modern computational tools such as SPPS (X. Liu et al., 2012) and ProteinPrompt (Canzler et al., 2022) can predict PPIs with high accuracy. However, the main limitation of *in silico* methods is their high dependence on the quality and completeness of input data for the training of machine learning algorithms. Therefore, they may have limited power in predicting PPIs for unknown or less-studied proteins.

- *PPI data collection*

Data for constructing a PPIN can be gathered from above-mentioned detection techniques or derived from curated PPI databases, which are produced on the needs of storing, organising and interpreting the vast amount of PPI data generated by high-throughput detection techniques. Based on the data source and processing procedure, PPI databases can be categorised into 2 types: the primary and secondary repositories.

- Primary PPI databases

Primary databases are the foundational repositories that directly compile experimental data on protein-protein interactions from peer-reviewed literature. Information such as the identifies of protein interactors, detection methods, publications where the PPIs were reported, and other experimental details are curated manually and uploaded into repositories. Depending on different purposes, these databases can be general and comprehensive (i.e., IntAct, MINT and BioGrid), organism-specific (i.e., FlyBase and WormBase) or subject-specific (i.e., immunology, InnateDB). Apart from experimental data, some primary PPI databases are based on predictive PPIs via *in silico* method, such as (PESCADOR and iHOP). However, it is worth to note that none of the primary database is complete as they cover a limited potation of the literature. Therefore, in PPIN, in order to construct a comprehensive network as possible, it is always important to merge information derived from multiple primary databases to obtain the maximum literature coverage.

- Secondary PPI databases

By contrast, secondary databases aggregate and integrate interaction data from multiple primary sources, often providing additional features, such as network analysis, functional annotation, and integration with other types of biological data, such as gene expression or pathway information to facilitate PPI data analysis. Of note, many secondary PPI databases also collect interaction data from literature, thereby forming some commonly used hybrid tools/repositories such as STRING (Search Tool for the Retrieval of Interacting Genes/Proteins)(von Mering et al., 2005), HIPPIE (Human Integrated Protein-Protein Interaction Reference) (Alanis-Lobato et al., 2017) and APID (Agile Protein Interactomes Data Server) (Alonso-López et al., 2019).

- *PPIN construction and network analysis*

After deriving PPIs from multiple types of resources, the next steps are to construct a PPIN, investigate the features of its structure and organisation and probably integrate annotation data from other datasets such as functional/pathway enrichment knowledgebases, gene-disease association catalogues, etc. Manually connecting and annotating each PPI in the network is impractical, especially for large scale PPIN analysis. Therefore, some user-friendly tools have been developed to visualise and analyse PPINs.

  - Cytoscape

Cytoscape is one of the most commonly used open-source tools for biological network analysis. It is a multi-platform desktop application for network visualisation, integration, annotation and analysis (Shannon et al., 2003). The most pronounced feature of Cytoscape is the extensive selection of add-in apps that provide offering specialized features to complement its core functionality. These apps cover the diverse analytical needs across various knowledge domains, such as topological analysis (e.g., NetworkAnalyzer), cluster detection (e.g., MCODE, clusterMaker2) and Gene set enrichment analysis (e.g., BiNGO, ClueGO, EnrichmentMap). In addition, some Cytoscape apps are connected to annotation databases (e.g., ReactomeFIPlugin and KEGGscape) and empower the users to interpret the network from multiple aspects. The major limitation of Cytoscape is the high demand of computing resources, especially when it comes to large-scake networks.

  - Scripting packages

In terms of large-scale network analysis, some scripting packages have been developed via different programming languages, such as igraph for R, Python and C/C++ (https://igraph.org/) and NetworkC (https://networkx.org/) for python. These programmatic packages are less computing-resource-demanding and can be implemented as part of bioinformatics analysing pipelines.

- *Applications of PPIN analysis*

  - Predicting biological functions of uncharacterised proteins

PPIN analysis aids in understanding and predicting biological functions of unknown proteins by examining the interactions they have with well-studies proteins. Grounded in the principle that interacting proteins often participate in similar pathways or processes, these proteins

may share the established roles in specific functions as their protein interactors, which is known as “guilt by association principle”. Moreover, pathway enrichment analysis on the neighbourhood of the unknown protein helps identify overrepresented cellular pathways or functional categories, which refines the predictions by highlighting the most relevant biological contexts. However, it is important to note that the accuracy of functional predictions depends on the quality and comprehensiveness of the interaction data and thereby should be interpreted with caution and validated through experimentation to confirm the true function of the unknown protein.

- Mapping a complicated cellular cascade

PPINs can be used to map pathways characterised by only several confirmed hub proteins. Starting from these hubs, a PPIN can be constructed by investigating the direct interactors of these key proteins, which are often referred as “neighbour proteins”. By examining the network, neighbour proteins with higher topological centrality are prioritised as potential novel hub proteins of the pathway. Though further validation is required, these proteins may unveil the missing links in the pathway. In addition, topological clustering on the neighbour proteins may reveal functional units, in which the proteins closely interact with each other and participate in the pathway as form of protein complexes.

- Investigating disease mechanisms and guiding drug development

A disease-specific PPI network can be constructed based on proteins of disease-causing genes and potential or established therapeutic targets. By expanding and analysing the networks through investigating interactors of these proteins, a more comprehensive view of the molecular mechanism is obtained for the disease of study. Additionally, pathway enriched within these proteins can reveal critical pathways and processes that are probably altered in the disease state. Furthermore, such disease-specific PPINs are instrumental in drug discovery. For example, novel hub proteins identified in the expanded PPIN probably play a central role in the disease-associated signalling pathways and thereby are potential drug targets for therapeutic intervention. After further computational analysis and experimental validation, researchers can design drugs that modulate the interactions or activity of disease-related proteins, ultimately working to protect or restore normal cellular function. Moreover, the analysis of disease-specific PPIN aids in the repurposing of existing drugs. By identifying connections between known drugs and disease-associated proteins within the network,

researchers can explore the potential of these drugs for new therapeutic applications, which accelerates the drug development process to a large extent.

- *Limitations of PPIN analysis*

Despite the advantages and wide application of PPIN analysis, it is important to recognise the limitations the come with it. First of all, the accuracy of PPIN analysis largely replies on comprehensive and reliable data resources. However, while most of PPI databases are well-organised and precisely-annotated, some problems cannot be avoided. For example, a large proportion of PPIs provided by current online resources were produced by high-throughput techniques, which comes with risk of false positives or false negatives (Tarassov et al., 2008; von Mering et al., 2002). These inaccuracies can impact the quality of interaction data. Also, it is challenging to keep PPI data up-to-date if a database relies on manual annotation. Outdated data within these databases can introduce inaccuracies into PPIN analyses and leading to potentially misleading conclusions. In addition, literature-based PPI databases are generally incomplete (Manconi et al., 2012). Proteins that lack of study possess insufficient PPIs and thereby lack of centrality in typical topological analysis of PPIN. However, these proteins may actually have more extensive set of interactors and play an essential role in the PPIN in the real case (Tomkins & Manzoni, 2021). A combinative use of multiple PPI data resources can help alleviate this bias. Additionally, combining PPI data with other types of information such as expression levels or binding sites may help fill the gaps in the incomplete PPINs. Moreover, for predicted PPIN network, although the predictions are made on the base of experimental data, the PPIN may not represent the true dynamic and multifaceted nature of cellular PPI systems. For example, some PPIs may only occur in a particular cell compartment or only happen in a certain stimuli circumstance. Therefore, the need for experimental validation should never be ignored despite the development of advanced network analysis algorithms. Finally, PPIs lack in tissue specificity as they are generally defined in few cancer cell line experiments or in completely ex-vivo systems, while in reality they have biological meaning within specific tissues and cell types.

1.2.2.4.2 Weighted gene co-expression network (WGCN) analysis

Gene co-expression refers to the tendency of certain genes to exhibit similar expression patterns across diverse experimental conditions or biological samples. In the molecular biology context, co-expressed genes often participate in common biological processes. WGCNs are built upon this principle by creating a weighted network on co-expressed genes, where genes are represented as nodes while the connections between the co-expressed genes are represented as edges. Different weights are assigned to their connections based on the strength of their co-expression. The main aim of WGCN analysis is to identify co-expression modules based on weighted network analysis, define hub genes within each module and study the relationships between co-expression modules. These modules usually reveal the coordinated functioning of genes and the underlying regulatory mechanisms governing various physiological and pathological processes. Based on the consideration of the directionality of gene correlations, WGCN analysis can be divided into 3 types: signed, unsigned and signed hybrid forms. In the signed approach, the analysis incorporates both positive and negative correlations between genes, while the unsigned approach focuses solely on the strength of correlations, regardless of their direction. In the signed hybrid approach, positive correlations are kept while negative correlations are defined as 0.

- *Data source for gene co-expression analysis*

In co-expression analysis, mRNA level is often used as a reflection of gene. Microarray and RNA-Sequencing (RNA-Seq) are 2 primarily used high-throughput techniques for measuring mRNA levels, in which RNA-Seq has higher sensitivity and throughput than microarray but it is more expensive to run, especially for large-scale studies. In recent years, several data repositories have been established to store well-formatted and curated RNA-Seq data. These resources can be a catalogue of a range of individual studies such as the Expression Atlas (https://www.ebi.ac.uk/gxa/home) and GEO (Gene Expression Omniubs, https://www.ncbi.nlm.nih.gov/geo/), or based on a large-scale study based on a single cohort, such as GTEx (Genotype-Tissue Expression project, https://www.gtexportal.org/home/).

- *Network construction and module detection*

After getting the expression data, the first step of WGCN construction is the calculation of pairwise gene correlations. Common correlation measures include Pearson correlation, Spearman rank correlation, or biweight midvariance. Pairwise co-expression coefficients form the "correlation matrix". Afterwards, a soft thresholding power (β) is applied to the correlation matrix to generate the "adjacency matrix", in which strong correlations are emphasize out of weaker ones. The way β is implemented on the correlation matrix varies across different types of WGCN analyses. For signed networks, $a_{ij} = (\frac{1+cor(i,j)}{2})^{\beta}$ in which a_{ij} refers to the adjacency between gene *I* and *j*; while for unsigned networks, $a_{ij} = |cor(i,j)|^{\beta}$; and in terms of signed hybrid networks, for positive correlations, $a_{ij} = |cor(i,j)|^{\beta}$, while for negative correlations, $a_{ij} = 0$. Next, the adjacency matrix is converted into a Topological Overlap Matrix (TOM), where the correlation information between genes is transformed into topological similarities among nodes in the network. Afterwards, hierarchical clustering is performed on the TOM to group genes with similar co-expression patterns. This results in a dendrogram, to which a dynamic tree cutting algorithms is applied to identify gene modules. Furthermore, an eigengene is calculated for each module, representing the first principal component of the module's overall expression profiles. Similarities between these module eigengenes can be further examined and modules with highly correlated eigengenes can be merged. In addition, the eigengenes can be correlated with clinical traits or phenotypes to identify disease or function-related gene modules.

- *Limitations of WGCN analysis*

Although WGCN analysis is a powerful tool in systems biology that provides insight into the complicated gene regulation and connection network, it has some limitations. First of all, this analysis is highly sensitive to the choice of β, the soft thresholding parameter. A high β leads to a more stringent threshold, resulting in a sparser network with stronger connections, while a low β yields a denser network with weaker connections. There are multiple algorithms that help in β selection. For example, the R package "WGCNA" contains a function named as "pickSoftThreshold", which optimise the value of β to ensure scale-free topology for the WGCN network (Langfelder & Horvath, 2008). An appropriate β is crucial for maintaining reliability and biological interpretability of the co-expression modules identified through

WGCNA. In the meantime, as for other *in silico* network models in systems biology, WGCNs are not able to fully represent the detailed signalling flow of in a biological process. For example, since they only model the co-expression patterns between genes, there is lack of directionality, they are unable to distinguish between upstream and downstream regulatory interactions. Therefore, it is always important to interpret these computational models with experimental validation.

1.2.2.4.3 Differential expression analysis (DEA)

- *Data preprocessing*

DEA identifies genes exhibiting significant changes in their expression level under different experimental conditions. Data source for DEA include proteomics and transcriptomics data produced by high-throughput techniques such as Mass Spectrometry and RNA-Seq. As compared to co-expression analysis, DEA is more sensitive to data quality since it focuses on the absolute value of expression. Therefore, the data preparation process is more complex for DEA. For example, for RNA-Seq data, quality control (QC) should be performed at sample and gene-level (Anders & Huber, 2010; Qi et al., 2017). Sample-level QC is aimed to i) observe whether the experimental condition explains the major source of variation among samples and ii) remove any possible outliers caused by batch effect or technical errors. In comparison, Gene-level QC involves omitting genes with missing, zero or extremely low counts in most of samples or the genes with count outlier, to increase the statistical power of further analysis. Principal component analysis (PCA) and hierarchical clustering are commonly-used QC method to visualise clustering pattern and detect outliers in large datasets (Tang et al., 2015). Moreover, read count normalisation is another crucial step in data preprocessing for DEA performed with RNA-Seq data. By far, plenty of normalisation methods have been suggested for RNA-Seq data, such as RPKM (Reads Per Kilobase Million) or FPKM (Fragments Per Kilobase Million). These methods normalise the raw read counts for gene length and sequencing depth (i.e., to the total number of reads or fragments obtained from an RNA-seq experiment) (Zhao et al., 2021). However, these methods do not take consideration of RNA library composition (i.e., the relative abundance and diversity of different RNA molecules present in a sample), though it is an important concept since a few highly differentially expressed genes or the presence of contamination can perturb mRNA expression profile in the examined samples

and thereby cause bias (Maza et al., 2013). Therefore, some new normalisation methods have been implemented and are applied like DESeq and TMM (Trimmed Mean of M-values) which are robust against high-count genes and differences in RNA composition(Love et al., 2014; Robinson et al., 2010).

- *Statistical testing*

Multiple types of statistical tests can be applied for DEA, such as t-tests, ANOVA, or specialized statistical methods within algorithms like DESeq2 and edgeR for RNA-seq data (Love et al., 2014; Robinson et al., 2010). Of note, multiple test correction, such as Bonferroni correction or false discovery correction, is necessary for controlling the Type II error rate especially in large-scale DEA pipeline.

- *Data visualisation*

Comparison results of DEA are normally presented via a volcano plot, in which the x-axis pepresents the fold change in gene expression between two conditions: a positive value indicates upregulation, and a negative value indicates downregulation, while the y-axis prepresents the statistical significance of the observed changes (W. Li, 2012). Typically, the negative logarithm (base 10) of the adjusted p-value is used. The higher the point on the Y-axis, the more statistically significant the change in expression.

1.2.2.4.4 Functional enrichment analysis

Functional enrichment analysis is used to identify functional categories or biological pathways that are overrepresented in a set of genes or proteins. It provides functional annotation for queried proteins or genes that are shortlisted from experimental evidence or by other systems biology approaches such as network analysis. Functional enrichment analysis plays a crucial role in associating *in silico* models with their potential biological significance.

- *Algorithm*

There are a wide range of analysing tools and platforms for functional enrichment analysis, which follow similar algorithms. The analysis begins with a list of genes or proteins of interest (test set), a larger set of genes or proteins to which the test set is compared (reference set or

"universe", this is often the entire genome or proteome of reference for the species under analysis) and by selecting a relevant functional annotation database, which contains information that link genes or gene products to functional terms such as cellular pathways, molecular functions and biological processes, in a standardised way. Annotations are curated through a combination of automated methods (such as text mining) and manual curation of peer-reviewed publications by domain curators. Within a functional annotation database, each functional term possesses a "complete" list of entities that were annotated as its contributors. Statistical methods, such as hypergeometric or Fisher's exact tests, are then applied to assess overrepresentation, i.e., to evaluate whether the observed number of genes in a particular functional category in the input set is significantly higher than in the universe. At last, significant enrichments are further adjusted via multiple testing correction.

- *Functional annotation databases*
    - Gene Ontology (GO)

Gene Ontology (GO) is one of the most widely utilised functional annotation databases (Consortium et al., 2023). Developed by the Gene Ontology Consortium, GO serves as a standardized vocabulary that classifies genes into three major categories: biological processes, molecular functions, and cellular components. This ontology provides a structured and hierarchical representation of the relationships between these terms, offering a comprehensive and organized view of the functional landscape of biological systems. The Biological Process (GO-BP) category encompasses broad biological activities, Molecular Function (GO-MF) involves specific biochemical activities, and Cellular Component (GO-CC) describes the locations within the cell where gene products are active. The ontology forms a directed acyclic graph (DAG), where terms are represented as nodes while the relations between the terms as edges. Some of the commonly used relations in GO include is a (is a subtype of); part of; has part; regulates, negatively regulates and positively regulates. The nodes and edges are organized in a hierarchical manner, reflecting the complicated "ontology tree". There are plenty of tools are available for GO enrichment analysis, such as DAVID (Database for Annotation, Visualization, and Integrated Discovery) (Sherman et al., 2022), Enrichr (Kuleshov et al., 2016) and g:GOSt from the g:Profiler tool set (Kolberg et al., 2023).

- Reactome

Reactome offers a freely available, open-source relational database of human-centric cellular pathways. The core unit of the Reactome data model is the reaction, forming a network of biological pathways with high density of interaction (Gillespie et al., 2022). The reactions, pathways and the interactions in-between in Reactome are manually curated from peer-reviewed literatures. In addition, the tool provides comparative analysis among different organisms, using human pathways to compute equivalent pathways in 20 other species. Moreover, Reactome provides users with graphical maps of cellular pathways through its website, offering an interface to access detailed information. Apart from its website, Reactome database can be accessed from many other analysing platforms which allows enrichment analysis on multiple annotation data recourses. It can also be accessed via API through scripting R package ReactomePA or Python package reactome2py, functioning as part of bioinformatics analysing pipeline.

- *Limitations*

The main drawback of functional enrichment analysis comes from the variability in the quality and depth of annotations across different biological domains. Certain areas of biology are more extensively and precisely annotated, resulting in a potential bias in the statistical analysis. For example, well-characterised proteins receive more thorough annotations as compared to less popular or newly identified proteins. Similarly, some pathways/functions receive more attention than others with the consequence that their ontology is a larger tree (i.e. containing a large array of terms) in comparison with less popular pathways/functions. Additionally, the criteria may vary when a gene or gene product is annotated either by curators or computational algorithms, which can be another source of bias. Another limitation of functional enrichment analysis is the complexity of results especially for large input sets. This is due to the generation of overwhelmingly detailed ontologies with complicated structure and a large proportion of similar terms. To alleviate this problem, many tools (such as GOslims) offer simplified version of enrichment analysis which combines semantically similar terms and assign fine detailed terms to more general parent terms, making the enrichment results more interpretable. However, due to the variety of simplifying pipelines, outputs produced by different tools for the same query set may differ.

1.3. Machine Learning (ML) in bioinformatics

Machine learning (ML) is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform tasks without explicit programming (Jordan & Mitchell, 2015). The central idea behind machine learning is to allow machines to learn from data, recognize patterns, and make decisions or predictions based on that learning. The sheer volume and complexity of biomedical data, from the intricate details encoded in DNA sequences to the dynamic interactions within protein structures, from the comprehensive records encapsulated in electronic health records and the visual narratives provided by medical imaging, demands sophisticated computational approaches for analysis and interpretation. This is where ML steps in, offering algorithms and tools to process, analyse, and derive meaningful patterns from these datasets.

The ML process is a structured sequence involving distinct stages (Alpaydin, 2021). It begins with data collection, where relevant information is gathered, encompassing examples, experiences, or observations that serve as the foundation for the machine's learning. Following this, data preprocessing takes centre stage, involving the meticulous cleaning and organization of data to render it suitable for training machine learning models. This step encompasses handling missing values, normalizing data, and converting it into a format conducive to analysis. Feature extraction then comes into play, focusing on the selection or transformation of key features (attributes) crucial for the learning task at hand. The heart of the machine learning process lies in model training, wherein an algorithm is employed to enable the machine to "learn" from the provided data. During this phase, the model adjusts its parameters to minimize errors or precisely predict outcomes. Subsequently, the trained model undergoes evaluation to gauge its performance on new, unseen data, ensuring its ability to generalize effectively and make accurate predictions. Finally, in model deployment, the trained model is applied to make predictions on real-world data, completing the iterative cycle of the machine learning process. In general, ML models can be divided into 2 types: supervised and unsupervised ML. Both of them are widely used in bioinformatics studies (El Naqa & Murphy, 2015).

1.3.1 Supervised ML algorithms

Supervised ML involves training algorithms on labelled datasets, where the input data is paired with corresponding output or target labels (Ghosh & Dasgupta, 2022b). With clear labels for training data, supervised learning is well-suited for tasks like classification and regression, making it applicable in various scenarios, especially in disease diagnosis and prediction. For example, in the context of Parkinson's disease, a prediction model can be trained on a dataset incorporating PD patients and healthy controls with their demographic features like age and gender, genetic features like with/without LRRK2 variants and motor & non-motor clinical features such as tremor severity, motor scores and cognitive scores. Following meticulous data preprocessing steps, supervised ML algorithms, such as logistic regression, is employed to train a model capable of discerning the intricate relationship between these features and the likelihood of an individual having Parkinson's disease and optimising the set of predictors with the best predicting performance. Rigorous evaluation, including metrics like accuracy and precision, ensures the model's reliability using a distinct testing set (Ghosh & Dasgupta, 2022c). Once validated, the logistic regression model is deployed for predicting the likelihood of PD in new individuals.

1.3.2 Unsupervised ML algorithms

Unsupervised ML algorithms explore and uncover inherent relationships within complex biological datasets (Jordan & Mitchell, 2015). It is commonly used for clustering analysis in complex biological networks, where similar entities are grouped together based on shared characteristics. For example, the above-mentioned module detection algorithms in PPIN and WGCN are all based on unsupervised ML approaches. In addition, unsupervised ML algorithms are also widely used for dimensionality reduction, which aids in simplifying while retaining essential features in high-dimensional data. The commonly used unsupervised ML algorithms include Principal Component Analysis (PCA), k-means and hierarchical clustering.

1.3.2.1 PCA

PCA is a dimensionality reduction approach widely used in nearly all sorts of high-dimensional data. For example, in Genome-Wide Association Studies (GWAS), PCA helps in condensing a range of potential cofounders such as ethnicity, gender, age, etc. into a set of linearly

uncorrelated variables known as principal components. By including the top principal components as covariates in the statistical models of association testing, researchers can effectively control for population stratification and minimise the risk of false-positive associations due to differences in genetic ancestry (Kurita, 2020). Moreover, in the context of gene expression studies, variations in sample quality, batch effects, or other sources of unwanted variability can impact the accuracy of results. In this scenario, PCA can be used as a valuable tool for sample quality control by capturing the dominant patterns of sample variability(Maćkiewicz & Ratajczak, 1993) . By visually inspecting the PCA plots or examining the proportion of variance explained by each principal component, researchers can identify outliers or clusters of samples with distinct expression patterns. Samples deviating from the main cluster in the PCA plot may indicate issues with sample quality or batch effects and thereby need to be re-examined or excluded from subsequent analysis.

1.3.2.2 K-means

K-means clustering is a partitioning algorithm widely used in bioinformatics (Lloyd, 1982). The algorithm operates by iteratively assigning data points to clusters based on their similarity to the mean of each cluster. The goal is to minimize the sum of squared distances between data points and their respective cluster centroids. One crucial aspect in implementing k-means is defining the optimal number of clusters, often denoted as 'k.' This can be achieved through techniques like the elbow method or silhouette analysis. The elbow method involves plotting the cost function (sum of squared distances) against different values of k and choosing the point where the rate of decrease sharply changes, resembling an elbow (Thorndike, 1953). Silhouette analysis quantifies how well-separated clusters are, with higher silhouette scores indicating better-defined clusters (Rousseeuw, 1987). In bioinformatics, k-means clustering is employed for various tasks, such as classifying gene expression patterns, grouping similar biological samples, or identifying distinct subtypes within a population. It aids in uncovering hidden structures within large biological datasets, providing valuable insights into the underlying biological processes and facilitating further analysis and interpretation.

1.3.2.3 Hierarchical clustering

Unlike k-means clustering, hierarchical clustering does not require specifying the number of clusters beforehand (Ghosh & Dasgupta, 2022a). The algorithm works by successively merging or dividing clusters based on the similarity of data points. The result is a hierarchical tree-like structure, or dendrogram, illustrating the relationships between data points and their groupings. There are 2 main types of hierarchical clustering: agglomerative, where each data point starts as its own cluster and is merged iteratively, and divisive, which begins with all data points in a single cluster and splits them (Nielsen, 2016). In both of the 2 types, deciding where to cut the dendrogram to form distinct clusters is a critical step. Similar approaches used in k-means can be applied in hierarchical models to define the optimal number of clusters and the "cutting height" for the dendrogram.

In terms of its application, hierarchical clustering finds extensive use in tasks such as classifying gene expression patterns, categorizing biological samples, and identifying relationships between different species or genomic features. The dendrogram generated by hierarchical clustering helps visualize the inherent structure in complex biological datasets, enabling researchers to explore and interpret the underlying patterns and relationships within biological systems.

1.3.3 Common issues in ML models

1.3.3.1 Overfitting

One of the main challenges in ML model construction is the risk of overfitting, which occurs when the model fits too closely to the training dataset but cannot generalize well to testing datasets (Altman & Krzywinski, 2018). It happens due to several reasons, such as small and noisy training set or a too complex model that involve an overwhelming number of parameters. Therefore, to reduce overfitting, one common approach is to use a more straightforward model with fewer parameters and train the model on a clean dataset with larger size (Ghosh & Dasgupta, 2022a). In the meantime, some algorithms have been developed to further decrease the risk of model overfitting, such as k-fold cross-validation, in which the training set is partitioned into k subsets or folds (R Kohavi 1995). The model is trained k times, each time using k-1 folds for training and the remaining fold for validation. This process is repeated k times, with each fold used exactly once as the validation data. The performance metrics obtained from each run are then averaged to provide a comprehensive

evaluation of the model's performance. K-fold cross-validation is particularly valuable when dealing with data of limited sample size, as it maximizes the use of available information for both training and validation. This approach helps mitigate overfitting and provides a more reliable estimate of how well a model is likely to perform on testing data.

1.3.3.2 Multicollinearity

Multicollinearity, arises when two or more predictor variables in a regression model are highly correlated (Dormann et al., 2013). This correlation makes it challenging for the ML model (especially supervised ML models) to distinguish the individual effects of each variable on the response variable, leading to instability in estimating regression coefficients. LASSO (Least Absolute Shrinkage and Selection Operator) is a regularisation technique introduced by Robert Tibshirani in 1996 aimed to mitigate multicollinearity (Tibshirani, 1996). The key characteristic of LASSO is its ability to shrink the coefficients of less influential variables all the way to zero which helps in feature selection by identifying and keeping only the most relevant predictors, reduce multicollinearity when predictor variables are highly correlated. The regularization parameter, often denoted as λ (lambda), controls the strength of the penalty. Larger values of λ result in more aggressive shrinkage and greater sparsity in the model. Therefore, the optimisation of λ is a crucial step in LASSO algorithm, involving a k-fold cross-validation process to assess the model performance for different λ values. The λ value minimises the Mean Squared Error (MSE) is selected as the optimal regularisation parameter, and the predictors with shrinked coefficients are excluded from the mode. These predictors are either highly correlated with each other or lack of predictive capability on the outcome.

1.4 Hypothesis testing

The work in this thesis will try and answer the following questions:

1. Is it possible to categorize the LRRK2 interactome into functional and topological units to better understand the various functions associated with LRRK2?
2. Do the function and expression of the LRRK2 interactome remain consistent across all tissues, or is there a tissue-specificity that influences both interacting proteins and interactions?

3. Do interactions within the LRRK2 interactome exhibit similar patterns in sporadic and familial Parkinson's disease, or are there distinct alterations that are indicative of one disease form over the other?

Databases, analysing tools and software used in this study

2.1 Online resources

2.1.1. PPI data

2.1.1.1 HIPPIE (v2.3)

HIPPIE (the Human Integrated Protein-Protein Interaction Reference) is a web resource that provides QC-ed PPIs with context-specific annotation (Alanis-Lobato et al., 2017). The PPI repository of HIPPIE is extracted and merged from IntAct, MINT, BioGRID, HRPD, DIP, BIND and MIPS, which is annually updated via the PSICQUIC interface. Afterwards, a confidence score is computed for each PPI in a semi-computational manner, examining the amount and the quality of the experimental evidence behind it, as well as the number of non-human organisms in which a PPI was reproduced, aiming to reduce the number of false positives. However, since this process involves manual evaluation on the reliability and accuracy of PPI detection methods, it may induce potential bias since it is difficult to quantify the quality of different techniques. For example, “gst pull down (MI:0059)” and “peptide array (MI:0081)” obtain same score in HIPPIE Experiment Based Quality Scoring system. However, in a real experimental case, the performance of these 2 techniques can be affected to a different extent by the properties of protein interactors or detecting environment. Therefore, a hard threshold may not be appropriate while performing QC with HIPPIE PPI scores. Different from PINOT, HIPPIE provides PPI annotation from multiple aspects using **i)** tissue-specific mRNA-level expression data derived from GTEx, **ii)** gene functions retrieved from Gene Ontology terms (biological process and cellular compartment) and **iii)** relatedness to disease extracted from MeSH ontologies. In addition, HIPPIE also provides predicted and KEGG-curated PPI directionality, i.e., whether a PPIs is activating or prohibiting the function of protein interactors. Moreover, HIPPIE integrates third-party-driven disease and functional gene set enrichment analysis, supported by GS2D(Gene Set to Diseases) (Andrade-Navarro et al., 2016)and PANTHER. These annotations help in interpreting the output PPI network.

2.1.1.2 MIST (v5.0)

MIST (Molecular Interaction Search Tool) is an online tool that integrate PPI data from curated public data resources, including BioGrid, IntAct, MINT, DIP, DroID, mentha, HPRD,

HumanMAPPK, PomBase and FlyBase, covering 9 prominent model organisms including *Mus musculus and Homo sapiens* (Hu et al., 2018). In terms of QC pipeline, MIST uses a similar algorithm as HIPPIE, which ranks PPIs based on the number of detections approaches and publications, as well as the comparison to other organisms, though without manually scoring PPI detection methods. In addition, unlike the quantitative approaches used by HIPPIE, in MIST, interactions obtain a high ranking when they were reported in multiple references and/or with diverse approaches. In contrast, interactions falling short of these criteria but backed by support from various species receive a moderate ranking, while all other PPIs are ranked as low. This categorical classification alleviate bias from PPI scoring system but may cause false positive in uni-organism PPI research. Moreover, MIST provide searching filters on evidence type, which helps distinguish direct interactions from indirect or uncertain interactions based on the detection approaches. The unique feature of MIST is that it supports cross-species interaction analysis by mapping PPIs of one species to another via the DRSC Integrative Ortholog Prediction Tool (DIOPT), which largely increase the number of candidate interactors. However, further experimental validation is especially important for such analysis due to an expected high false positive rate.

2.1.1.3 PINOT (v1.0)

PINOT (the Protein-Protein Interaction Query Tool) is an online server designed to efficiently search curated literature and extract the most recent PPI data associated with specific proteins or genes of interest (Tomkins et al., 2020). The unique feature of PINOT is that human PPIs are directly downloaded from the PSICQUIC (Proteomics Standard Initiative Common Query Interface) platform at every query. PSICQUIC is founded by the HUPO Proteomics Standard Initiative (HUPO-PSI) to standardise the access to multiple molecular interaction databases programmatically, facilitating the integration of PPI data across various sources (Aranda et al., 2011). So far PSICQUIC includes a total of 34 members, including the prominent PPI databases such as IntAct. As compared to the built-in repositories, which most of other PPI searching tools rely on, retrieving PPIs from PSICQUIC ensures the most up-to-date PPI output of larger coverage. In addition, PINOT incorporates multiple QC steps, which is mainly based on counting the number of different methods by which a PPI was detected (the "method score") and the number of literatures where a PPI was reported (the "publication score"). Of note, due to the variety of curation protocols utilised by different databases,

simply adding up the number of records of detection methods which merged from primary data sources can be biased. To dilute the bias, PINOT includes a unique and crucial QC step involving grouping technically similar PPI detection methods based on an in-house dictionary. For example, “Two Hybrid - MI:0018”, “Two Hybrid Array - MI:0397” and “Two Hybrid Pooling Approach - MI:0398” are grouped together into the “Two Hybrid (2Hyb)”. In this way, the “method score” based on clustered and reassigned method groups is more accurate. Moreover, users are allowed to choose from 2 different method grouping algorithms (“Lenient” or “Stringent”) based on the query preference, which slightly change the way how PINOT group general detection methods such as “biophycial (MI:0013)” and “docking (MI:0035)”. However, it is worth to mention that PINOT’s method dictionary is not yet complete, which thereby limits its capability in reducing method redundance bias.

2.1.2 Protein list annotation: g:Profiler

g:Profiler contains a collection of tools covering functional enrichment analysis (g:GOSt), protein/gene identifier conversion (g:convert and g:Orth) and human SNP-gene mapping (g:SNPense). In this project, all enrichment analyses were performed via g:GOSt which performs functional enrichment analysis calculating overrepresentation via a standard Fisher's one-tailed test (cumulative hypergeometric probability) using the most commonly annotation tools such as Gene Ontology (molecular function, biological process and cellular component), KEGG, Reactome and WikiPathways. In terms of multiple testing correction, g:GOSt provides options for Bonferroni’s correction, Benjamini–Hochberg False Discovery Rate and g:SCS, which is a specific p-value correction approach developed for functional enrichment analysis, taking into consideration that functional terms are not completely independent from each other due to the way functional ontologies are organized with terms linked in parent/child relationships. By default, the analysing background is set as all annotated protein-coding genes. The background can be customised by the user, for example, a set of disease-related genes can be used as background while performing disease association enrichment analysis. In addition, g:GOSt support multiquery to compare enrichment results among different protein lists.

2.1.3 RNA-Seq data

2.1.3.1 GTEx

The Genotype-Tissue Expression (GTEx, https://www.gtexportal.org/) project is aimed at establishing a comprehensive public resource for molecular assays such as WGS, WES and RNA-Seq, which gathers samples from 54 tissue sites across nearly 1000 healthy individuals (Aguet et al., 2020). In addition, GTEx also uses the internal dataset to provide information regarding expression quantitative trait loci (eQTL) or splicing quantitative trait loci (sQTL) that were identified from genetic variation that are highly correlated with the expression levels or alternative splicing. This project utilised RNA-Seq expression data (in read counts) obtained from GTEx for different brain regions and peripheral tissues.

2.1.3.2 PPMI

The Parkinson's Progression Markers Initiative (PPMI, https://www.ppmi-info.org/) is a multi-central study that aims at identifying biological markers of Parkinson's risk, onset and progression. It provides a comprehensive, standardised, longitudinal dataset that contains clinical features, bioimaging and genetics data of 3 major cohorts, including "Parkinson's disease confirmed cases", "Prodromal cases", "Healthy controls", in which the "Parkinson's disease" cohort contains sporadic untreated PD participants and genetic PD participants with pathogenic mutations in *LRRK2, GBA, SNCA, PRKN and PINK1*, while the prodromal cohort are individuals that are at risk of PD based on clinical features (such as imaging), presence of genetic variants or first-degree family history. This project used whole blood RNA-Seq data from the PPMI initiative. Of note, access to this initiative is restricted to researchers who have received permission upon signing a user agreement.

2.2 R packages

2.2.1 UniprotR

UniprotR is an R package that accesses protein-related details by connecting to Uniprot (https://www.uniprot.org/) and retrieve information like the name or taxonomy details of a protein (Soudy et al., 2020). In this study, it was used to extract protein family information for LRRK2 interactors

2.2.2 GO.db

GO.db is an R package that encompasses a collection of annotation maps providing a comprehensive description of the entire Gene Ontology. These maps are constructed using data sourced from the Gene Ontology database. DOI: 10.18129/B9.bioc.GO.db. In this study, it was used to extract hierarchical relation between Gene Ontology terms enriched for LRRK2 interactors.

2.2.3 Wordcloud

Wordcloud is an R package that helps create word clouds, visualize differences and similarity between documents, and avoid over-plotting in scatter plots with text. URL: http://www.fellstat.com/. In this study, it was used to extract and present the key words of enriched Gene Ontology terms for LRRK2 interactors.

2.2.4 WGCNA

WGCNA is an R package that performs Weighted Correlation Network Analysis (WGCNA) on high-dimensional data (Langfelder & Horvath, 2008). Includes functions for data cleaning, construction of correlation networks, module detection, summarisation, and relating modules to sample traits. In this study, it was used to perform WGCNA on expression data of LRRK2 interactors in healthy tissues and the whole blood mRNA level of healthy controls vs. patients with sPD and LRRK2-PD.

2.2.5 DESeq2

DESeq2 is an R package that estimates variance-mean dependence in read count data from high-throughput sequencing assays and test for differential expression based on a model using the negative binomial distribution. It also provides functions for data normalisation and visualisation. URL: https://github.com/thelovelab/DESeq2. In this study, it was used to perform differential expression analysis on expression data of LRRK2 interactors in healthy tissues and the whole blood mRNA level of healthy controls vs. patients with sPD and LRRK2-PD.

2.2.6 SKAT

SKAT is an R package that performs kernel-regression-based association tests including Burden test, SKAT and SKAT-O. These methods aggregate individual SNP score statistics in a SNP set and efficiently compute SNP-set level p-values. URL: https://cran.r-project.org/web/packages/SKAT/index.html. In this study, it was used to perform genetic burden analysis of LRRK2 interactors against sPD and LRRK2-PD

2.3 Software

2.3.1 Cytoscape 3.13.2

Cytoscape is an open-source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data (Shannon et al., 2003). Cytoscape by default provides a basic set of features for data integration, topological analysis, and network visualization. Additional features are available as Apps, which aid in a wide range of bioinformatics analysis such as network and molecular profiling analyses, provide new layouts and support scripting, and connection with databases. Most of the Apps are freely available from Cytoscape App Store (https://apps.cytoscape.org/). In this study, it was used to perform topological and clustering analysis on the PPI network of LRRK2, as well as generate network graphs.

2.3.2 GraphPad Prism 10.1.1

GraphPad Prism combines functions of scientific graphing, comprehensive curve fitting (nonlinear regression), statistics, and data organization. It was used to perform One-way ANOVA with post hoc Tukey's test and data visualisation.

2.3.3 StataSE 15.0

Stata, a versatile statistical software package designed for tasks such as data manipulation, visualization, statistical analysis, and automated reporting. Researchers across various disciplines, including biomedicine, economics, epidemiology, and sociology, utilize Stata for their analytical needs. In this study, it was used to perform data manipulation, t-tests and chi-square tests.

2.3.4 R 4.1.2 & RStudio

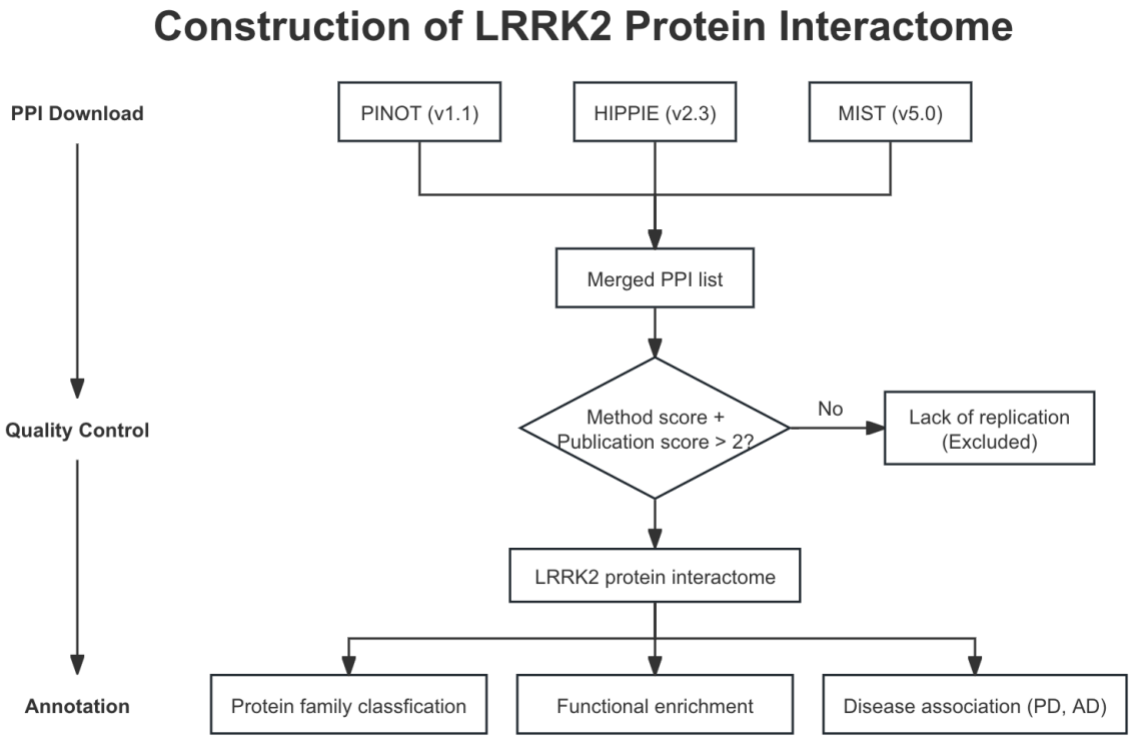
RStudio is an integrated development environment tailored for the R programming language, specializing in statistical computing and graphics. In this study, all systems biology analyses were based on R and performed via RStudio.

Chapter 1. Construction and analysis of the LRRK2 protein interactome

Objectives

- To construct a protein interactome of LRRK2 ($LRRK2_{int}$) with peer-reviewed PPI data
- To functionally annotate the $LRRK2_{int}$ via functional enrichment analysis
- To investigate the association between the $LRRK2_{int}$ and neurodegenerative disease (Alzheimer's disease, AD) and PD via Gene Set enrichment analysis

Analysis pipeline



Note: Figures and text in Chapter 1 are adapted from the following publication: Zhao et al.; "Tissue specific LRRK2 interactomes reveal a distinct striatal functional unit"; PLoS Comput Biol. 2023 Jan 30;19(1):e1010847. doi: 10.1371/journal.pcbi.1010847.

Methods

- Construction of LRRK2 protein interactome ($LRRK2_{int}$)

PINOT (v1.1), HIPPIE (v2.3) and MIST (v5.0) were queried to download "*homo sapiens*" PPIs for LRRK2 (UniProt ID: Q5S007, 21 October 2020). To access the broadest possible set of PPI data, "Lenient" filter level was applied in PINOT; for MIST, "Networks to Search" was set as "protein-protein interactions" only with "no filer by rank".

PPIs derived from the 3 web resources were quality-controlled via an in-house pipeline:

1) The identifiers (IDs) of LRRK2 protein interactors derived from the 3 repositories were converted to HUGO Gene Nomenclature Committee (HGNC) gene symbols. Interactors marked as "Unreviewed" in UniProtKB were removed;
2) Converted protein lists and the corresponding records of detection methods and publications were merged after removing duplicates;
3) Interaction detection method IDs were reassigned referring to the PINOT Method Classification Dictionary (https://www.reading.ac.uk/bioinf/PINOT/FILE2.xlsx), which clusters similar detection methods annotated in PSI-MI ontology (e.g. "Enzymatic Study MI:0415" and "Polymerisation MI:0953" are allocated in the same category: "Enzyme") and merged after removing duplications;
4) LRRK2 interactors were then scored (Final Score, FS) by adding the number of reassigned detection methods (Method Score, MS) and the number of reporting publications (Publication Score, PS);
5) LRRK2 interactors with FS ≤ 2 were removed from further analysis due to their low reliability (either were not replicated in multiple experiments or with missing publication identifier or with missing record of detection method).

- Functional annotation

Protein family classification

Family domains for LRRK2 interactors were extracted from UniProtKB via the R package "UniprotR". Retrieved family domains were further classified based on semantic similarity and biological functions.

Gene Ontology (GO) Enrichment Analysis

GO Biological Processes (GO-BP) enrichment analysis were performed on the $LRRK2_{int}$ via g:GOSt (https://biit.cs.ut.ee/gprofiler/gost) on 18th March 2023. Query parameters were set as following: Organism: *Homo sapiens* (Human); Statistical domain scope: Only annotated genes (only genes with at least one annotation); Significance threshold: Bonferroni correction. Of note, for the interactors with multiple Ensembl IDs, only the ID with the most GO annotations was kept. Of note, GO-BPs with term size (the total number of genes associated with a given functional term) ≥ 2500 were considered as "general terms" and thereby removed from further analysis. Enriched GO-BP terms were then grouped based on their hierarchical relation in the GO ontology tree. Relations among terms (considered only "part of" and "is a") were retrieved via R package "GO.db". Text mining was performed on each GO-BP group to highlight the most re-occurring key words among the functional terms via the R package "wordcloud". Text cleansing was conducted manually to remove general words from the key word list ("regulation", "positive", "of", "activity", "pathway", "negative", "to", "process", "cell", "cellular", "in", "factor", "function", "involved", "biological", "molecular", "cascade", "from", "by", "or", "into", "via", "class", "compound", "changes", "mechanism", "component", "complex", "part"). Additionally, semantically similar key words were manually combined, e.g., "apoptosis", "apoptotic" and "programmed cell death" were considered as the same functional key word "cell death" and the frequencies were thereby combined. LRRK2 interactors contributing to the same GO-BP functional group were allocated into one functional unit (FU), and those contribute to all GO-BP groups were defined as the "functional core" of the $LRRK2_{int}$. Moreover, GO Cellular Component (GO-CC) enrichment analysis was performed on each FU to investigate the intracellular locations of included LRRK2 interactors.

Gene set enrichment analysis (GSEA)

GSEA was performed to examine the association between the functional core of the $LRRK2_{int}$ and the 2 most common neurodegenerative diseases, namely Alzheimer's disease (AD) and Parkinson's disease (PD). The pipeline of GSEA was designed as following:

1) AD and PD-related gene lists were downloaded from the Open Target Platform (Ochoa et al., 2023);
2) The number of overlapping genes between $LRRK2_{int}$'s functional core and the 2 disease-related gene lists was counted respectively (the "test_intersection");

3) 10000 randomly sampled gene lists at same size of the functional core of the $LRRK2_{int}$ was generated from the Ensembl gene annotation (N = 19831 coding genes). The overlap sizes between each random gene list and the 2 disease-related gene lists were counted (the "ref_intersection);
4) A significant association with AD or PD was defined as when the "test_intersection" > 95% of the "ref_intersection".

Results

- The $LRRK2_{int}$

A total of 1448, 1539 and 1850 human LRRK2 interactors were retrieved from PINOT, HIPPIE and MIST respectively. The 3 sets of proteins were then merged into a unique list of 1921 interactors using HGNC gene symbols (hereby referred as "the Merged List", **Table S1**). Among the 1921 interactors, 1414 (73.5%) were found in all the 3 databases, 86 (4.4%) were found in 2 of 3 databases while 421 (21.9%) were found in only 1 database (**Figure 3A**). The Merged List was then passed to the QC pipeline, 3 proteins (RPL17-C18orf32, TPTEP2-CSNK1E and BUB1B-PAK6) were removed due to their "unreviewed" profile in the UniProtKB, and 1500 (78.0%) proteins were removed due to their low reliability (FS < 2), hence generating the final list of LRRK2 interactors (N = 418, the $LRRK2_{int}$, **Figure 3C**). Among the 418 interactors, 375 (89.7%) were scored FS ≤ 5; 28 (6.7%) were scored between 6 and 8; 15 (3.6%) were scored FS ≥ 9. Of note, LRRK2 itself exhibited the highest FS = 52, suggesting that the self-interactions of LRRK2 have been well-studied. Other robust LRRK2 interactors were HSP90AA1 (FS = 18); MSN, YWHAQ/14-3-3T, YWHAZ/14-3-3Z (FS = 13); HSPA8 (FS = 12); CDC37, DNM1L, STUB1, TUBB (FS = 10); followed by GAK, MAP1B, RAB5B, YWHAG and TUBA1A (FS = 9) (**Figure 3B**).

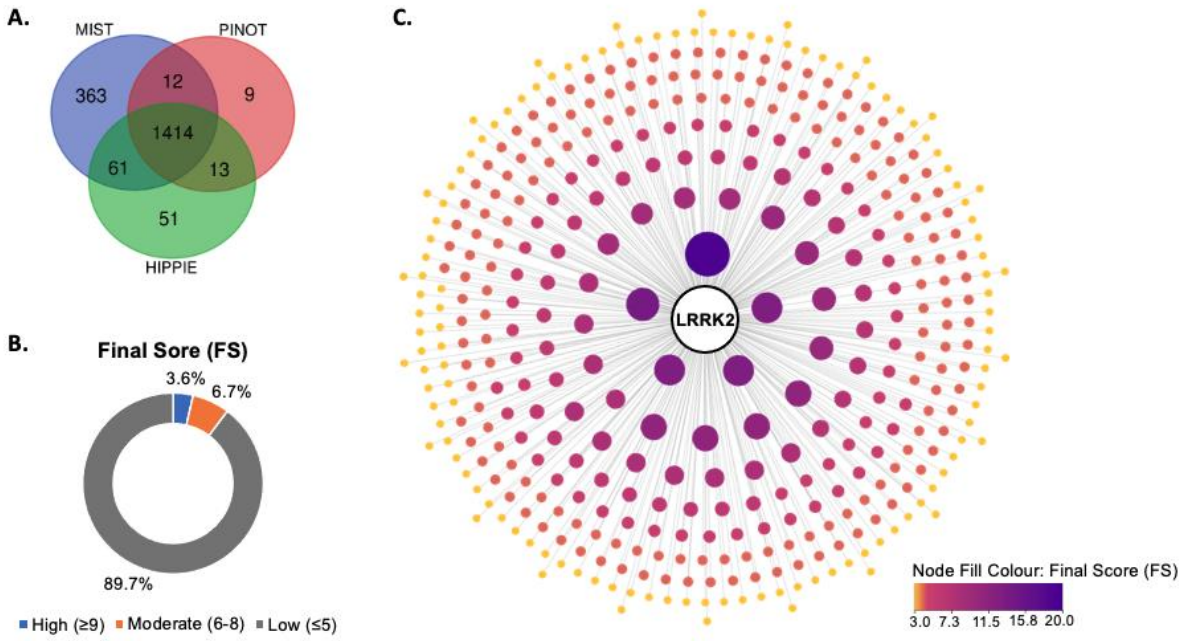


Figure 3. Construction of the LRRK2$_{int}$

A) *The Venn graph shows the overlap among the 3 lists of LRRK2 protein interactors downloaded from PINOT, MIST and HIPPIE;* ***B)*** *The pie graph shows the distribution of final score (FS) of QC-ed LRRK2 interactors (N = 418);* ***C)*** *The network graph shows the LRRK2$_{int}$. Both node fill colour and node size represent the FS: the darker the node colour, the larger the node size, the higher the FS. Of note, LRRK2 itself exhibited the highest FS = 52, which is beyond the colour code range.*

- Protein family classification for LRRK2 interactors

Protein family classification showed that a large proportion of LRRK2 interactors are protein kinases (N = 39, 9.3%), ribosomal proteins (N = 38, 9.0%) and cytoskeleton-related protein families (N = 36, 8.6%) such as proteins from the actin family (N = 5) and the tubulin family (N = 10). In addition, 12 (2.8%) LRRK2 interactors are from membrane trafficking related ATPase families. Furthermore, the LRRK2$_{int}$ contains seven 14-3-3 proteins (YWHAB/14-3-3β, YWHAG/14-3-3γ, YWHAZ/14-3-3ζ, YWHAE/14-3-3ε, YWHAH/14-3-3η, SFN/14-3-3б, YWHAQ/ɵ) and 13 Rab GTPases (RAB5B, RAB29, RAB32, RAB8A, RAB10, RAB1A, RAB1B, RAB7A, RAB11A, RAB11B, RAB11FIP2, RAB38, RAB5A).

- Biological processes enriched for LRRK2 interactors

G:GOst returned a total of 504 GO-BP terms that were significantly enriched for the LRRK2$_{int}$, of which 83 “general terms” with term size ≥ 2500 were excluded from further analysis.

Hierarchical relations ("part of" or "is a") were retrieved for the 415 remaining terms (98.6%), among which 7 major GO-BP groups (contained ≥ 10 terms) were identified, involving a total of 341 GO-BPs. Among the 7 GO-BP groups, the largest group contained 87 terms (GO-BP group 1, G1), involving 233/418 (55.7%) LRRK2 interactors (Functional Unit 1, FU1, **Table 1**). Text mining identified keywords of G1 as: "response", "apoptotic", "mitochondrial", "catabolic", "stimulus", "protein", "stress", "autophagy", "signalling", and "depolarisation", suggesting that FU1 is involved in response to stimulus and stress, apoptosis, protein catabolic, autophagy and mitochondrial functions (**Figure 4, Table B1**). GO-CC analysis on FU1 showed that these interactors were mainly enriched in intracellular and extracellular vesicle, synapse, microtubule cytoskeleton and mitochondrion (**Table B2**).

Table 1. Functional Unit 1 of the LRRK2$_{int}$

ACTB	PSMD11	AFG3L2	MBP	DVL1
ACTR2	PSMD2	AP3B1	SYNJ1	GNA12
AIFM1	PSMD6	CALM1	PPP2R2A	GSK3B
AKT1	RACK1	CNP	RGS2	DDX5
APEX2	RPL11	HSPA4	ARPC1B	DIS3
ATRX	RPL23	HSPA9	EMD	HDAC6
AURKB	RPS7	MFN2	IQGAP1	LATS1
BAX	SQSTM1	OPA1	SH3GL2	LRRK2
BRCA2	TRAF2	SLC25A4	RPS15	PSMC6
CCT3	UBXN10	SLC25A5	RPS20	STN1
CCT5	UFD1	SLC25A6	SCEL	STUB1
CDC42	VIM	CDC37	SCFD1	TAOK3
CDK2	VPS4A	DNM1L	ANKS4B	TK1
CHD1L	BAG3	VDAC1	HACD3	TP53
CYREN	CASP8	DDIT4	NFATC2	ADRM1
DBF4B	DNAJA1	DYRK2	SEC16A	AGO1
DDB1	DNAJB6	SFN	STK3	AGO2
DFFA	HSPB1	TXNDC12	CLTC	BAG5
FANCM	MAPT	PYGB	GOLGA2	C18ORF25
HELLS	PRKACA	ACTG1	LGALS8	DAPK1
HERC2	STAC	ACTR3	RAB1A	PRDX3
HSP90AA1	TPR	AHCYL1	RAB1B	STK24
HSP90AB1	YWHAE	AKAP8	RAB5A	STK25
HSPD1	EEF1D	ATP5F1A	SNAPIN	TPM1
KDM4D	FADD	CBX3	SPTLC1	TRAP1

MAP2K4	PIK3R1	CXCL11	AHNAK	ECHS1
MAP2K7	PLEC	CYFIP1	AP2M1	HMMR
MAPK3	RAB11B	DIAPH1	DNM1	CSNK1A1
MMS19	RHBDD1	EPRS1	FCHSD1	CSNK1D
MRGBP	APP	ESRRG	GAK	RPL3
MSH2	CDK5	GNAI2	LMNB1	RPS16
MTA1	HIF1A	LRP6	LMNB2	STIP1
NPM1	PPP1CA	MAP1B	RAC1	TRADD
PCNA	RO60	MAP2K3	TOR1AIP1	YWHAG
PHB	BAG2	MAP2K6	ACIN1	YWHAH
POLD1	CBLB	MPC2	ARHGEF7	HSPH1
POLE	EEF1A1	MYO1C	BAG1	HSPA1A
PPP2R1A	EEF1A2	NCL	CAMK1D	ITCH
PRKDC	HSPA8	PFKP	DIDO1	KHSRP
RAD51AP1	LAMP2	PKM	ITGB3BP	LDHA
RBBP7	MSN	PRKCZ	PAK6	TUFM
RIF1	RAB7A	RAB10	RAI14	RAB29
RPS3	RIPK1	RAB11FIP2	RPS3A	MFN1
RUVBL2	SNCA	RAB8A	STK40	NDUFAF7
SAMHD1	VGLL4	RHOA	YWHAZ	RAB32
SF3B3	ATP2A2	RIPK2	HK1	
PLK1	PPP1R8	PRKN	RAB38	

G2 contained 71 GO-BPs and was associated with 247/418 (59.1%) interactors (**Table B3**). Keywords identified in G2 were: “trafficking”, “localisation”, “vesicle”, “protein”, “microtubule”, “organisation”, “synaptic”, “establishment” and “organelle”, suggesting the corresponding functional unit of G2 (FU2, **Table 2**) was related to biological functions of membrane trafficking, protein localisation, microtubule and organelle organisation. The FU2 was found enriched in cytoskeleton, extracellular vesicle, and cell junction (**Table B4**).

Table 2. Functional Unit 2 of the LRRK2$_{int}$

AHCYL1	HERC2	SMTNL2	TUBB4B	ARPC5
AIFM1	HIF1A	TMOD3	TUBB6	ATP2A2
AKAP8	HK1	TPM1	MMS19	CAPZA1
AKT1	HSP90AA1	TPM2	SRPK1	CAPZA2
AP2M1	HSP90AB1	TPM3	LARP7	CAPZB
AP3B1	HSPA4	XIRP2	MSH2	CDC42EP3
ARFGAP1	HSPA8	AURKB	TOR1AIP1	CYFIP1

BAG3	HSPA9	CDK2	KLC2	TUBB2A
BAX	HSPB1	CSNK1D	MFN1	TUBB3
BRCA2	HSPD1	DYNC1H1	DBF4B	TUBB4A
CALM1	KPNB1	E2F4	DDB1	DBN1
CBLB	LAMP2	HSPA1A	EIF4EBP1	DIAPH1
CCT3	LRRK2	KIF2A	PCNA	DNAJB6
CCT5	MAPT	MDN1	POLE	FCHSD1
CDC37	MFN2	MSN	TAOK3	IQGAP1
CDC42	MPC2	PRKDC	RIPK1	LATS1
CDK5	MYO1C	RPL23A	VIM	RHOA
CETN3	MYO1D	RPL24	DIS3	YWHAG
CLTC	NPM1	RPLP0	LAS1L	RPL11
COPG2	NUP107	RPS14	RBIS	RPL23
CSE1L	NUP133	RPS15	RPL10A	RUVBL2
DNAJA1	NUP160	RPS19	RPL14	SCFD1
DNM1L	PIK3R1	RPS23	RPS16	SEC16A
EMD	PLK1	RPS27	RPS7	SFN
GOLGA2	PRKACA	RPS3	RPS8	SLC25A22
GSK3B	PRKCZ	RPS5	CFAP20	SNAPIN
HDAC6	PRKN	RPSA	ENKUR	SQSTM1
YWHAH	RAB10	TUBB	KHSRP	STK3
YWHAQ	RAB11A	DNM1	RGS2	TNPO1
YWHAZ	RAB11B	SNCA	SFXN1	TP53
ACTA2	RAB11FIP2	SYNJ1	CAMK1D	TPR
ACTB	RAB1A	ABCE1	GNAI2	UFD1
ACTBL2	RAB1B	APP	HMMR	VPS4A
LIMA1	RAB29	CNP	APEX2	YWHAB
MPRIP	RAB32	DVL1	CDKL3	YWHAE
MYL6	RAB38	GAK	CSNK1A1	ACTR3
MYL9	RAB5A	JHY	HELLS	POLDIP3
MYO1B	RAB5B	LMNB1	ITGB3BP	PPP1R12A
MYO1F	RAB7A	LMNB2	MAP2K6	RANBP1
PLEC	RAB8A	LRP6	MRGBP	SH3GL1
RAC1	RACK1	MAP1B	NEK1	SH3GL2
ARPC1B	RANBP2	MAPK3	PPP1CA	SH3GL3
ARPC2	RAPGEF4	NCBP3	RAD51AP1	SLC25A4
ARPC4	RHBDD1	OPA1	RIF1	SLC25A5
FANCM	ABLIM1	CHD1L	ARHGEF7	SLC25A6
ACTG1	LARP4	MAP2K7	MARK1	STK25

ACTR2	PAK6	STN1	SKA3	TUBA1A
TTLL1	RO60	NSL1	CKAP5	TUBA1C
RPS15A	UBXN10	PPP2R1A	ATRX	
USP39	ACIN1	TUBG1		

G3 contained 62 GO-BPs, associated with 179 LRRK2 interactors (FU3, **Table 3**). Text analysis showed the key words of G3 as: “protein”, “phosphorylation”, “metabolic”, “ubiquitination”, “catabolic”, “transferase”, “kinase”, “binding” and “modification”, suggesting that FU3 was related to protein metabolism (**Table B5**). GO-CC analysis showed that the interactors in FU3 were enriched in cytoskeleton and cell conjunction (**Table B6**).

Table 3. Functional Unit 3 of the LRRK2$_{int}$

AKT1	BAX	DNAJB6	GOLGA2	AGO1
BAG2	CALM1	DVL1	PPP1CA	AGO2
BAG5	CASP8	HDAC6	SNAPIN	AP3B1
CDK5	CDC42	HSPB1	ACIN1	DDX5
DNAJA1	DAPK1	LATS1	BAG3	DIS3
HSPA1A	DBF4B	LRP6	CBX3	E2F4
RPL11	DVL2	PLEC	EIF4EBP1	ESRRG
RPL23	DVL3	PPP1R8	EPRS1	GTF2I
RPS15	EEF1A2	PRDX3	MAGED2	HSPA8
RPS20	FADD	PRKN	RPS13	KHSRP
RPS3	GNA12	RGS2	RPS14	LARP7
RPS7	GNA13	SFN	SCFD1	MLLT3
SQSTM1	GNAI2	TP53	TCF25	MYO1C
ADRM1	GSK3B	YWHAG	TPR	NCL
AIFM1	HIF1A	ACTB	TRAP1	NFATC2
APP	HSP90AA1	CDC37	VDAC1	NKRF
ARHGEF7	HSP90AB1	DNM1L	VGLL4	POU5F1
AURKB	HSPD1	EEF1A1	YWHAE	PRPF6
IQGAP1	RACK1	PIK3R1	YWHAQ	RBBP7
LRRK2	RANBP2	CSNK1A1	YWHAZ	SF3B1
MAP2K3	RAPGEF4	CSNK1D	ZMYM5	SUPT4H1
MAP2K6	RHOA	PSMC6	RAB38	TCEA2
MAP2K7	RIPK2	CDK2	AKAP8	YWHAH
MAPK3	RPS2	PRKACA	ARFGAP1	ATP2A2
MAPT	SNCA	DDIT4	EEF1D	CYREN
MBP	STK3	LRRK1	HACD3	SF3B3
MSH2	STUB1	PPP2R1A	HERC2	ACTR3
NPM1	TAOK3	PRKDC	PPP2R2A	
PCNA	TBC1D22B	SH3GL2	PSMD2	

PHB	TRAF2	ACTR2	PSMD6	
PLK1	YWHAB	ATRX	RANBP1	
PPP1R12A	ABCE1	CCT3	CAMK1D	
PRKCZ	CBLB	CCT5	STAC	
RAB11FIP2	DFFA	KDM4D	TOR1AIP1	
MAP2K4	RUVBL2	RIF1	TPM1	
MMS19	STN1	FANCM	TRADD	
MRGBP	RAC1	MTA1	RHBDD1	
RAD51AP1	RIPK1	ITCH	BRCA2	

G4 contained 50 terms, involving 179 LRRK2 interactors (FU4, **Table 4**), with functional key words: “organisation”, “microtubule”, “protein”, “actin”, “catabolic” and “polymerisation”, “assembly”, “projection” and “filament”, suggesting that FU4 contributed to cytoskeleton organisation (**Table B7**). In addition, FU4 was enriched in cytoskeleton and synapse (**Table B8**).

Table 4. Functional Unit 4 of the LRRK2$_{int}$

ARHGEF7	TMOD3	CASP8	RACK1	FCHSD1
CKAP5	TPM1	CBLB	RHBDD1	IQGAP1
CLTC	DIAPH1	CSNK1A1	RIF1	LATS1
DYNC1H1	MAPK3	CSNK1D	RIPK1	LIMA1
GSK3B	NPM1	DAPK1	RIPK2	PIK3R1
HDAC6	ATRX	DBF4B	RPS2	PRKN
HSPA1A	AURKB	DDB1	RPS6KB2	CDKL3
MAP1B	BAX	DVL1	SF3B1	GAK
MAPT	CCT3	DVL2	STK3	MARK1
PLK1	CCT5	DVL3	STUB1	MLLT3
RAC1	CDK2	EEF2	TAOK3	NFATC2
RHOA	DNM1L	FADD	TRAF2	PPP2R1A
RPS3	HIF1A	FANCM	VGLL4	RAB29
SKA3	MAP2K7	GNA12	VIM	SFN
SNCA	MSN	GOLGA2	BAG5	RBBP7
TPR	RAD51AP1	HSP90AA1	EEF1A1	ADRM1
TUBB4A	RUVBL2	HSP90AB1	EEF1A2	AIFM1
ACTG1	SLC25A4	HSPD1	PRKACA	AKAP8
ARPC1B	SLC25A5	ITCH	PSMD2	APP
ARPC2	SYNJ1	LARP4	RPL11	BAG2
ARPC4	TP53	LRRK1	SF3B3	CALM1
ARPC5	VDAC1	MAP2K3	ACTR2	SH3GL3

CAPZA1	AKT1	MAP2K4	ACTR3	SNAPIN
CAPZA2	LRRK2	MAP2K6	DNAJB6	STK24
CAPZB	OPA1	MBP	DNAJB8	STK25
CDC42	RAB7A	MTA1	HSPA8	TUBA1A
CDC42EP3	SCFD1	MYO1C	NCLN	TUBB
CDK5	SLC25A6	PHB	RAB5A	YWHAH
CYFIP1	STN1	PKM	VPS4A	AGO2
DBN1	SQSTM1	POLDIP3	BAG3	HSPB1
RAB1A	RPL23	SMTNL2	CFAP20	PPP1R12A
RAB1B	RGS2	ACTA2	RAB5B	
DIS3	SPTLC1	ATP2A2	CAMK1D	
KHSRP	ACTB	DDIT4	RPS7	
E2F4	AHNAK	DFFA	PRKCZ	
PPP1CA	AP2M1	YWHAZ	PRKDC	
RAB11A	ARFGAP1	RAB8A	PSMC6	

G5 contained 35 terms and related to 147 LRRK2 interactors (FU5, **Table 5**), associated with key words: “apoptotic”, “signalling” “stress”, “response”, “transduction”, “oxidative”, “programmed”, “intracellular”, “communication”, “intrinsic”, suggesting FU5 was related to apoptosis and response to stress (**Table B9)**. GO-CC analysis showed that FU5 was enriched for cytoskeleton and mitochondria (**Table B10).**

Table 5. Functional Unit 5 of the LRRK2$_{int}$

SH3GL2	HSPA8	RPS14	AKT1	PHB
ACTR2	KHSRP	SCFD1	BAG2	PLK1
ATRX	LARP7	TCF25	BAG5	PPP1R12A
CCT3	MLLT3	TPR	CDK5	PRKCZ
CCT5	MYO1C	TRAP1	DNAJA1	RAB11FIP2
KDM4D	NCL	VDAC1	HSPA1A	RACK1
MAP2K4	NFATC2	VGLL4	RPL11	RANBP2
MMS19	NKRF	YWHAE	RPL23	RAPGEF4
MRGBP	POU5F1	YWHAQ	RPS15	RHOA
RAD51AP1	PRPF6	YWHAZ	RPS20	RIPK2
RIF1	RBBP7	ZMYM5	RPS3	RPS2
RUVBL2	SF3B1	RAB38	RPS7	SNCA
STN1	SUPT4H1	AKAP8	SQSTM1	STK3

RAC1	TCEA2	ARFGAP1	ADRM1	STUB1
RIPK1	YWHAH	EEF1D	AIFM1	TAOK3
FANCM	ATP2A2	HACD3	APP	TBC1D22B
MTA1	GOLGA2	HERC2	ARHGEF7	TRAF2
ITCH	PPP1CA	PPP2R2A	AURKB	YWHAB
RHBDD1	SNAPIN	PSMD2	BAX	ABCE1
BRCA2	ACIN1	PSMD6	CALM1	CBLB
CYREN	BAG3	RANBP1	CASP8	DFFA
SF3B3	CBX3	CAMK1D	CDC42	DNAJB6
ACTR3	EIF4EBP1	STAC	DAPK1	DVL1
AGO1	EPRS1	TOR1AIP1	DBF4B	HDAC6
AGO2	MAGED2	TPM1	DVL2	HSPB1
AP3B1	RPS13	TRADD	DVL3	LATS1
DDX5	MAP2K3	MSH2	EEF1A2	LRP6
DIS3	MAP2K6	NPM1	FADD	PLEC
E2F4	MAP2K7	PCNA	GNA12	PPP1R8
ESRRG	MAPK3	CSNK1A1	GNA13	PRDX3
GTF2I	MAPT	CSNK1D	GNAI2	PRKN
LRRK1	MBP	PSMC6	GSK3B	RGS2
PPP2R1A	HSPD1	CDK2	HIF1A	SFN
PRKDC	IQGAP1	PRKACA	HSP90AA1	TP53
DNM1L	LRRK2	DDIT4	HSP90AB1	YWHAG
EEF1A1	PIK3R1	CDC37	ACTB	

G6 consisted of 20 GO-BPs, involving 110 interactors (FU6, **Table 6**) which linked to functional key words of “morphogenesis”, “projection”, “neuron”, “development”, “growth”, “bounded”, “membrane”, “lamellipodium”, “plasma” and “organisation”, suggesting FU6 contributed to cell development and membrane organisation (**Table B11**). In addition, GO-CC analysis showed that FU6 was enriched for axon, synapse, vesicle and microtubule cytoskeleton (**Table B12**).

Table 6. Functional Unit 6 of the LRRK2$_{int}$

YWHAG	ACTR2	RAB29	CDK5	MAPK3
YWHAH	AFG3L2	RAB8A	CDKL3	NUP133
YWHAZ	AIFM1	RAC1	CYFIP1	SYNJ1
AP3B1	AKT1	RGS2	DBN1	DNM1L

ARHGEF7	APP	RHOA	DVL1	HSPA1A
CAPZB	BAG5	RPL24	GSK3B	ITCH
CDC42EP3	BAX	SH3GL3	HDAC6	MAP2K4
DIAPH1	CAMK1D	SNAPIN	HSP90AA1	NPM1
GNA12	CDC42	STK24	HSP90AB1	PHB
GNA13	CNP	STK25	IQGAP1	PPP2R1A
LARP4	CSNK1D	TAOK3	MAP1B	RACK1
LATS1	DDIT4	TP53	MAPT	RBBP7
MSN	DVL2	TTLL1	PAK6	SFN
MYL12B	DVL3	TUBB2A	PRKCZ	VGLL4
PLEC	GAK	TUBB3	PRKN	CFAP20
PRKDC	HIF1A	VASH2	SH3GL2	E2F4
SCFD1	LRP6	VIM	ACTB	JHY
TPM1	LRRK2	YWHAE	ACTBL2	LIMA1
ABLIM1	MARK1	BRCA2	MYL9	NEK1
ACTR3	MBP	STK3	TMOD3	RAB5A
ARPC2	RAB10	STK40	PDCL	RO60
PIK3R1	RAB11A	ACTG1	ATRX	UBXN10

G7 contained 16 terms and in volved 92 LRRK2 interactors (FU7, **Table 7**), related to key words: “transport”, “protein”, “localisation”, “vesicle”, “intracellular”, “establishment”, “nucleus”, “export” and “nucleocytoplasmic”, suggesting that FU7 contributed to intracellular protein transport and localisation (**Table B13**). In addition, FU7 was enriched for intracellular vesicle, synapse and cell projection (**Table B14**).

Table 7. Functional Unit 7 of the LRRK2$_{int}$

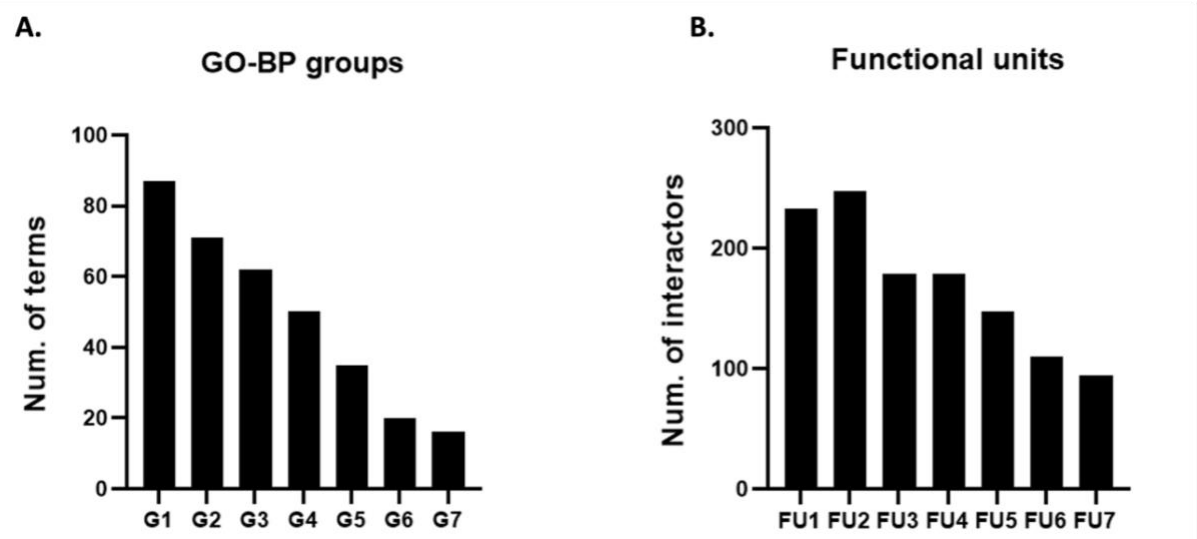
ACTB	CLTC	HSP90AB1	PPP1R12A	RHOA
ACTG1	CXCL11	HSPA1A	PRKACA	RIPK1
AHCYL1	CYFIP1	HSPA8	PRKCZ	RUVBL2
AHNAK	DAPK1	LARP7	PRKN	SCFD1
AKT1	DBN1	LATS1	RAB10	SFN
AP2M1	DIAPH1	LRRK2	RAB11A	SH3GL3
APP	DNAJA1	MAP1B	RAB11B	SLC25A22
ARFGAP1	DNAJB6	MAP2K6	RAB29	SLC25A4
ARHGEF7	DNM1L	MAPK3	RAB38	SLC25A5
ATP2A2	DVL1	MAPT	RAB5A	SLC25A6

AURKB	DVL3	MDN1	RAB5B	SNAPIN
BAG3	DYNC1H1	MPC2	RAB7A	SNCA
BAX	EMD	MSN	RAB8A	SQSTM1
CALM1	EPRS1	MYO1C	RAC1	STAC
CAMK1D	GNAI2	NPM1	RACK1	SYNJ1
CCT3	GSK3B	PIK3R1	RAPGEF4	TPR
CCT5	HDAC6	PLK1	RGS2	VDAC1
CDC42	HIF1A	YWHAE	YWHAH	VPS4A
CDK5	HSP90AA1			

Of note, a total of 29 LRRK2 interactors exhibited in all the 7 functional units, thereby forming the functional core of the $LRRK2_{int}$. These interactors were multifunctional and cooperated with LRRK2 in a range of biological processes (**Figure 4C**). Of note, the 29 interactors presented a significantly higher mean FS as compared to the overall average value (6.62 vs. 4.24, t-test $p < 0.001$), suggesting that the functional core is composed of robust LRRK2 interactors with high reproducibility.

- GSEA

In addition, Gene Set Enrichment Analysis (GSEA) showed that the 29 LRRK2 interactors were represented in all the 7 functional units significantly associated with PD and Alzheimer's disease AD ($p < 0.001$), in which LRRK2, PRKN, MAPT and GSK3B were related to PD while MAPK3, CDK5, MAPT, APP, HSPA1A, HDAC6 and GSK3B were related to AD, suggesting that the functional core of $LRRK2_{int}$ is potentially closely related to neurodegeneration (**Figure 4C**).



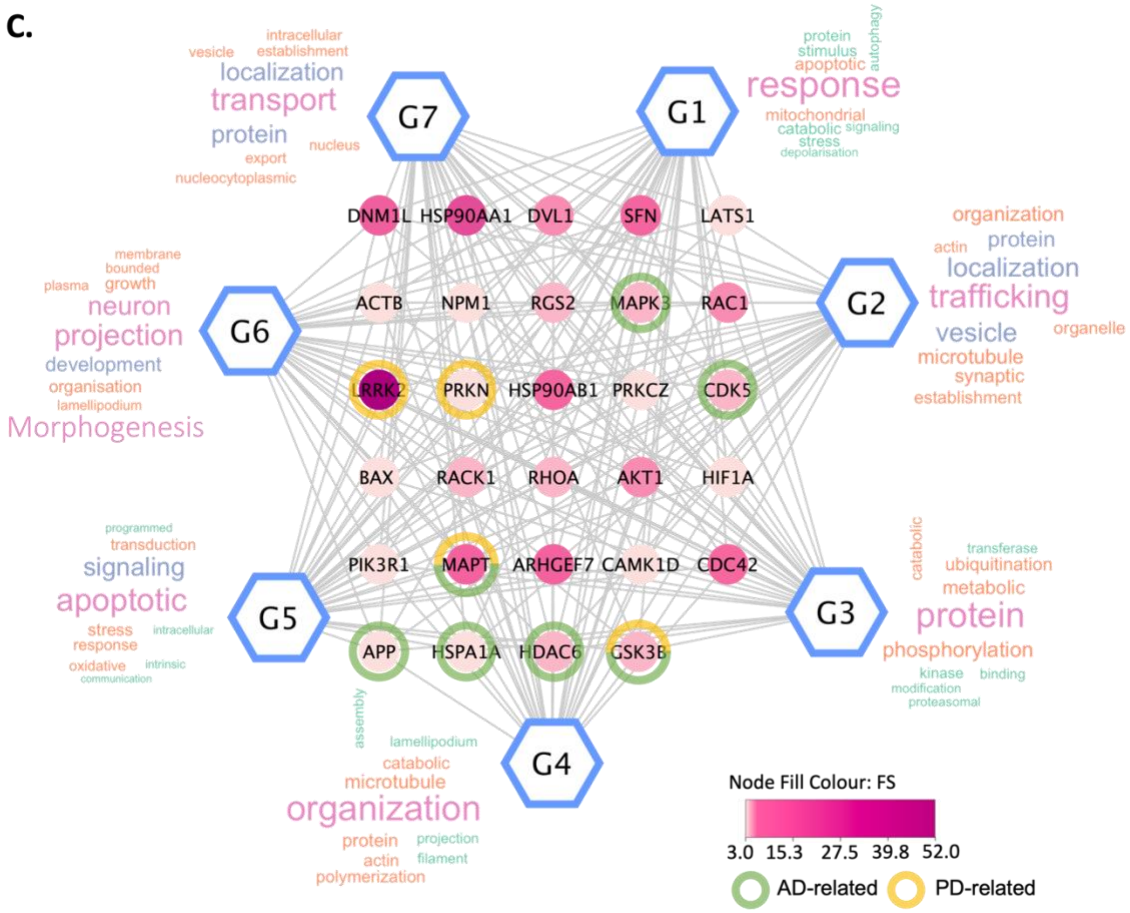


Figure 4. Functional Enrichment Analysis on the LRRK2$_{int}$

A) *The bar graph shows the number of functional terms in the 7 major GO-BP groups (G1-7, containing ≥ 10 terms) identified based on hierarchy relations;* ***B)*** *The bar graph shows the number of LRRK2 interactors in the 7 functional unit (FU1-7), i.e., interactors related to the functional terms of each GO-BP group; C) The network graph shows the functional core of the LRRK2int, i.e., the 29 interactors that appeared in all the 7 functional units. . Hexagonal nodes represent GO-BP groups G1-7. Word clouds show the key words identified via text mining for each GO-BP group. The larger the word, the higher the frequency it shows in term names. Round nodes represent LRRK2 interactors. Round node fill colour represents the FS of interactors. The darker the round node, the higher the FS. AD and PD-related interactors identified in GSEA are highlighted with green circles and yellow circles, respectively.*

Main findings

1. The LRRK2$_{int}$ consists of 418 interactors, including protein kinases, ribosomal proteins, cytoskeletal proteins, RAB GTPases, 14-3-3 proteins and others.

2. Self-interaction is the most robust/validated PPI of LRRK2.
3. The $LRRK2_{int}$ can be divided into 7 functional units each of them related to different biological functions (GO:BPs) such as: response to stress, vesicular trafficking, protein metabolism, cytoskeleton organisation, apoptotic signalling, cell development and intracellular protein localisation.
4. Functional units of LRRK2 interactors were mainly enriched in cellular components (GO:CCs) such as: microtubule and actin cytoskeleton, vesicle, synapse, mitochondria and cell junction.
5. The $LRRK2_{int}$ has a functional core composed of 29 interactors that participate in multiple biological processes and is significantly enriched with AD and PD proteins/genes.

Discussion

This section collected a total of 1921 proteins as potential interactors of LRRK2 from peer-reviewed literatures via 3 online PPI querying tools, PINOT, MIST and HIPPIE, without adding any type of filtering. These tools extract PPIs from multiple manually curated primary PPI databases using different pipelines. Out of the 1921 interactors, 73.5% were returned by all the 3 searching tools, 4.4% were returned by 2 tools, while 21.9% were found in only 1 tool. This suggested that although the 3 tools presented a relatively high overlap (thus were similar in performance), a combined use of these tools is essential to obtain a comprehensive LRRK2 protein interactome with the maximised literature coverage.

Considering the large dependence of *in vitro* high-throughput PPI detection techniques (such as yeast-2-hybrid screening) as well as ex-vivo experiments (such as proteo-arrays) which may produce false positives, PPIs derived from primary databases often require further QC. In this study, a simple but effective QC pipeline was established (as suggested by the PINOT tool), involving calculating the “final score” from the “method score” and “publication score”, which refers to the number of detection methods and independent publications where a PPI was reported. The obtained final score is therefore a direct reflection of how many times a certain PPI has been replicated in literature and as such it gives an indication of the reproducibility/reliability of the information. Compared to the QC systems utilised by other PPI tools, such as MIST and HIPPIE, which involves ranking different detection methods based

on “detection strength”, this pipeline avoids potential bias induce by subjective curation of PPI detection techniques and increases the replicability of the construction of LRRK2 interactome. Of note, regarding the method score, a method reassignment and grouping approach was adapted from PINOT, which ensures a more accurate calculation system for the method score by counting only the technically different methods. By applying the QC pipeline, only the most reliable interactors that have been reported in multiple studies and/or with technically different detection methods were kept for further analysis, which reduced the size of LRRK2 interactome from 1921 to 418, with 1503 (78.2%) proteins removed from the list. This shows that one of the problems faced by LRRK2 PPI research is that most of the PPI data are not reproduced in literature and thereby not directly reliable. This may result from delays in literature curation in the primary databases, lack of interest in wet-lab research in reproducing PPIs that have already been published, or false interactions published in literature (Type I errors) produced for example by certain PPI detection methods such as tandem affinity purification (TAP) (Edwards et al., 2002).

Among the 418 proteins, LRRK2 itself exhibited the highest score for interaction (final score = 52), confirming LRRK2 self-interaction as the best known and reproducible PPI of LRRK2. This replicates previous 2 previous LRRK2 interactome studies (Manzoni et al., 2015; Porras et al., 2015b). In the cell, LRRK2’s self-interactions occur at a range of sites within different domains of the protein structure and form dimer-sized structures with high molecular weight, which has been highly associated with its autophosphorylation activity (Greggio et al., 2008). Moreover, LRRK2 self-interaction is essential for autophosphorylation and multiple pathogenic mutations in *LRRK2* have been reported to increase autophosphorylation both *in vivo* and *in vitro* and may contribute to the alterations in PPIs between LRRK2 and other proteins such as the 14-3-3 family and with downstream pathways such as endocytosis (Manschwetus et al., 2020b; Stevers et al., 2017). In addition, similar alterations were observed in sporadic PD patients too (Sheng et al., 2012). Together with these experimental data, this study confirms the importance of LRRK2 self-interactions in its protein interactome.

Apart from LRRK2 itself, interactors of LRRK2 belong to a variety of families, among which there are 7 14-3-3 proteins and 13 Rab GTPases in the interactome. These 2 protein families are the most widely recognised LRRK2 interactors (Jeong et al., 2018b; Stevers et al., 2017).

Among the 14-3-3 proteins, YWHAQ, YWHAZ and YWHAG are the mostly reported LRRK2 interactors based on their final score (FS = 13, 13 and 9, respectively), while for the Rab GTPases, RAB5B (FS = 9) is the protein most robustly associated with LRRK2. It was reported in 4 publications via 5 types of detection methods (D. H. Ho et al., 2016; Imai et al., 2015; Shin et al., 2008; Yun et al., 2015). In addition, 36 cytoskeleton proteins were included in the LRRK2 interactome, among which 5 are from the actin family while 10 from the tubulin family. Previous studies have found that LRRK2 binds directly with actin proteins and affect its polymerisation and depolymerisation (Meixner et al., 2011; Parisiadou & Cai, 2010; Tombesi et al., 2022). Moreover, LRRK2 has been shown to phosphorylate tubulin and thereby enhance microtubule stability (Bonet-Ponce et al., 2020; Gillardon, 2009; Law et al., 2014). PPIs with cytoskeletal proteins form the foundation of LRRK2's role in cytoskeleton organisation and microtubule dynamics, which is crucial for neuronal morphology, axonal transport, synaptic formation and maintenance. Impairment or dysregulation of these PPIs may lead to neuronal dysfunction and neuron loss. Additionally, there is a large proportion of ribosomal proteins in the $LRRK2_{int}$ (N = 38). Multiple lines of evidence have shown that LRRK2 regulates ribosomal function especially in the neurons (Juli et al., 2016; Martin et al., 2014). Additionally, ribosomal function impairment and reduced protein synthesis have been found in striatal and substantia nigral neurons in mouse models of LRRK2-G2019S PD and other non-genetic PD, as well as in fibroblast cells isolated from sporadic and LRRK2-G2019S PD patients (Deshpande et al., 2020; J. W. Kim et al., 2020; R. Wallings et al., 2015). Therefore, understanding LRRK2:Ribosome interaction will bring further insights for PD pathology and provide novel drug targets for PD therapy. Moreover, the $LRRK2_{int}$ includes a large number of protein kinases of various types. These proteins might either phosphorylate LRRK2 or function as LRRK2's substrate, forming a complex kinase network that maintains the LRRK2-centred signalling cascade.

A total of 578 GO-BP terms were returned from functional enrichment analysis on the entire $LRRK2_{int}$. The large number of enriched terms made the interpretation process complicated, which is, in fact, a very common problem faced by any large query in functional/pathway enrichment analysis. In this study, a special dimension reduction pipeline was set-up and applied to gain a better understanding of the functional enrichment results. GO terms were firstly filtered by the "term size". Terms size refers to the number of genes (in the entire

human genome) that have been annotated with a given GO term. A larger term size normally indicates a more general term. For example, the GO-BP term “autophagy (GO:0006914)” has a term size of 571, while another more specific term “chaperone-mediated autophagy (GO:0061684)” has a term size of 15. In this first part of the study, a hard threshold of 2500 was set to define “general terms” (term size >2500) and “specific terms” (term size <2500), which was adapted from a previous study. This approach is easy to perform and assists in downsizing the enrichment results, though with a possibility of information loss, especially for the terms referring to complex cellular pathways that involve a wide range of proteins. However, the impact of losing several terms as such is expected to be small in this study since only 83/504 (16.5%) terms were removed.

The remaining GO-BP terms were grouped based on their hierarchical relation in the GO ontology tree. Compared to semantic clustering, which is more commonly used, and simpler to perform, grouping by GO relations keeps the connection between terms, avoid variation induced by different (and subjective) semantic mappings and thereby increasing the replicability of the dimension reduction approach. The grouping process resulted in a total of 7 GO-BP groups, among which 3 were related to microtubule dynamics and actin cytoskeleton organisation (G2, G4, G6), emphasizing a pivotal role of the $LRRK2_{int}$ in modulating the structural dynamics of the cell. The intricate orchestration of these cytoskeletal elements is crucial for maintaining cellular shape and facilitating intracellular transport. In addition, 2 GO-BP groups (G1 and G5) were related to response to stress, autophagy, mitochondrial organisation and apoptotic signalling. These processes highlighted the potential role of LRRK2 signalling in regulating cellular response to stimuli. In addition, there are 2 GO-BP groups related to protein synthesis, metabolism and modification. These functions play critical roles in regulating protein function, stability, and interactions and could have downstream effects on the above-mentioned cellular functions. Functional units contributing to each GO-BP groups were extracted, in which a total of 29 interactors were found in all of the 7 units. These proteins include LRRK2 (FS = 52), HSP90AA1 (FS = 18), MAPT (FS = 8), ARHGEF7 (FS = 8), SFN (FS = 7), HSP90AB1 (FS = 7), CDC42 (FS = 6), DNM1L (FS = 5), RAC1 (FS = 5), AKT1 (FS = 5), DVL1 (FS = 5), RGS2 (FS = 4), MAPK3 (FS = 4), RACK1 (FS = 4), RHOA (FS = 4), CDK5 (FS = 4), HDAC6 (FS = 4), GSK3B (FS = 4), LATS1 (FS = 3), ACTB (FS = 3), NPM1 (FS = 3), PRKN (FS = 3), PRKCZ (FS = 3), BAX (FS = 3), HIF1A (FS = 3), PIK3R1 (FS = 3), CAMK1D (FS = 3), APP (FS = 3), HSPA1A (FS

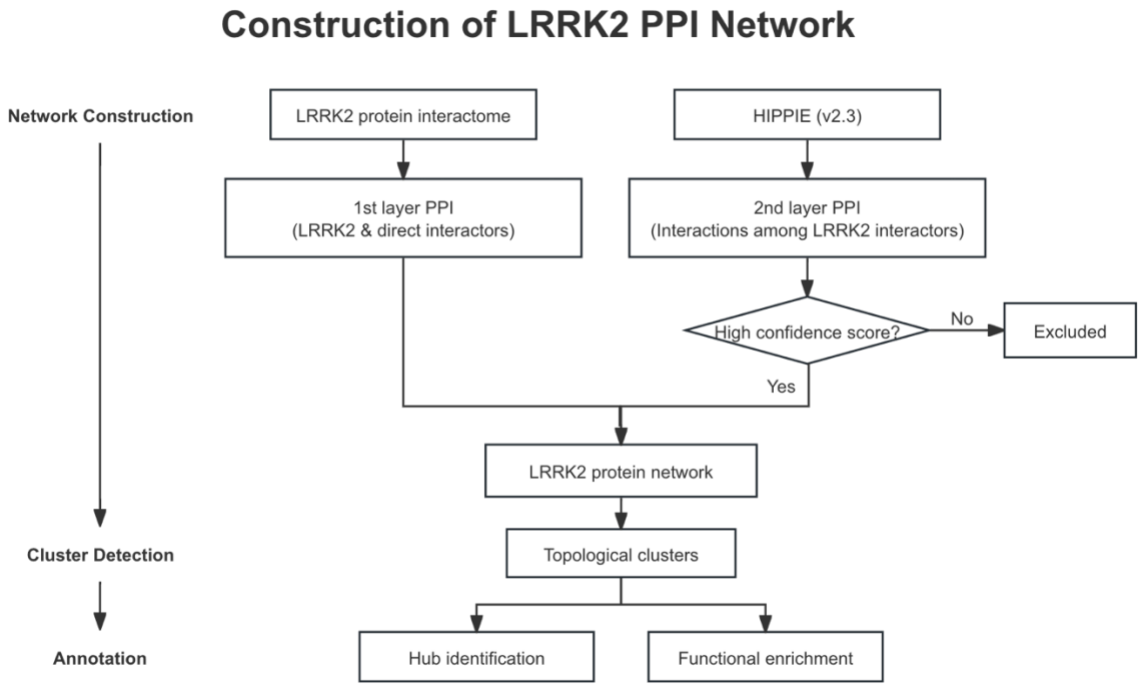
= 3), forming the functional core of LRRK2 signalling network. In addition, the functional core was significantly related to PD and AD, thereby linking the $LRRK2_{int}$ with neurodegenerative diseases.

Chapter 2. Construction and analysis of the LRRK2 PPI network ($LRRK2_{net}$)

Objectives

- Add the "2nd layer" PPIs among LRRK2 interactors to construct a PPI network ($LRRK2_{net}$)
- Investigate the topological features of the $LRRK2_{net}$
- Detect potential topological clusters in the $LRRK2_{net}$ and perform functional annotation on these clusters

Analysis pipeline



Note: Some of the figures and text in Chapter 2 are adapted from the following manuscript deposited in bioRxiv: Zhao et al.; "Transcriptomics analyses of the LRRK2 protein interactome reveal distinct molecular signatures for sporadic and LRRK2 Parkinson's Disease" https://doi.org/10.1101/2023.09.12.557373. Some of the results were further elaborated in a collaborative project in bioRxiv: Tombeisi et al.; "LRRK2 regulates synaptic function through BDNF signalling and actin cytoskeleton." https://doi.org/10.1101/2022.10.31.514622.

Methods

- Construction of the $LRRK2_{net}$

PPIs between LRRK2 interactors were defined as the "1st-layer" interactions (downloaded and QC-ed in Chapter 1), while the "2nd-layer" interactions were defined as the PPIs among LRRK2

interactors (apart from LRRK2 itself). The "2nd-layer" PPIs were downloaded from HIPPIE (v2.3) on 5th May 2022. In order to keep the most robust PPIs and avoid overloading the network, only non-self-interaction PPIs with high confidence score (≥ 0.72) in HIPPIE were kept for further analyses. Of note, the confidence score in HIPPIE was calculated directly by the tool via a semi-computational manner, examining the number and the quality of the methods by which a PPI was detected, as well as the number of publications where a PPI was reported, as well as the number of non-human organisms in which a PPI was reproduced, aiming to reduce the number of false positives. The PPI network of LRRK2 ($LRRK2_{net}$) was constructed by combining the 1st-layer and 2nd-layer PPIs (edges). Degree and betweenness centrality (BC) were calculated for each node (interactor) to identify "semi-seed" interactors in the $LRRK2_{net}$, i.e., the interactors with high degree and/or high BC in the network (i.e., interactors with the most interactions and/or located on the "bridges" connecting different parts of network).

- Topological cluster detection

The Fast Greedy algorithm was used to identify topological clusters of LRRK2 interactors based on Edge Betweenness via Cytoscape (v3.9.1). Of note, LRRK2 itself was excluded from topological clustering to avoid its disturbance on the clustering algorithm due to its high degree residual of the "1st-layer" construction. Within each topological cluster, nodes with highest degree and/or highest BC were defined as the "hub proteins" within the cluster. These proteins may play a central role in mediating and maintaining the connection of the subnetwork within each cluster. Additionally, the variance of node degree and BC was compared across the clusters via F-test. In fact, clusters with lower variance on node betweenness are less centralised and are thereby more stable from a topological perspective (Valente, 2010).

- Functional annotation of clusters

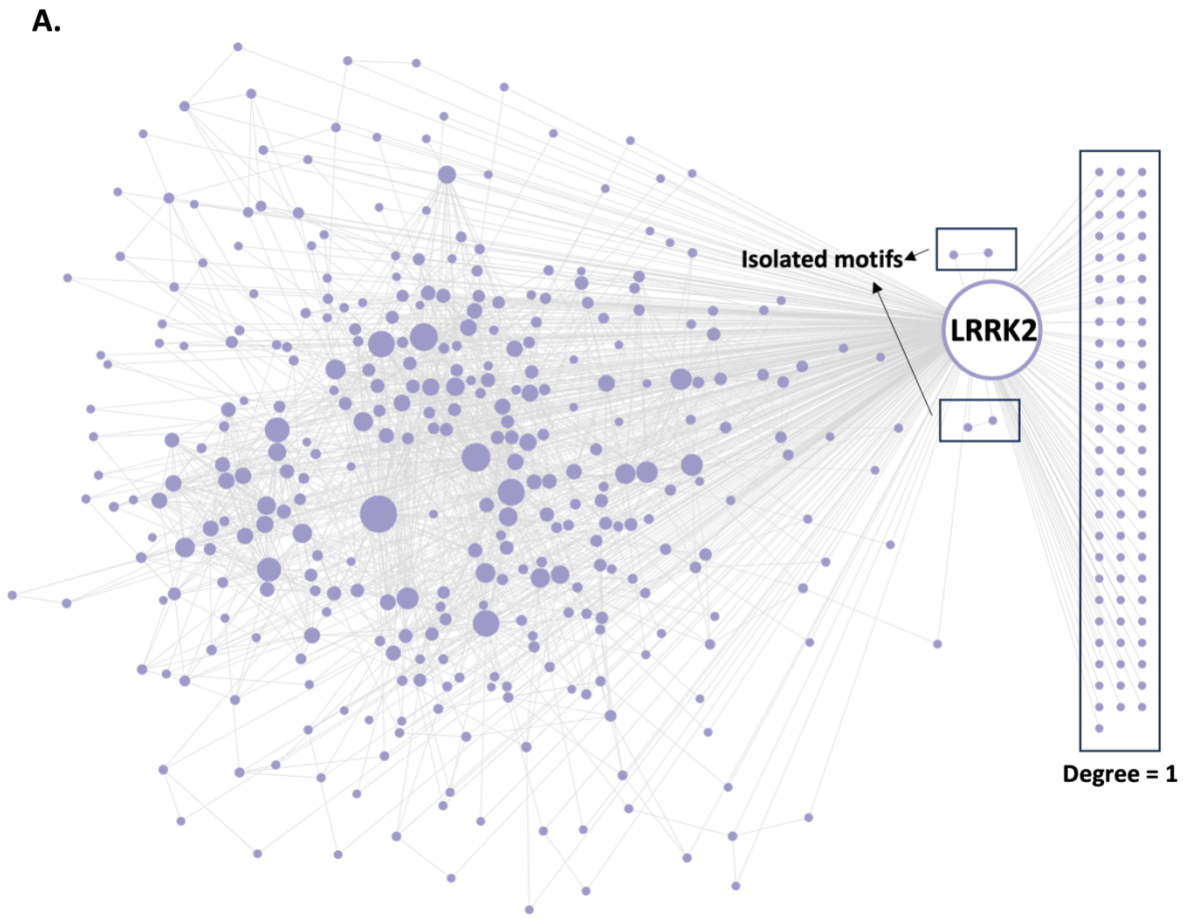
GO Biological Processes (GO-BP) enrichment analysis were performed via g:GOSt (https://biit.cs.ut.ee/gprofiler/gost) on the topological clusters For each cluster, the Enrichment Score (ES) was defined as the total number of GO-BP terms returned from the enrichment analysis, and the ES was used as proxy to evaluate the biological significance of each topological cluster. Only clusters with high ES were kept for further analysis.

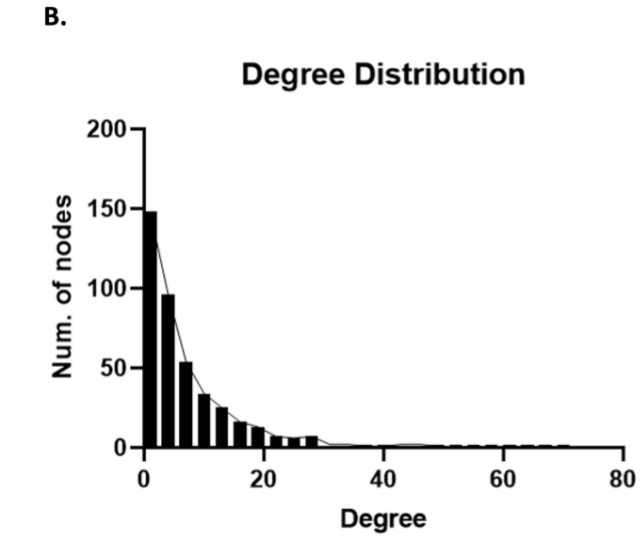
Results

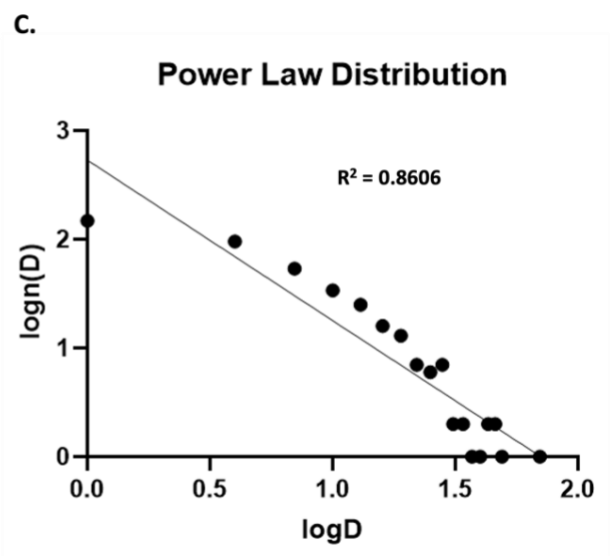
- $LRRK2_{net}$

A total of 4860 "2^{nd}-layer" PPIs were extracted from the HIPPIE database (v2.3), among which 1466 (30.2%) were scored as "high confidence" (HIPPIE confidence score ≥ 0.72), out of which 121 self-interactions were removed from the list, thereby leaving 1345 "2^{nd}-layer" PPIs for 338 LRRK2 interactors. Combined with the "1^{st}-layer" PPIs (N = 417), a final network of 418 nodes and 1762 edges was constructed around LRRK2 and its interactors ($LRRK2_{net}$, **Figure 5A**).

Degree distribution analysis showed that N = 216 interactors (51.7%) had degrees ≤ 4; N = 141 interactors (33.7%) presented degrees between 5 and 14; N = 41 interactors (9.8%) presented degrees between 15 and 24; N = 24 interactors with degree ≥ 24, suggesting that the $LRRK2_{net}$ follows the Power Law distribution (log-log plot R-square = 0.8606) (**Figure 5B-C, Table S2**). Interactors with degree ≥ 24 (the top 5% of all) were defined as "sub seed" proteins in the $LRRK2_{net}$, with TP53 (degree = 68), CDK2 (degree = 48), HSPA8 (degree = 46), HSP90AB1 (degree = 44), HSP90AA1 (degree = 43), YWHAZ (degree = 43), LAPR7 (degree = 39), NPM1 (degree = 37), TRAF2 (degree = 32), IQGAP1 (degree = 32), LIMA1 (degree =31), CAPZA2 (degree = 31), PRKN (degree = 28), DBN1 (degree = 28), YWHAQ (degree = 27), RPS8 (degree = 27), YWHAG (degree = 26), TRADD (degree = 26), RPS3 (degree = 26), AKT1 (degree = 25), YWHAB (degree = 24), HSPA1A (degree = 24), RPS3A (degree = 24) presenting the highest degree, suggesting that these proteins may play an essential role in maintaining the local connectivity of $LRRK2_{net}$. Among the 25 "sub seed" proteins, 2 were AD-related (DBN1 and HSPA1A) while 1 was PD-related (PRKN). In terms of BC analysis, a total of 17 interactors presented high BC (top 5% of all), including TP53 (BC = 0.23), CDK2 (BC = 0.10), YWHAZ (BC = 0.09), HSPA8 (BC = 0.07), HSP90AB1 (BC = 0.06), HSP90AA1 (BC = 0.05), LRRK1 (BC = 0.05), PRKN (BC = 0.05), TRAF2 (BC = 0.04), NPM1 (BC = 0.04), TUBA1C (BC = 0.04), LIMA1 (BC = 0.03), IQGAP1 (BC = 0.03), VIM (BC = 0.03), YWHAG (BC = 0.03), PPP1CA (BC = 0.03), AKT1 (BC = 0.03) (**Figure 5D**). These interactors are potentially mediators responsible for maintaining the overall connectivity of $LRRK2_{net}$. These interactors formed the topological backbone of the LRRK2 (N = 27) (**Figure 5E**)


A.
Isolated motifs
LRRK2
Degree = 1


B.
Degree Distribution
Num. of nodes
200
150
100
50
0
0
20
40
60
80
Degree


C.
Power Law Distribution
R² = 0.8606
logn(D)
3
2
1
0
0.0
0.5
1.0
1.5
2.0
logD

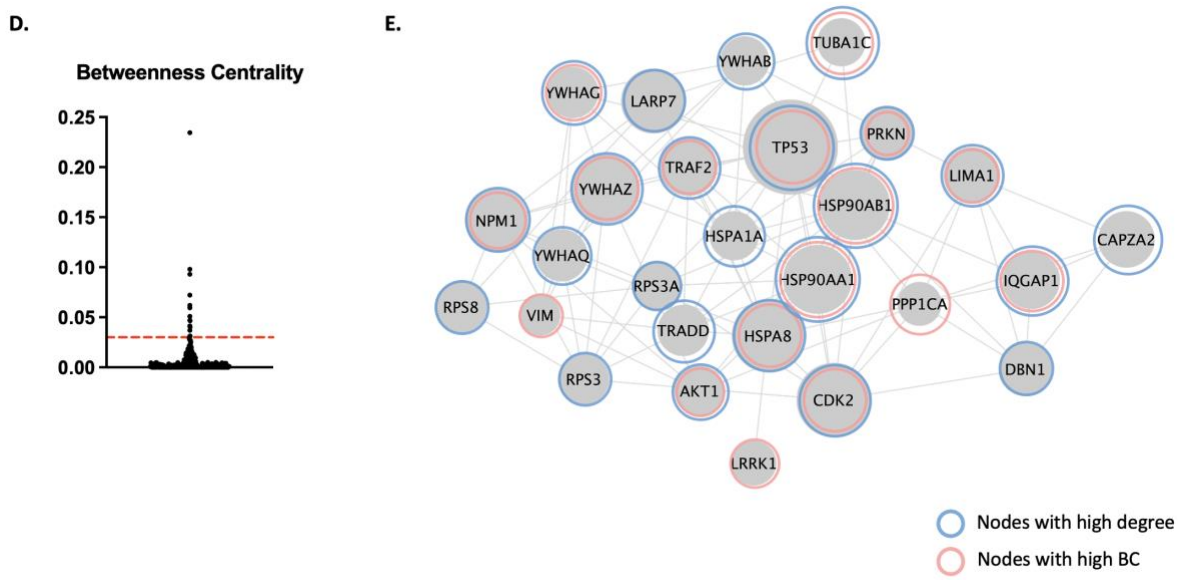


Figure 5. LRRK2 PPI network

A) *The network graph shows the PPI network of LRRK2. Nodes represent interactors (N = 418) while edges represent PPIs (N = 1762). Node size represents the degree centrality. The larger the node, the higher the degree. A total of 79 interactors exhibited a degree of 1 (highlighted with rectangular). Two isolated motifs containing 4 LRRK2 interactors were highlighted as well. These interactors have no connection with the rest of the network;* ***B)*** *The bar graph shows the distribution of degree of LRRK2 interactors in the network. Interactors with degree ≥ 24 (top 5%) were defined as "sub seed" proteins;* ***C)*** *The log-log plot shows that LRRK2$_{net}$ follows power law, in which X-axis represents log-transformed degree (logD), while Y-axis represents log-transformed frequency of LRRK2 interactor with a certain degree level (log(n(D))). Regression analysis shows that the plot fits a linear regression with R-square = 0.8606;* ***D)*** *The scatter plot shows the distribution of betweenness centrality (BC) of LRRK2 interactors. LRRK2 was excluded from this graph. The rest of LRRK2 interactors with top 5% BC (above the red line) was selected as "sub seed" protein;* ***E)*** *The network graph shows LRRK2 interactors with highest (top 5% of all interactors) degrees (marked with pink circles) and/or betweenness centrality (marked with blue circles). These interactors formed the topological core of the LRRK2$_{net}$.*

Cluster detection within the LRRK2$_{net}$

In order to perform topological clustering on the LRRK2$_{net}$, interactors with degree = 1 (i.e. those that only possessed connection with LRRK2 but not with the other members of the LRRK2 interactome) were discarded (N = 79). In addition, 2 isolated triangle motifs were identified in the LRRK2$_{net}$. Each of them comprised LRRK2 and other 2 interactors with degree of 2, which only interacted with each other but not with any other interactors in the network

and were thereby excluded from the clustering analysis as well (**Figure 6A**). Finally, LRRK2 itself was removed from the network, thereby generating the "trimmed-LRRK2$_{net}$", containing 338 nodes and 1345 edges. The trimmed-LRRK2$_{net}$ was then subset into 14 topological clusters via the Fast Greedy algorithm based on Edge Betweenness (Cluster A-N), among which Cluster L, M, N contained ≤ 5 interactors and were thereby excluded from further analysis (**Figure 6A**, **Table S3**). Node centrality (degree and betweenness) analysis was performed for each remaining cluster (N = 7) to identify "hub" proteins of top 5% degree or betweenness (**Figure 6C-D**). In addition, GO-BP enrichment analyses excluded 4 more clusters (Cluster H, I, J, K) due to their low ES < 10 (i.e., less than 10 terms were returned for the 3 clusters) (**Figure 6B**).

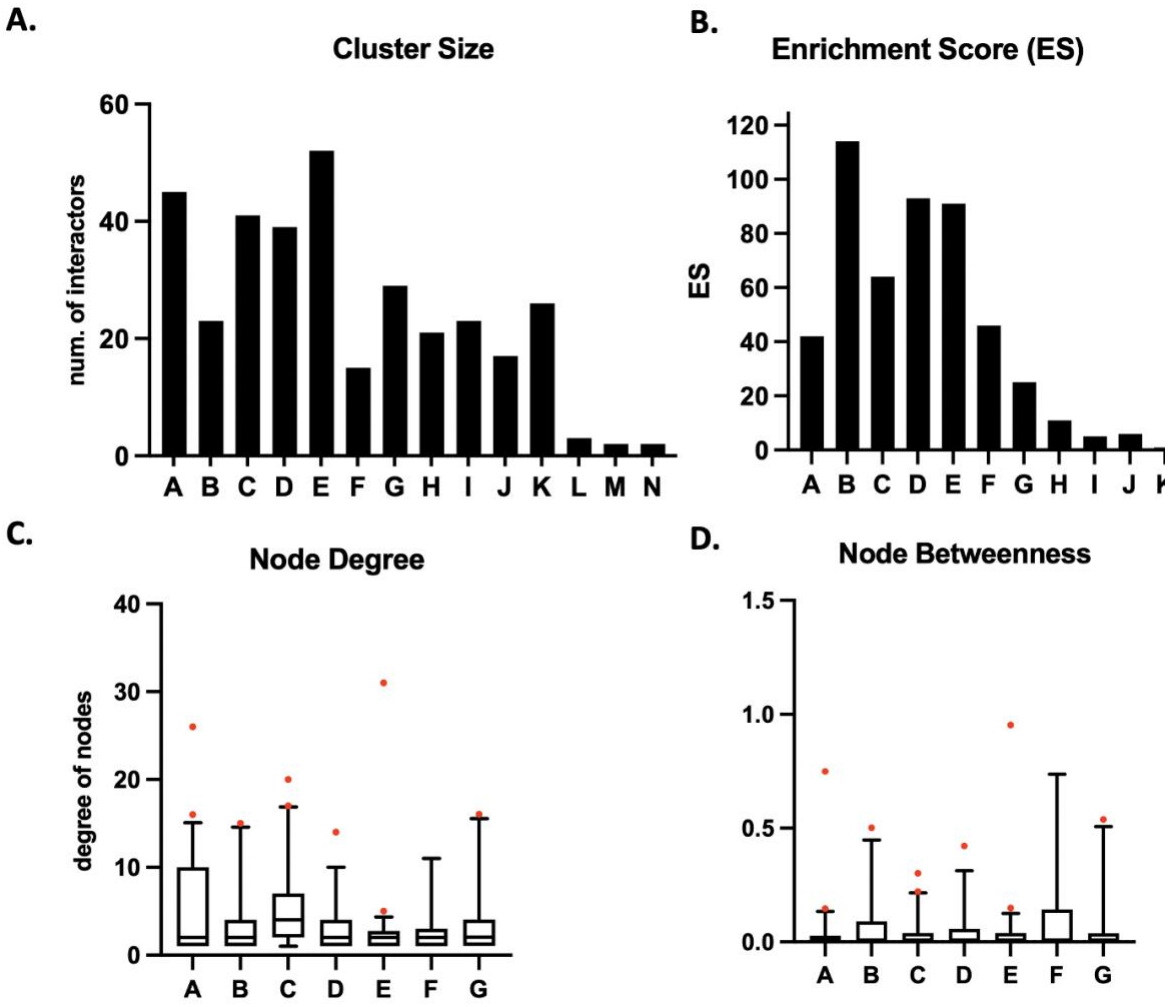


Figure 6. Characterisation of topological clusters in the LRRK2$_{net}$

A) *The bar graph shows the size of the 14 topological clusters identified in the LRRK2$_{net}$ via the Fast Greedy algorithm. Cluster L, M, N were excluded due to their small cluster size (N ≤ 5);* ***B)*** *The bar graph shows the number of GO-BP terms returned from enrichment analyses (Enrichment Score, ES) for the*

remaining 11 clusters. Cluster H, I, J, K were excluded due to their low ES ≤ 10; ***C, D)*** *The box plots show the distribution of node degree and betweenness in the remaining 7 clusters. The whiskers represent 95% percentiles. Nodes with degree or betweenness above 95% percentiles within each cluster were marked as red dots F-test was preformed to compare the variance of node betweenness, in which Cluster C presented significant low SD as compared to other clusters apart from Cluster E (p < 0.05; *).*

Among the remaining 11 clusters, Cluster A comprised N = 45 interactors, among which CDK2 (Degree = 26; Betweenness = 0.75) and RPS8 (Degree = 16; Betweenness = 0.10) were defined as "hub" proteins, among which 40 contributed to the functional enrichment of 42 GO-BPs (**Figure 7A**). A total of 28 GO-BPs were with a term size < 2500, from which GO hierarchy analysis identified 1 group which was related to translation, involving 40/45 (88.9%) interactors in the cluster (**Figure 7B**, **Table C1**). Of note, a robust ribosomal protein unit of 16 RPs were identified in Cluster A, in which the RPs presented a condense within-unit connection, accounting for 71/115 (61.7%) edges in Cluster A. Among the 16 RPs, RPS8 exhibited the highest centrality and connected with all other proteins, it also linked to 2 mitochondrial RPs (MRPL19 and MRPL28), suggesting that the RB unit is potentially associated with mitochondrial function. In addition, the RB unit was connected with the rest of the cluster by CDK2, thereby suggesting a potentially crucial role of CDK2 as a "hub" in linking ribosomal biosynthesis and other stages of gene translation.

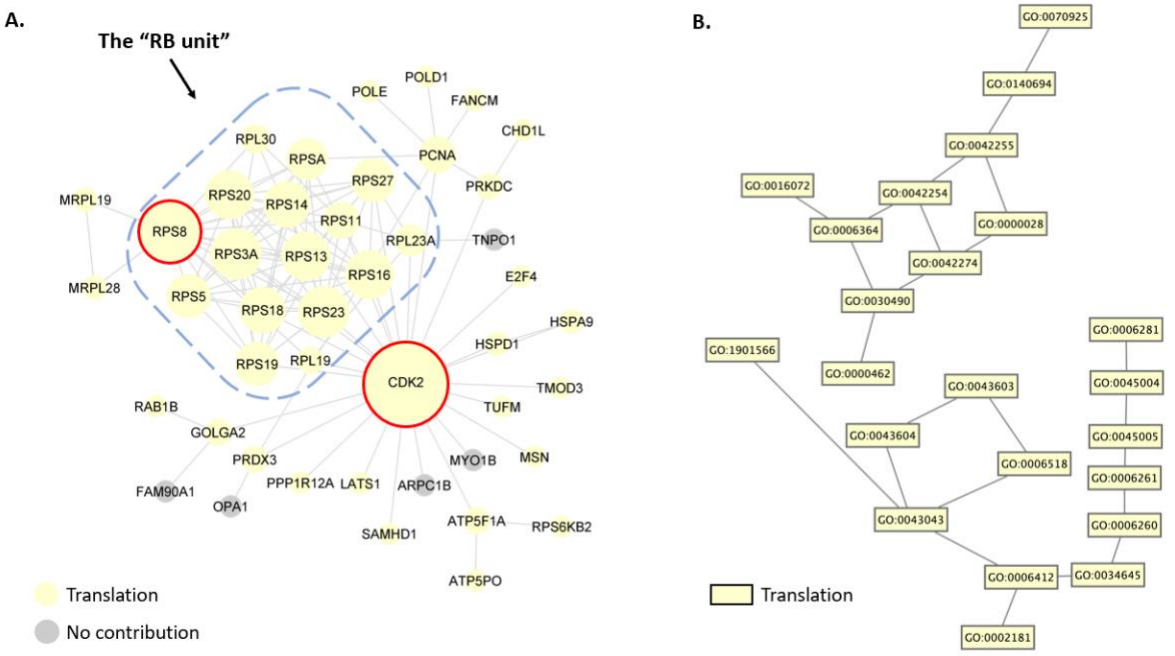


Figure 7. Analysis of Cluster A of the LRRK2$_{net}$

A) *The network graph shows Cluster A in the LRRK2$_{net}$, containing a total of 45 interactors (represented with round nodes). The "hub" proteins (CDK2 and RPS8) were highlighted with red circles. Node fill colour represents biological functions that a certain interactor is engaged in based on the GO-BP enrichment analysis, while node size represents the degree centrality (the larger the node, the higher the degree). Interactors that were not included in any functions were filled in grey. Of note, a condense unit of ribosomal proteins (the "RB unit") was identified in Cluster A (highlighted in blue square), accounting for 71/115 (61.7%) of all edges.* ***B)*** *The network graph shows the hierarchical groups of GO-BPs (term size < 2500, represented as rectangular nodes) returned from enrichment analysis for Cluster A. Only groups with ≥ 3 terms were presented. Node fill colour represents the hierarchical group. Edges represents hierarchy relations between GO-BP terms in Gene Ontology. Details regarding GO-BPs in the graph were shown in* ***Table C1****.*

In comparison, Cluster B contained N = 23 interactors, among which TRAF2 (Degree = 15, Betweenness = 0.50) was defined as "hub" protein (**Figure 8A**), though another 2 proteins TRADD (Degree = 13, Betweenness = 0.23) and RIPK1 (Degree = 9, Betweenness = 0.18) also exhibited high centralities in the cluster. These 3 proteins were responsible for 31/42 (73.8%) edges in the cluster, indicating their potential central roles in maintaining the connectivity of the subnetwork. GO-BP enrichment returned a total of 96 GO-BPs with term size < 2500, within which 3 hierarchical groups were identified: "Cell death", "Protein metabolism", "Response to stress" (**Figure 8B**, **Table C2**). These 3 groups involved 17, 14 and 15 interactors, respectively. Of note, a total of 14 interactors were engaged in all the 3 groups, suggesting that these 3 cellular pathways were closely connected with each other.

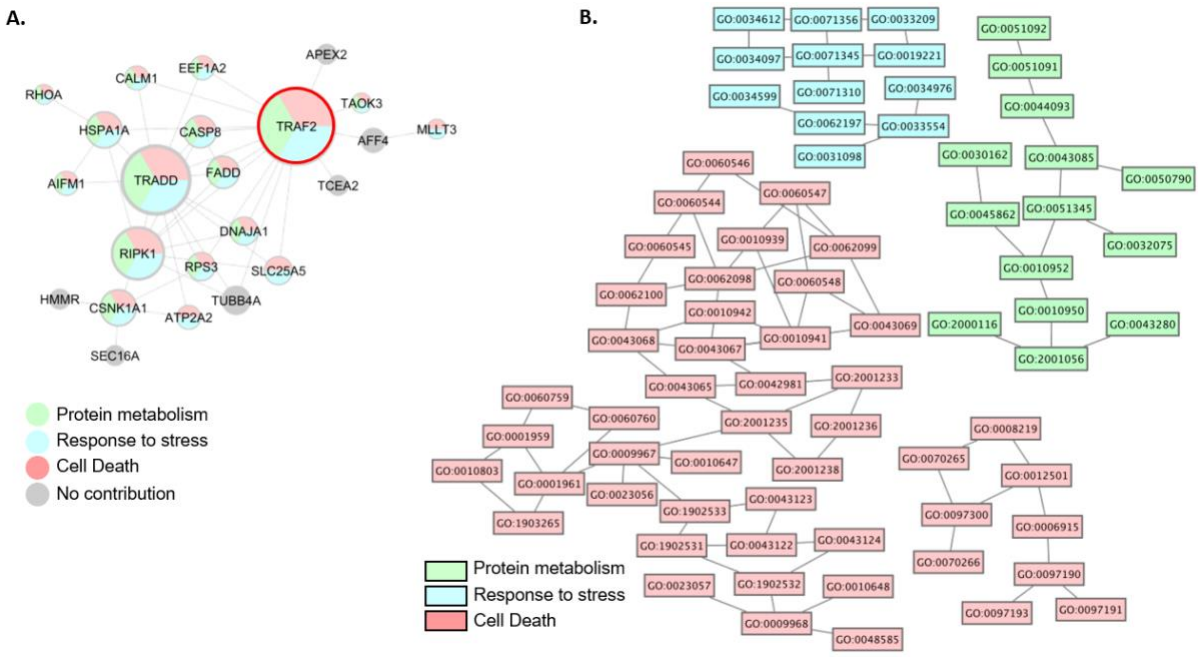


Figure 8. Analysis of Cluster B in the LRRK2$_{net}$

A) *The network graph shows Cluster B in the LRRK2$_{net}$, containing a total of 23 interactors (represented with round nodes). The "hub" proteins (TRAF2) was highlighted with a red circle. Node fill colour represents biological functions that a certain interactor is engaged in based on the GO-BP enrichment analysis, while node size represents the degree centrality (the larger the node, the higher the degree). Interactors that were not included in any functions were filled in grey.* ***B)*** *The network graph shows the hierarchical groups of GO-BPs (term size < 2500, represented as rectangular nodes) returned from enrichment analysis for Cluster B. Only groups with ≥ 3 terms were presented. Node fill colour represents the hierarchical groups. Edges represents hierarchy relations between GO-BP terms in Gene Ontology. Details regarding GO-BPs in the graph were shown in* ***Table C2****.*

Cluster C included a total of 41 interactors, among which IQGAP1 (Degree = 20; Betweenness = 0.30), CAPZA2 (Degree = 17; Betweenness = 0.12) and LIMA1 (Degree = 16; Betweenness = 0.22) were "hub" proteins (**Figure 9A**). Of note, pair-wise F-test showed that the variance of node betweenness in Cluster C was significantly lower than all other clusters apart from Cluster E (**Figure 6D**, 1-tailed p-value < 0.05), suggesting that interactors in Cluster C were well connected and thereby formed a stable PPI unit in the LRRK2$_{net}$. GO-BP enrichment analysis returned 64 terms, in which 59 were with term size < 2500, involving 29 interactors. A total of 3 hierarchical groups of GO-BPs were identified in the 59 terms, related to cytoskeleton organisation, vesicular transport and translation (**Figure 9B**, **Table C3**). A total of

25, 9 and 4 interactors contributed to the 3 functions, respectively, suggesting that the dominate role of Cluster C was maintaining cytoskeleton dynamics. Of note, only 1 protein (ACTB) was shared by the 3 GO-BP groups, suggesting its potentially important role in connecting the 3 functions together.

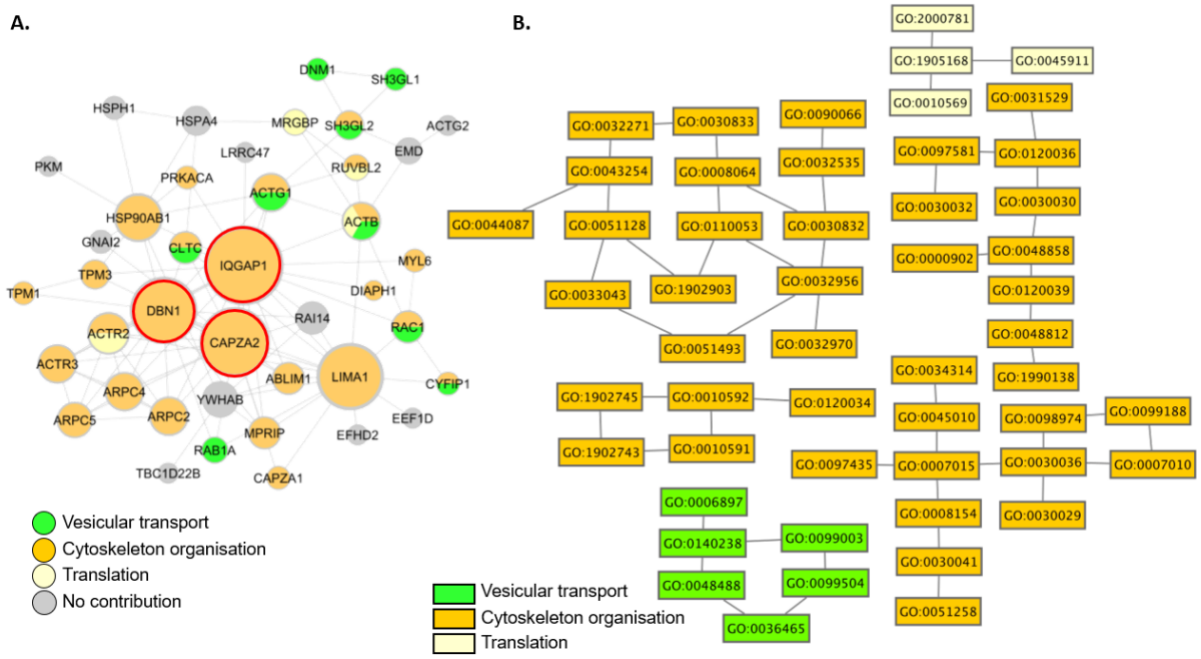


Figure 9. Analysis of Cluster C of the LRRK2$_{net}$

A) *The network graph shows Cluster C in the LRRK2$_{net}$, containing a total of 45 interactors (represented with round nodes). The "hub" proteins (IQGAP1, CAPZA2 and DBN1) were highlighted with red circles. Node fill colour represents biological functions that a certain interactor is engaged in based on the GO-BP enrichment analysis, while node size represents the degree centrality (the larger the node, the higher the degree). Interactors that were not included in any functions were filled in grey.* ***B)*** *The network graph shows the hierarchical groups of GO-BPs (term size < 2500, represented as rectangular nodes) returned from enrichment analysis for Cluster C. Only groups with ≥ 3 terms were presented. Node fill colour represents the hierarchical groups. Edges represents hierarchy relations between GO-BP terms in Gene Ontology. Details regarding GO-BPs in the graph were shown in* ***Table C3****.*

Cluster D contained 39 LRRK2 interactors, in which NPM1 (Degree = 14; Betweenness = 0.42) and VIM (Degree = 10; Betweenness = 0.31) were "hub" proteins (**Figure 10A**). Of note, Cluster D exhibited a significantly low node betweenness variance as compared to other clusters apart from Cluster C (F-test 1-tailed p-value < 0.05), thereby revealing another robust PPI unit in the LRRK2$_{net}$ (**Figure 6D**). GO-BP enrichment analysis associated Cluster D with 93

terms, in which 70 were with term size < 2500, among which a total of 4 hierarchical groups were identified: protein metabolism, protein localisation, translation and cell death, involving a total of 38, 18, 15 and 17 LRRK2 interactors, respectively (**Figure 10B**, **Table C4**). NPM1, AKT1, and RACK1 contributed to all the functions, thereby probably functioning as mediators of the 4 cellular processes.

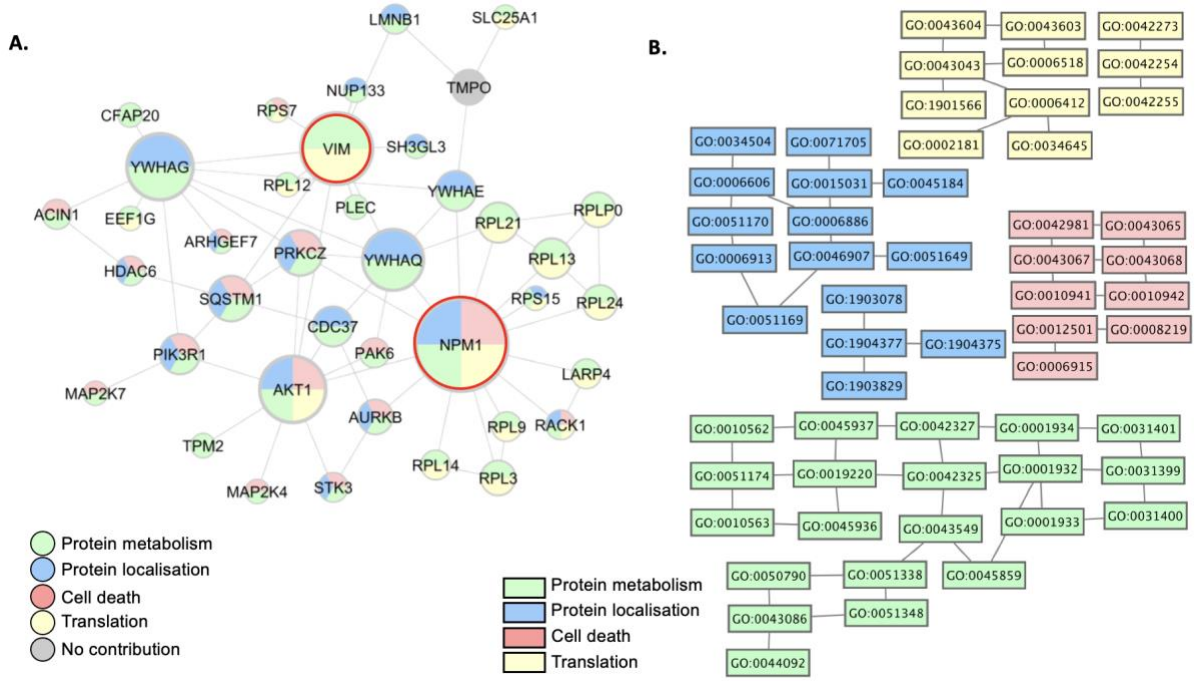


Figure 10. Analysis of Cluster D of the LRRK2$_{net}$

A) *The network graph shows Cluster D in the LRRK2$_{net}$, containing a total of 39 interactors (represented with round nodes). The "hub" proteins (NPM1 and VIM) were highlighted with red circles. Node fill colour represents biological functions that a certain interactor is engaged in based on the GO-BP enrichment analysis, while node size represents the degree centrality (the larger the node, the higher the degree). Interactors that were not included in any functions were filled in grey.* ***B)*** *The network graph shows the hierarchical groups of GO-BPs (term size < 2500, represented as rectangular nodes) returned from enrichment analysis for Cluster D. Only groups with ≥ 3 terms were presented. Node fill colour represents the hierarchical groups. Edges represents hierarchy relations between GO-BP terms in Gene Ontology. Details regarding GO-BPs in the graph were shown in* ***Table C4****.*

Cluster E was composed of a total of 52 interactors, within which TP53 was the "hub" (Degree = 31; Betweenness = 0.95). Of note, TP53 exhibited dominating centrality, accounting for

31/65 (47.7%) of all edges in the cluster, suggesting that TP53 plays a crucial role in maintaining the function of the cluster (**Figure 11A**). GO-BP enrichment analysis returned a total of 91 terms for Cluster E, within which 64 were with term size < 2500, involving 77/91 (84.6%) interactors. GO-BP hierarchy clustering identified 4 groups associated with: protein localisation, cell cycle, protein metabolism and response to stress, contributed by a total of 25, 21, 16 and 5 interactors, respectively (**Figure 11B**, **Table C5**).

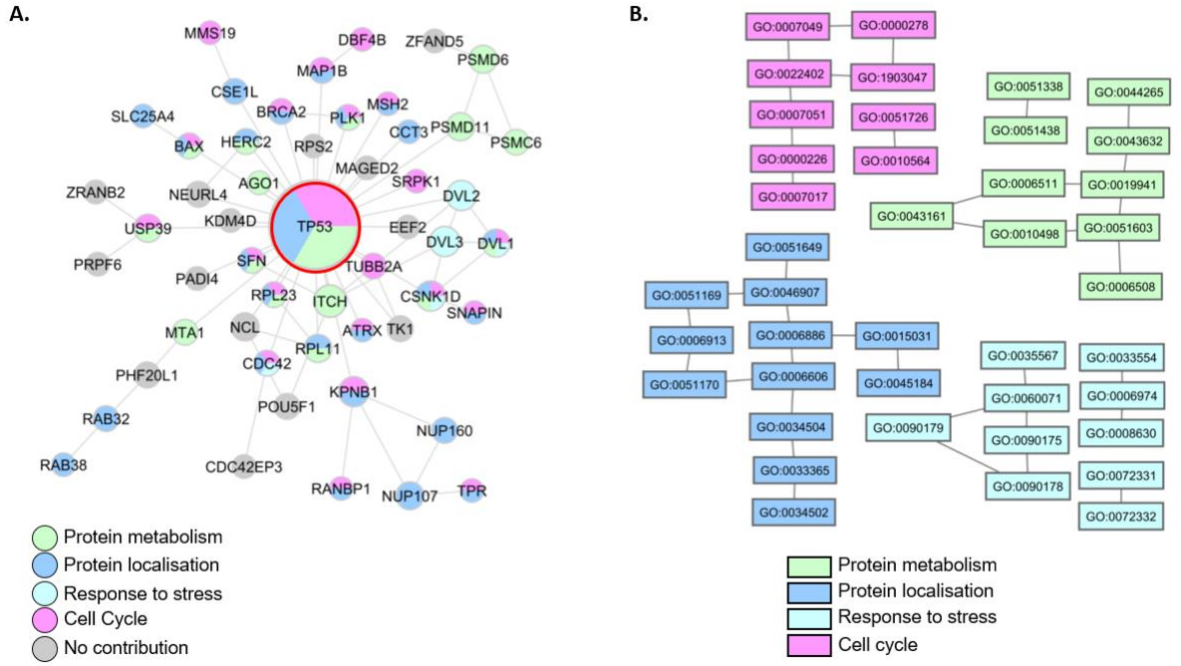


Figure 11. Analysis of Cluster E of the LRRK2$_{net}$

A) *The network graph shows Cluster E in the LRRK2$_{net}$, containing a total of 52 interactors (represented with round nodes). The "hub" protein (TP53) was highlighted with a red circle. Node fill colour represents biological functions that a certain interactor is engaged in based on the GO-BP enrichment analysis, while node size represents the degree centrality (the larger the node, the higher the degree). Interactors that were not included in any functions were filled in grey.* ***B)*** *The network graph shows the hierarchical groups of GO-BPs (term size < 2500, represented as rectangular nodes) returned from enrichment analysis for Cluster E. Only groups with ≥ 3 terms were presented. Node fill colour represents the hierarchical groups. Edges represents hierarchy relations between GO-BP terms in Gene Ontology. Details regarding GO-BPs in the graph were shown in* ***Table C5****.*

In comparison, Cluster F contained 15 LRRK2 interactors, in which PRKN (Degree = 12; Betweenness = 0.74) was defined as the "hub" protein (**Figure 12A**). Topological analysis

showed that Cluster F was highly centralised, in which the node betweenness variance was significantly higher than all other clusters (F-test 1-tailed p-value < 0.05), suggesting that the stability of Cluster F was highly dependent on the normal function of its "hub" PRKN, i.e., pathological alterations (e.g., pathogenic mutations or expression level changes) on PRKN may lead to an overall dysfunction of the cluster (**Figure 6D**). A total of 26 GO-BPs were returned by functional enrichment analysis for Cluster F, among which 23 were with term size < 2500. These terms were related to 1 hierarchy group of Autophagy, involving 6/15, 40%) interactors (**Figure 12B**, **Table C6**).

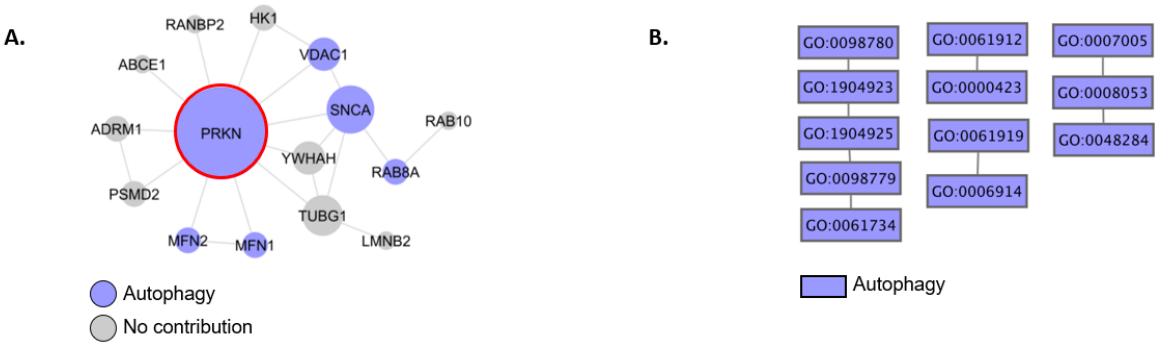


Figure 12. Analysis of Cluster F of the LRRK2$_{net}$

A) *The network graph shows Cluster F in the LRRK2$_{net}$, containing a total of 52 interactors (represented with round nodes). The "hub" protein (PRKN) was highlighted with a red circle. Node fill colour represents biological functions that a certain interactor is engaged in based on the GO-BP enrichment analysis, while node size represents the degree centrality (the larger the node, the higher the degree). Interactors that were not included in any functions were filled in grey.* ***B)*** *The network graph shows the hierarchical groups of GO-BPs (term size < 2500, represented as rectangular nodes) returned from enrichment analysis for Cluster F. Only groups with ≥ 3 terms were presented. Node fill colour represents the hierarchical groups. Edges represents hierarchy relations between GO-BP terms in Gene Ontology. Details regarding GO-BPs in the graph were shown in* ***Table C6****.*

As for Cluster G, a total of 29 interactors were included in the cluster, among which LRRK1 (Degree = 15; Betweenness = 0.54) and HSPA8 (Degree = 16; Betweenness = 0.48) were defined as "hub" proteins (**Figure 13A**). The 29 interactors were related to 25 GO-BPs, among which 23 were with term size < 2500. A total of 2 hierarchical groups were identified from these functional terms, related to autophagy (involving 7/29 (24.1%) interactors) and protein metabolism (involving 17/29 (58.6%) interactors) (**Figure 13B**, **Table C7)**.

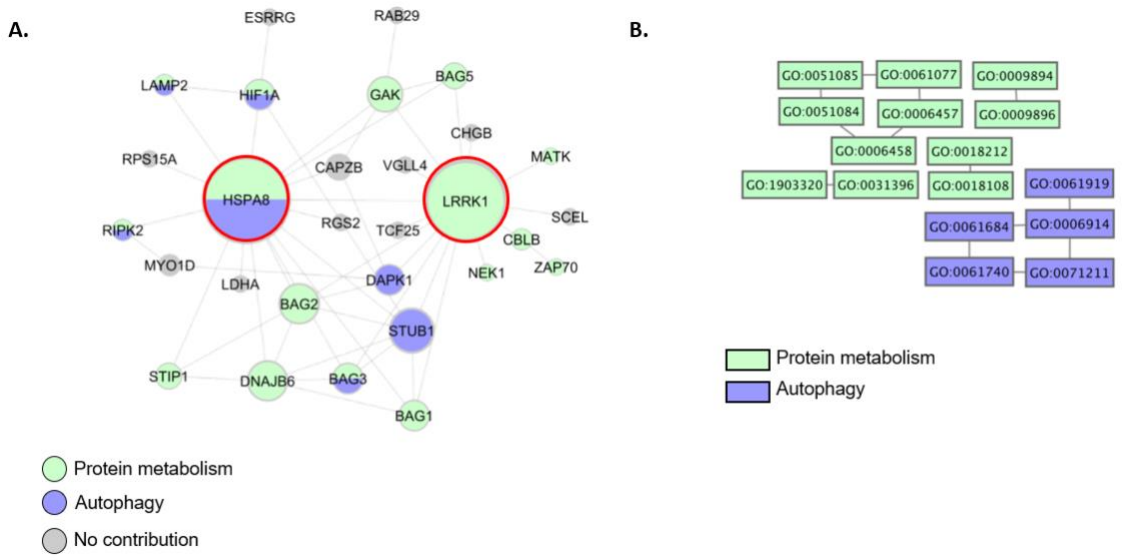


Figure 13. Analysis of Cluster G of the LRRK2$_{net}$

A) *The network graph shows Cluster G in the LRRK2$_{net}$, containing a total of 29 interactors (represented with round nodes). The "hub" proteins (HSPA8 and LRRK1) was highlighted with red circles. Node fill colour represents biological functions that a certain interactor is engaged in based on the GO-BP enrichment analysis, while node size represents the degree centrality (the larger the node, the higher the degree). Interactors that were not included in any functions were filled in grey.* ***B)*** *The network graph shows the hierarchical groups of GO-BPs (term size < 2500, represented as rectangular nodes) returned from enrichment analysis for Cluster G. Only groups with ≥ 3 terms were presented. Node fill colour represents the hierarchical groups. Edges represents hierarchy relations between GO-BP terms in Gene Ontology. Details regarding GO-BPs in the graph were shown in* ***Table C7****.*

Main findings

1. The topological core of the LRRK2$_{net}$ (defined as the most highly connected interactors within the LRRK2net) contained 27 proteins, including 14-3-3 proteins, ribosomal proteins, heat shock proteins and cytoskeleton-related proteins. Together with the functional core identified in Chapter 1, these proteins formed the fundamental set of proteins interacting with LRRK2.
2. A total of 11 topological clusters containing N ≥ 5 interactors were identified in the LRRK2$_{net}$, and 7 of them were found with high biological significance, suggesting a good overlap between topological models of LRRK2 interactors and the real functional modules.

3. The 7 biological significant clusters re-capitulated the primary functions of LRRK2 in: the regulation of ribosomal functionality, protein localisation, protein metabolism, protein transport, cellular response to stress, cell death and autophagy.
4. An interesting ribosomal unit was identified in the $LRRK2_{net}$, suggesting LRRK2's potential role in mediating/regulating the process of protein translation (Topological Cluster A).
5. An interesting cytoskeletal unit was identified in the $LRRK2_{net}$, suggesting LRRK2's potential role in mediating/regulating the maintenance of the cellular structure and the regulation of vesicle dynamics (Topological Cluster C).
6. An interesting mitochondrial unit was identified in the $LRRK2_{net}$, suggesting LRRK2's potential role in regulating the maintenance of mitochondrial integrity is association with other PD related proteins (Topological Cluster F).

Discussion

In this section, PPIs among LRRK2 interactors were retrieved from literature to increase the connectivity of the simple $LRRK2_{int}$ and effectively construct an interaction network. As compared to the download pipeline for 1^{st}-layer PPI, which included the combination of 3 PPI tools (PINOT, HIPPIE and MIST), retrieval of the 2^{nd}-layer PPIs was conducted via HIPPIE (v2.3) solely. This is due to the following reasons: 1) HIPPIE (v2.3) offers time-efficient network query which allows for the construction of a subnetwork from an input query set of proteins, thereby avoiding the significant amount of time and computing power required by retrieving, filtering and merging of PPIs for individual LRRK2 interactors, while no such function is provided in the other 2 PPI tools; 2) As discussed in Chapter 1, the 3 tools share a relatively large overlap rate especially on the most replicated PPIs, suggesting that using 1 PPI tool solely is able to cover the most robust PPIs among LRRK2 interactors. In addition, with the score filter provided by HIPPIE (v2.3), the pipeline returns adequate 2^{nd}-layer PPIs with high confidence for $LRRK2_{net}$ construction. The $LRRK2_{net}$ turned out to be a scale-free network that follows the power law, in which only a few nodes are highly connected while most of the nodes are characterized by a relatively low degree (i.e., they possess few connections). This might indicate that there are some "hubs" in the LRRK2's protein network sustaining the LRRK2 functions that are exerted via its interactome. These hubs were identified according to their degree or betweenness centrality via topological analysis. In graph theory, degree of node represents the number of edges connected to a given nodes, while betweenness

centrality reflects the number of shortest paths a node lies on, a proxy for the influence a node has over the flow of information or resources in the network. In scale-free networks, the degree of a node and its betweenness centrality are often correlated, but they do not always agree. Therefore, a combination of nodes with high degree or high betweenness ensures the coverage of the "topological core" with both centralised nodes and bridge nodes in the network. In this way, a total of 27 interactors were identified as hubs, including TP53, CDK2, HSPA8, HSP90AB1, HSP90AA1, YWHAZ, LAPR7, NPM1, TRAF2, IQGAP1, LIMA1, CAPZA2, PRKN, DBN1, YWHAQ, RPS8, YWHAG, TRADD, RPS3, AKT1, YWHAB, HSPA1A, RPS3A, TUBA1C, RPS2, PPP1CA, LRRK1. Of note, caution should be exercised when interpreting topological analyses in the context of protein interaction networks since these networks are inherently incomplete. Consequently, there is a risk of underestimating or overlooking central nodes and connectivity. Hence, these analyses require periodic repetition every few years to ensure the continuous incorporation of updated data and gradually enhance the completeness of the model over time. Among the hub proteins, AKT1, HSP90AB1, HSPA1A, NPM1 also presented in the functional core identified in Chapter 3, suggesting their central roles in supporting LRRK2 functionality.

AKT1 is a central protein that participate in signal transduction for cell cycle progression, cell survival and prevention of apoptosis. Previous studies have found that AKT1 is a direct substrate of LRRK2 kinase. Phosphorylation of AKT1 negatively regulates apoptosis signalling molecule and activates some cascades for cell survival. PD-associated *LRRK2* mutations, including LRRK2-R1441C, G2019S, and I2020T, were found to interrupt LRRK2:AKT1 interaction (Ohta et al., 2011). Therefore, neurons with mutated LRRK2 are potentially more vulnerable to apoptotic stress, which could be a possible mechanism for the neurodegeneration in the LRRK2-PD.

Heat shock proteins (HSPs) constitute a vital family of cellular components that respond to various stressors. Functioning as molecular chaperones, HSPs play a fundamental role in the maintenance of protein structure, preventing misfolding, aggregation, and degradation. In the meanwhile, heat shock proteins have been highly linked to PD. Previous studies found lower expression levels in a range of heat shock proteins in PD brains as compared to healthy control (Zhu et al., 2022). In addition, HSP90AB1 exhibits significantly negative correlation

with Lewy pathology by inhibiting the aggregation of α-synuclein (Gao et al., 2015). Also, HSPA1A (a member of HSP70 family) disaggregate complex is responsible for the disassembly of α-synuclein aggregates (Daturpalli et al., 2013). In this study, heat shock proteins present high final score (HSP90AA1, HSPA8 (a member of HSP70 family)) in the $LRRK2_{int}$, high degree & betweenness centrality (HSP90AB1, HSP90AA1, HSPA8) in the $LRRK2_{net}$, and high engagement in all enriched biological functions in the functional core (HSP90AB1, HSP90AA1, HSPA1A), suggesting an essential role of this protein family in LRRK2 signalling pathway. These findings are in accordance with previous studies which suggested a substantial association between LRRK2 and HSP90, HSP70 proteins. For example, it has been reported that pharmacological inhibition of HSP90 rescued cortical neurons in a LRRK2-G2019 mouse model from axonal growth retardation (Hurtado-Lorenzo & Anand, 2008). This may result from the fact that LRRK2:HSP90 interaction increase the stability of G2019S LRRK2 protein, while inhibition of HSP90 increases its proteasomal degradation (L. Wang et al., 2008). Hence, HSP90 inhibitors may thereby be a potential drug target for LRRK2-PD. However, on the other hand, some studies suggested that HSP70 overexpression enhanced neuronal survival against oxidative stress induced by increased LRRK2 kinase activity from pathogenic mutations (Jang et al., 2018). Therefore, HSP70 induction can also be a potential therapeutic route for LRRK2-PD. These findings suggest a complicated signalling network between LRRK2 and heat shock proteins. Future studies can extract functional, topological, transcriptomic and genetics data of heat shock proteins and establish a LRRK2:HSPs PPI network. By combining the network with experimental data, the edges can be directed based on inhibition/activation effect of each LRRK2:HSPs binding, thereby turning the PPI network to a pathway model. The model can be tested in the health condition as well as in the presence of pathogenic mutations of *LRRK2*, which will provide insights in the LRRK2:HSPs signalling system and prioritise "hub" heat shock proteins as drug target candidates.

NPM1 (Nucleophosmin 1) is an abundant nucleolar protein that plays a crucial role in ribosome assembly, regulation of cell division, and response to cellular stress. Like HSP90 proteins, NPM1 is also a type of chaperon that regulates protein misfolding and aggregating process, presenting a neuroprotective role in neurodegeneration process. One previous study found upregulated NPM1 expression level in Human dopaminergic SH-SY5Y cells under 1-methyl-4-phenyl-pyridinium ion (MPP+) treatment, which is a commonly used model for

sporadic PD (Xie et al., 2016). However, not much research has been done in exploring the functional role of LRRK2:NPM1 interaction, either in healthy or PD cases. Considering the neuroprotective role of NPM1, it is worth to explore this PPI in further research to evaluate NPM1's potential as a novel drug target for PD.

Apart from the topological unit of hub proteins, a total of 14 topological clusters were identified in the LRRK2 network via Fast Greedy Clustering Algorithm. Fast Greedy is one of the simplest community detection approaches in topological analysis, involving a hierarchical clustering based on edge betweenness to detect the mostly connected nodes as the "core" of a subnetwork. Then the algorithm iteratively selects random edges that improve the modularity of the subnetwork and adds them in, until the modularity stops improving (Newman, 2004). One of the primary benefits of Fast Greedy algorithm lies in its high efficiency and low memory occupancy. Moreover, unlike Partitioning-based community detection, such as k-means, which is another commonly used clustering method, the Fast Greedy focus on optimising the local connectivity rather than the global optimum, which fits the goal of clustering analysis in this study, i.e., to identify dense interaction units within the $LRRK2_{net}$. Among the 14 clusters, only half was linked with high biological significance (enriched for ≥ 10 GO-BPs), suggesting that in-silico models established via bioinformatics methods should always be interpreted in the biological context.

Among the 7 biological significant topological clusters, Cluster A was highly associated with translation. It contains a "ribosomal unit" of 16 ribosomal proteins with 71 PPIs among them, which accounting for 35% nodes and 61% edges the whole subnetwork. This is in accordance with previous findings that LRRK2 plays an important role in regulating ribosomal functions in neurons. Wild type LRRK2 was found to repress protein synthesis in mouse neurons, while pharmacological inhibition of LRRK2 kinase or LRRK2 knockout rescued translation (Deshpande et al., 2020). Similar alterations have been found in fibroblasts from patients with sporadic PD and G2019S-PD, and LRRK2 inhibition restores normal protein synthesis (Flinkman et al., 2023). Interestingly, such changes were not observed in fibroblasts from patients with other neurodegenerative disease such as multiple system atrophy, suggesting that LRRK2's impact on ribosomal functions is potentially disease-specific. Moreover, the ribosomal unit is connected to the other part of Cluster A via the hub protein CDK2, indicating

that CDK2 potentially functions as a mediator of ribosomal function. This is in accordance with previous findings which suggested that CDK2 positively modulates the assembly of the transcription initiation complex of ribosomes (Voit & Grummt, 2001).In addition, it has been suggested that inhibiting CDK2 activity leading to downregulation of rRNA synthesis (Iadevaia et al., 2010; Juli et al., 2016). Therefore, it can be hypothesised that LRRK2 and CDK2 form a balanced mediating system on ribosomal functions, though the detailed mechanism requires further research. Using Cluster A as a starting point, future research can focus on linking LRRK2, CDK2 and the LRRK2 ribosomal unit with other interactors in the cluster, which may function as signalling transductors in the regulation process.

Cluster C was highly enriched for GO-BP terms related to cytoskeleton organisation and vesicular transport. It contained cytoskeleton proteins including actin, endophilin, myosin, tropomyosin, and vesicular transport regulators such as RAB1A and DNM1 (Mukhopadhyay et al., 2011; Von Spiczak et al., 2017; Zhuang et al., 2010). It also contained signal transductors such as HSP90AB1. Therefore, Cluster C can be considered as a model for the LRRK2-mediated regulating cascade for the regulation of cytoskeleton dynamics. There are 3 hub proteins in this cluster: IQGAP, CAPZA2 and DBN1. IQGAP is a conserved scaffold protein that facilitates the formation of protein complexes that regulate a range of cellular processes (Hedman et al., 2015). For example, it has been reported to interact with F-actin and small GTPase and promote actin polymerisation. In addition, IQGAP also scaffold signal transducing molecules such as participants of MAPK pathway, which is responsible for response to stimuli. CAPZA2 regulates actin dynamics by capping the barbed ends of actin filaments. The capping of actin filaments helps control the polymerization and depolymerization of actin, which is essential for processes such as cell motility, muscle contraction, and maintenance of cell structure. DBN1 codes for Drebrin, a protein that binds actin and regulates actin cytoskeleton organisation and dynamics. Drebrin is particularly abundant in the brain, where it is involved in the regulation of neuronal morphology and synaptic function. To date, not much evidence has been reported regarding the impacts of LRRK2 interactions on the functions of the 3 proteins. Considering their essential roles in regulating the cytoskeleton dynamics and the same downstream effectors, it is worth to investigate the connections among these mediators to obtain a more comprehensive understanding in the mechanism of cytoskeleton organisation. Of note, the PPIs in Cluster C have been selected and evaluated functionally via

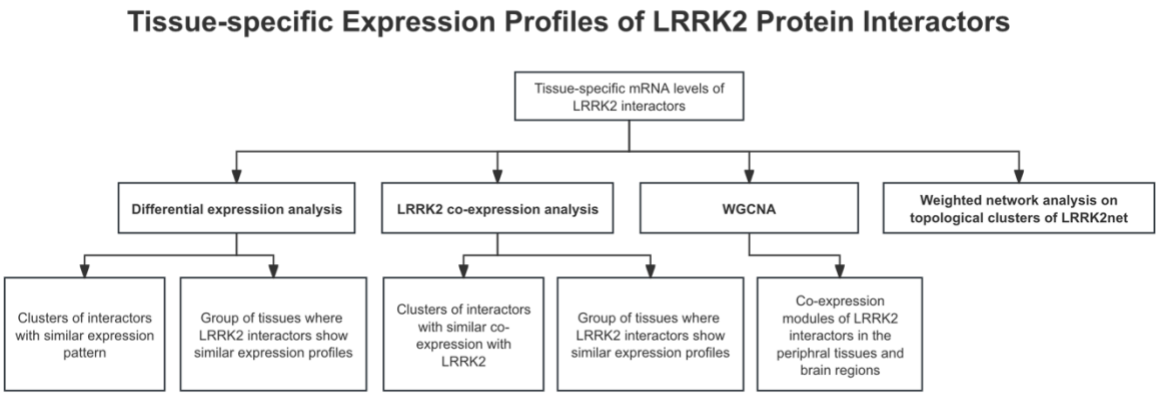
affinity purification coupled with mass spectrometry (AP-MS/MS) in a collaborative study with Prof. Elisa Greggio's group in University of Padova. The results showed that the interactions between LRRK2 and DBN1, ARPC2, ACTR2 and ACTR3 increased under brain derived neurotrophic factor (BDNF) stimulation, suggesting a potential role of LRRK2 and its interactors in Cluster C in regulating actin cytoskeleton dynamics in complementary neuronal models (Tombesi et al., 2023). Cluster F is highly correlated with mitophagy. It contains one hub protein PRKN, which is a well-established player in the ubiquitin-proteasome system and is intricately linked to mitophagy. Through its E3 ubiquitin ligase activity, PRKN targets damaged or dysfunctional mitochondria for degradation, thereby preventing the accumulation of compromised organelles and ensuring cellular health. In the meanwhile, cluster also include SNCA (α-synuclein), which has been reported to influence mitochondrial dynamics and potentially affecting the initiation of mitophagy. Moreover, RAB8A and RAB10 also presented in Cluster F, which are the primary contributors to vesicular trafficking and membrane dynamics. Additionally, the presence of MFN1 and MFN2 underscores the significance of mitochondrial dynamics within this cluster considering their pivotal roles in mitochondrial fusion. Therefore, considering the complexity of LRRK2-associated mitophagy pathway, it is worth to analyses these interactions in 1 system, as how they were connected in Cluster F.. This can be applied to multiple research context: for example, since *LRRK2, PRKN* and *SCNA* are all confirmed familial PD-causing genes, it would be interesting to investigate how they work together in maintaining normal mitophagy process in the healthy cells. Also, in the disease scenario, it would be informative to dissect how pathogenic mutations of each of these PD genes affect the mitophagy network. These will provide valuable insights for PD pathology and mitophagy-targeted drug development for genetic PD. The remaining 4 topological clusters obtained from the $LRRK2_{net}$ were associated with a combination of different biological processes. For example, Cluster B was related to protein metabolism, response to stress and cell death, while Cluster E was associated with protein metabolism, protein localisation, response to stress and cell cycle. These clusters may function as bridges that connect each functional block of the $LRRK2_{net}$ together. In conclusion, the in-silico model of the $LRRK2_{net}$ and its subnetworks well recapitulated the primary biological functions of LRRK2 described in literature for PD and LRRK2 with the advantage of showing the actual component of the LRRK2 interactome that coordinate those functions with LRRK2 and describing their connectivity and the flux of information sustaining those biological pathways.

Chapter 3. Tissue-specific expression profiles of LRRK2 protein interactors

Objectives

- Investigate the mRNA expression patterns of LRRK2 interactors in 15 different healthy human tissues
- Compare the LRRK2:interactor co-expression features across the 15 healthy human tissues
- Identify co-expression modules of LRRK2 interactors among the 15 healthy human tissues
- Evaluate weighted connectivity of topological clusters of the $LRRK2_{net}$ detected in Chapter 4 in different tissues

Analysis pipeline



Note: Figures and text in Chapter 3 are adapted from the following publication: Zhao et al.; "Tissue specific LRRK2 interactomes reveal a distinct striatal functional unit"; PLoS Compute Biol. 2023 Jan 30;19(1):e1010847. doi: 10.1371/journal.pcbi.1010847.

Methods

- RNA-Seq data download and quality control (QC)

The Genotype-Tissues Expression (GTEx) Portal (GTEx Analysis V8) was queried for mRNA expression levels (read counts) of LRRK2 interactors in different tissues on 19 August 2021 (https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEx_Analysis_2015-06-05_v8_RNASeQCv1.1.9_gene_reads.gct.gz). A total of 15 tissue sites were included in this section: 11 brain regions: amygdala, anterior cingulate cortex, caudate (basal ganglia),

cerebellum/cerebellar hemisphere, cortex/frontal cortex (BA9), hippocampus, hypothalamus, nuclues accumbens (basal ganglia), putamen (basal ganglia), spinal cord (cervical c-1) and substantia nigra (basal ganglia); as well as 4 peripheral tissues: whole blood, lung, liver and kidney(cortex). Of note, in GTEx Portal, "cerebellum/cerebellar hemisphere" and "cortex/frontal cortex (BA9)" are duplicated pairs (https://www.gtexportal.org/home/faq – brainCortexAndCerebellum). The sets of "cerebellum" and "cortex" were chosen since they contain a larger sample size as compared to their duplicates (209 vs. 155; 205 vs. 155). Expression data were extracted for LRRK2 interactors using their HUGO/HGNC gene symbols. Samples quality control (QC) was performed via hierarchical clustering for each tissue expression dataset. Samples that fell outside the major cluster in the dendrogram were considered as outliers and were thereby excluded from analysis. Interactor-level QC involved excluding LRRK2 interactors with read counts < 15 in more than 75% of samples within each tissue. QC-ed tissue expression datasets were then merged into the "Tissue Expression Matrix", in which rows were LRRK2 protein interactors while columns were all samples for the 15 tissues. Read counts normalisation was conducted via the R package "DESeq2", obtaining the "Normalised Tissue Expression Matrix".

- Tissue-specific expression signature of the $LRRK2_{int}$

Hierarchical clustering was then performed on the "Normalised Tissue Expression Matrix" to:

1) identify clusters of tissues in which LRRK2 interactors showed similar co-expression patterns.
2) identify clusters of interactors that presented similar expression profiles with LRRK2 across the 15 tissues (i.e., interactors that presented in the same cluster with LRRK2 in the hierarchical dendrogram, hereby referred as "Exp_Cluster").

The optimal cutting height (h) for each dendrogram tree was selected by the Elbow method. The average expression levels of each Exp_Cluster in the 15 tissues were compared via One-way ANOVA followed by post hoc Tukey's test.

- Pair-wise Differential Expression Analysis (DEA)

The mRNA expression levels of each LRRK2 interactor in the 15 tissues were compared via pair-wise DEA using the R package "DESeq2". Log2 fold change (log2FC) and p-value (adjusted via the Benjamini-Hochberg procedure) for each comparison were automatically calculated

by “DESeq2”. Of note, the p-values were further adjusted via the Bonferroni’s method to minimise the overall type II error and were utilised to calculate the “Tissue Scores” (TS) for the 15 tissues via the following approach:

1) for each pair of tissues, if the expression level of interactor *I* is significantly higher in Tissue *A* than in Tissue *B* (log2FC > 1 and adjusted p-value < 0.05), then $TS^{E}_{I,A} = TS^{E}_{I,A} + 1$, while $TS^{E}_{I,B}$ remains unchanged, and vice versa;
2) if the comparison between Tissue *A* and Tissue *B* is insignificant, both $TS^{E}_{I,A}$ and $TS^{E}_{I,B}$ remain unchanged.

In this way, each tissue was scored multiple times based on the expression levels of each LRRK2 interactor. The higher the score ($TS^{E}_{I,T}$), the higher the interactor *I* was expressed in tissue *T*. If $TS^{E}_{I,T} \geq 12$, i.e., the expression level of *I* in tissue *T* is significantly higher than 12/15 (80%) of tissues included in this study, *I* is then defined as a significant tissue-specifically expressed interactor of LRRK2. These tissue-specific interactors were then annotated via GO-BP enrichment.

- Tissue-specific LRRK2-Co-expression Analysis (L-CEA) on LRRK2 interactors

Pearson’s Correlation test was performed on the expression level of LRRK2 and each of its interactor to examine their co-expression behaviours in the 15 tissues, thereby generating the “LRRK2-cor matrix”. The “LRRK2-cor matrix” was then used to perform hierarchical clustering on LRRK2 interactors and the 15 tissues in order to:

i. identify tissues where similar LRRK2:interactor co-expression was observed
ii. identify clusters of interactors presented similar co-expression behaviours with LRRK2 in the 15 tissues (Co-ex_Cluster);

In addition, the average LRRK2:interactor co-expression level of each Co-ex_Cluster was compared across the 15 tissues via One-way ANOVA followed by post-hoc Tukey HSD Test. Tissues were ranked based on the significant comparison results via the following approach:

1) if the average LRRK2:interactor co-expression levels of Co-ex_Cluster *X* are significantly higher in Tissue *A* than in Tissue *B* (|log2FC| > 1 and adjusted p-value < 0.05), then $TS^{C}_{X,A} = TS^{C}_{X,A} + 1$, while $TS^{C}_{X,B}$ remains unchanged, and vice versa;
2) if the comparison between Tissue *A* and Tissue *B* is insignificant (p-value > 0.05), both $TS^{C}_{X,A}$ and $TS^{C}_{X,B}$ remain unchanged.

In this way, for a given tissue, the higher the TS_T^C, the higher the LRRK2:interactor co-expression was observed in the tissue.

- Weighted Gene Co-expression Network Analysis (WGCNA)

WGCNA was performed on the "Tissue-matrix" to construct signed co-expression networks for the 15 tissues via the R package "WGCNA" through the following steps:

1) Co-expression levels among LRRK2 interactors were calculated via Pearson's correlation test ($s_{(i,j)}$), forming the similarity matrix.
2) The adjacency between each pair of interactors ($a_{(i,j)}$) was calculated as $a_{(i,j)} = (\frac{1+s_{(i,j)}}{2})^{\beta}$, in which β represents the soft thresholding power ($0 \leq \beta \leq 30$). The value of β was selected via the "WGCNA" R package, by selecting the minimum β that achieves the largest Scale-free Fit Index and the lowest Mean Connectivity, forming the adjacency matrix.
3) The Topological Overlap Matrix (TOM) was then calculated based on the adjacency matrix automatically the "WGCNA" package, thereby forming the weighted co-expression network of LRRK2 interactors

After the network construction, hierarchical clustering was then performed on the co-expression network. Co-expression modules of LRRK2 interactors were identified automatically by cutting the hierarchical dendrogram. Optimal cutting height was automatically selected via the "WGCNA" package. Detected modules were then labelled with different colours for identification. Next, module eigengene (ME) was calculated as the first principal component of the module expression matrix. MEs were then associated with each tissue via the function "corPvalueStudent" in the same R package. A significant correlation was defined as with **a)** |correlation coefficient| > 0.50 and **b)** p-value < 0.05. Modules with their MEs significantly correlated with certain tissue(s) were defined as tissue-specific and were annotated by GO-BP enrichment analysis.

- Weighted network analysis on topological clusters of the $LRRK2_{net}$

Subnetwork for each topological cluster identified in Chapter 4 was extracted from the $LRRK2_{net}$. Within the subnetwork in tissue *X*, interactors were weighted based on the degree of node (*D*) and the expression score (ES), which is equal to $TS_{I,X}^E$ ($node\ weight = D_I \times ES_I$),

while edges were weighted based on the edge betweenness ($EB_{(i,j)}$) and the co-expression levels of each pair of connected interactors ($C_{(i,j)}$)($edge\ weight = EB_{(i,j)} \times C_{(i,j)}$). In this way, each topological cluster was weighted 15 times for the 15 tissues based on the expression and co-expression behaviours of LRRK2 interactors, thereby forming 15 weighted subnetworks. Node weights and edge weights were compared across the 15 weighted subnetworks via Kruskal-Wallis test followed by post hoc Dunn's test.

Results

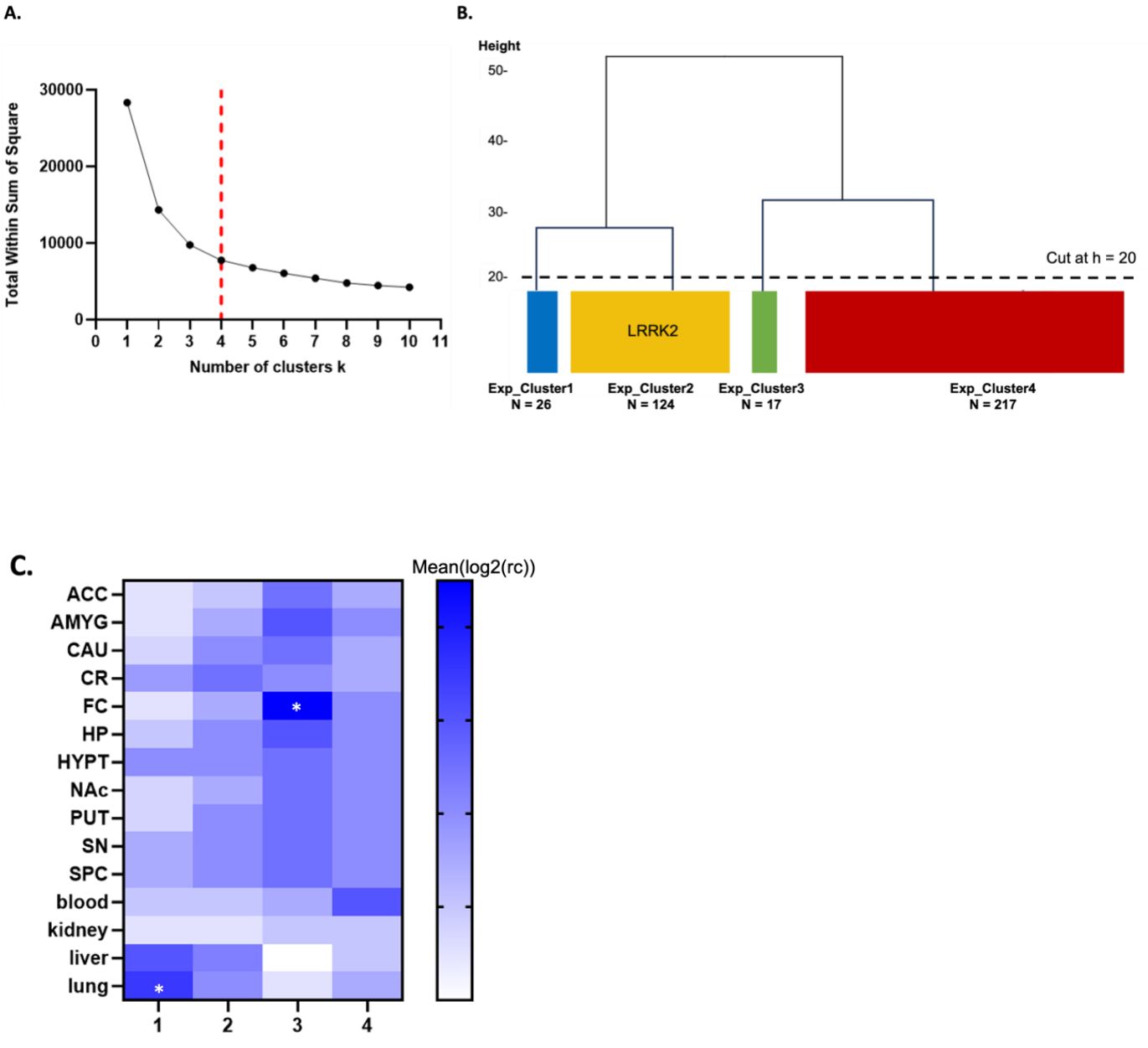
- Expression data preparation

RNA-Seq read counts were extracted for 404 out of 418 (97.4%) LRRK2 interactors from GTEx (for 14 LRRK2 interactors no data was available). Out of 3954 samples, a total of 280 outlier samples was identified by hierarchical clustering and thereby excluded (**Figure S1**). In addition, interactor QC removed 20 proteins with low counts or missing values, including "ACTBL2", "ANKS4B", "AURKB", "C1orf87", "CXCL11", "DNAJB8", "ESRRG", "FAM47B", "GTSF1", "KDM4D", "PADI4", "RAB38", "SCEL", "SH3GL3", "SKA3", "SMTNL2", "TAS2R60", "TGIF2LX", "WTI1-AS" and "XIRP2". The QC-ed read count matrix was further normalised via the R package "DESeq2", generating the "Normalised Tissue Expression Matrix" of 384 LRRK2 interactors for 3674 samples in 15 tissues.

- Expression profile of the $LRRK2_{int}$ in different tissues

Expression levels of each LRRK2 interactor in the 15 tissues were calculated and log2-transformed. Hierarchical clustering was performed on the transformed matrix to identify clusters of interactors that exhibited similar expression signatures across the 15 tissues. The Elbow method suggested 4 clusters in the dendrogram (**Figure 14A**). Therefore, the tree was cut at the height of 20 (h = 20) to obtain the 4 clusters (Exp_Cluster1-4, containing N = 26, 125, 17 and 217 interactors) (**Figure 14B**). Of note, LRRK2 presented in the Exp_Cluster2, suggesting that the 102 interactors in this cluster exhibited similar expression profiles as LRRK2 across the 15 tissues. One-way ANOVA showed that Exp_Cluster1 and Exp_Cluster3 presented significantly higher expression in the lung and frontal cortex (adjusted p-value < 0.05) (**Figure 14C**). Upon functional enrichment, no GO-BP term was found associated with Exp_Cluster1, while Exp_Cluster 3 was related to "negative regulation of calcium ion export

across plasma membrane" (N = 28/133 GO-BP terms) and "regulation of kinase activity and protein phosphorylation" (N = 34/133 GO-BP terms), suggesting that Exp_Cluster 3 plays an important role in maintaining calcium homeostasis and protein modification. In addition, hierarchical clustering was performed on the 15 tissues based on the similarities of expression profiles of the $LRRK2_{int}$ (**Figure 14D**). The Elbow method suggested a total of 2 tissue groups in the dendrogram (cutting at h = 36) (**Figure 14E**). Therefore, the tree was cut at the height of 36 to obtain the 2 clusters, which contained N = 11 and 4 tissues, respectively. Of note, the 11 brain regions and 4 peripheral tissues were well separated into the 2 groups, suggesting the $LRRK2_{int}$ possesses distinct mRNA expression signatures in the CNS and the periphery.



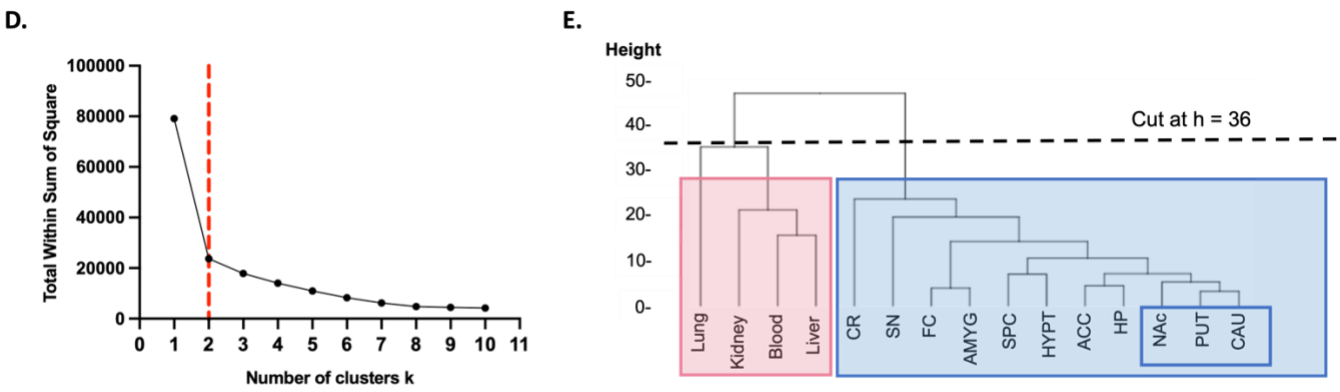


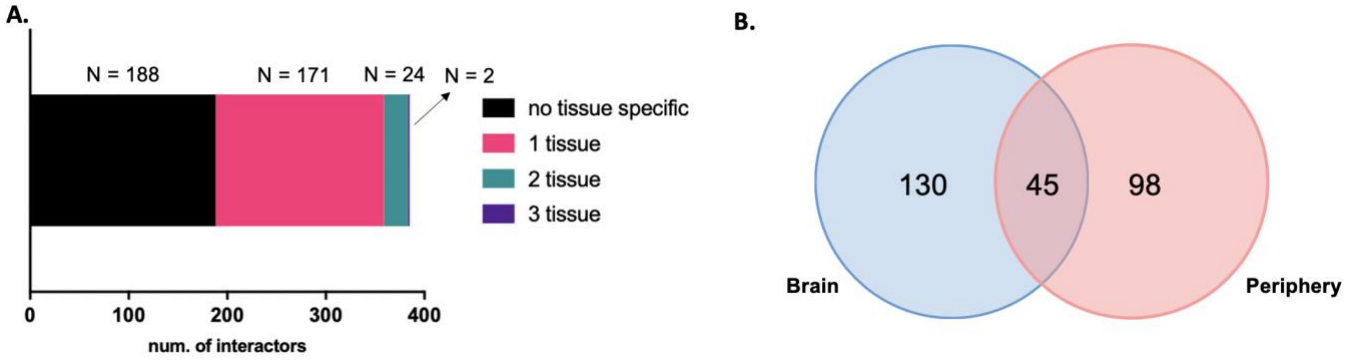
Figure 14. Expression profiles of the LRRK2$_{int}$ in the 15 tissues

A) *The line graph shows the optimal number of clusters (k) among LRRK2 interactors selected via the Elbow method. The elbow of the curve situated at k = 4 (marked with red dash lines), suggesting 4 clusters in the dendrogram;* ***B)*** *The dendrogram shows the hierarchical clustering for LRRK2 interactors based on the mRNA expression patterns in the 15 tissues. The tree was cut at h = 20 to obtain 4 clusters (Exp_Cluster1-4), in which Exp_Cluster2 contains LRRK2;* ***C)*** *The heatmap shows the comparison of mean expression levels of each Exp_Cluster among the 15 tissues. Cell colour represent the mean (log2(rc)) of each Exp_Cluster in a certain tissue. The darker the colour, the higher the expression level. Significant comparisons were marked with *.* ***D)*** *The line graph shows the optimal number of clusters (k) among the 15 tissues selected via the Elbow method. The elbow of the curve situated at k = 2 (marked with red dash lines), suggesting 2 clusters in the dendrogram;* ***E)*** *The dendrogram shows the hierarchical clustering of 15 tissues based on the expression profiles of the LRRK2$_{int}$. By cutting the tree at h = 36, 2 clusters were obtained: 1 contained the 11 brain regions while the other contained 4 peripheral tissues. The 3 striatal regions, namely putamen, nucleus accumbens and caudate was allocated in 1 subgroup (marked in blue rectangular);* ***Abbreviations****: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

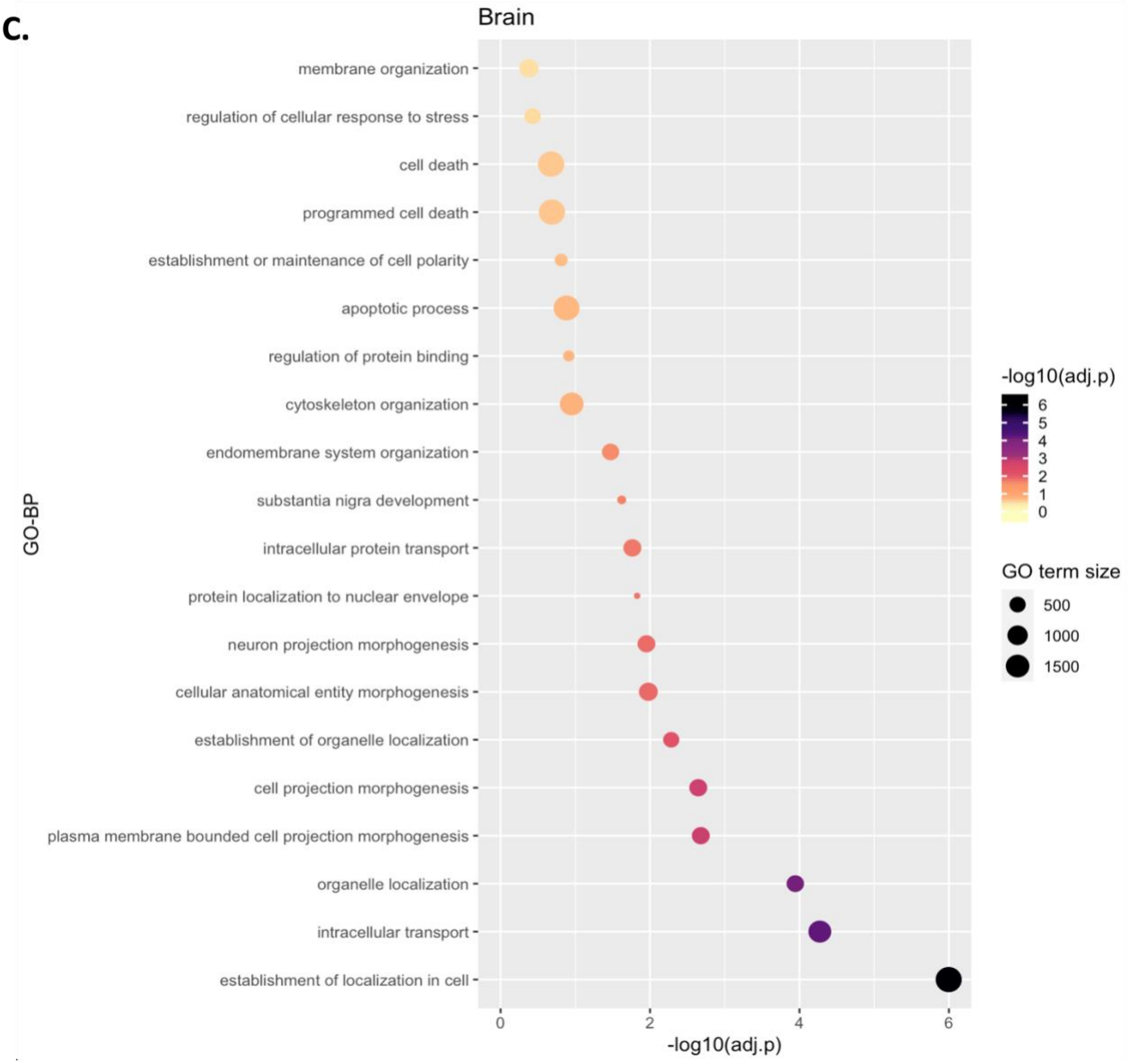
- Pair-wise DEA

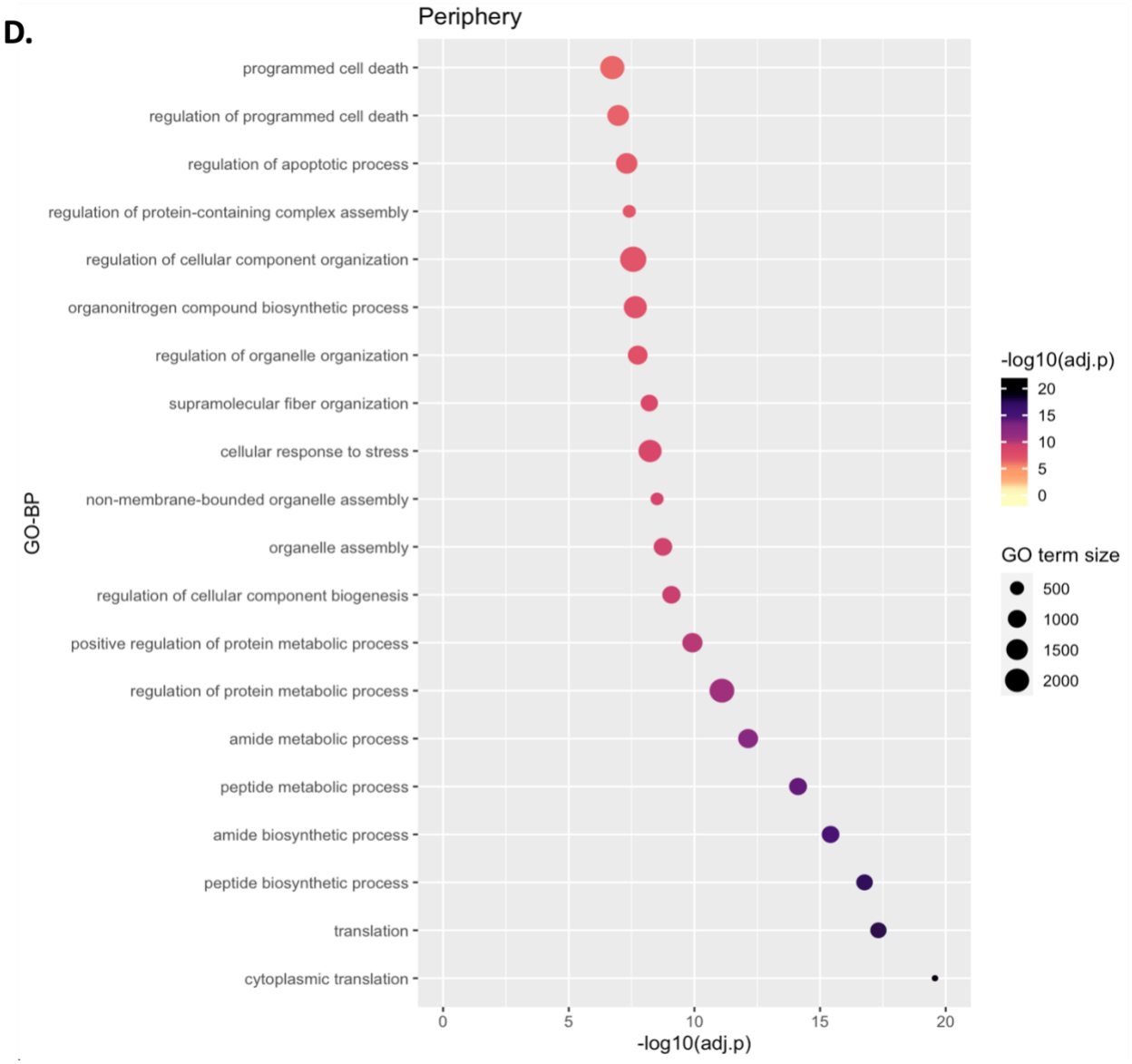
The mRNA levels of each LRRK2 interactor were compared across the 15 tissues via pair-wise DEA. Tissues were scored based on the pairwise comparison results of each interactor (**Table S3**). The results showed that a total of 197/384 (51.3%) interactors presented significant tissue-specificity ($TS^{E}_{I,X} \geq 12$), among which 171 (86.8%) interactors exhibited significantly high expression in only 1/15 tissue; 24 (12.1%) showed high expression in 2/15 tissues while

2 (1.0%) showed high expression in 3/15 tissues, suggesting that the tissue-specific expression pattern varies largely among LRRK2 interactors (**Figure 15A**). A total of 71 and 142 interactors presented significantly high expression levels in the brain and the periphery, respectively (hereby referred as the "brain list" and the "periphery list"), with an overlap of 45 interactors (**Figure 15B**). Functional enrichment analysis showed that the brain list was associated with "cell death", "cell projection" and "intracellular transport", while the peripheral list was related to "cell death", "translation" and "protein metabolism" (**Figure 15C,D**). In specific, among the 15 tissues, a largest number of LRRK2 interactors were highly expressed in the whole blood ($TS^{E}_{I,blood} \geq 12$; N = 142) and cerebellum ($TS^{E}_{I,CR} \geq 12$; N = 92), followed by the frontal cortex and spinal cord c-1 (N = 41 and 39, respectively). A total of 15 and 14 interactors exhibited significantly high expression in anterior cingulate cortex and hypothalamus, respectively, while < 10 proteins were highly expressed in the rest of 9 tissues (**Figure 15E**). These findings suggest that although the expression level of LRRK2 is lower in the brain, some of its interactors are highly expressed in certain brain regions.



C.


Brain
membrane organization
regulation of cellular response to stress
cell death
programmed cell death
establishment or maintenance of cell polarity
apoptotic process
regulation of protein binding
cytoskeleton organization
endomembrane system organization
substantia nigra development
intracellular protein transport
protein localization to nuclear envelope
neuron projection morphogenesis
cellular anatomical entity morphogenesis
establishment of organelle localization
cell projection morphogenesis
plasma membrane bounded cell projection morphogenesis
organelle localization
intracellular transport
establishment of localization in cell
GO-BP
0
2
4
6
-log10(adj.p)
-log10(adj.p)
6
5
4
3
2
1
0
GO term size
500
1000
1500


D.
Periphery
programmed cell death
regulation of programmed cell death
regulation of apoptotic process
regulation of protein-containing complex assembly
regulation of cellular component organization
organonitrogen compound biosynthetic process
regulation of organelle organization
supramolecular fiber organization
cellular response to stress
non-membrane-bounded organelle assembly
organelle assembly
regulation of cellular component biogenesis
positive regulation of protein metabolic process
regulation of protein metabolic process
amide metabolic process
peptide metabolic process
amide biosynthetic process
peptide biosynthetic process
translation
cytoplasmic translation
GO-BP
0
5
10
15
20
-log10(adj.p)
-log10(adj.p)
20
15
10
5
0
GO term size
500
1000
1500
2000

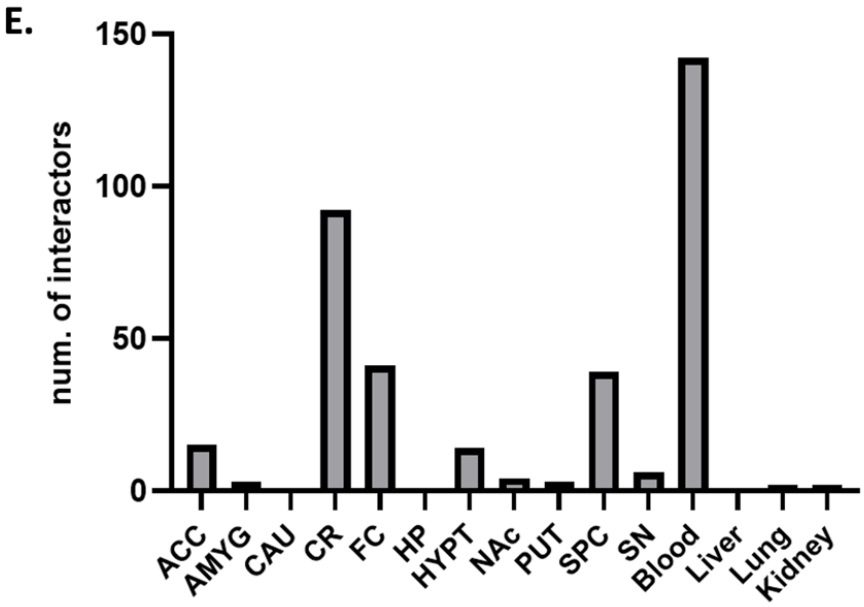
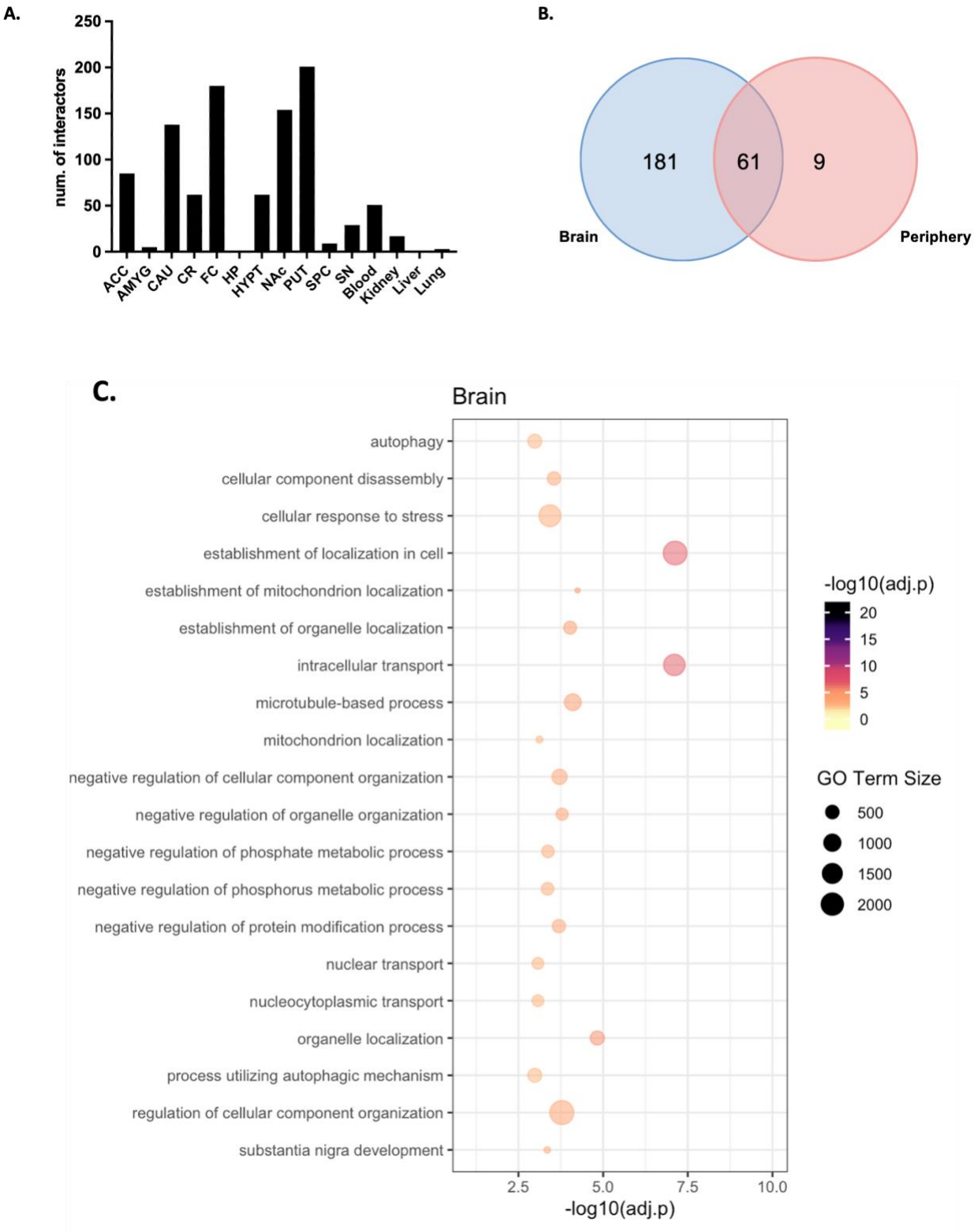

E.
num. of interactors
150
100
50
0
ACC
AMYG
CAU
CR
FC
HP
HYPT
NAc
PUT
SPC
SN
Blood
Liver
Lung
Kidney

Figure 15. LRRK2 interactors with highly tissue-specific expression pattern

A) *The bar chart shows the percentages of LRRK2 interactors that presented significantly high expression levels in i.) 1 tissue, ii) 2 tissues and iii) 3 tissues;* ***B)*** *The Veen graph shows the overlap between highly-expressed LRRK2 interactors in the 11 brain regions ("brain", marked in blue) and 4 peripheral tissues ("periphery", marked in red);* ***C,D)*** *The bubble graphs show the top 20 GO-BP terms (ordered by adjusted p-values) returned for the highly-expressed LRRK2 interactors in the brain and the periphery. Bubble size represent the term size (i.e., the number of genes annotated with a certain functional term), the larger the bubble, the larger the term size. Bubble colour represents the -log10(adjusted-p), the darker the colour, the lower the adjusted-p and thereby the higher the significance of enrichment for a certain functional term.* ***E)*** *The bar graph shows the distribution of LRRK2 interactors with significantly higher expression level ($TS^{E}_{I,T} \geq 12$) in the 15 tissues.* ***Abbreviations****: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

- Co-expression between LRRK2 and its interactors in 15 tissues

Pearson's correlation test showed that a total of 201, 180, 154 and 138 interactors presented high co-expression level (Pearson's coefficient > 0.7) with LRRK2 in the putamen, frontal cortex, nucleus accumbens, and caudate, respectively, followed by anterior cingulate cortex (N = 85), cerebellum (N = 62), hypothalamus (N = 62) and blood (N = 51) (**Figure 16A, Table S4**). The lowest LRRK2:interactor co-expression was observed in the *substantia nigra*, amygdala, lung, liver and hippocampus (N < 10). Functional enrichment analysis was performed on the interactors that highly co-expressed with LRRK2 in the brain and the periphery. There was an overlap of 61 interactors between the 2 query lists (**Figure 16B**). The enrichment results showed that interactors presenting high co-expression with LRRK2 in the brain were associated with autophagy, mitochondrial localisation and negative regulation of protein metabolism, while the interactors co-expressed with LRRK2 in the periphery were related to actin cytoskeleton organisation, apoptosis and positive regulation of protein metabolism (**Figure 16C-D**).


A.
num. of interactors
250
200
150
100
50
0
ACC
AMYG
CAU
CR
FC
HP
HYPT
NAc
PUT
SPC
SN
Blood
Kidney
Liver
Lung
B.
181
61
9
Brain
Periphery
C.
Brain
autophagy
cellular component disassembly
cellular response to stress
establishment of localization in cell
establishment of mitochondrion localization
establishment of organelle localization
intracellular transport
microtubule-based process
mitochondrion localization
negative regulation of cellular component organization
negative regulation of organelle organization
negative regulation of phosphate metabolic process
negative regulation of phosphorus metabolic process
negative regulation of protein modification process
nuclear transport
nucleocytoplasmic transport
organelle localization
process utilizing autophagic mechanism
regulation of cellular component organization
substantia nigra development
2.5
5.0
7.5
10.0
-log10(adj.p)
-log10(adj.p)
20
15
10
5
0
GO Term Size
500
1000
1500
2000

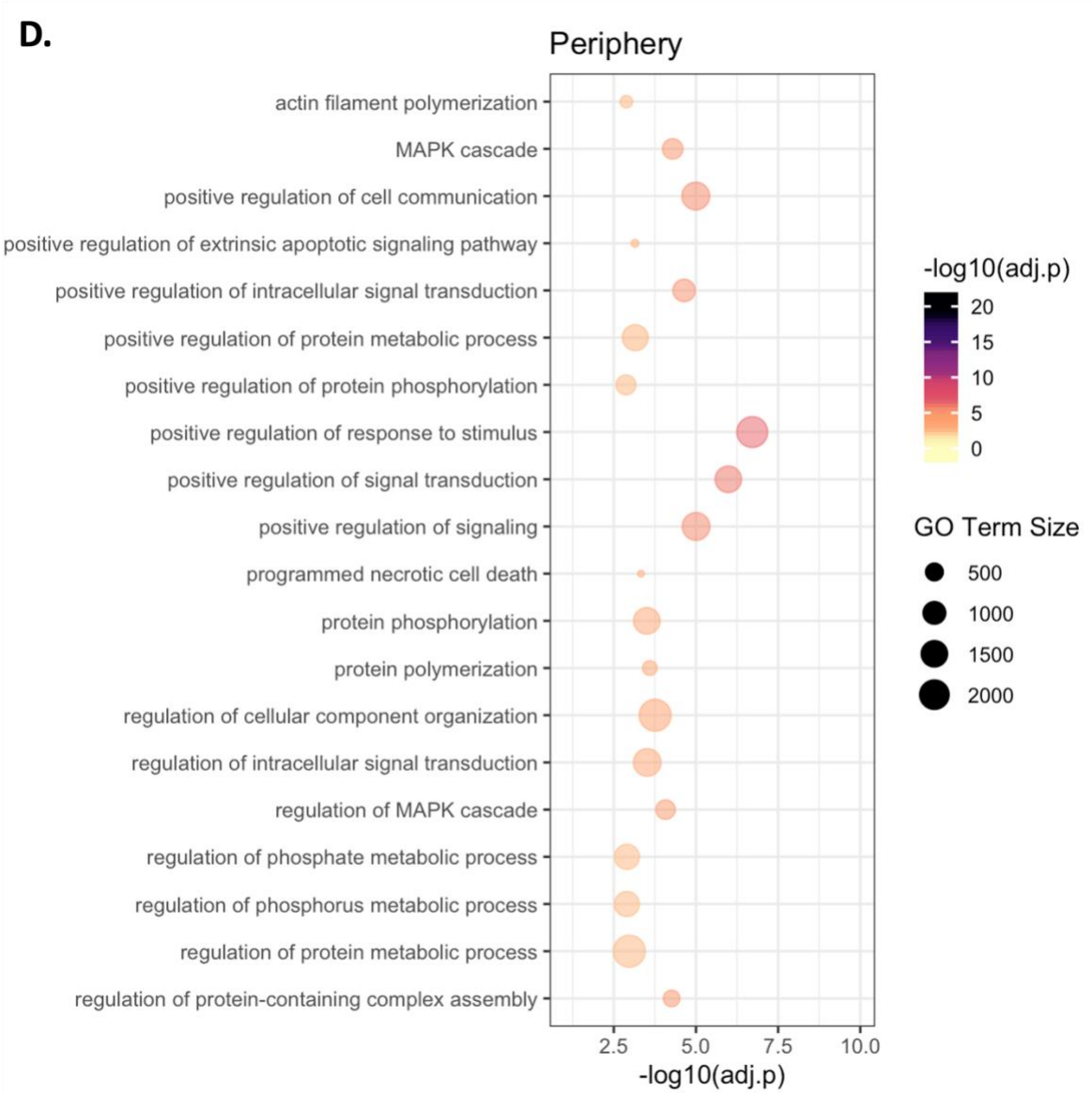
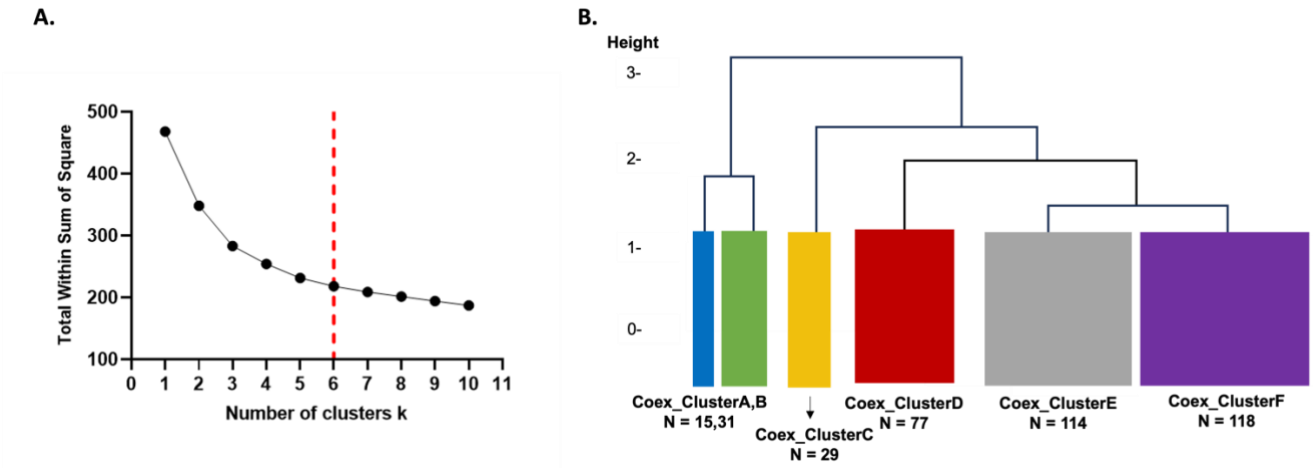


Figure 16. Tissue-specific co-expression profiles between LRRK2 and its interactors

A) *The bar graph shows the number of interactors with high co-expression level with LRRK2 (Pearson's coefficient > 0.7) in the 15 tissues;* ***B)*** *The Veen graph shows the co-expressed interactors queried for functional enrichment analysis for the brain and the periphery. In order to keep similar query size, interactors with co-expression coefficients > 0.8 in brain regions and with co-expression coefficients > 0.7 in peripheral tissues were included in the enrichment analysis;* ***C,D)*** *The bubble graphs show the top 20 GO-BP terms (ordered by adjusted p-values) returned for the highly co-expressed interactors with LRRK2 in the brain and the periphery. Bubble size represent the term size (i.e., the number of genes annotated with a certain functional term), the larger the bubble, the larger the term size. Bubble colour represents the -log10(adjusted-p), the darker the colour, the lower the adjusted-p and thereby the higher the significance of enrichment for a certain functional term.* ***Abbreviations****: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP:*

hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.

Hierarchical clustering was then performed on the matrix of LRRK2:interactor co-expression coefficients to identify clusters of interactors that exhibited similar co-expression patterns with LRRK2 across the 15 tissues. The Elbow method suggested a total of 6 clusters among the interactors (**Figure 17A**), therefore, the dendrogram was cut at the height of 1.2 to obtain Co-ex_ClusterA-F (**Figure 17B**). For each of the 6 clusters, the average LRRK2:interactor co-expression levels were compared across the 15 tissues. The result showed that: Co-ex_ClusterA and B presented higher co-expression level with LRRK2 in blood ($TS^{C}_{A,\,\text{blood}} = 8$) and kidney cortex $TS^{C}_{A,\,\text{kidney}} = 7$, respectively, while Co-ex_ClusterC-F presented higher co-expression level in brain regions especially in caudate, frontal cortex, nucleus accumbens and putamen, suggesting a tissue specific profile of LRRK2:interactor co-expression (**Figure 17C**). In addition, hierarchical clustering was performed on the 15 tissues based on the similarity of LRRK2:interactor co-expression profiles. The Elbow method showed 7 groups of tissues, therefore the dendrogram was cut at the height of 2.6, obtaining 3 groups of brain regions: Group 1 comprises of the *substantia nigra*, hippocampus and amygdala; Group 2 contains the spinal cord c-1, cerebellum, anterior cingulate cortex and hypothalamus, while Group 3 includes frontal cortex, putamen, caudate and nucleus accumbens (**Figure 17D-E**). No group was identified among the 4 peripheral tissues.



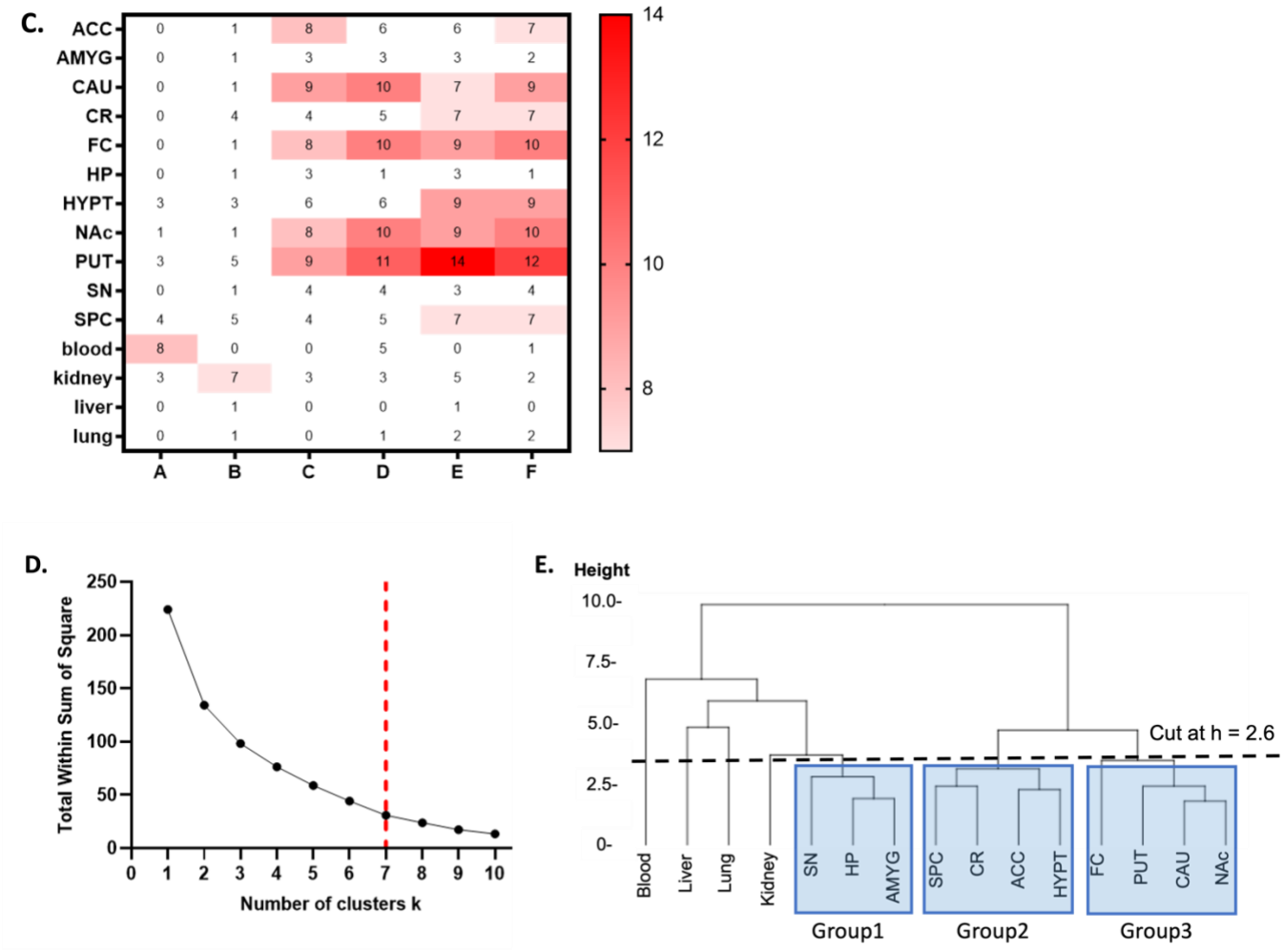


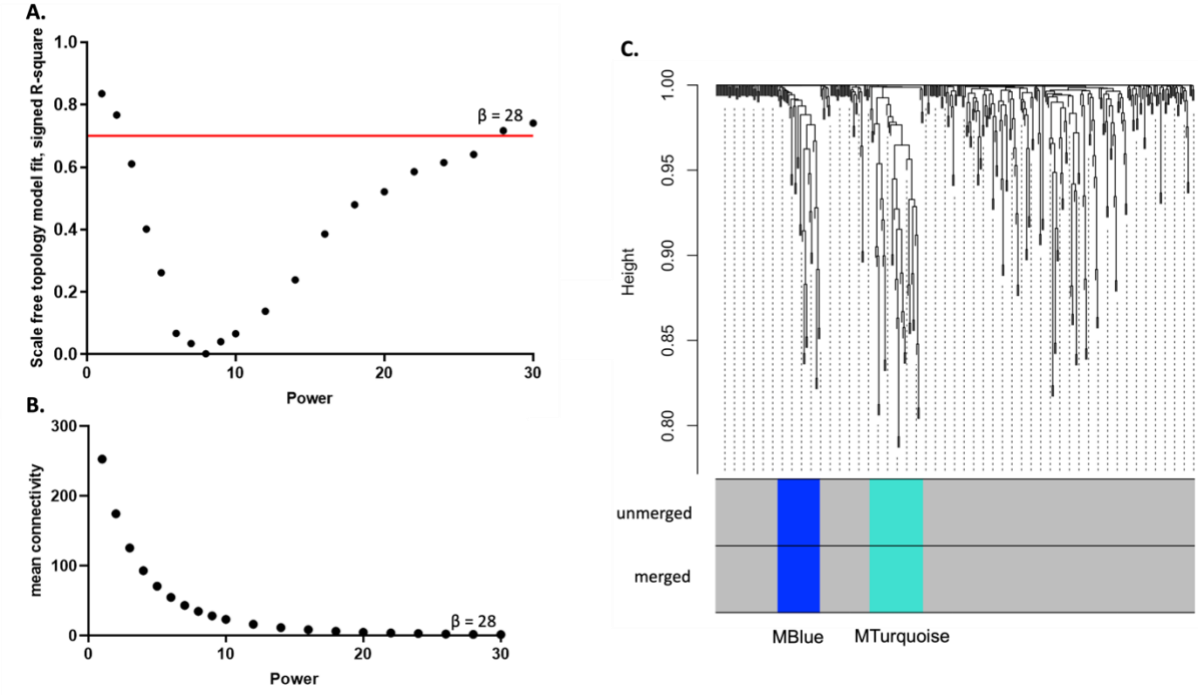
Figure 17. Hierarchical clustering on LRRK2:interactor co-expression feature

A) *The line graph shows the optimal number of clusters (k) among LRRK2 interactors selected via the Elbow method based on the co-expression levels with LRRK2 in 15 different tissues. The elbow of the curve situated at k = 6 (marked with red dash lines), suggesting 6 clusters in the dendrogram* ***B)*** *The dendrogram shows the hierarchical clustering for LRRK2 interactors based on their co-expression levels with LRRK2 in 15 different tissues. The tree was cut at h = 1.2 to obtain 6 clusters (Co-ex_ClusterA-F);* ***C)*** *The heatmap shows the 15 tissues (rows in the heatmap) were scored based on the comparison of LRRK2:interactor co-expression levels of each Co-ex_Cluster (columns in the heatmap). Scores were represented by the numbers in the cells. The higher the score is, the higher the co-expression with LRRK2 for a given Co-ex_Cluster in a given tissue (adjusted p-value < 0.05);* ***D)*** *The line graph shows the optimal number of clusters (k) among the 15 tissues selected via the Elbow method based on the overall LRRK2:interactor co-expression features. The elbow of the curve situated at k = 7 (marked with red dash lines), suggesting 7 clusters in the dendrogram;* ***E)*** *The dendrogram shows the hierarchical clustering for 15 tissues based on the overall LRRK2:interactor co-expression features. The tree was cut at h = 2.6 to obtain 7 clusters, among which 3 clusters contains ≥ 3 tissues (highlighted in rectangles).*

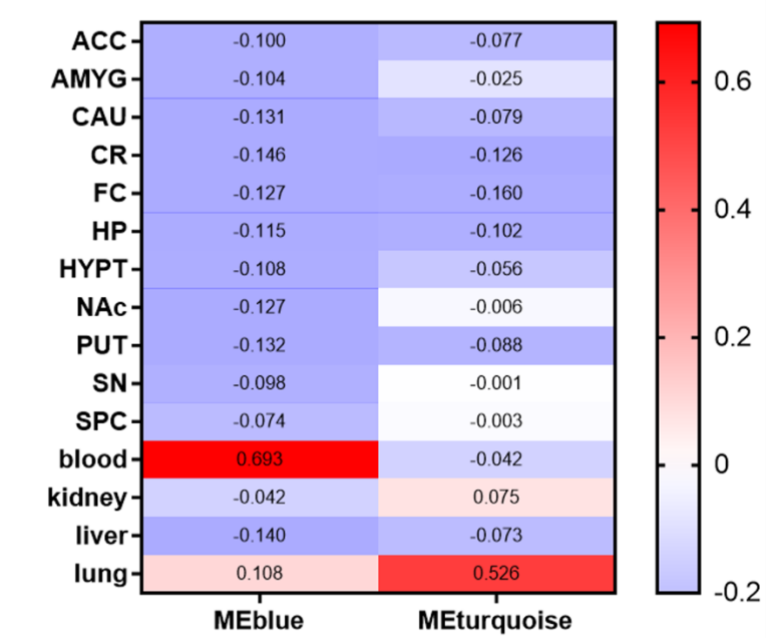
***Abbreviations**: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

- Weight Gene Co-expression Network Analysis (WGCNA)

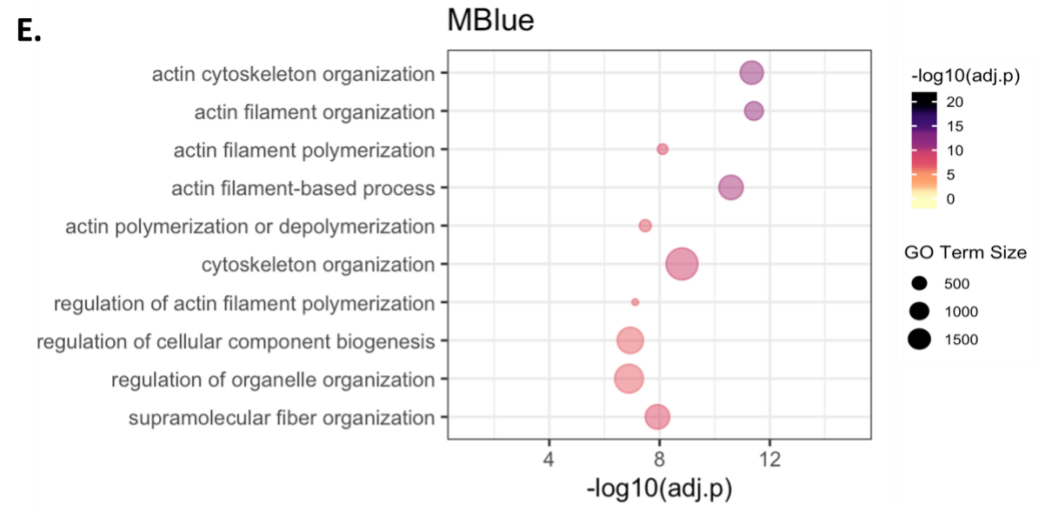
WGCNA was performed on the "Normalised Tissue Matrix" to identify co-expression modules of LRRK2 interactors across the 15 tissues. A power of β = 28 was selected as the soft-threshold to ensure a scale-free network (SFT R-square = 0.72) as well as to maintain a low mean connectivity in the signed co-expression network (mean connectivity = 2.17) (**Figure 18A**- **B**). After merging submodules at the height of 0.20, a total of 2 co-expression modules were identified within LRRK2 interactors ("MBlue", N = 34; "MTurquoise", N = 43) (**Figure 18C**). Module-Trait correlation analysis showed that ME of" MBlue" was highly expressed in blood (Pearson's coefficient = 0.69), while ME of "MTurquoise" was highly expressed in lung (Pearson's coefficient = 0.53), suggesting that "MBlue" possessed high expression level in blood while "MTurquoise" showed high expression level in lung (**Figure 18D**). Functional enrichment analysis showed that "MBlue" was associated with actin cytoskeleton organisation and cell morphology, while "MTurquoise" was related to ribosomal function and gene translation (**Figure 18E-F**).



D.


ACC
AMYG
CAU
CR
FC
HP
HYPT
NAc
PUT
SN
SPC
blood
kidney
liver
lung
-0.100
-0.077
-0.104
-0.025
-0.131
-0.079
-0.146
-0.126
-0.127
-0.160
-0.115
-0.102
-0.108
-0.056
-0.127
-0.006
-0.132
-0.088
-0.098
-0.001
-0.074
-0.003
0.693
-0.042
-0.042
0.075
-0.140
-0.073
0.108
0.526
MEblue
MEturquoise
0.6
0.4
0.2
0
-0.2

E.


MBlue
actin cytoskeleton organization
actin filament organization
actin filament polymerization
actin filament-based process
actin polymerization or depolymerization
cytoskeleton organization
regulation of actin filament polymerization
regulation of cellular component biogenesis
regulation of organelle organization
supramolecular fiber organization
4
8
12
-log10(adj.p)
-log10(adj.p)
20
15
10
5
0
GO Term Size
500
1000
1500

F.

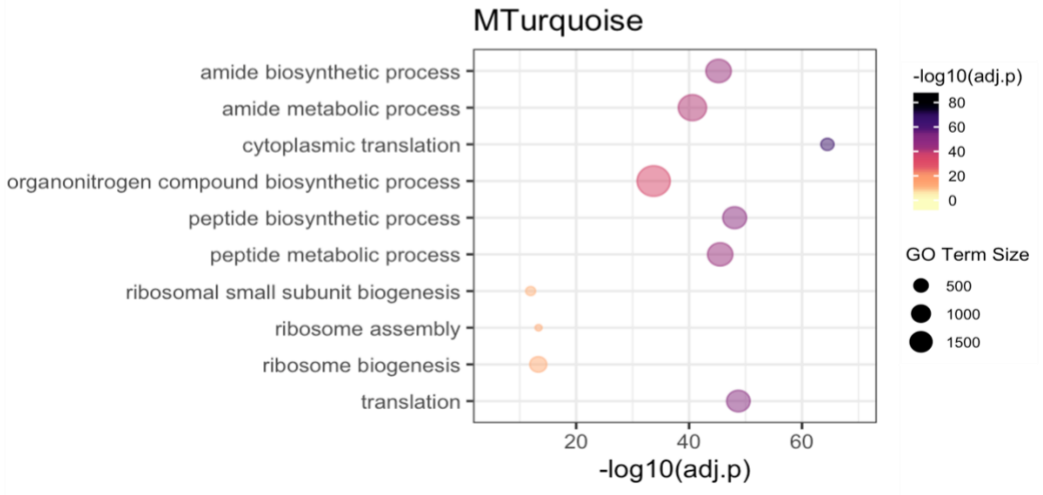

MTurquoise
amide biosynthetic process
amide metabolic process
cytoplasmic translation
organonitrogen compound biosynthetic process
peptide biosynthetic process
peptide metabolic process
ribosomal small subunit biogenesis
ribosome assembly
ribosome biogenesis
translation
20
40
60
-log10(adj.p)
-log10(adj.p)
80
60
40
20
0
GO Term Size
500
1000
1500

Figure 18. WGCNA on $LRRK2_{int}$ across the 15 tissues

A,B) *The scatter plots show the selection of the proper power β for WGCNA. The optimal power β = 28 was selected on 1) achieving a high (> 0.7) scale-free topology fit index (showed in A) and low mean connectivity in the network (showed in B);* ***C)*** *The dendrogram shows hierarchical clustering of LRRK2 interactors based on the dissimilarity matrix generated by WGCNA. A total of 2 modules of interactors were obtained ("MBlue" and "MTurquoise") in both preliminary (unmerged) and final (merged) module detection process;* ***D)*** *The heatmap shows the results of Module-Tissue correlation analysis, in which module eigengene of MBlue (MEblue) was positively correlated with blood while the module eigengene of MTurquoise (MEturquoise) was positively correlated with lung;* ***E,F)*** *The bubble graphs show the top 10 GO-BP terms (ordered by adjusted p-values) returned for the 2 co-expression modules obtained for WGCNA. Bubble size represent the term size (i.e., the number of genes annotated with a certain functional term), the larger the bubble, the larger the term size. Bubble colour represents the -log10(adjusted-p), the darker the colour, the lower the adjusted-p and thereby the higher the significance of enrichment for a certain functional term.*

- Tissue-specific weighted topological clusters of the $LRRK2_{net}$

The subnetwork for Topological Cluster A consisted of 45 nodes and 115 edges, which was mainly enriched for ribosomal biosynthesis (**Figure 19A**). Node weight analysis found no significant differences across the 15 tissues apart from the frontal cortex and cerebellum, in which the median node weight was significantly lower than 4 and 5 of other tissues, respectively, suggesting that the median expression level of LRRK2 interactors in Cluster 1 was lower in these 2 brain regions (**Figure 20A**). No significant difference was found in edge weight comparison among the 15 tissues, suggesting the co-expression levels among the interactors in this topological cluster were similar in the brain and the periphery (**Figure 20C**). Of note, there is a high-density unit of interactors consisting of 16 ribosomal proteins (the "ribosomal unit"). These proteins also exhibited high co-expressions in all of the tissues analysed (median Pearson's coefficient > 0.88). Median node weight of the ribosomal unit is highest in amygdala and blood, followed by substantia nigra and spinal cord, but lower in cerebellum and frontal cortex (**Figure 20B**). These findings suggest that ribosomal function is maintained by a unit of highly connected and highly co-expressed LRRK2 interactors existing in both of the brain and the periphery. Of note, the hub protein of Cluster 1 (CDK2) presented the lowest node weight among all the 45 nodes in 8/15 of the tissues. It presented low co-expression levels with its first neighbours (average Pearson's coefficients < 0.5) in 14/15

tissues apart from liver (average Pearson's coefficients = 0.6), suggesting that CDK2 may function as a regulator of ribosomal synthesis in these tissues rather than a direct participator in the process.

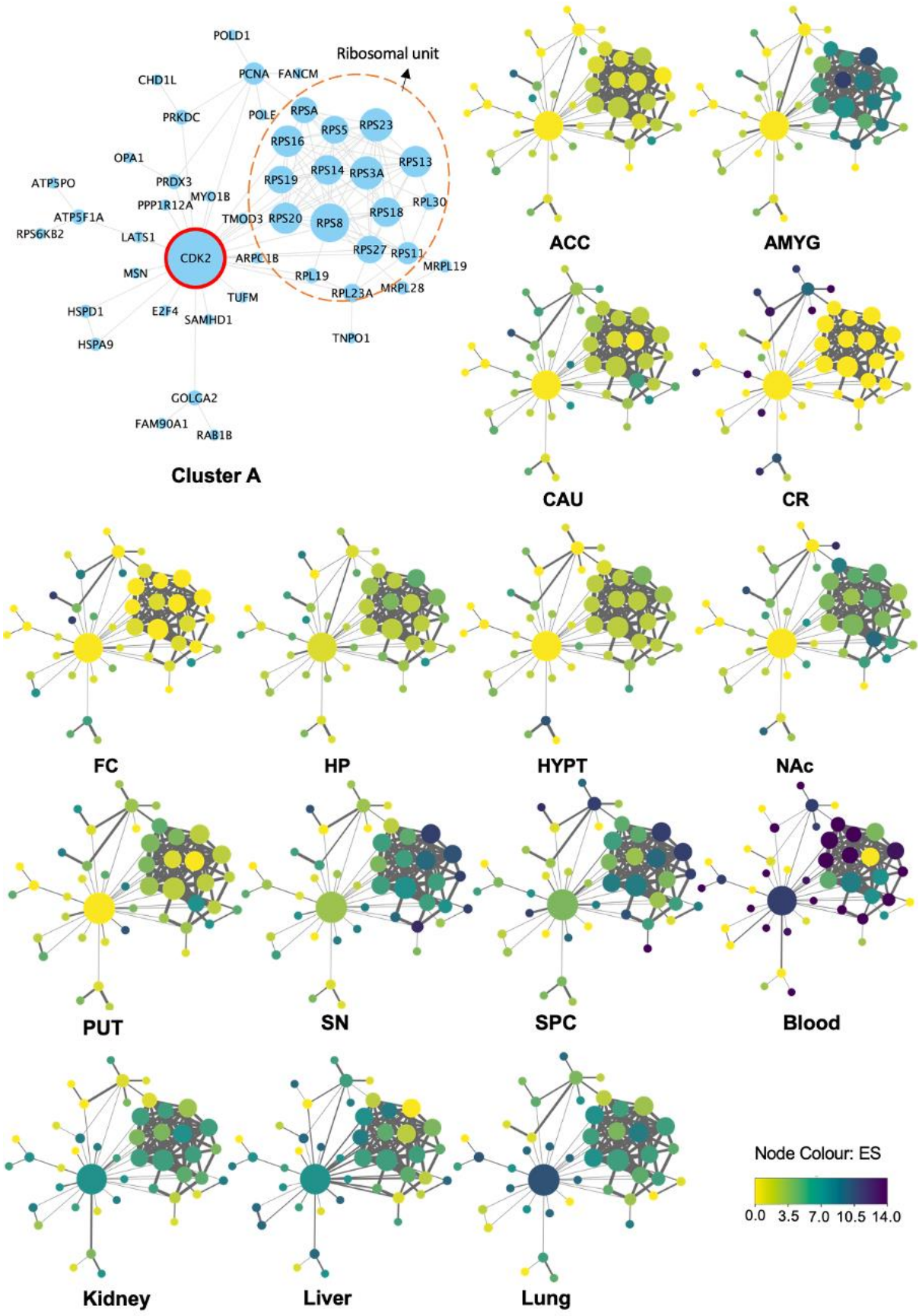


Figure 19. Expression profiles of Topological Cluster A in the 15 tissues

The network graphs show the tissue-specific expression and co-expression profiles of interactors in Cluster A, in which LRRK2 interactors are represented as nodes while PPIs were represented as edges. Node size refers to degree of nodes, i.e., the number of edges connected to each node. The large the node, the higher the degree. Node colour refers to the expression level of each interactor in different tissues. The higher the expression level, the darker the node. Edge width refers to co-expression levels between each pair of connected proteins. The higher the co-expression level, the thicker the edge. A high-density unit of 16 ribosomal proteins was found in Cluster 1 (circled).

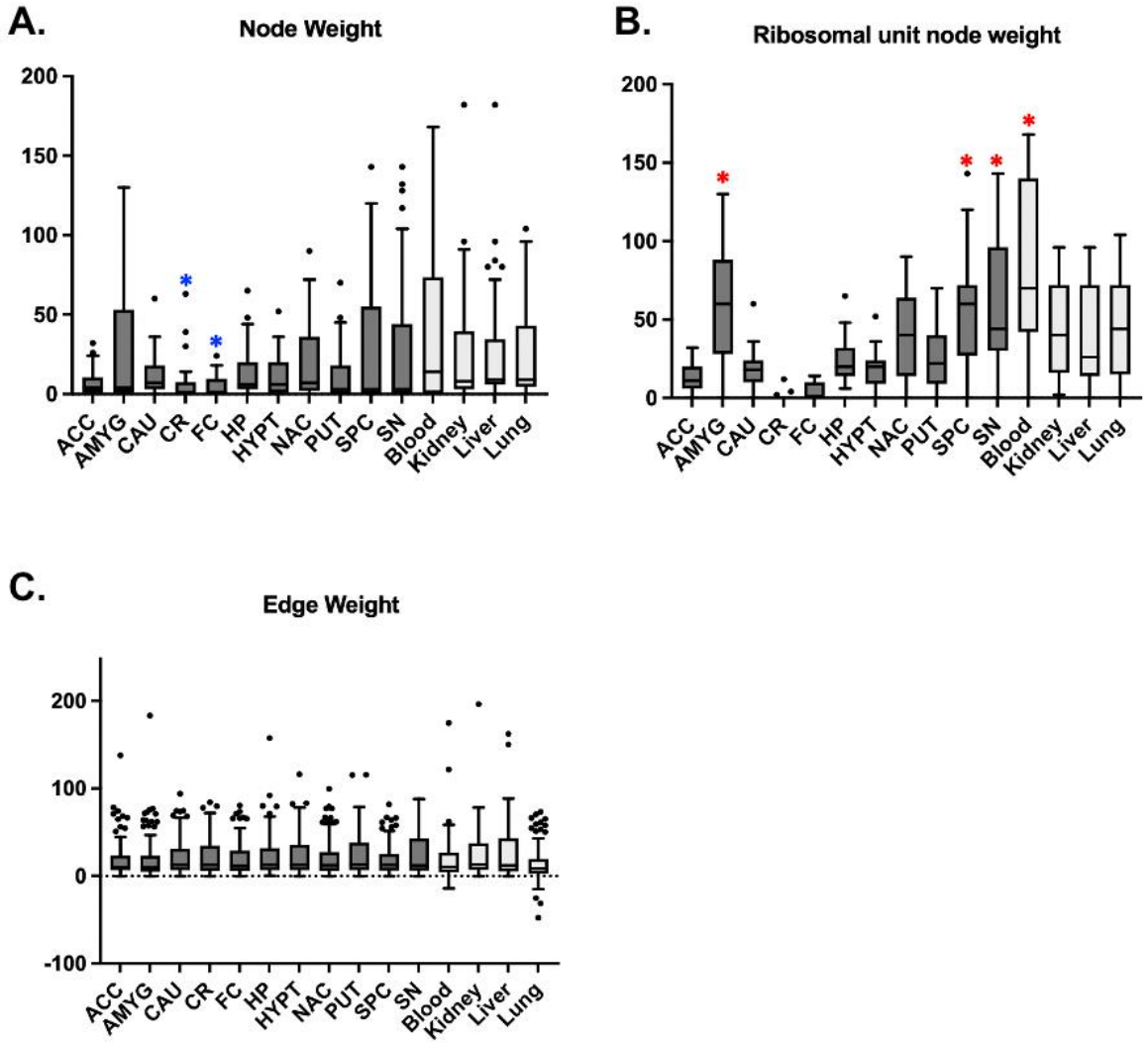


Figure 20. Tissue-specific weighted network analysis on Topological Cluster A

***A)** The box plot shows the pairwise comparison of median node weight of Cluster 1 across the 15 tissues, in which cerebellum and frontal cortex presented the lowest level, with median node weight significantly lower than 6 and 4 of other tissues of analysis (adjusted $p < 0.05$, *). **B)** The box plot shows the pairwise comparison of median node weight of the ribosomal unit in Cluster 1 across the 15 tissues, in which cerebellum and frontal cortex presented the lowest level, while blood, substantia nigra,*

*amygdala and spinal cord presented the highest level (adjusted p < 0.05, *).* ***C)*** *The box plot shows the pairwise comparison of median edge weight between each pair of connected proteins. No significant difference was found;* ***Abbreviations****: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

Topological Cluster B contained 23 interactors and 42 PPIs, associated with cell death, protein metabolism and response to stress (**Figure 21**). There was no significant difference in terms of the median node weight or median edge weight among the 15 tissue-specific subnetworks (apart from putamen vs. liver, adjusted $p < 0.05$) (**Figure 22A,B**). Of note, the hub protein TRAF2 presented high node weight and moderate co-expression levels with its first neighbours in the networks of cerebellum, caudate, putamen, and nucleus accumbens (Pearson's coefficients > 0.55), suggesting a potentially higher activity of TRAF2 in mediating the function of Cluster 2 in these brain regions.

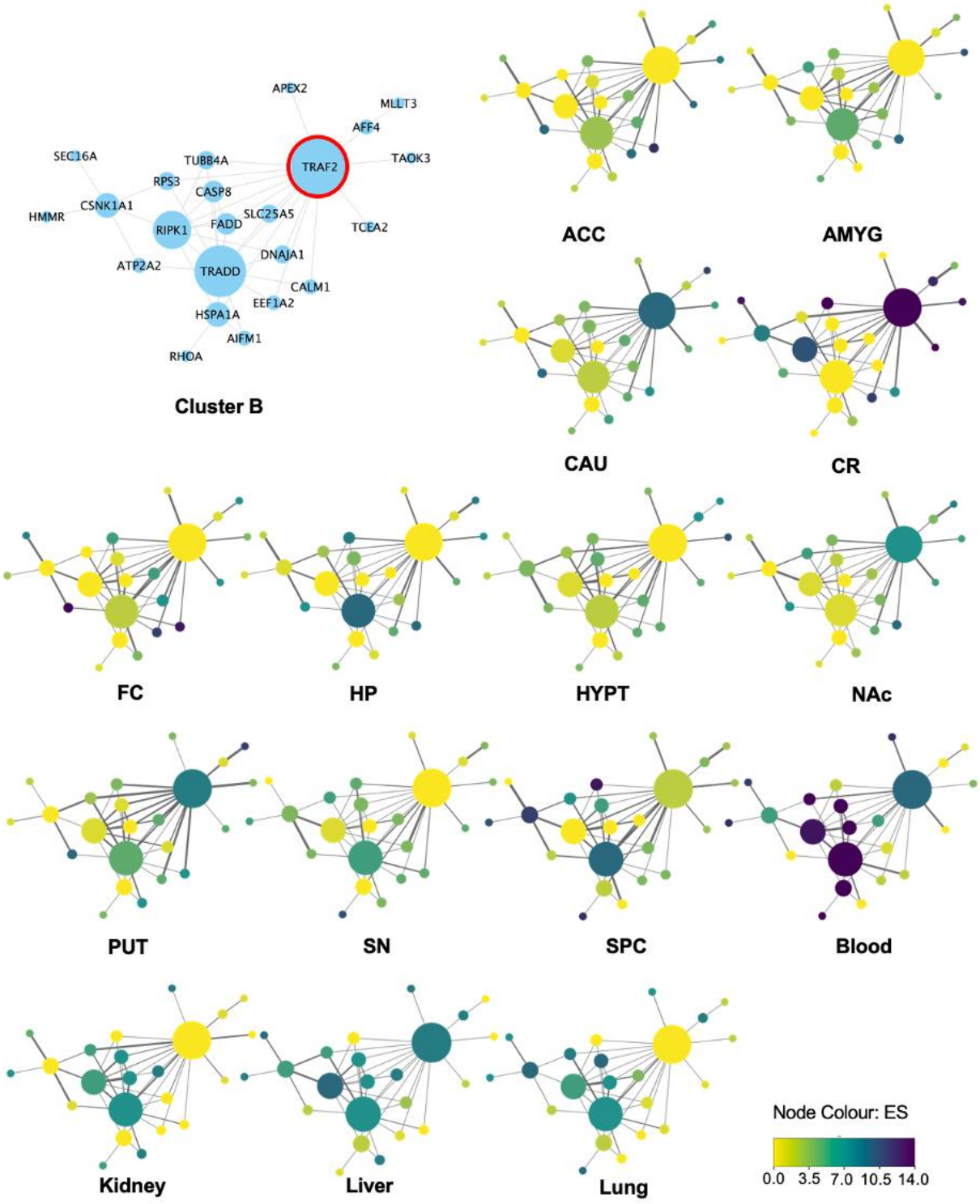


Figure 21. Expression profiles of Topological Cluster B in the 15 tissues

The network graphs show the tissue-specific expression and co-expression profiles of interactors in Cluster B, in which LRRK2 interactors are represented as nodes while PPIs were represented as edges. Node size refers to degree of nodes, i.e., the number of edges connected to each node. The large the node, the higher the degree. Node colour refers to the expression level of each interactor in different

tissues. The higher the expression level, the darker the node. Edge width refers to co-expression levels between each pair of connected proteins. The higher the co-expression level, the thicker the edge.

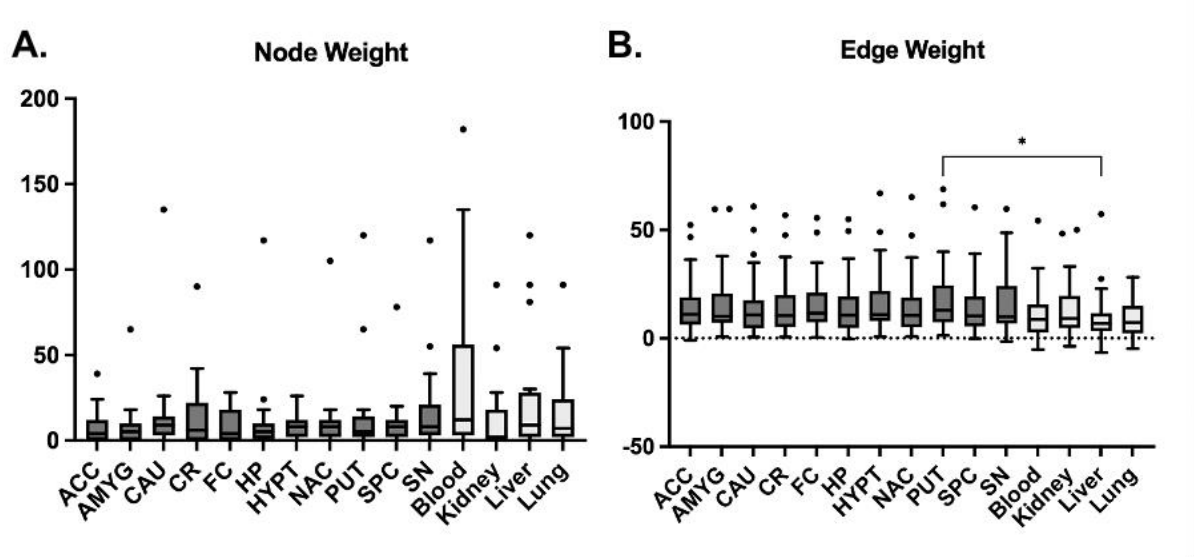


Figure 22. Tissue-specific weighted network analysis on Topological Cluster B

***A)** The box plot shows the pairwise comparison of median node weight of Cluster B across the 15 tissues.* ***B)** The box plot shows the pairwise comparison of median edge weight between each pair of connected proteins. Significant comparisons was found between putamen and liver (adjusted $p < 0.05$); Abbreviations: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

Topological Cluster C was associated with vesicular transport, cytoskeleton organisation and translation, consisting of a total of 41 LRRK2 interactors and 104 PPIs (**Figure 23**). Weighted network analysis found no significant difference in median node weight across the 15 tissues. Lowest median edge weight was found in the subnetwork of blood as compared to all the other tissues except for liver, suggesting that the connection of the LRRK2 interactors in Cluster C possess weaker collaboration in the 2 tissues (adjusted $p < 0.05$, **Figure 24A,B**). There are 3 hub proteins in Cluster C: IQGAP1, CAP2A2 and DBN1, among which CAP2A2 exhibited the highest co-expression levels with its first neighbour in the network across the 15 tissues (average Pearson's coefficient = 0.69, t-test $p < 0.05$), suggesting it may function as the primary participator in the cytoskeleton dynamics mediated by Cluster C, while the other 2 hub proteins may function as regulators.

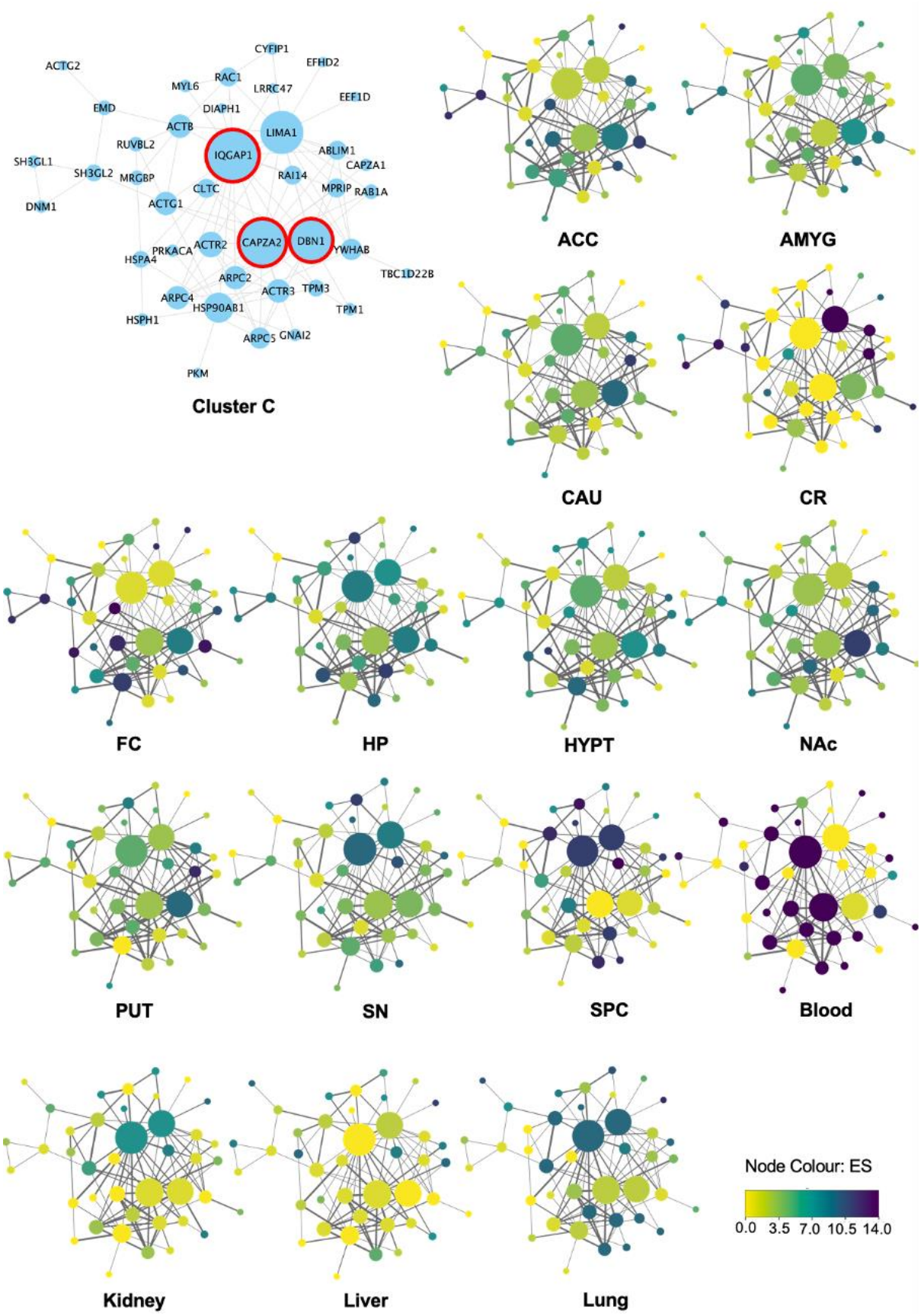

CYFIP1
EFHD2
ACTC2
RAC1
MYL6
LRRC47
EEF1D
EMD
DIAPH1
ACTB
LIMA1
RUVBL2
IQGAP1
ABLIM1
SH3GL1
CAPZA1
SH3GL2
MRGBP
RAI14
CLTC
MPRIP
RAB1A
DNM1
ACTG1
CAPZA2
DBN1
PRKACA
ACTR2
YWHAB
HSPA4
TBC1D22B
ARPC2
ACTR3
TPM3
ARPC4
HSP90AB1
TPM1
HSPH1
ARPC5
GNAI2
PKM
Cluster C
ACC
AMYG
CAU
CR
FC
HP
HYPT
NAc
PUT
SN
SPC
Blood
Kidney
Liver
Lung
Node Colour: ES
0.0 3.5 7.0 10.5 14.0

Figure 23. Expression profiles of Topological Cluster C in the 15 tissues

The network graphs show the tissue-specific expression and co-expression profiles of interactors in Cluster C, in which LRRK2 interactors are represented as nodes while PPIs were represented as edges. Node size refers to degree of nodes, i.e., the number of edges connected to each node. The large the node, the higher the degree. Node colour refers to the expression level of each interactor in different tissues. The higher the expression level, the darker the node. Edge width refers to co-expression levels between each pair of connected proteins. The higher the co-expression level, the thicker the edge.

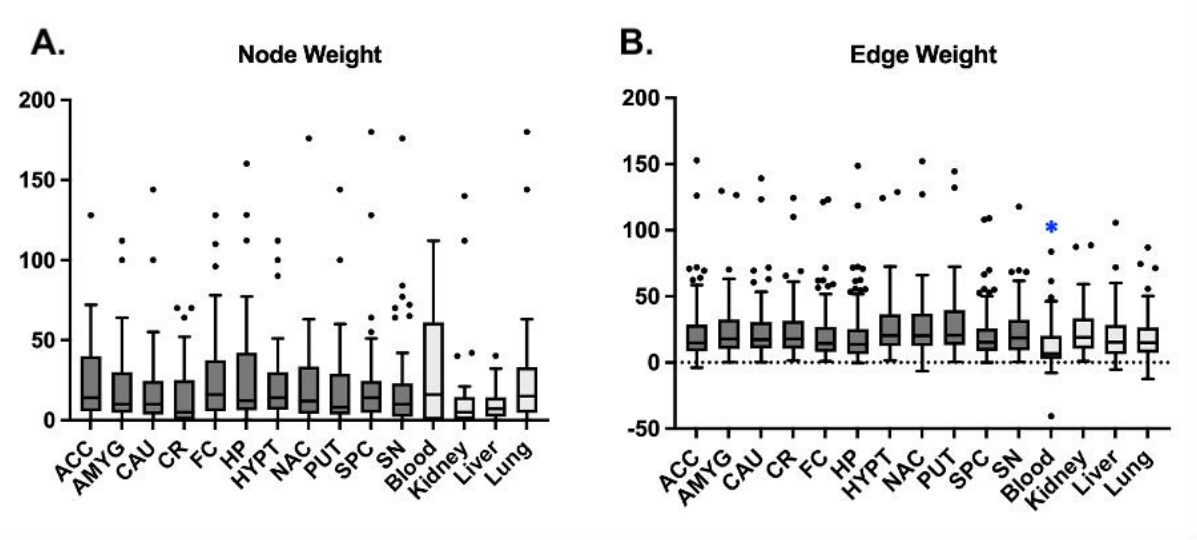


Figure 24. Tissue-specific weighted network analysis on Topological Cluster C

***A)** The box plot shows the pairwise comparison of median node weight of Cluster C across the 15 tissues. No significant difference was observed; **B)** The box plot shows the pairwise comparison of median edge weight between each pair of connected proteins. **Abbreviations**: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

Topological Cluster D was comprised of 39 LRRK2 interactors connected by 63 PPIs, which was enriched for protein metabolism, protein localisation and translation (**Figure 25**). No significant difference was found in the comparison of average node weight across the 15 tissues, apart from blood vs. caudate (5.60 vs. 2.72, adjusted $p < 0.05$, **Figure 26A**). In terms of the average edge weight, higher values were observed in the subnetworks of putamen, hypothalamus, substantia nigra and nucleus accumbens as compared to other brain regions but with no statistical significance (adjusted $p > 0.05$, **Figure 26B**). Subnetworks of blood and liver presented the lowest edge weight as compared to the 4 above-mentioned brain regions

(adjusted $p < 0.05$). These findings suggest that although the mean expression level of Cluster D was similar across brain regions and peripheral tissues, the co-expression behaviours among these interactors are higher in the brain regions.

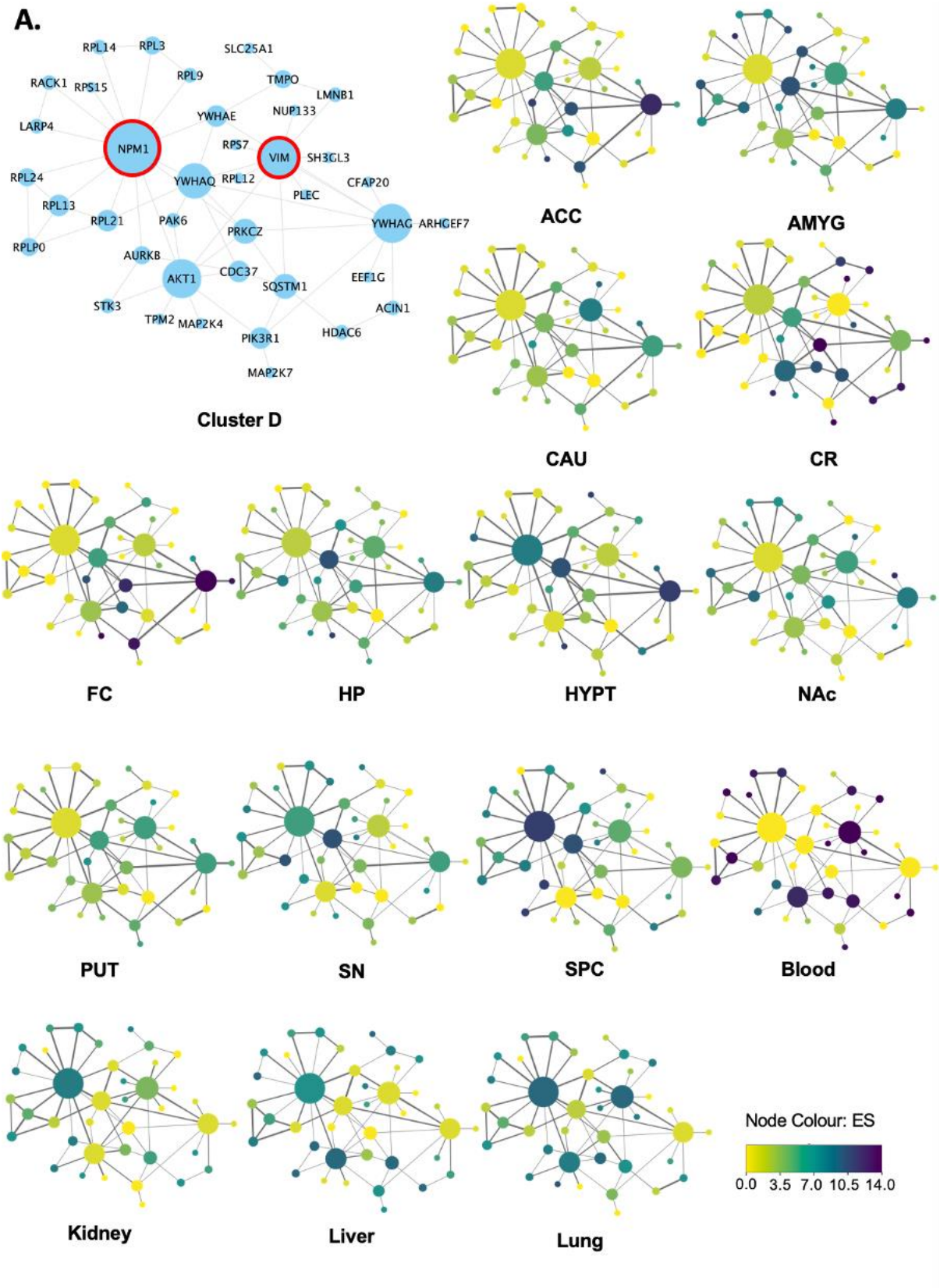


Figure 25. Expression profiles of Topological Cluster D in the 15 tissues

The network graphs show the tissue-specific expression and co-expression profiles of interactors in Cluster D, in which LRRK2 interactors are represented as nodes while PPIs were represented as edges. Node size refers to degree of nodes, i.e., the number of edges connected to each node. The large the node, the higher the degree. Node colour refers to the expression level of each interactor in different tissues. The higher the expression level, the darker the node. Edge width refers to co-expression levels between each pair of connected proteins. The higher the co-expression level, the thicker the edge.

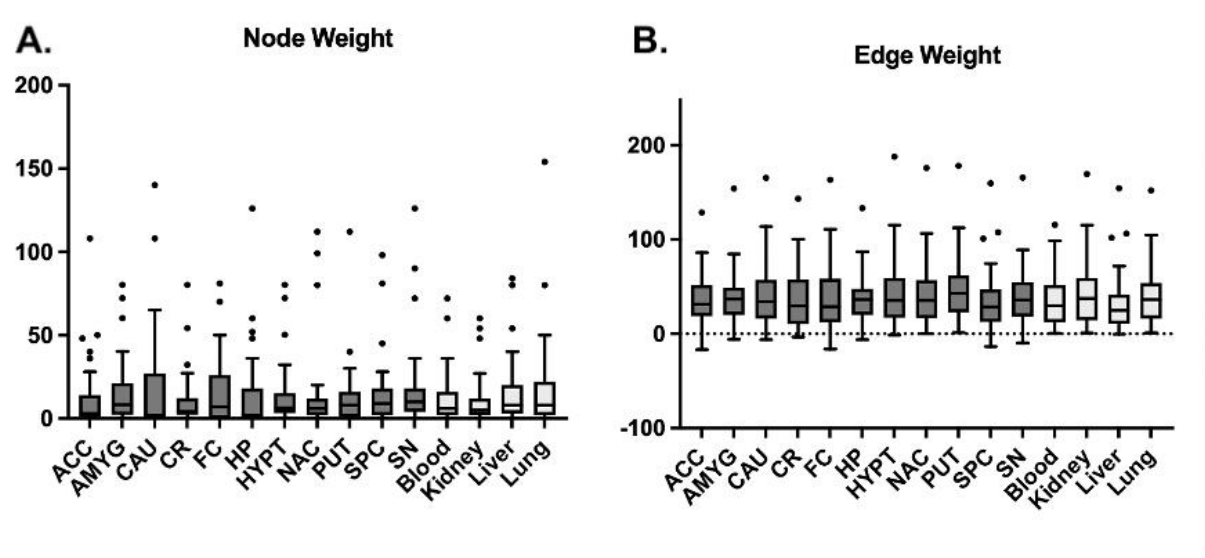


Figure 26. Tissue-specific weighted network analysis on Topological Cluster D

A) *The box plot shows the pairwise comparison of median node weight of Cluster D across the 15 tissues. No significant difference was observed;* ***B)*** *The box plot shows the pairwise comparison of median edge weight between each pair of connected proteins. No significant difference was observed;* ***Abbreviations****: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

Topological Cluster E consisted of 52 interactors and 65 PPIs, which were associated with protein localisation, cell cycle, protein metabolism and response to stress (**Figure 27**). No significant difference was found in node weight analysis or edge weight analysis (**Figure 27A,B**). The hub protein TP53 presented high node weight and high co-expression levels with its first neighbours in the subnetwork of lung, liver and kidney, suggesting its potentially important role in the periphery.

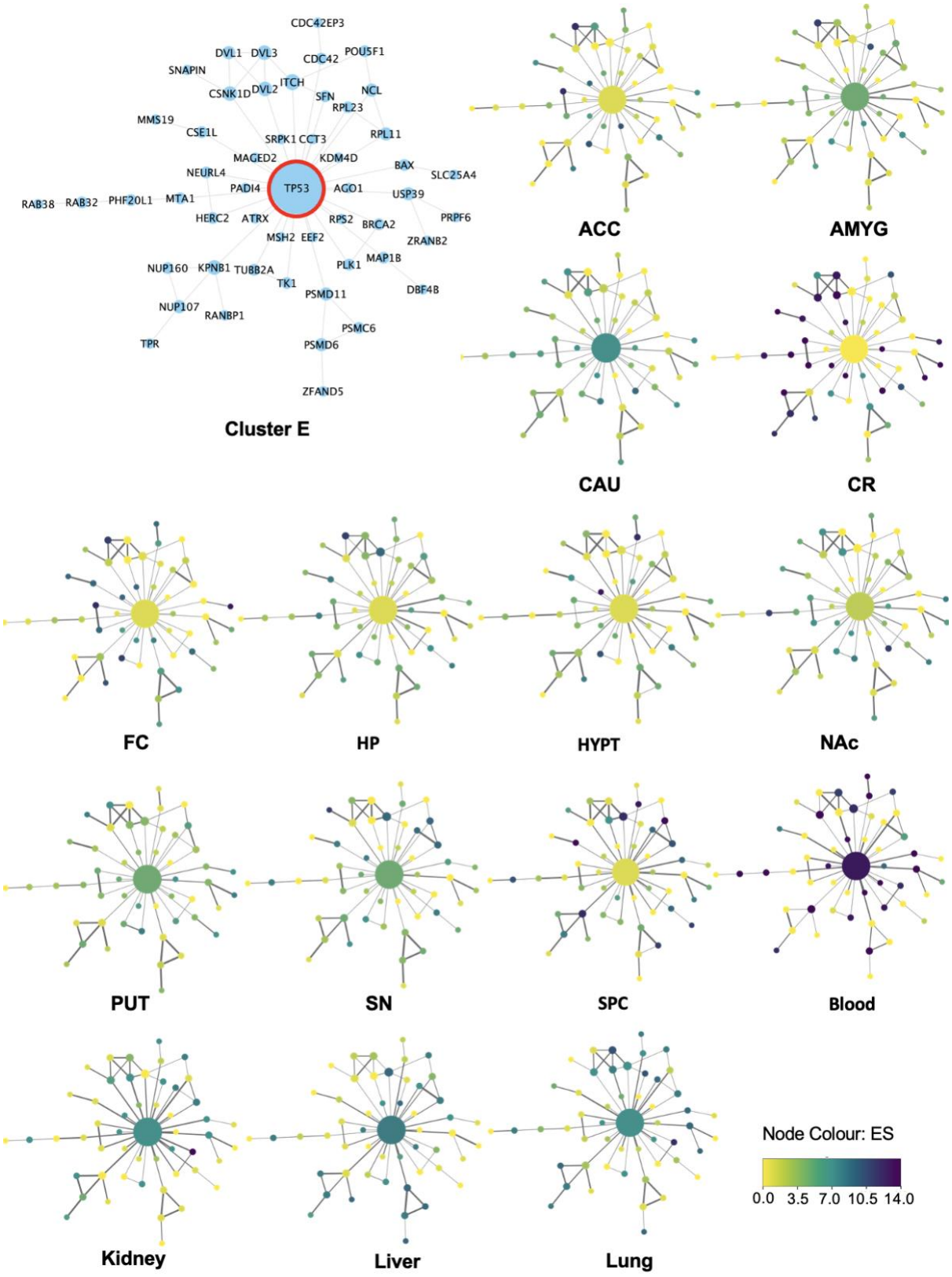


Figure 27. Expression profiles of Topological Cluster E in the 15 tissues

The network graphs show the tissue-specific expression and co-expression profiles of interactors in Cluster E, in which LRRK2 interactors are represented as nodes while PPIs were represented as edges. Node size refers to degree of nodes, i.e., the number of edges connected to each node. The large the

node, the higher the degree. Node colour refers to the expression level of each interactor in different tissues. The higher the expression level, the darker the node. Edge width refers to co-expression levels between each pair of connected proteins. The higher the co-expression level, the thicker the edge.

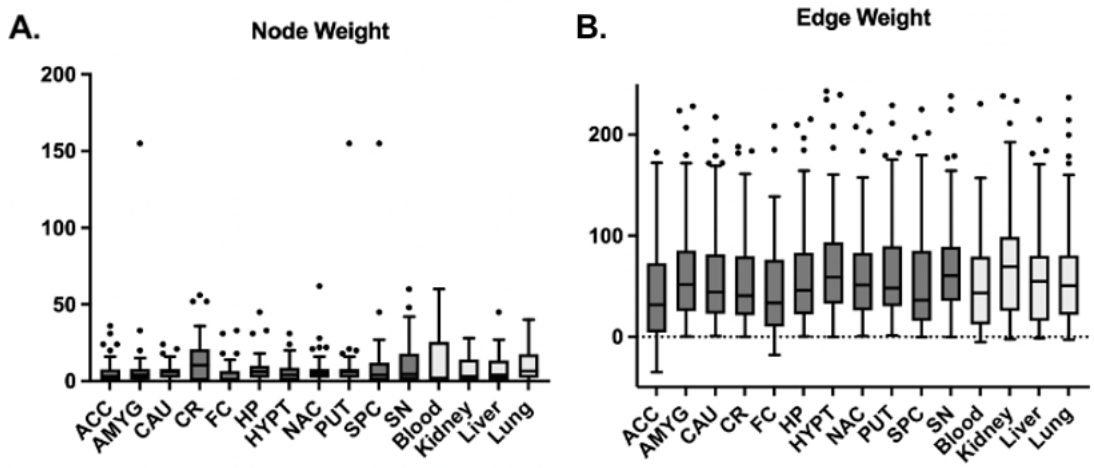


Figure 28. Tissue-specific network analysis on Topological Cluster E

***A)** The box plot shows the pairwise comparison of median node weight of Cluster E across the 15 tissues. No significant difference was observed; **B)** The box plot shows the pairwise comparison of median edge weight between each pair of connected proteins. No significant difference was observed; **Abbreviations**: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

Topological Cluster F contained 15 LRRK2 interactors and 20 edges associated with autophagy (**Figure 29**). The highest median node weight was observed in the cerebellum and frontal cortex (adjusted $p < 0.05$, **Figure 30A**). No significant difference was found in median edge weight (**Figure 30B)**. The hub protein PRKN showed low co-expression level with its first neighbours, while a motif comprised by RAB8A, RAB10, SNCA, VDAC1, HK1, YWHAH, TUBG1 and LMNB2 presented high connection in all brain regions but not in the peripheral tissues. GO-BP and GO-CC enrichment showed that this motif was highly enriched in transport vesicle and associated with regulation of neuronal synaptic plasticity.

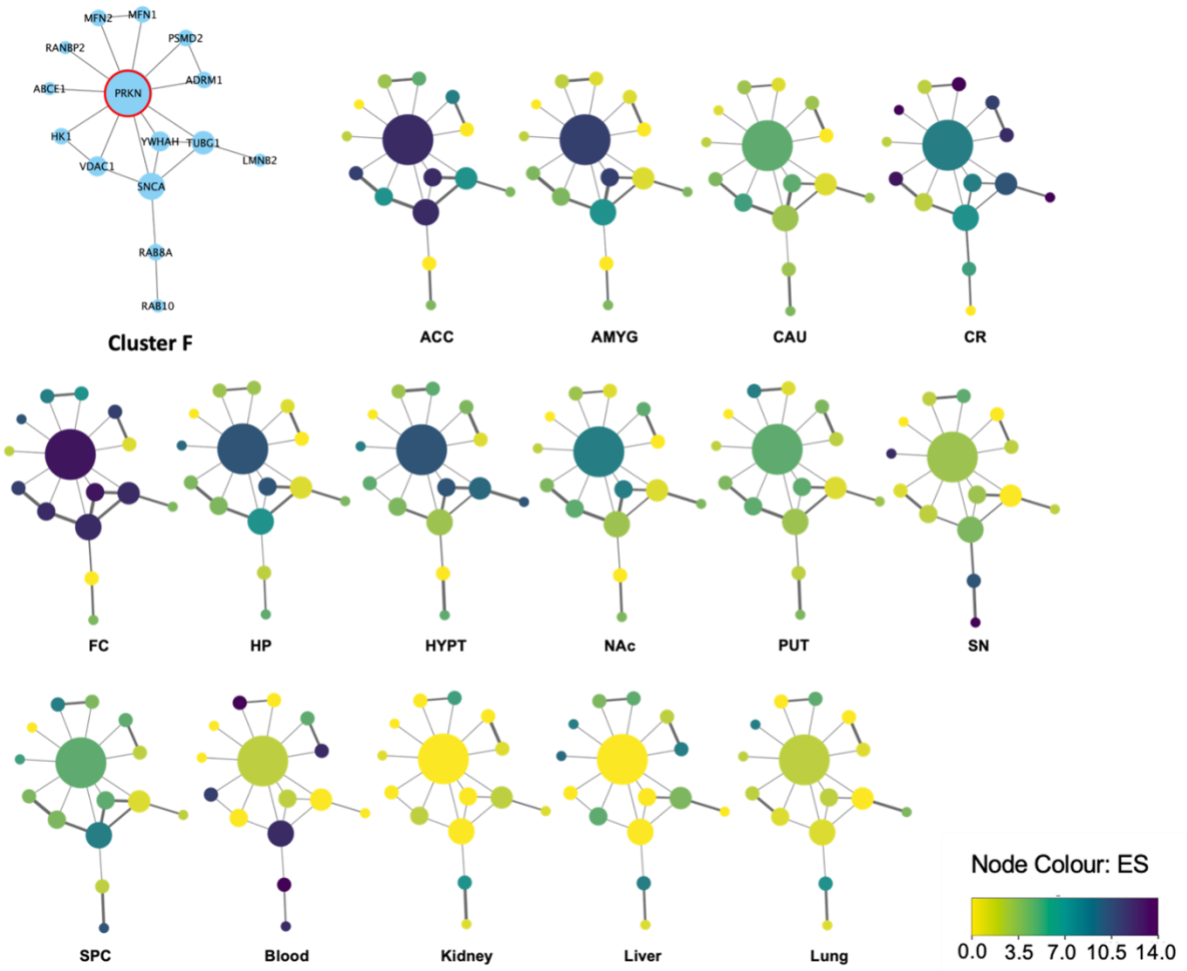


Figure 29. Expression profiles of Topological Cluster F in the 15 tissues

The network graphs show the tissue-specific expression and co-expression profiles of interactors in Cluster F, in which LRRK2 interactors are represented as nodes while PPIs were represented as edges. Node size refers to degree of nodes, i.e., the number of edges connected to each node. The large the node, the higher the degree. Node colour refers to the expression level of each interactor in different tissues. The higher the expression level, the darker the node. Edge width refers to co-expression levels between each pair of connected proteins. The higher the co-expression level, the thicker the edge.

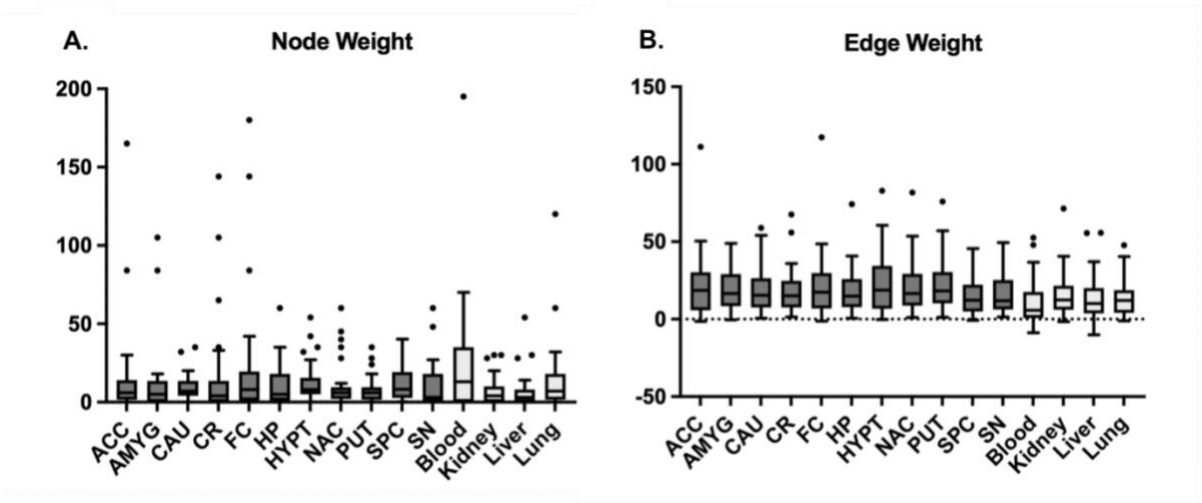


Figure 30. Tissue-specific weighted network analysis on Topological Cluster F

A) *The box plot shows the pairwise comparison of median node weight of Cluster F across the 15 tissues. Highest values were observed for Cerebellum and frontal cortex (adjusted p < 0.05);* ***B)*** *The box plot shows the pairwise comparison of median edge weight between each pair of connected proteins. No significant difference was observed;* ***Abbreviations****: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

Topological Cluster G contained 29 interactors and 50 PPIs, which were associated with protein metabolism and autophagy (**Figure 31**). No significant difference was observed in the median node weight among the 15 tissues, while subnetwork of blood presented the lowest median edge weight as compared to other tissues (adjusted $p < 0.05$, **Figure 32A-B**). There were 2 hub proteins in Cluster G, LRRK1 and HSPA8. Both of them presented high node weight and high co-expression level with its neighbours in frontal cortex, while LRRK1 exhibited these features in anterior cingulate cortex as well, suggesting that these proteins possessed higher activity in mediating the functions of Cluster G in these 2 tissues.

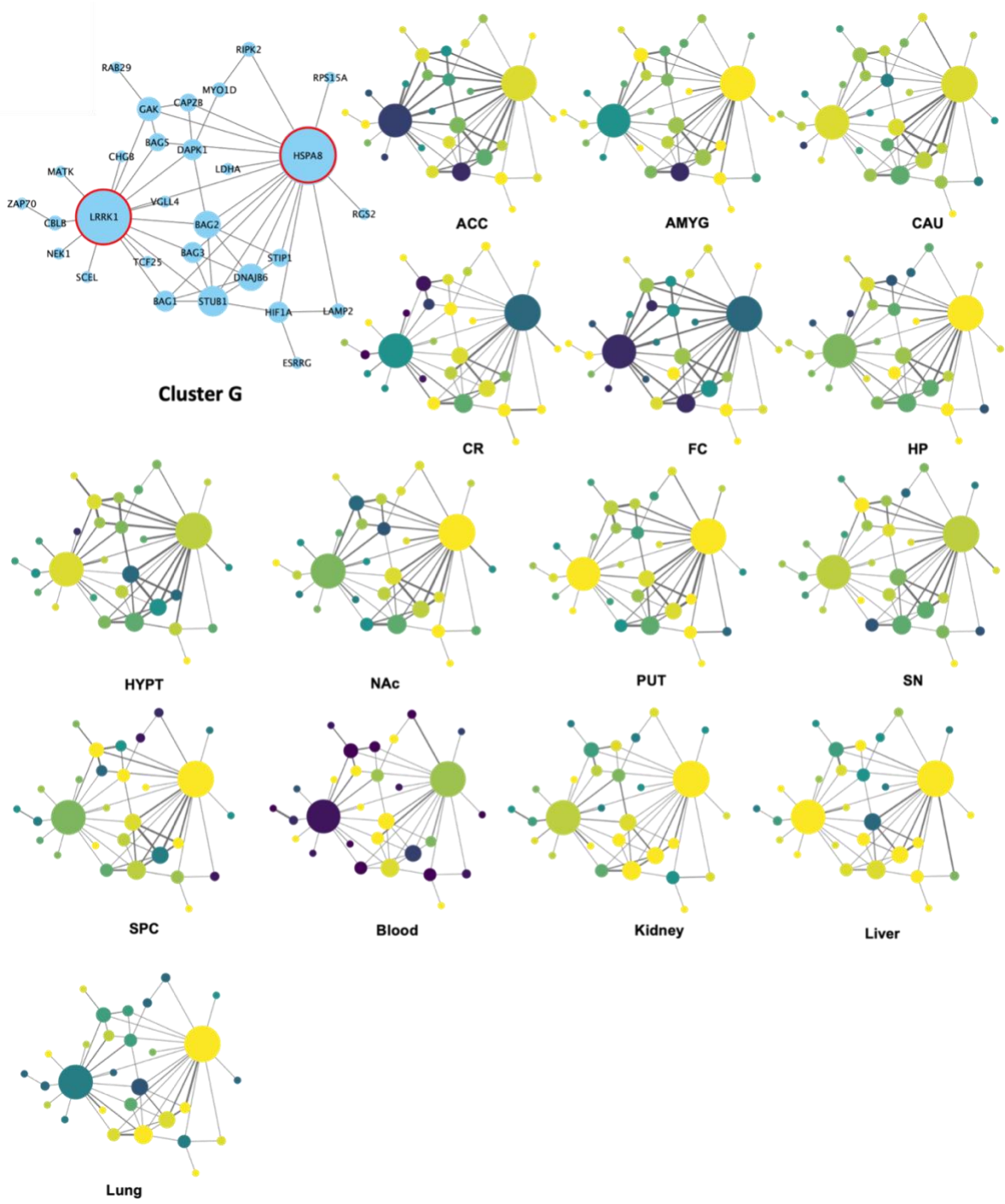


Figure 31. Expression profiles of Topological Cluster G in the 15 tissues

The network graphs show the tissue-specific expression and co-expression profiles of interactors in Cluster G, in which LRRK2 interactors are represented as nodes while PPIs were represented as edges. Node size refers to degree of nodes, i.e., the number of edges connected to each node. The large the node, the higher the degree. Node colour refers to the expression level of each interactor in different tissues. The higher the expression level, the darker the node. Edge width refers to co-expression levels

between each pair of connected proteins. The higher the co-expression level, the thicker the edge. The hub proteins LRRK1 and HSPA8 are marked with red circle;

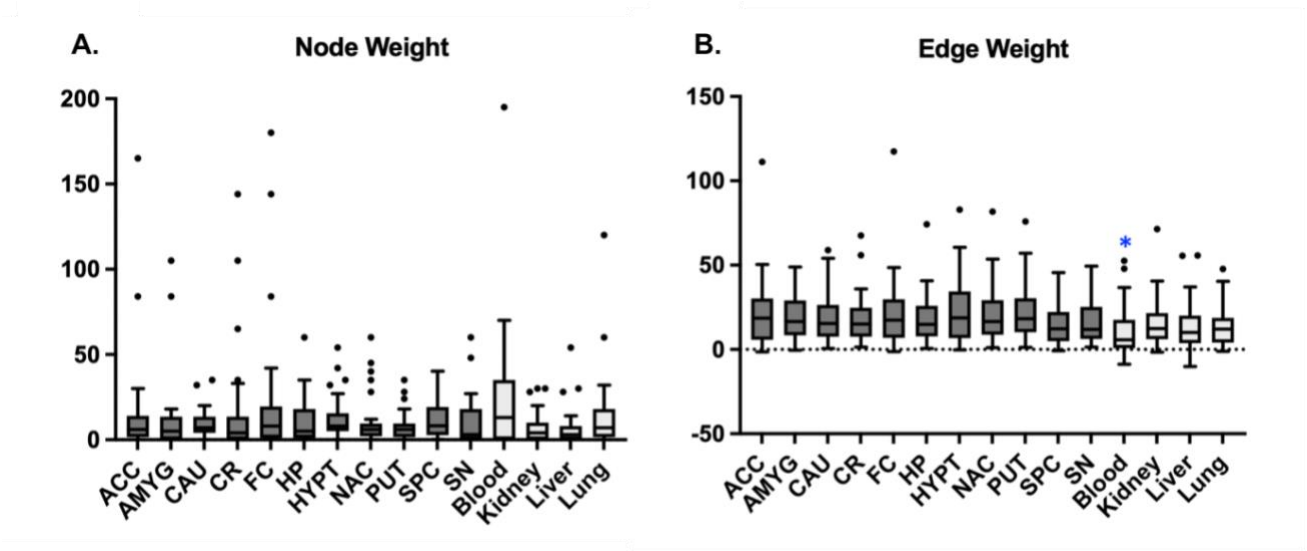


Figure 32. Tissue-specific weighted network analysis on Topological Cluster G

***A)** The box plot shows the pairwise comparison of median node weight of Cluster G across the 15 tissues. No significant difference was observed; **B)** The box plot shows the pairwise comparison of median edge weight between each pair of connected proteins, in which lowest value was observed in blood (adjusted p-value < 0.05); **Abbreviations**: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.*

Main findings (summarised in Figure 33)

1. The $LRRK2_{int}$ possesses distinct mRNA expression profiles in the brain regions and peripheral tissues based both on analysis of DEA and co-expression.
2. In general, highly expressed interactors in the brain regions are related with cell projection, cell death and protein transport, while those highly expresses in the peripheral tissues are associated with apoptosis, protein metabolism and translation.
3. Interactors that are highly co-expressed with LRRK2 in the brain regions are associated with negative regulation of protein modification and organelle localisation (especially for mitochondria), while those in peripheral tissues are related to positive regulation of protein metabolism and response to stimuli.

4. LRRK2 interactors that associated with calcium membrane transport exhibits particularly high expression levels in the frontal cortex.
5. LRRK2:interactor co-expression level is higher in the brain regions than the periphery, especially in the frontal cortex, putamen, caudate and nucleus accumbens.
6. LRKK2 interactors within the putamen, caudate and nucleus accumbens also show similar differential expression profiles.
7. A total of 2 tissue-specific co-expression modules were identified in the $LRRK2_{int}$, highly expressed in the whole blood and lung, respectively. These 2 modules are associated with cell morphology and protein synthesis, respectively.
8. Topological clusters of the $LRRK2_{net}$ shows similar connectivity across the 15 tissues, suggesting that the PPI network of LRRK2 is equally maintained in different healthy tissues regardless being central or peripheral.
9. The "ribosomal unit" in the $LRRK2_{int}$ presented significantly higher expression levels in the amygdala, spinal cord, substantia nigra and whole blood.

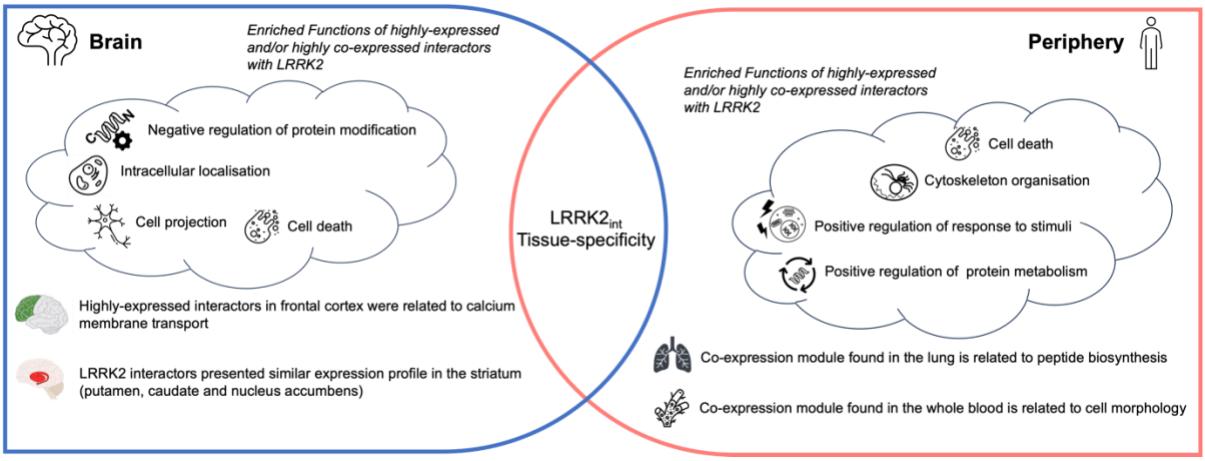


Figure 33. Summary of tissue-specificity of the $LRRK2_{int}$

Discussion

This section investigated the tissue specificity of expression patterns (mRNA level) of LRRK2 interactors by comparing 1) mRNA levels of each interactor, 2) co-expression levels between LRRK2 and its interactor and 3) co-expression modules of LRRK2 interactors among 11 brain regions and 4 peripheral tissues of healthy individuals.

It is worth to mention that the level of LRRK2 was relatively low in the brain; however, the expression levels of LRRK2 passed the QC test (genes with read counts lower than 15 in ≥ 75% of all samples) used by the R package DESeq2, which is one of the most commonly used RNA-Seq analysing pipelines. On the one hand, this very low expression of LRRK2 in the brain might affect the co-expression analysis due to the mean-correlation relationship in RNA-seq data, both bulk and single-cell (Y. Wang et al., 2020). The mean-correlation refers to the fact that the distribution of gene correlations depend on the expression level of the involved genes, i.e., highly expressed genes are more likely to be highly correlated and vice versa. However, it is also worth remembering that there is no golden standard for the acceptable range of expression level for co-expression analysis, therefore it was still worth trying the co-expression analysis on the $LRR2_{int}$ to gain an insight into the potential collaboration between LRRK2 and its interactors in different tissues. On the other hand, it was interesting to notice that such an important protein in terms of disease involvement had such a low expression. One possible explanation for this is that the tissues analysed were collected from healthy individuals who experienced sudden death – it might be that LRRK2 levels are tightly modulated and elevated only in specific circumstances such as infections and immune system activation (Herbst & Gutierrez, 2019; R. L. Wallings & Tansey, 2019a).

The results of the hierarchical clustering on mRNA levels showed that a total of 4 clusters could be identified based on their expression patterns across the 15 tissues, in which 2 presented significant tissue specificity (Exp_Cluster1 and Exp_Cluster3). Exp_Cluster1 exhibited significantly high expression level in the lungs. GO-BP enrichment analysis associated Exp_Cluster1 with peptide biosynthesis, apoptotic signalling and cytoskeleton organisation. Similarly, 1 of the 3 co-expression modules identified in WGCNA (MTurquoise) was also highly expressed in lung and enriched for ribosomal biogenesis. These are in accordance with previous studies showing that LRRK2 plays a crucial role in preventing pulmonary fibrosis (Tian et al., 2021). In addition, an increased LRRK2 level was observed in lung cancer cells, while disrupting LRRK2 expression induce apoptosis of these cancer cells and increase the secretion of proinflammatory factors (J. Wu et al., 2023). Moreover, LRRK2 kinase inhibitors have been found to induce abnormal cytoplasmic accumulation of secondary lysosomes known as lamellar bodies in type II pneumocytes of the lung in mouse models (Fuji

et al., 2015). The highly expressed LRRK2 interactors identified in the Exp_Cluster1 may help in understanding the molecular mechanism behind these lung-specific processes.

In comparison, Exp_Cluster3 exhibited significantly high expression in the frontal cortex. This cluster was related to the regulation of calcium ion export and regulation of protein phosphorylation. The role of LRRK2 in regulating calcium homeostasis has been revealed in previous studies and it might be related to the function of LRRK2 in the modulation of cellular signalling cascade such as the Wnt pathway and the MAPK pathway (Berwick et al., 2017; Boon et al., 2014; R. L. Wallings & Tansey, 2019b). In addition, LRRK2 knock-out was found to enhance cellular response to calcium ion by upregulating the expression levels of associated proteins (Chen et al., 2020). In addition, pathogenic mutations of LRRK2 (including *LRRK2-G2019S* and *LRRK2-R1441C*) perturbed calcium buffering capability of mitochondria in mouse cortex neurons and increased depolarization-induced mitochondrial calcium uptake (Verma et al., 2017, 2022). Interactors in Exp_Cluster3 thus provide a shortlist of candidate LRRK2 interaction partners that could be used as a starting point to investigate the LRRK2 signalling network particularly in the frontal cortex.

The DEA found that a large number of LRRK2 interactors were highly expressed in the whole blood and the cerebellum as compared to other tissues. High LRRK2 expression has been widely reported in peripheral blood mononuclear cells (PBMCs), including B cells, monocytes, and dendritic cells and closely related to immune response pathways (Gardet et al., 2010b; Hakimi et al., 2011c; Kubo et al., 2010; Thévenet et al., 2011b). In accordance with these findings, the authors found that highly-expressed LRRK2 interactors in the whole blood were significantly related to response to cytokine stress, apoptotic signalling, and actin cytoskeleton organisation. The regulatory role of LRRK2 in peripheral immune response has been closely linked to the pathology of PD and CD. For example, a positive correlation between cytokine level and LRRK2 expression was observed in PD patients but not in healthy controls (Cook et al., 2017d). Taken together, these results suggest a crucial role of LRRK2 and its interactors in the peripheral immune system.

In comparison, the highly expressed LRRK2 interactors in the cerebellum were related to intracellular transport of protein and RNA, gene expression and nucleus organisation. The role

of LRRK2 in cerebellum has not been widely studied since it is regarded as unaffected brain region in PD progression. Considering the fact that the cerebellum is responsible for motor control and that the Purkinje neurons are vulnerable to undergo conspicuous degeneration, it could be worth to pay more attention on this brain region in neurological pathology (Mandemakers et al., 2012). Indeed, some lines of evidence have shown that LRRK2 may play an important role in cerebellar Purkinje neurons. For example, overexpression of LRRK2 or kinase-enhancing LRRK2 mutations (*LRRK2-G2019S* and *I2020T*) have been liked to increase mRNA translation in cerebellum (Friedman et al., 2012b). In addition, accumulation of LRRK2 and increased LRRK2 mRNA level were observed in the autophagy-impaired neurons in cerebellar nuclei (Plotegher & Civiero, 2012). Together with the highly-expressed LRRK2 interactors identified in this study, LRRK2 interactors may participate in the regulation of protein synthesis and localisation in the cerebellum (Friedman et al., 2012a).

Finally, functional enrichment analysis was performed on the interactors with either high expression levels or co-expression with LRRK2 in the brain and the periphery separately. The results showed that the portion of the LRRK2 interactors with a better expression profile in the brain were associated with intracellular organisation, organelle localisation, cell projection and negative regulation of protein modification. In comparison, those with higher expression level in the peripheral tissues were more related to positive regulation of protein metabolism, response to stimuli, cytoskeleton organisation and apoptotic signalling suggesting that it might be indeed possible to differentiate (considering both composition and functional association) a central or brain-related LRRK2int and a peripheral $LRRK2_{int}$.

Similar results were found when considering the $LRRK2_{int}$ expression behaviour. When tissues were grouped via hierarchical clustering based on the similarity of overall expression profiles, the 11 brain regions and 4 peripheral tissues were allocated into 2 distinct groups. Similarly, LRRK2 and its interactors presented distinct co-expression behaviours in the brain and the periphery, in which the co-expression levels were higher in the brain regions, especially in frontal cortex, putamen, nucleus accumbens and caudate. These findings suggest the $LRRK2_{int}$ exhibits distinct functional patterns in the brain in comparison with the periphery.

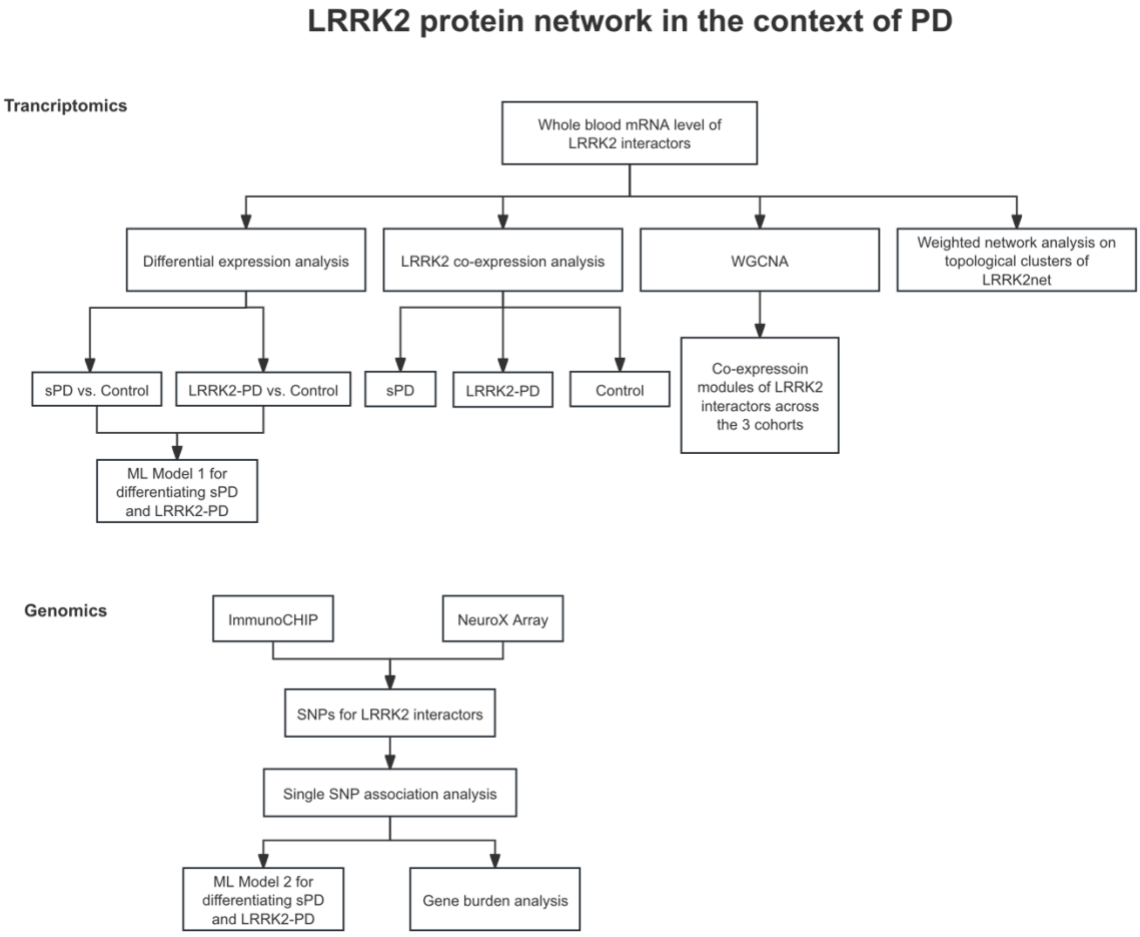
Finally, tissue grouping based on mRNA level and LRRK2:interactor co-expression level suggested that putamen, caudate and nucleus accumbens shares similar transcriptomic patterns of LRRK2 interactors, indicating that the striatum could be a functional unit for the LRRK2 interactome, in which LRRK2 interactors presented harmonious expression patterns and high co-expression levels with LRRK2. Of note, the striatum is the target for the projection of dopaminergic neurons and one of the most affected brain regions during Parkinson's disease (PD) progression. In fact, multiple studies have associated the degeneration of putamen and caudate with motor and non-motor PD symptoms (Manes et al., 2018; Playford et al., 1992; J. Wang et al., 2018); while nucleus accumbens, involved in mediating emotional and motivational processes such as rewarding experiences, impulsive and compulsive behaviours, might be implicated in the neuropsychiatric symptoms of PD (Barbosa et al., 2019; Hammes et al., 2019b).

Chapter 4. LRRK2 protein interactome in the context of PD

Objectives

- Compare the expression patterns of LRRK2 interactors and LRRK2:interactor co-expression features in the sPD and LRRK2-PD cohorts as compared to healthy controls based on the whole blood mRNA-Seq data.
- Identify co-expression modules of LRRK2 interactors among the sPD, LRRK2-PD and healthy control cohorts.
- Evaluate weighted connectivity of topological clusters of the $LRRK2_{net}$ detected in Chapter 4 in the sPD and LRRK2-PD cases
- Identify LRRK2 interactors with SNPs significantly related to the disease status
- Construct Machine Learning (ML) classification models on expression levels of LRRK2 interactors and genetic variants

Analysis pipeline



Note: Some of the figures and text in this chapter are adapted from the following manuscript deposited in bioRxiv: Zhao et al.; "Transcriptomics analyses of the LRRK2 protein interactome reveal distinct molecular signatures for sporadic and LRRK2 Parkinson's Disease" https://doi.org/10.1101/2023.09.12.557373.

Methods

- mRNA data download and QC

Whole blood mRNA levels (in read counts) at baseline patient screening (time at enrolment) were extracted for the PPMI cohorts "de novo PD" and "healthy control" on 24th January 2023. The 2 cohorts were firstly examined on the robustness of genetic status record for each individual via the following criteria:

i.) The genetic status needed to be confirmed by at least 3 out of 6 detection techniques (WGS, WES, RNA-Seq, GWAS, CLIA, SANGER sequencing);

ii.) The genetic status needed to be confirmed by at least 1 next generation sequencing technique (WGS, WES, RNA-Seq) and 1 screening technique (GWAS, CLIA, SANGER sequencing)

Only individuals with robust genetic status record proceeded to the next QC step, involving filtering out subjects with non-*LRRK2* mutations, including variants in *GBA*, *PINK*, *Parkin* and *SNCA*. QC-ed subjects were then allocated into 3 groups: Healthy control, Sporadic PD (with not known *LRRK2* mutations in record, hereby referred to as sPD) and PD patients with pathogenic *LRRK2-G2019S* or *LRRK2-R1441C/G/H* variants (hereby referred to as LRRK2-PD). Interactor-level QC was performed by excluding transcripts with mRNA read counts ≤ 15 in more than 75% subjects. QC-ed read counts were then normalised using the R package "DESeq2", thereby generating the "PPMI_rc_Matrix".

- Cohort characterisation

Baseline characteristics of QC-ed subjects in the "PPMI_rc_Matrix" were downloaded on 21 Dec 2022, including gender, age-at-baseline, progression of PD as evaluated by Hoehn and Yahr Scale (H&Y); PD motor symptoms measured by MDS-Unified Parkinson's Disease Rating Scale III (MDS-UPDRS III); depression as evaluated by geriatric depression scale (GDS); state and trait anxiety as evaluated by State-Trait Anxiety Inventory (STAI); visuospatial function as evaluated by the 15-item Benton Judgment of Line Orientation Test (BJLOT); repaid eye movement behaviour disorder as evaluated using the REM Sleep Behaviour Disorder Screening Questionnaire (RBDSQ); daytime sleepiness as evaluated by the Epworth Sleepiness Scale (ESS); cognitive functions as evaluated by the 1) Montreal Cognitive Assessment (MoCA) in general, 2) the Semantic Fluency Test (SFT; total score of animal, fruit and vegetable) for semantic memory, 3) the Letter-Number Sequencing (LNS) test for executive function and working memory, 4) the Hopkins Verbal Learning Test (HVLT; immediate recall and delayed recall) for memory.

- Differential Expression Analysis (DEA)

DEA was performed on the "PPMI_rc_Matrix" to compare the baseline expression levels of LRRK2 interactors in healthy controls vs. the 2 PD cohorts respectively using the R package "DESeq2". Of note, all DEA results were adjusted for gender. Multiple test adjustment was

automatically conducted by the "DESeq2" package via Benjamini-Hochberg procedure. Interactors with significant differential expression figures in the 2 PD cohorts as compared to controls were queried in functional enrichment analysis. Significant changes were defined as with $|log2FC| > 0.1$ and adjusted p-value < 0.05 (threshold adapted from (Craig et al., 2021)).

- LRRK2 co-expression analysis

Baseline co-expression levels between LRRK2 and its interactors in the "PPMI_rc_Matrix" were calculated for the 3 PPMI cohorts via Pearson's correlation test. Significant co-expression was defined as with having Pearson's coefficient > 0.6. Interactors that presented different co-expression behaviours with LRRK2 in the 2 PD cohorts as compared to the healthy controls were annotated via functional enrichment analysis.

- Weighted Gene Co-expression Network Analysis (WGCNA)

Weighted Gene Co-expression Network Analyses (WGCNA) was performed on the "PPMI_rc_Matrix" to identify co-expression modules of LRRK2 interactors in healthy and PD conditions (healthy vs. sPD and healthy vs. LRRK2-PD). Co-expression levels between LRRK2 interactors were calculated via Pearson's correlation test. Soft power β was selected to ensure the gene co-expression networks follow the power law, i.e., reaching the largest Scale-free Fit Index and the lowest Mean Connectivity. Co-expression modules of interactors were detected automatically by R package "WGCNA". Eigengene for each module was calculated as the first principal component of the module's topological overlap matrix (TOM). The similarity between modules returned from WGCNA on control vs. sPD cases (M_sPD) and control vs. LRRK2-PD cases (M_LPD) were compared via permutation test. For each pair of M_sPD and M_LPD, the number of overlapping proteins (test_intersection) was compared to the overlap count distribution generated by 1000 pairs of randomly sampled protein lists from the $LRRK2_{net}$ at same size of M_sPD and M_LPD (random_intersection). Of note, the random_intersection distribution curve was considered as normal distributed. A significant overlap between 2 modules was defined as: 1) test_intersection > 95% of the points in the random overlap distribution curve and 2) the percentage of overlapped interactors in both of the 2 modules > 60%.

- SNP association analysis

In PPMI, genotyping was performed using the Illumina Immunochip and NeuroX arrays, which produced 3 datasets: PPMI_IMMUNOCHIP_Nov11th2013, PPMI_NEUROX_ Nov11th2013, and PPMI_Project_107_NeutoX_Genotyping, hereby referred as: "Immuno", "NeuroX1" and "NeuroX2", respectively. The 3 datasets were downloaded on 15th September 2022. SNPs for LRRK2 interactors were extracted based on gene positions on the chromosomes. These SNPs went through a QC pipeline as following: 1) Conversion of SNP ID to rsID; 2) removal of non-coding SNPs (based on the annotation retrieved from Ensembl Variant Effect Predictor, https://www.ensembl.org/info/docs/tools/vep/index.html); 3) removal of palindromic SNPs; 4) removal of SNPs with low genotyping rate (< 95%); 5) removal of SNPs with significant missingness in the cohort (Bonferroni adjusted-p < 0.05). Another QC pipeline was applied to the genotyped subjects: 1) subjects with abnormal heterozygosity rates were exclude (< Mean-4SD or > Mean+4SD) since an abnormal heterozygosity rate suggests potential DNA sample contamination or low sample quality; 2) subjects that exhibited high genetic relatedness (3rd relatives or beyond) were calculated to reduce familial biases via PI_HAT; 3) ethnicities of included PPMI subjects were investigated using cohorts of HapMap3 project as reference (Altshuler et al., 2010) via PCA. Of note, considering the small amount of available genotyping data for LRRK2-PD patients, prodromal participants with the *LRRK2-G2019S* or *LRRK2-R1441C/G/H* were also included in this section. Univariate logistic regression was used to identify $LRRK2_{int}$-SNPs that were associated with sPD and PD cases with LRRK2 mutation (LRRK2-PD + prodromal LRRK2-PD). All univariate logistic regression analyses were adjusted for gender and significant Principal Components (PCs). Multiple test correction was performed via Bonferroni's method. In addition, genetic burden analyses were performed to determine the contribution of SNPs in each LRRK2 interactor to sPD and LRRK2-PD via the R package "SKAT". Multiple test correction was performed via Bonferroni's method.

- Classification model of sPD and LRRK2-PD using transcriptomic and genetic features of LRRK2 interactors

Whole blood mRNA levels of the LRRK2 interactors which presented significant differential expression (DE) in the DEA test of the sPD and/or LRRK2-PD cases vs. the healthy controls were utilised as independent variables in a Least Absolute Shrinkage and Selection Operator

(LASSO) regression model to classify the 2 types of PD cases (Model 1). To reduce the dimensionalities, univariate logistic analyses was performed on each DE interactors. Only interactors with $p < 0.05$ were included in the model. In addition, the coding SNPs of LRRK2 interactors which showed significant association in univariate logistic regression analyses as well as the burden of SNPs at gene level which exhibited significant difference between controls and sPD and/or LRRK2-PD were utilised as independent variables in a second LASSO regression model (Model 2), with the phenotype (sPD/LRRK2-PD) as outcome. Of note, only LRRK2-PD and sPD patients were included in Model 2. For the 2 models, the train:test sets were split as 4:1. The lambdas (λ) were tuned by a 10-fold cross-validation (CV) method via the "cv.glmnet" function of the "glmnet" R package. The refined models were then assessed on the test cohort. Receiver Operating Characteristic (ROC) curves were generated via the "roc.glmnet" function of the "glmnet" R package.

Results

- PPMI cohort characterisation

After subject QC a total of 657 individuals with robust genetic status were retained (170 controls, 116 LRRK2-PD cases, 371 sPD cases). Among these individuals, 389 were male (59.2%). Average baseline age of the control, sPD and LRRK2-PD cohorts were 60.9Y, 61.3Y, 63.1Y, respectively. No significant different were observed in gender and age among the 3 cohorts (chi-square and One-way ANOVA $p > 0.05$, **Table 1**). Moreover, 95.1% (N = 625) of the 657 subjects were Caucasian, 2.5% (N = 16) were Africans (removed from the cohort), while 1.6% (N = 10) were Asian (removed from the cohort). Among the 116 LRRK2-PD patients, 102 carry the *LRRK2-G2019S* mutation while 12 carry the *LRRK2-R1441C* mutation. Comparisons on baseline clinical features of the sPD and LRRK2-PD showed that LRRK2-PD patients presented significantly more daytime sleepiness, depression, trait anxiety and worse visuospatial function but less REM sleep disorders and better cognitive functions. No difference was observed between the 2 PD cohort in terms of immediate and delayed memory, executive function, state anxiety, and motor symptoms.

Table 8. Cohort characterisation

	Control	sPD	LRRK2-PD	p-value

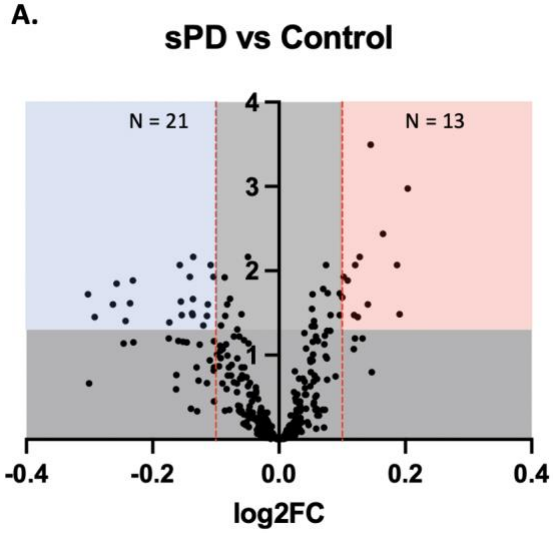
Male (%)	94 (55.3%)	224 (60.4%)	71 (61.2%)	0.875
Age	60.9 (0.8)	61.3 (0.5)	63.1 (0.8)	0.987
ESS	-	5.8 (0.2)	7.1 (0.4)	0.002*
GDS	-	2.5 (0.1)	3.1 (0.3)	0.009*
HVLT_recall	-	11.1 (0.1)	11.1 (0.1)	0.762
LNS	-	11.2 (0.1)	10.9 (0.2)	0.271
BJLOT	-	12.6 (0.1)	11.7 (0.3)	<0.001*
MoCA	-	26.9 (0.1)	26.2 (0.2)	0.013*
RBDSQ	-	3.3 (0.1)	2.6 (0.2)	0.014*
STAI_state	-	47.3 (0.3)	46.5 (0.5)	0.151
STAI_trait	-	46.1 (0.2)	43.9 (0.6)	<0.001*
MDS-UPDRS III	-	21.1 (0.5)	20.1 (0.9)	0.290

*Data is reported as Mean (SE). * p-value < 0.05.* ***Abbreviations:*** *ESS: Epworth Sleepiness Scale; GDS: Geriatric Depression Scale; HVLT: Hopkins Verbal Learning Test; LNS: Letter-Number Sequencing; BJLOT: 15-item Benton Judgment of Line Orientation Test; MoCA: Montreal Cognitive Assessment; RBDSQ: REM Sleep Behaviour Disorder Screening Questionnaire; STAI: State-Trait Anxiety Inventory; NP3: MDS-Unified Parkinson's Disease Rating Scale III.*

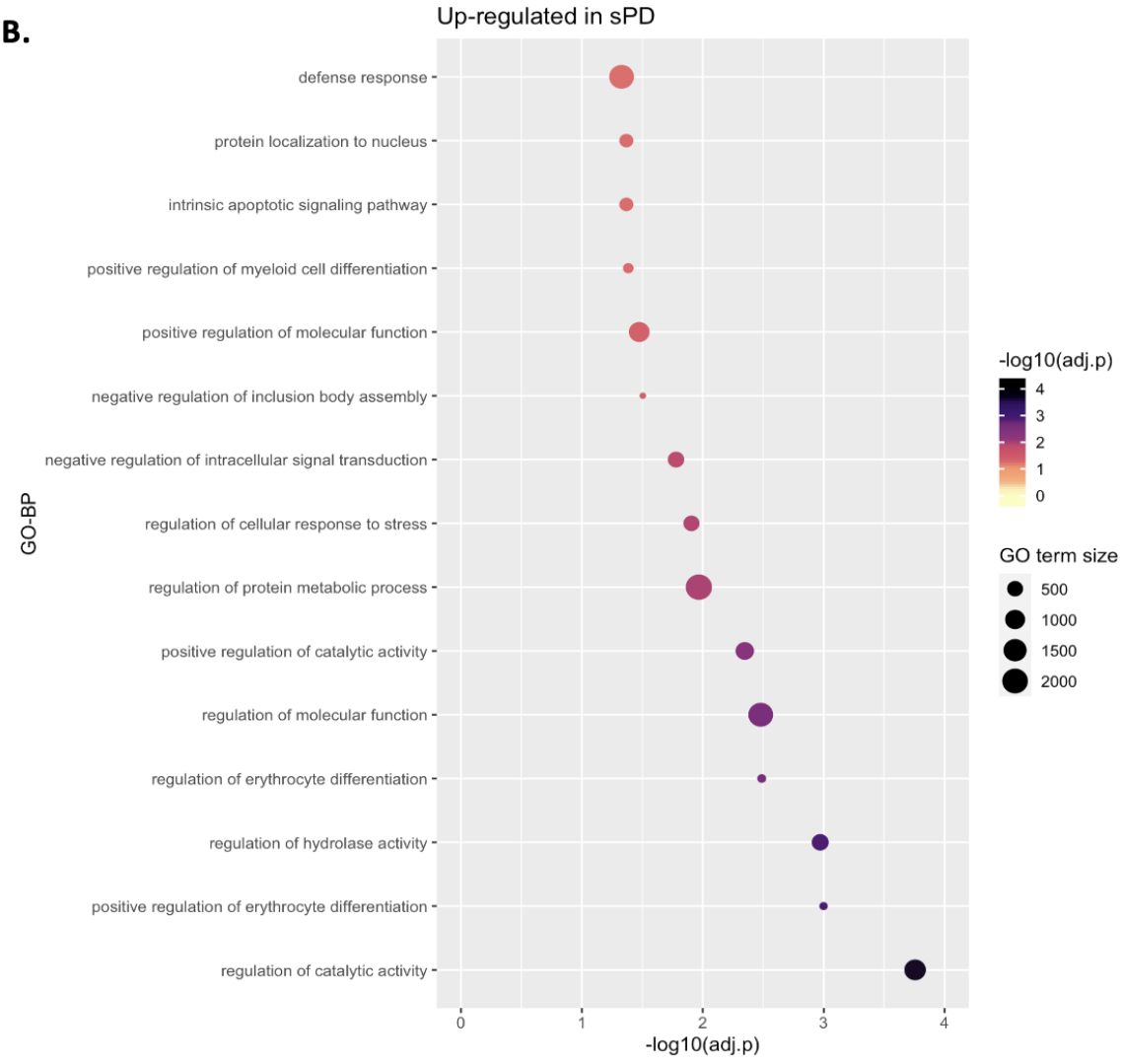
- PD-associated DEA

Baseline whole blood mRNA read counts were extracted for 416 out of 418 LRRK2 interactors (no data was available for GTF2I and CYFIP1). Among these 416 interactors, 38 with missing or low baseline mRNA levels (with < 15 read counts in > 75% samples) were excluded, including MAP1B, MAPT, SFN, HSPB1, CHGB, TUBB4A, ACTBL2, ACTG2, ANKS4B, C17orf53, EEF1A2, EEF1G, PAK6, RAI14, SH3GL2, TK1, AURKB, C1orf87, CXCL11, DNAJB8, ESRRG, FAM47B, FAM90A1, KDM4D, POU5F1, RAPGEF4, SCEL, SH3GL3, SMTNL2, STAC, TGIF2LX, TUBB3, VASH2, VN1R1, WT1-AS, XIRP2, SERF1 and MARK1, thereby leaving 378 interactors for further analysis. DEA identified 34 interactors with significantly differential expression in the sPD cohort as compared to healthy controls, in which 13 were up-regulated while 21 were down-regulated (**Figure 26A, Table S5**). Functional enrichment analysis showed that the up-regulated interactors were enriched for protein metabolism, response to stress and negative regulation of apoptotic signal transduction, while the down-regulated interactors were associated with translation and ribosomal function (**Figure 26B,C**).

A.


sPD vs Control
N = 21
N = 13
4
3
2
1
-0.4
-0.2
0.0
0.2
0.4
log2FC

B.


Up-regulated in sPD
defense response
protein localization to nucleus
intrinsic apoptotic signaling pathway
positive regulation of myeloid cell differentiation
positive regulation of molecular function
negative regulation of inclusion body assembly
negative regulation of intracellular signal transduction
regulation of cellular response to stress
regulation of protein metabolic process
positive regulation of catalytic activity
regulation of molecular function
regulation of erythrocyte differentiation
regulation of hydrolase activity
positive regulation of erythrocyte differentiation
regulation of catalytic activity
GO-BP
0
1
2
3
4
-log10(adj.p)
-log10(adj.p)
4
3
2
1
0
GO term size
500
1000
1500
2000

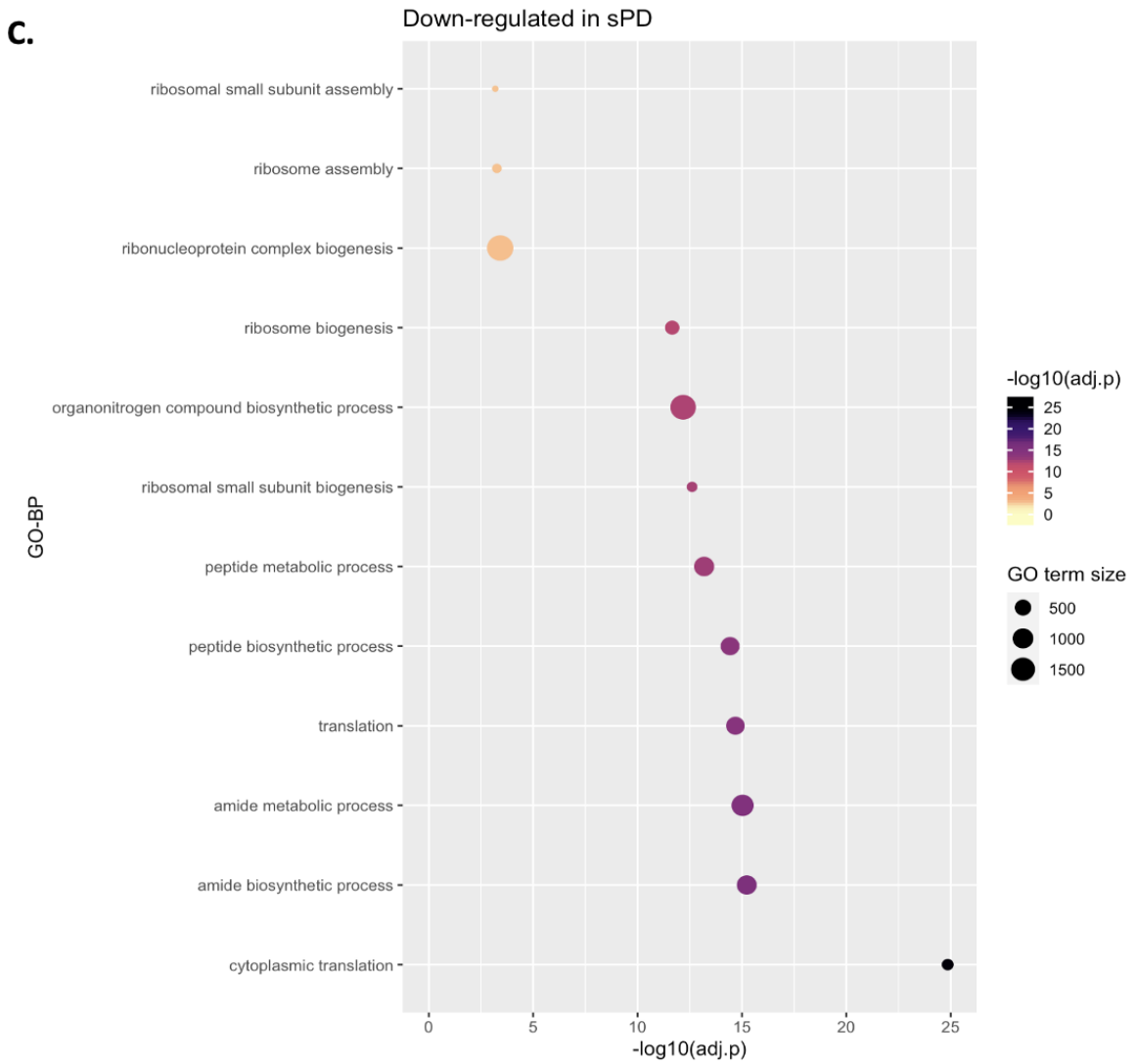
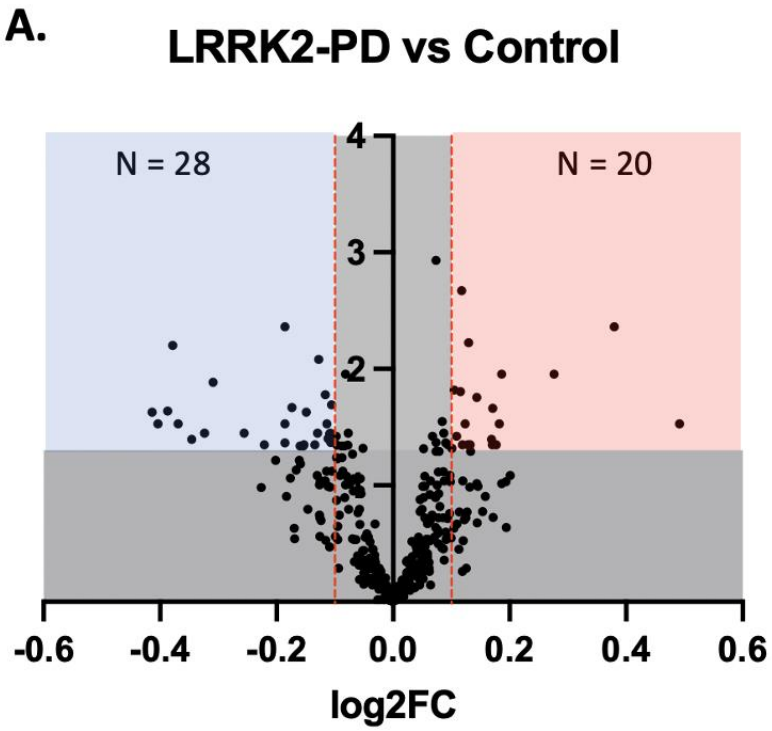


Figure 34. Differential expression of LRRK2 interactors in the sPD cohort vs. control

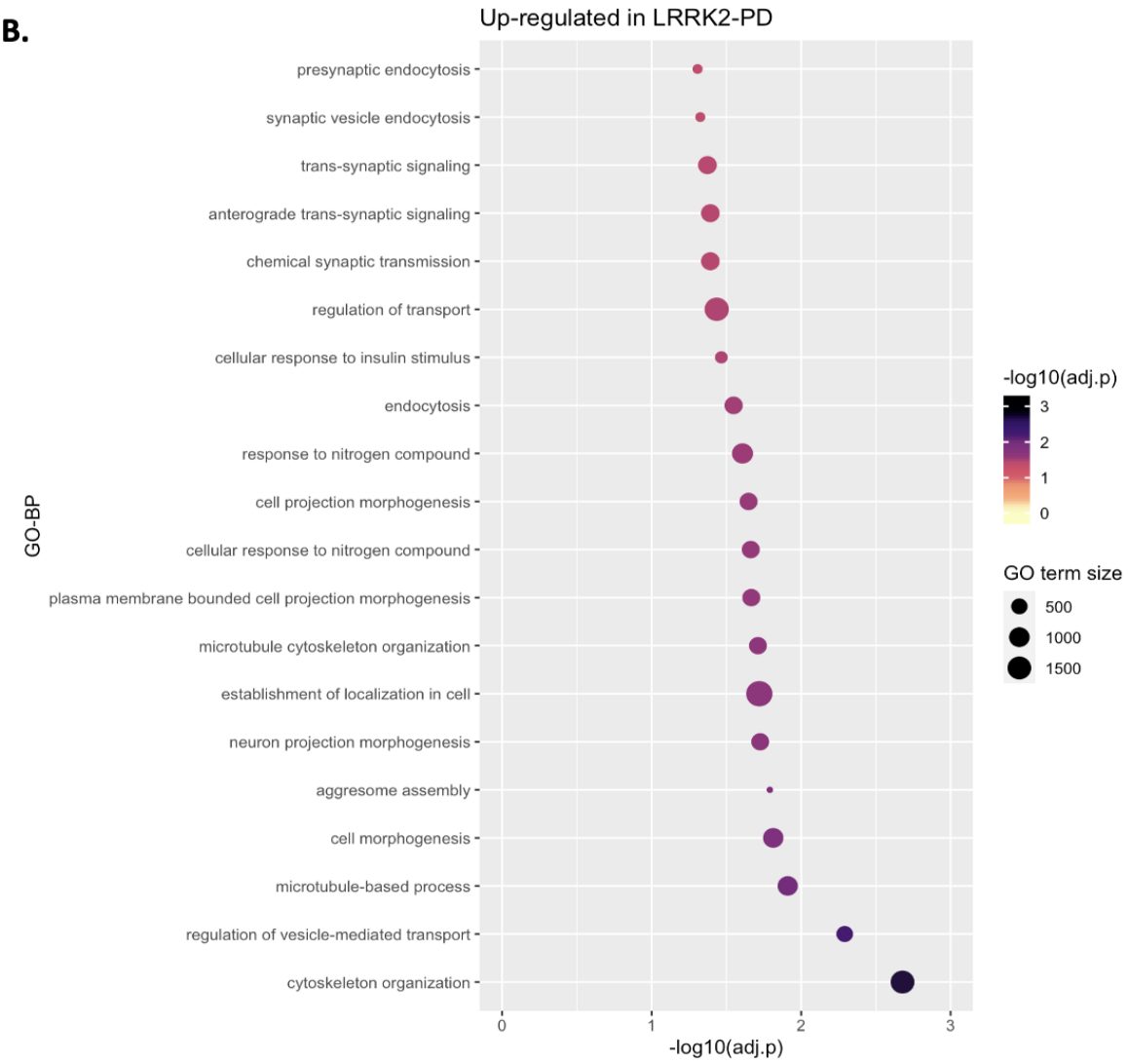
A) *The scatter plot shows differential expression profiles of LRRK2 interactors in the sPD cases and controls at baseline of diagnosis. Scatters located on left side of y-axis were considered as down-regulated (highlighted in the blue area), i.e., the interactor presented lower expression levels in the sPD cases as compared to controls; while scatters located on right side were considered as up-regulated (in the red area), i.e., the interactors presented significantly higher expression levels in the sPD cases as compared to controls. Scatters in the blue/red highlighted zones (adjusted $p < 0.05$, $|log2FC| > 0.1$) are defined as significant differential expression;* ***B,C)*** *The bubble graphs show the GO-BP terms associated with significantly up-regulated and down-regulated LRRK2 interactors in the sPD subjects vs. Controls. Bubble colour represents the enrichment significance, while bubble size represents the term size.*

There were 48 interactors with significantly altered expression levels in the LRRK2-PD cohort as compared to the controls, in which 20 were upregulated and 28 were down-regulated

(**Figure 27A, Table S5**). GO-BP enrichment analysis associated the up-regulated interactors at BL with synaptic signalling and cell morphogenesis while down-regulated proteins were related with protein biosynthesis, apoptotic signal transduction, negative regulation of ubiquitin-dependent protein catabolism (**Figure 27B,C**).



B.


Up-regulated in LRRK2-PD
presynaptic endocytosis
synaptic vesicle endocytosis
trans-synaptic signaling
anterograde trans-synaptic signaling
chemical synaptic transmission
regulation of transport
cellular response to insulin stimulus
endocytosis
response to nitrogen compound
cell projection morphogenesis
cellular response to nitrogen compound
plasma membrane bounded cell projection morphogenesis
microtubule cytoskeleton organization
establishment of localization in cell
neuron projection morphogenesis
aggresome assembly
cell morphogenesis
microtubule-based process
regulation of vesicle-mediated transport
cytoskeleton organization
GO-BP
0
1
2
3
-log10(adj.p)
-log10(adj.p)
3
2
1
0
GO term size
500
1000
1500

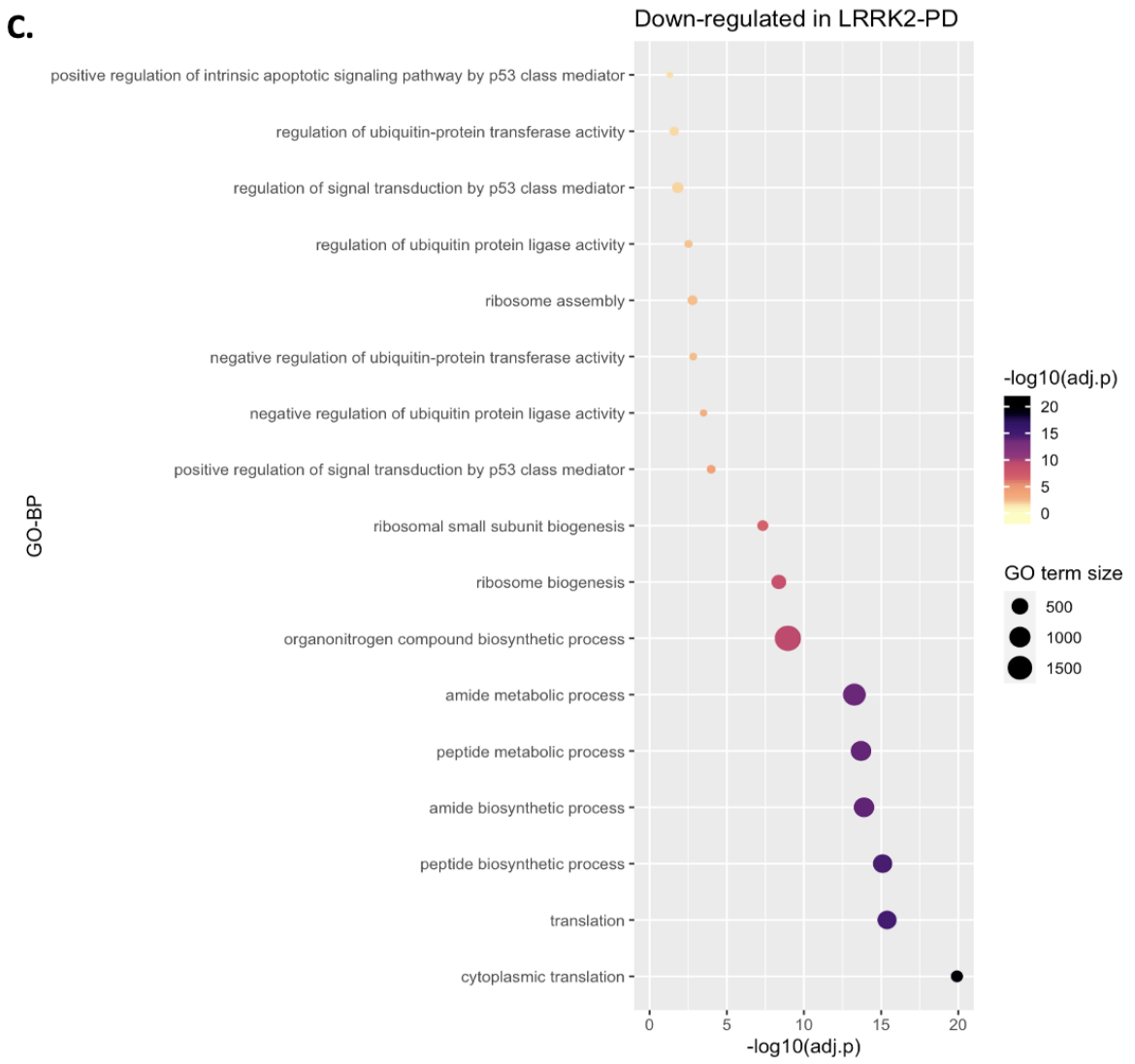
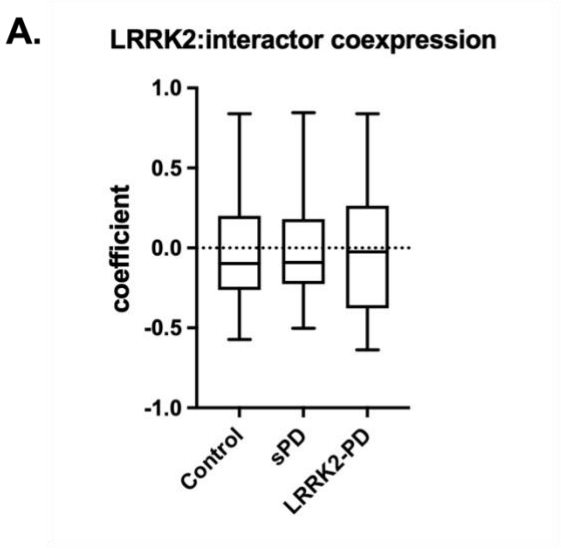


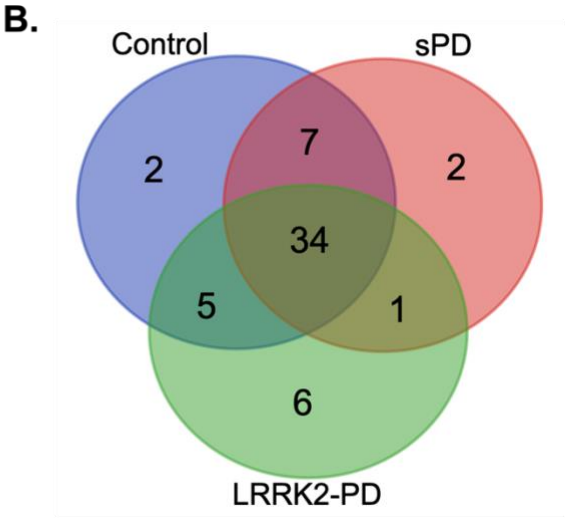
Figure 35. Differential expression of LRRK2 interactors in the LRRK2-PD cohort vs. Control

A) *The scatter plot shows differential expression profiles of LRRK2 interactors in the LRRK2-PD cases and controls at baseline of diagnosis. Scatters located on left side of y-axis were considered as down-regulated (highlighted in the blue area), i.e., the interactor presented lower expression levels in the LRRK2-PD cases as compared to controls; while scatters located on right side were considered as up-regulated (in the red area), i.e., the interactors presented significantly higher expression levels in the LRRK2-PD cases as compared to controls. Scatters in the blue/red highlighted zones (adjusted $p < 0.05$, $|log2FC| > 0.1$) are defined as significant differential expression;* ***B,C)*** *The bubble graphs show the GO-BP terms associated with significantly up-regulated and down-regulated LRRK2 interactors in the LRRK2-PD subjects vs. Controls. Bubble colour represents the enrichment significance, while bubble size represents the term size.*

- LRRK2 co-expression analysis in the 3 cohorts

In general, LRRK2:interactor co-expression was similarly low in the whole blood sample from the 3 cohorts, with median correlation coefficient < 0 (**Figure 29A, Table S6**). Pearson's correlation test found a total of 48, 44 and 46 interactors co-expressed with LRRK2 (Correlation Coefficients > 0.5) in the control, sPD and LRRK2-PD subjects, among which 34 presented co-expression with LRRK2 in all the 3 cohorts (**Figure 29B**). Functional enrichment analysis associated the 34 interactors with intracellular organisation and cell division (**Figure 29C**). These findings suggest that LRRK2:interactor co-expression is not altered by LRRK2 pathogenic variants and remain stable at the early stage of PD.





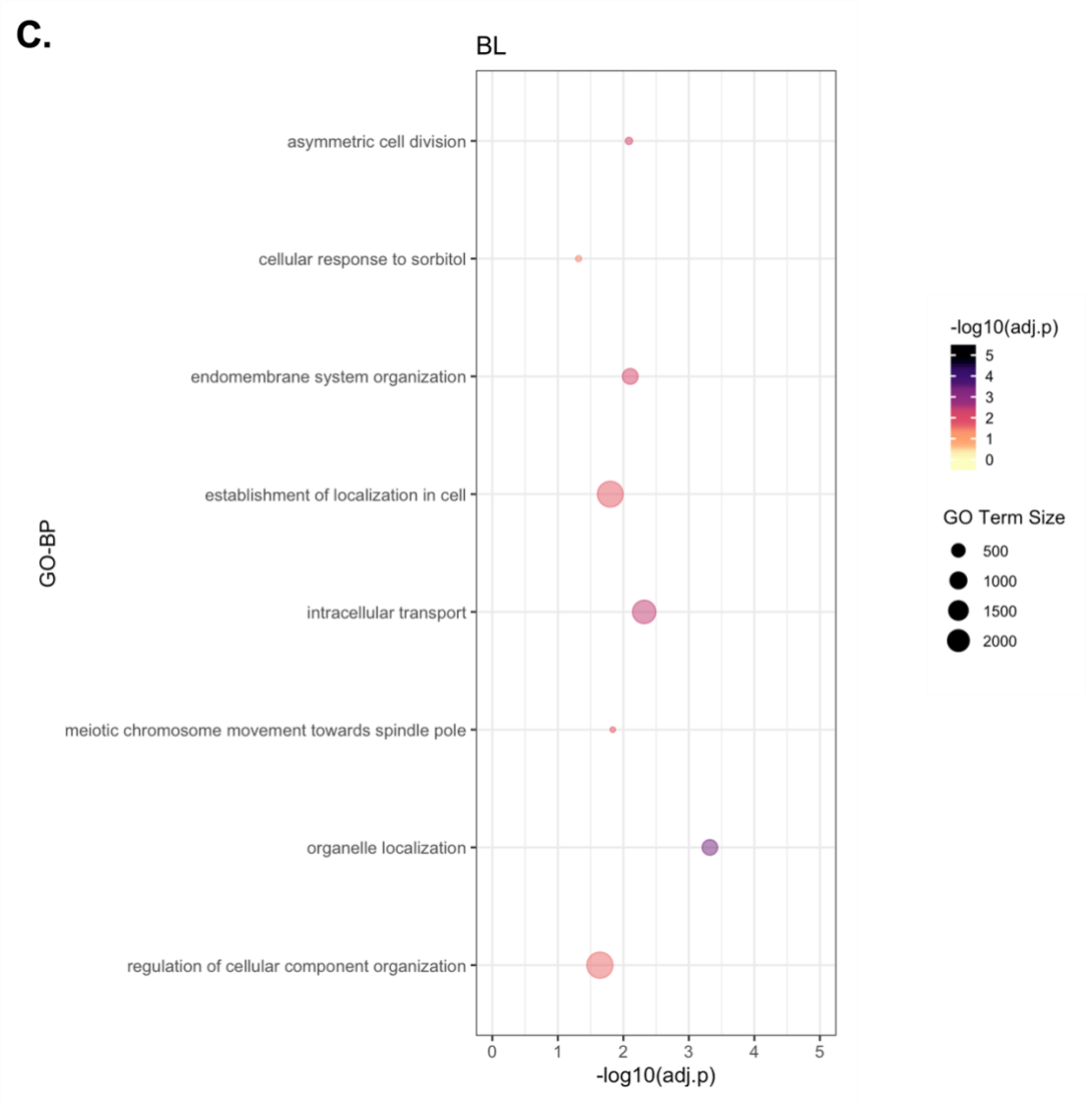


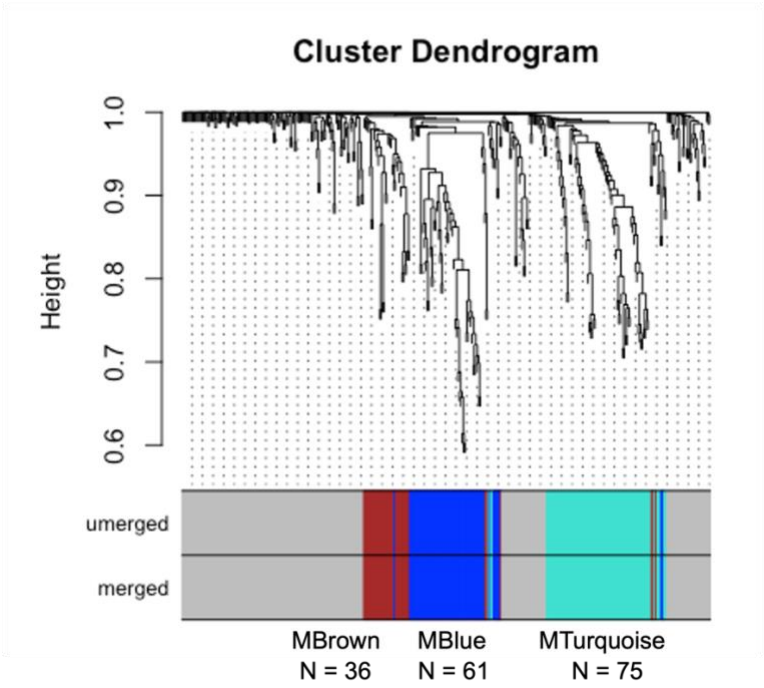
Figure 36. PD-associated LRRK2 co-expression analysis

A) *The box plot shows the distribution of LRRK2:interactor co-expression in the control, sPD and LRRK2-PD cohorts. One-way ANOVA found no significant difference in the LRRK2:interactor co-expression among the 3 cohorts;* ***B)*** *The Veen graph shows the intersection of interactors that were co-expressed with LRRK2 in the LRRK2-PD cases, sPD cases and controls at baseline of diagnosis;* ***C)*** *The bubble graph shows the GO-BPs enriched for interactors that exhibited co-expression with LRRK2 at BL , in which the bubble size represents term size: the larger the bubble, the larger the term size, the more genes were annotated with a given GO-BP term in Gene Ontology; while bubble colour represents the significance of enrichment: the darker the bubble, the lower the adjusted p-value, the more significant a GO-BP terms were associated with the given protein list.*

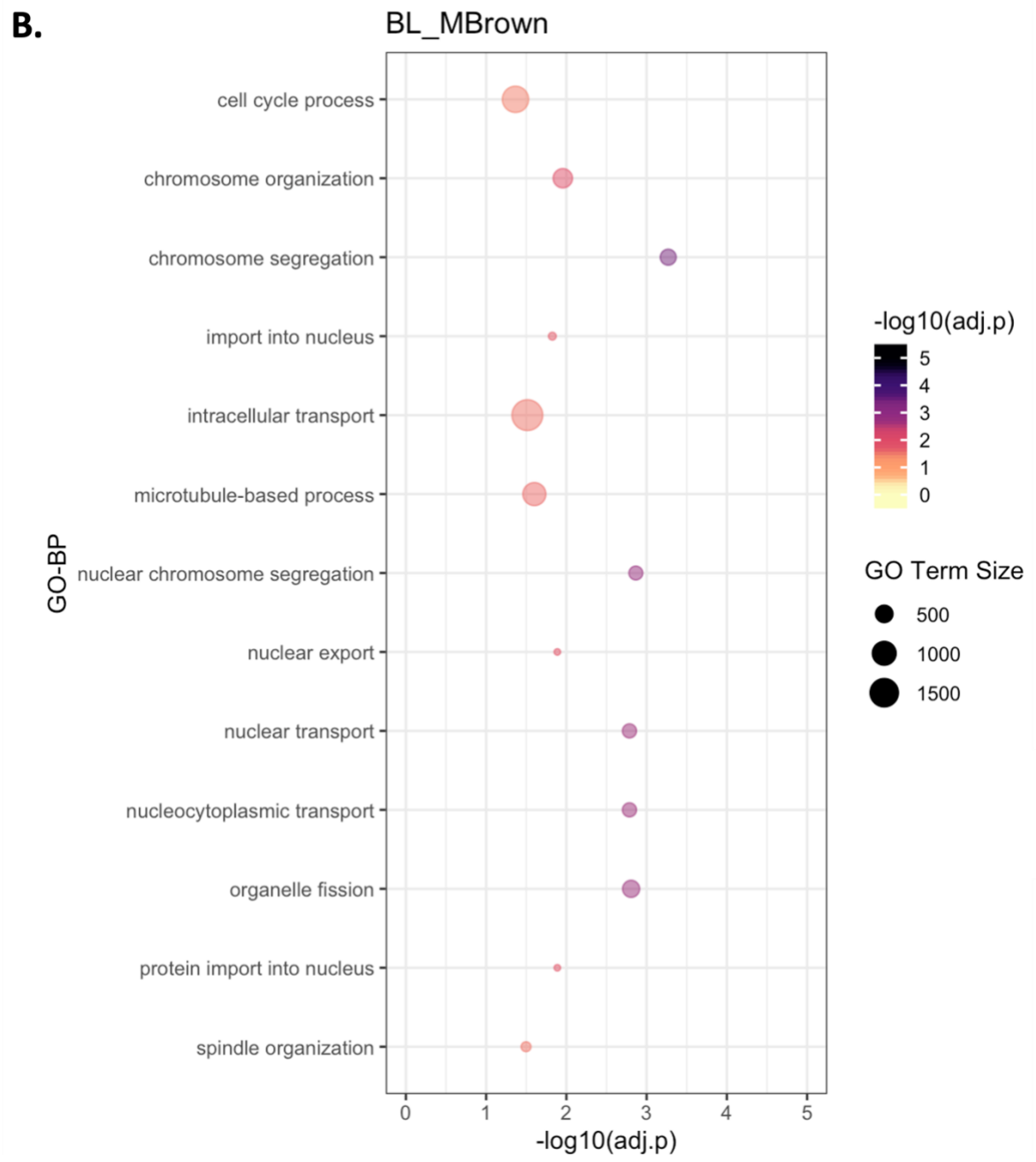
- PD-associated WGCNA on LRRK2 interactors

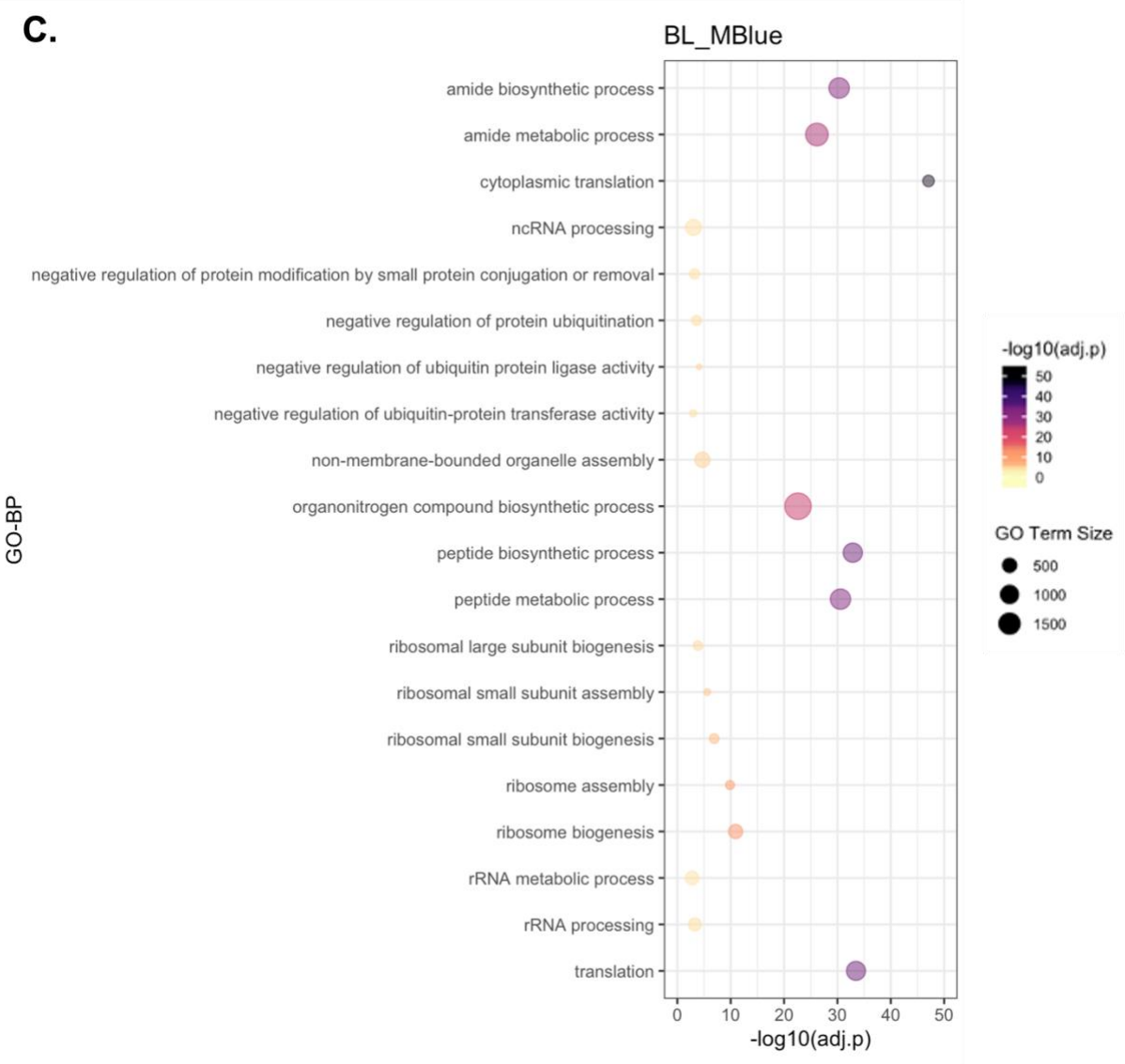
Co-expression levels among LRRK2 interactors at baseline were calculated via Pearson's correlation test. A signed gene co-expression network was constructed with the soft power β = 19, within which a total of 3 co-expression modules were identified: MBrown (N = 36 interactors), MBlue (N = 61 interactors), and MTurquoise (N = 75 interactors) (**Figure 30A**). Functional enrichment analysis associated MBrown with nucleus transport and cell cycle; while MBlue was related with protein biosynthesis, ribosomal functions and negative regulation of protein ubiquitination; while MTurquoise was enriched for regulation of protein metabolism (**Figure 30B-D**). These 3 co-expression modules were conserved in both healthy and PD status regardless of the presence of LRRK2 pathogenic mutations. Module-Trait correlation analysis showed that the eigengene of MBlue (MEBlue) was significantly down-regulated in the LRRK2-PD and sPD cases as compared to control cohort ($p < 0.05$), while MEBrown was only down-regulated in the LRRK2-PD cohort, suggesting that these 2 functional units were potentially negatively affected by PD progression and/or LRRK2 variants (**Figure 30E**). No significant alteration was observed in METurquoise in the PD cohorts as compared to controls, suggesting that the functional unit of protein metabolism maintained stable at the early stage of PD.

A.

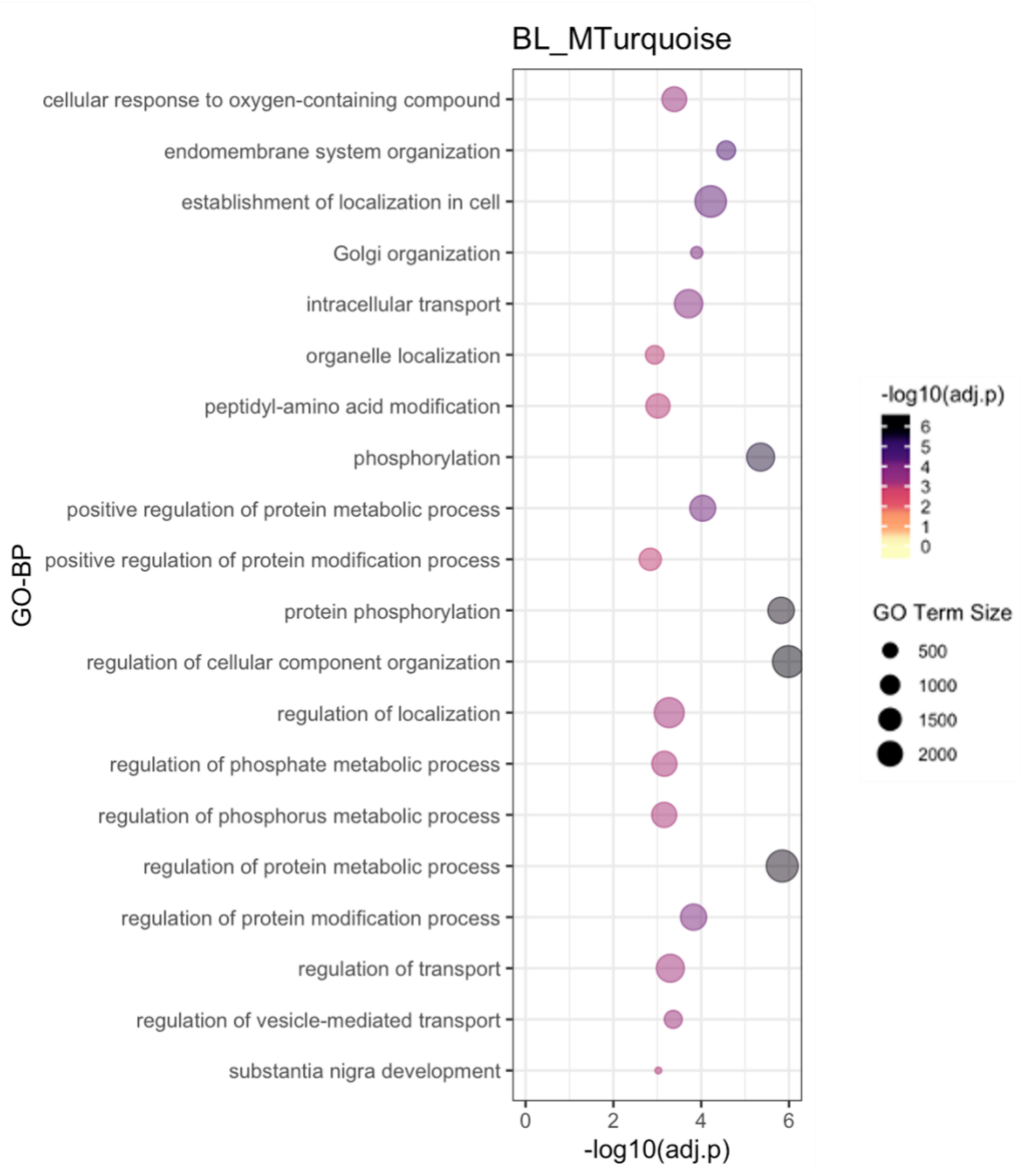


B.


BL_MBrown
cell cycle process
chromosome organization
chromosome segregation
import into nucleus
intracellular transport
microtubule-based process
nuclear chromosome segregation
nuclear export
nuclear transport
nucleocytoplasmic transport
organelle fission
protein import into nucleus
spindle organization
GO-BP
0
1
2
3
4
5
-log10(adj.p)
-log10(adj.p)
5
4
3
2
1
0
GO Term Size
500
1000
1500


C.
BL_MBlue
amide biosynthetic process
amide metabolic process
cytoplasmic translation
ncRNA processing
negative regulation of protein modification by small protein conjugation or removal
negative regulation of protein ubiquitination
negative regulation of ubiquitin protein ligase activity
negative regulation of ubiquitin-protein transferase activity
non-membrane-bounded organelle assembly
organonitrogen compound biosynthetic process
peptide biosynthetic process
peptide metabolic process
ribosomal large subunit biogenesis
ribosomal small subunit assembly
ribosomal small subunit biogenesis
ribosome assembly
ribosome biogenesis
rRNA metabolic process
rRNA processing
translation
GO-BP
0
10
20
30
40
50
-log10(adj.p)
-log10(adj.p)
50
40
30
20
10
0
GO Term Size
500
1000
1500

D.


BL_MTurquoise
cellular response to oxygen-containing compound
endomembrane system organization
establishment of localization in cell
Golgi organization
intracellular transport
organelle localization
peptidyl-amino acid modification
phosphorylation
positive regulation of protein metabolic process
positive regulation of protein modification process
protein phosphorylation
regulation of cellular component organization
regulation of localization
regulation of phosphate metabolic process
regulation of phosphorus metabolic process
regulation of protein metabolic process
regulation of protein modification process
regulation of transport
regulation of vesicle-mediated transport
substantia nigra development
GO-BP
0
2
4
6
-log10(adj.p)
-log10(adj.p)
6
5
4
3
2
1
0
GO Term Size
500
1000
1500
2000

E.

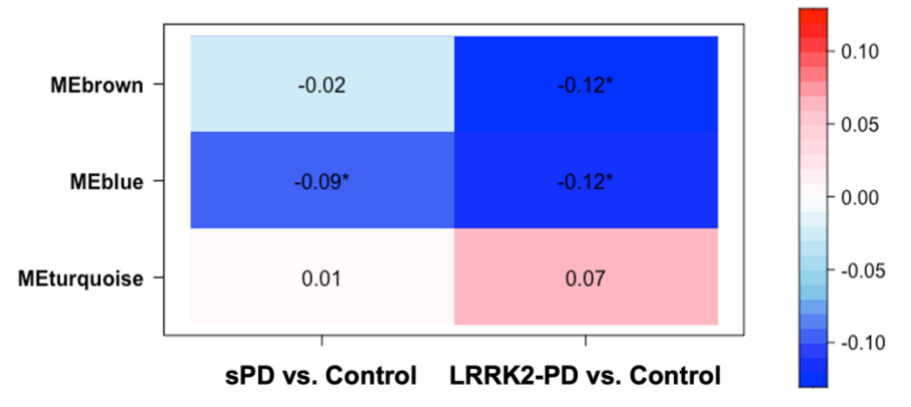

MEbrown
MEblue
MEturquoise
-0.02
-0.12*
-0.09*
-0.12*
0.01
0.07
sPD vs. Control
LRRK2-PD vs. Control
0.10
0.05
0.00
-0.05
-0.10

Figure 37. WGCNA on LRRK2 interactors across the 3 cohorts

A) *The dendrogram shows the co-expression modules identified in LRRK2 interactors among the 3 cohorts at BL. Modules are represented by colours (MBrown, MBlue and MTurquoise).* ***B-D)*** *The bubble graphs show GO-BPs enriched for MBrown, MBlue and MTurquoise; respectively. Bubble colour represents the significance of enrichment while bubble size represents term size;* ***E)*** *The heatmap shows the Module-Trait correlation between the eigengene of the 3 co-expression modules (MEbrown, MEblue and MEturquoise) and PD phenotype. The numbers in cells and cell colours represent Pearson's coefficient. Significant correlation was defined as Pearson's p-value < 0.05 (marked with *).*

- PD-associated edge weight analysis on $LRRK2_{net}$'s topological clusters

Weighted network analysis was performed on the 7 topological clusters identified in Chapter 2. In Cluster A, a total of 72 (62.6%) and 57 (49.5%) edges were down-regulated in the sPD and LRRK2-PD cohorts, respectively, which were significantly higher than the percentages of up-regulated and unchanged edges (**Figure 31A,B**). Of note, the ribosomal unit was highly affected in the sPD and LRRK2-PD cases, involving decrease in both co-expression and expression levels among ribosomal proteins. In addition, the expression level of the seed protein CDK2 significantly increased in these 2 PD conditions as compared to the control cohort. CDK2 serves as a bridge node connecting the ribosomal unit with the other half of the network. Therefore, increased CDK2 level may function as a compensation decreased ribosomal function.

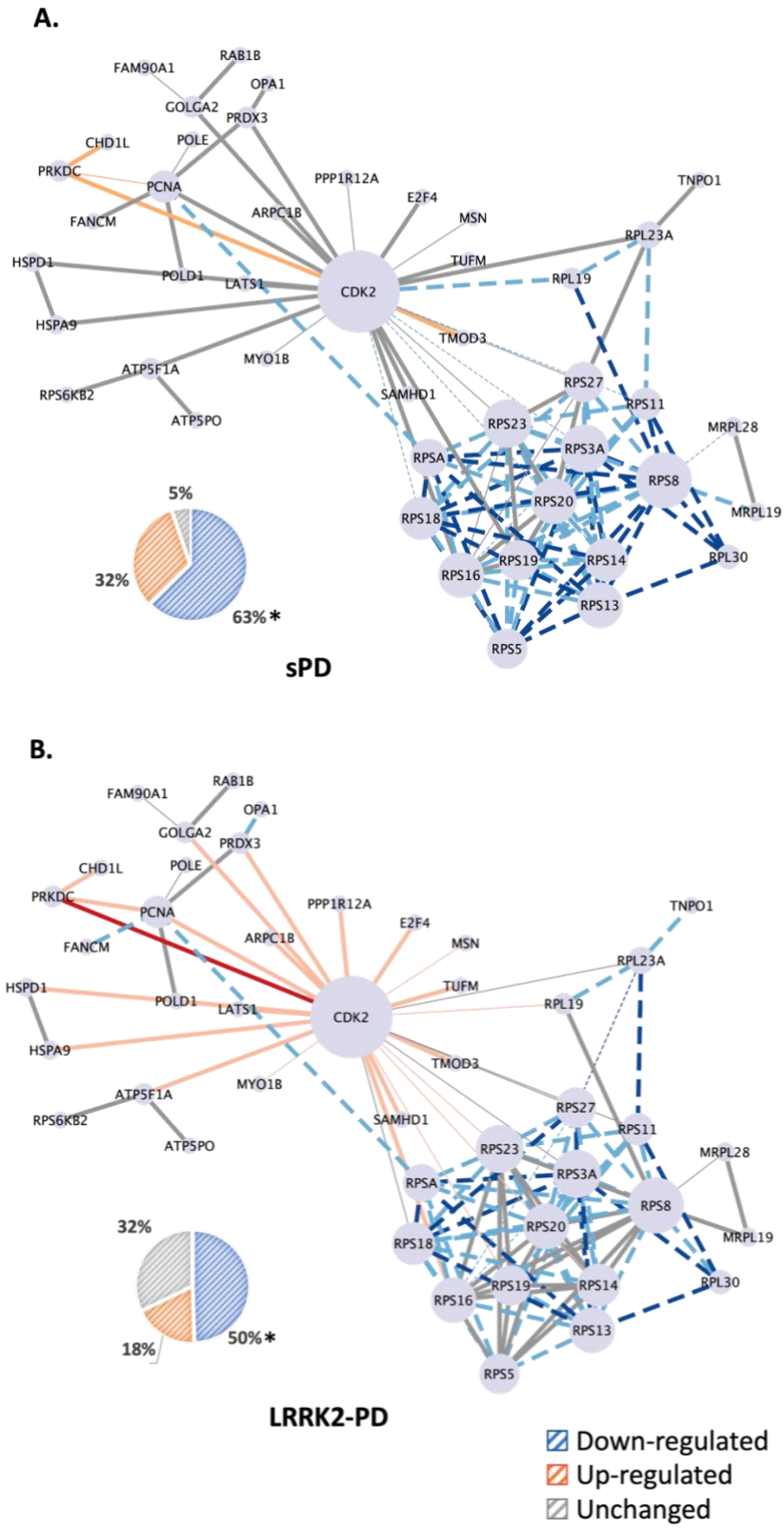


Figure 38. PD-associated edge weight analysis on Topological Cluster A

The network graphs show how Cluster A was affected by the sPD and LRRK2-PD. Edge width represents the co-expression between 2 proteins. Thicker edges represent significant co-expression (Pearson's correlation coefficient > 0.5). Edge colour represents alteration in expression levels of the connected

proteins, in which red edges represent that both of the 2 proteins were significantly up-regulated in PD condition as compared to the controls; orange edges represent that 1 of the 2 proteins were significantly up-regulated in PD conditions as compared to the controls; light blue edges represent that 1 of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls, while dark blue edges represent that both of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls. Grey edges means that none of the 2 connected proteins presented differential expression in the PD cases vs. controls. The pie graphs show the percentages of edges that were up-regulated and down-regulated in the 2 PD conditions.

In comparison, Cluster B remained stable in the sPD cases, with 35 (83%) edges with unchanged weight, 6 (14%) with up-regulated weight and 1 (3%) with down-regulated weight (**Figure 32A**). In the LRRK2-PD cohort, a total of 24 (57%) edges were up-regulated, while 18 (43%) edges were unchanged (**Figure 32B**). No edge was down-regulated in the LRRK2-PD cohort. Of note, the up-regulated edges in the LRRK2-PD condition result from the increased expression level of hub protein TRADD, suggesting that LRRK2 may function as a mediator of TRADD expression and the pathogenic variants in the *LRRK2* gene can cause dysregulation of TRADD level in the blood.

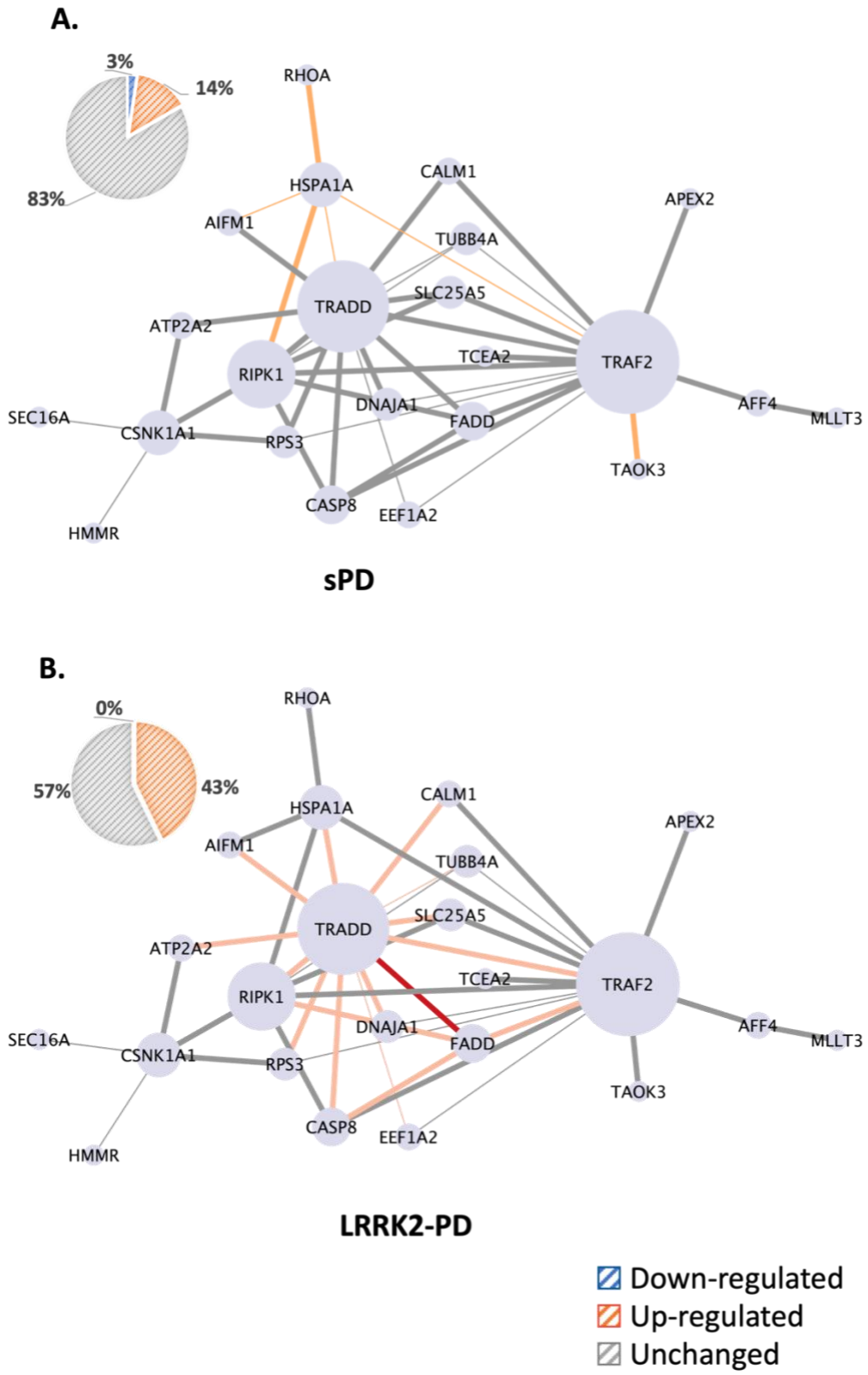


Figure 39. PD-associated edge weight analysis on Topological Cluster B

The network graphs show how Cluster B was affected by the sPD and LRRK2-PD. Edge width represents the co-expression between 2 proteins. Thicker edges represent significant co-expression (Pearson's correlation coefficient > 0.5). Edge colour represents alteration in expression levels of the connected

proteins, in which red edges represent that both of the 2 proteins were significantly up-regulated in PD condition as compared to the controls; orange edges represent that 1 of the 2 proteins were significantly up-regulated in PD conditions as compared to the controls; light blue edges represent that 1 of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls, while dark blue edges represent that both of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls. Grey edges means that none of the 2 connected proteins presented differential expression in the PD cases vs. controls. The pie graphs show the percentages of edges that were up-regulated and down-regulated in the 2 PD conditions.

Similarly, the edge weight of Cluster C remained stable, with a total of 66 (64%) and 75 (72%) edges unchanged in the 2 PD conditions (**Figure 33A,B**). In the sPD cases, 37 edges were up-regulated because of the increased expression of the seed protein IQGAP1, while such alteration was not observed in the LRRK2-PD condition, suggesting this pathological change may rely on the normal LRRK2 structure. In addition, there were 10 edges with decreased weight in the LRRK2-PD cases, which was due to the down-regulation of HSPA4 and HSP0AB1, suggesting that the expression level of these 2 proteins may be regulated by the activity of LRRK2 and can potentially be affected by its pathogenic mutations.

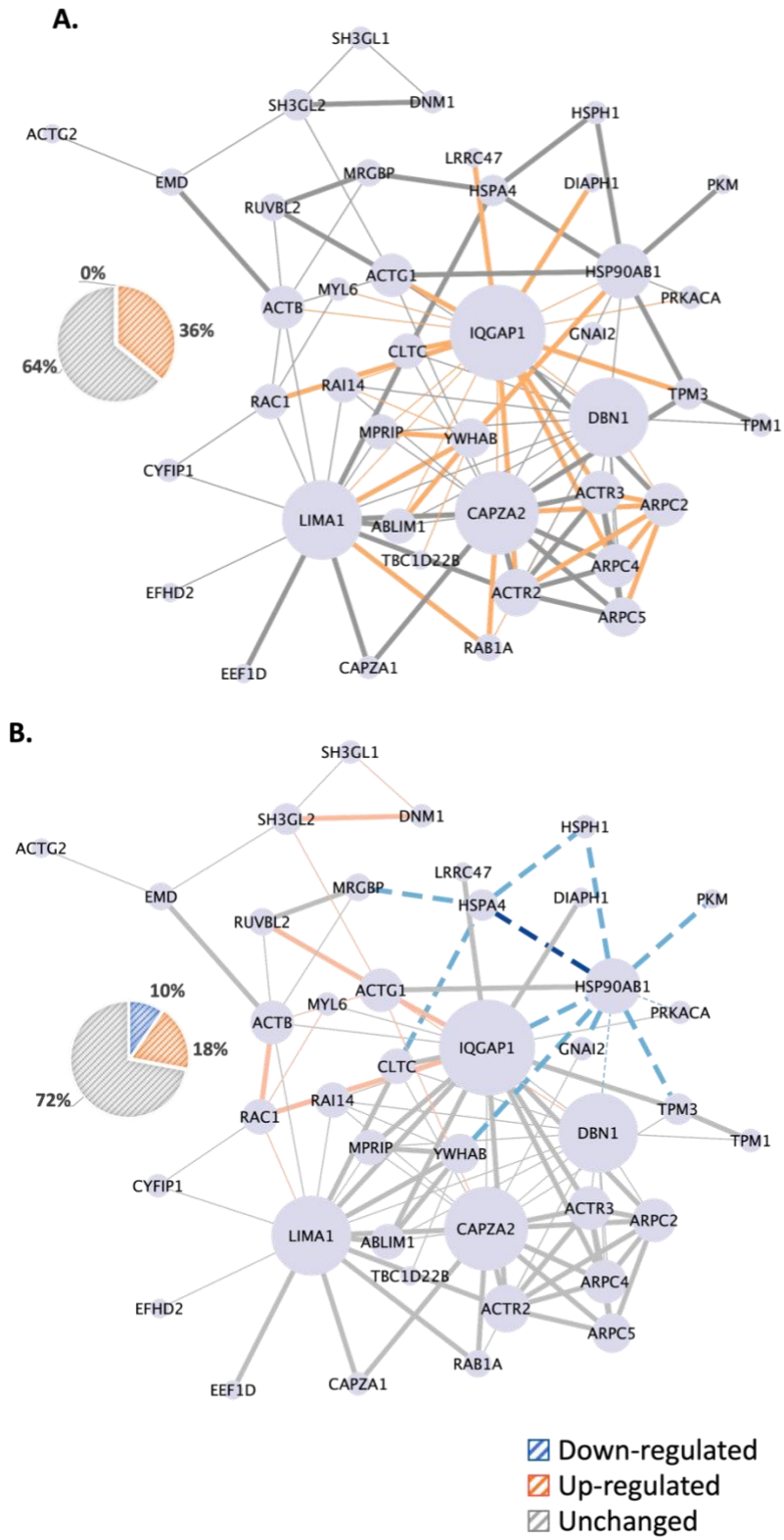


Figure 40. PD-associated edge weight analysis on Topological Cluster C

The network graphs show how C was affected by the sPD and LRRK2-PD. Edge width represents the co-expression between 2 proteins. Thicker edges represent significant co-expression (Pearson's correlation coefficient > 0.5). Edge colour represents alteration in expression levels of the connected

proteins, in which red edges represent that both of the 2 proteins were significantly up-regulated in PD condition as compared to the controls; orange edges represent that 1 of the 2 proteins were significantly up-regulated in PD conditions as compared to the controls; light blue edges represent that 1 of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls, while dark blue edges represent that both of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls. Grey edges means that none of the 2 connected proteins presented differential expression in the PD cases vs. controls. The pie graphs show the percentages of edges that were up-regulated and down-regulated in the 2 PD conditions.

In Cluster D, most of edges remained unchanged in the 2 PD conditions (88% in the sPD condition while 92% in the LRRK2-PD) (**Figure 34A,B**). Moreover, 7 edges were down-regulated in the LRRK2-PD condition due to decreased NPM1 expression. Of note, NPM1 is one of the hub proteins in Cluster D which connect to 8/10 ribosomal proteins in the network. Therefore, down-regulation of NPM1 may affect its interaction with these proteins and thereby negatively affect the ribosomal functions.

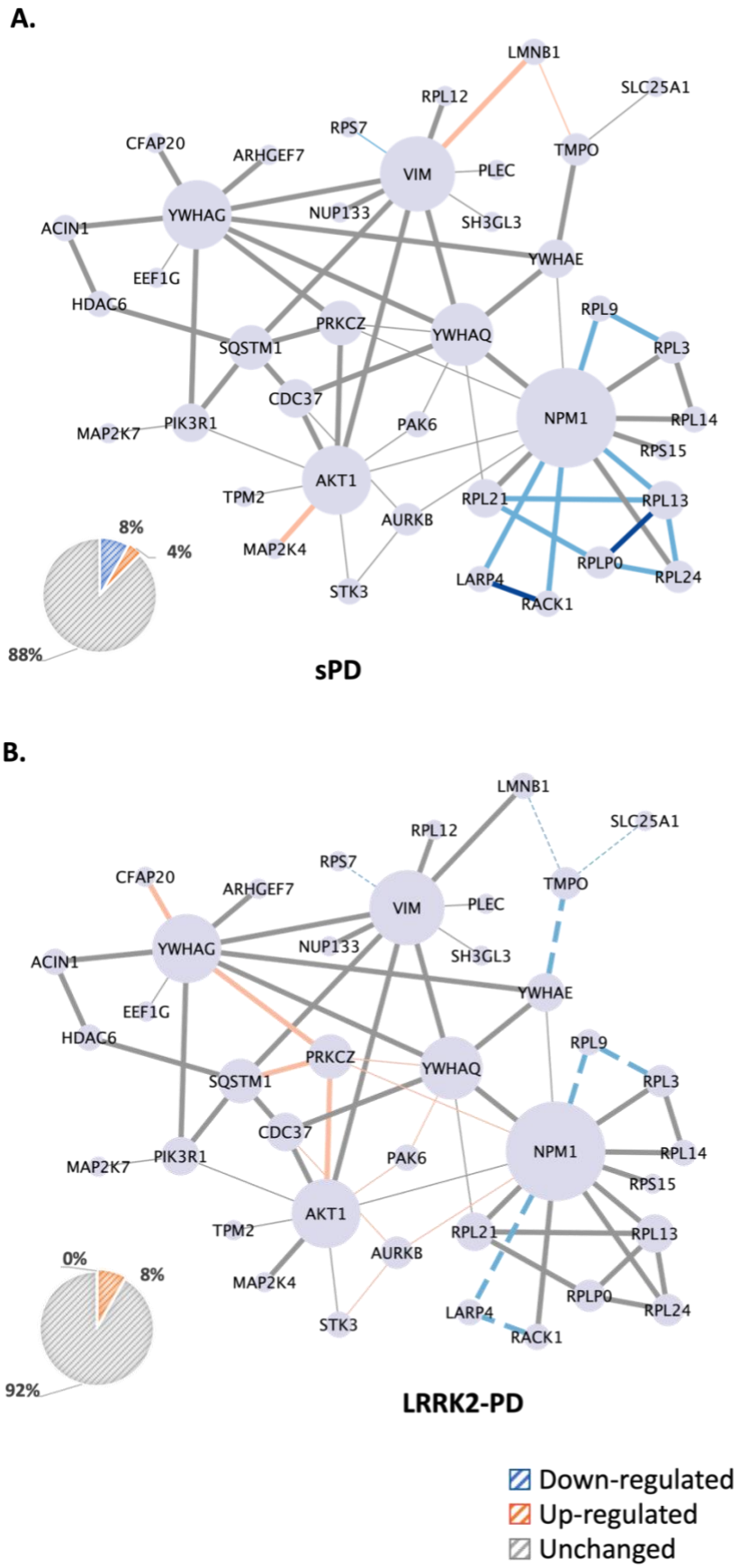

A.
LMNB1
SLC25A1
RPL12
RPS7
CFAP20
ARHGEF7
TMPO
VIM
PLEC
YWHAG
NUP133
SH3GL3
ACIN1
YWHAE
EEF1G
HDAC6
RPL9
PRKCZ
YWHAQ
SQSTM1
RPL3
CDC37
NPM1
RPL14
MAP2K7
PIK3R1
PAK6
RPS15
AKT1
TPM2
RPL21
RPL13
AURKB
8%
4%
MAP2K4
RPLP0
RPL24
LARP4
STK3
RACK1
88%
sPD
B.
LMNB1
SLC25A1
RPL12
RPS7
CFAP20
ARHGEF7
TMPO
VIM
PLEC
YWHAG
NUP133
SH3GL3
ACIN1
YWHAE
EEF1G
HDAC6
RPL9
PRKCZ
YWHAQ
SQSTM1
RPL3
CDC37
NPM1
RPL14
MAP2K7
PIK3R1
PAK6
RPS15
AKT1
TPM2
RPL21
RPL13
AURKB
0%
8%
MAP2K4
RPLP0
RPL24
LARP4
STK3
RACK1
92%
LRRK2-PD
Down-regulated
Up-regulated
Unchanged

Figure 41. PD-associated edge weight analysis on Topological Cluster D

The network graphs show how Cluster D was affected by the sPD and LRRK2-PD. Edge width represents the co-expression between 2 proteins. Thicker edges represent significant co-expression (Pearson's correlation coefficient > 0.5). Edge colour represents alteration in expression levels of the connected proteins, in which red edges represent that both of the 2 proteins were significantly up-regulated in PD condition as compared to the controls; orange edges represent that 1 of the 2 proteins were significantly up-regulated in PD conditions as compared to the controls; light blue edges represent that 1 of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls, while dark blue edges represent that both of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls. Grey edges means that none of the 2 connected proteins presented differential expression in the PD cases vs. controls. The pie graphs show the percentages of edges that were up-regulated and down-regulated in the 2 PD conditions.

In terms of Cluster E, a total of 50 (77%) and 41 (63%) edges remained unchanged in the 2 PD conditions (**Figure 35**). In the sPD cases, a unit formed by TP53, ITCH, SFN, POU5F1, DVL2 and DVL3 was up-regulated because of a significant increased ITCH expression. GO-BP enrichment analysis associated this unit with RNA biosynthesis and metabolism, suggesting that RNA transcription may increase in the early stage of sPD. However, such alteration was not observed in the LRRK2-PD cases. Instead, a unit of KPNB1, UNP160, UNP170 and TPR was down-regulated as compared to the controls, which result from the decreased level of NUP107. This unit presented high co-expression and was associated with intracellular transport of RNA and protein.These findings suggest that gene translation may be differently affected by the sPD and LRRK2-PD.

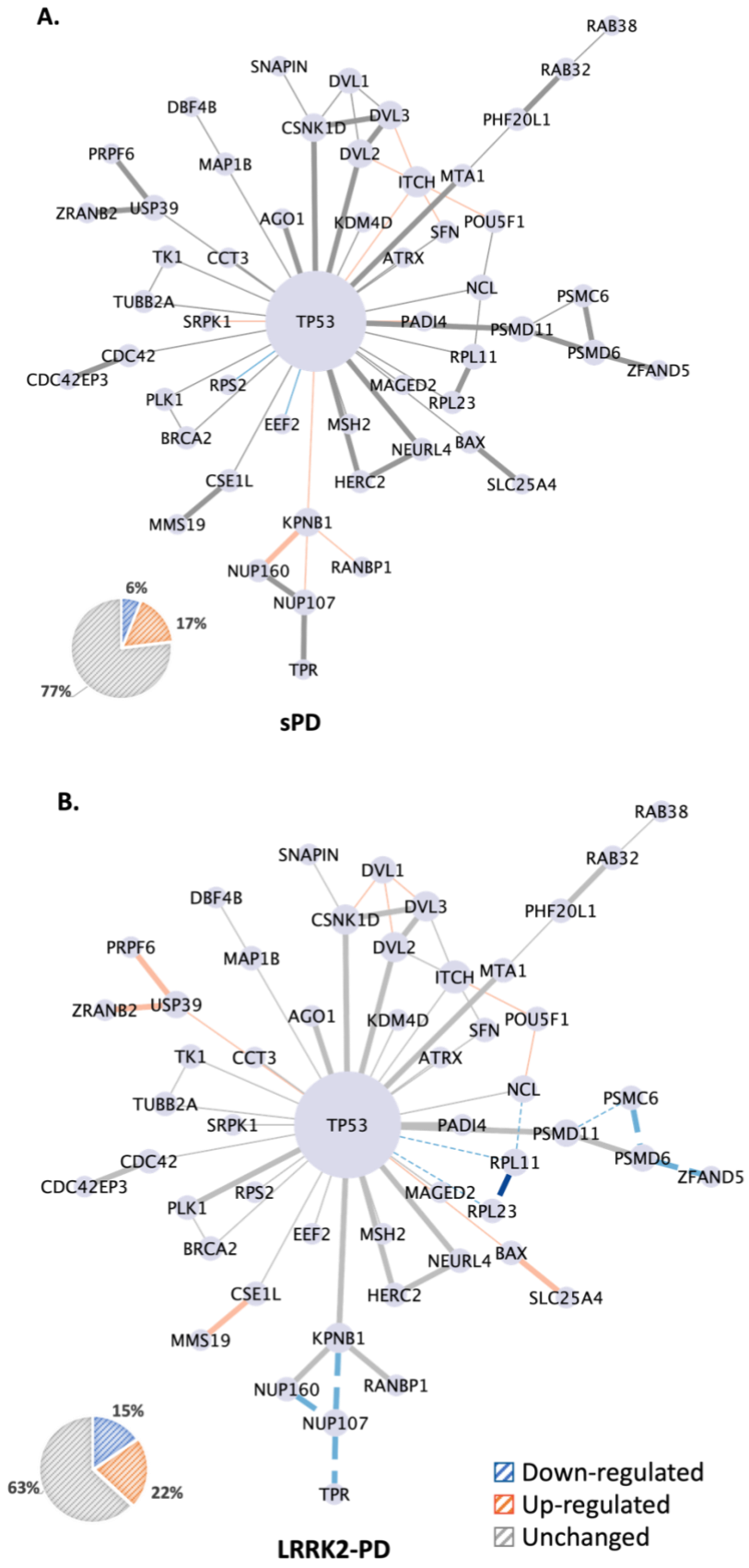


Figure 42. PD-associated edge weight analysis on Topological Cluster E

The network graphs show how Cluster E was affected by the sPD and LRRK2-PD. Edge width represents the co-expression between 2 proteins. Thicker edges represent significant co-expression (Pearson's correlation coefficient > 0.5). Edge colour represents alteration in expression levels of the connected proteins, in which red edges represent that both of the 2 proteins were significantly up-regulated in PD condition as compared to the controls; orange edges represent that 1 of the 2 proteins were significantly up-regulated in PD conditions as compared to the controls; light blue edges represent that 1 of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls, while dark blue edges represent that both of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls. Grey edges means that none of the 2 connected proteins presented differential expression in the PD cases vs. controls. The pie graphs show the percentages of edges that were up-regulated and down-regulated in the 2 PD conditions.

In comparison, edge weight of Cluster F increased significantly in the LRRK2-PD condition, with 76% of edges up-regulated and 24% unchanged (**Figure 36B**). The alteration results from increased expression levels of the hub protein PRKN, TUBG1, RAB8A and RAB10. These proteins are crucial contributors of autophagy, suggesting that pathogenic mutation is linked to up-regulated autophagy in the early stage of PD, while this alteration was not observed in the sPD situation, where 75% (N = 16) of the edges remained unchanged (**Figure 36A**).

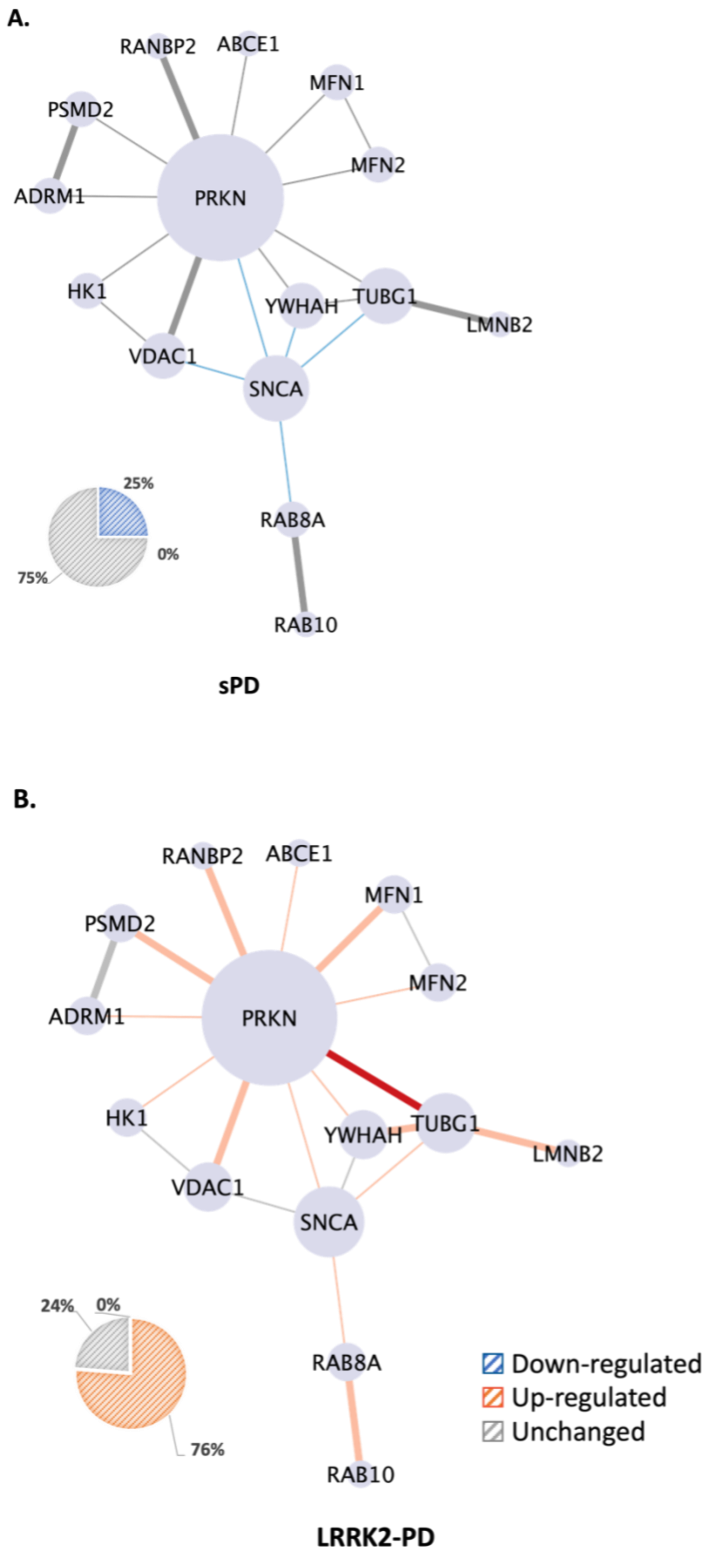


Figure 43. PD-associated edge weight analysis on Topological Cluster F

The network graphs show how Cluster F was affected by the sPD and LRRK2-PD. Edge width represents the co-expression between 2 proteins. Thicker edges represent significant co-expression (Pearson's correlation coefficient > 0.5). Edge colour represents alteration in expression levels of the connected

proteins, in which red edges represent that both of the 2 proteins were significantly up-regulated in PD condition as compared to the controls; orange edges represent that 1 of the 2 proteins were significantly up-regulated in PD conditions as compared to the controls; light blue edges represent that 1 of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls, while dark blue edges represent that both of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls. Grey edges means that none of the 2 connected proteins presented differential expression in the PD cases vs. controls. The pie graphs show the percentages of edges that were up-regulated and down-regulated in the 2 PD conditions.

A total of 37 (74%) and 40 (80%) edges in Cluster G remained unchanged in the 2 PD conditions. However, a unit of BAG1, BAG2, BAG3, STUB1, DNAJB6, HIF1A, LAMP2, RGS2 and the 2 hub proteins LRRK1, HSPA8 were un-regulated in both sPD and LRRK2-PD conditions, involving upregulation of 12 (24%) edges and 7 (14%) edges in the network. This unit is enriched for GO-BPs regarding protein metabolism, suggesting that protein dynamics is enhanced in the early stage of PD (**Figure 37**).

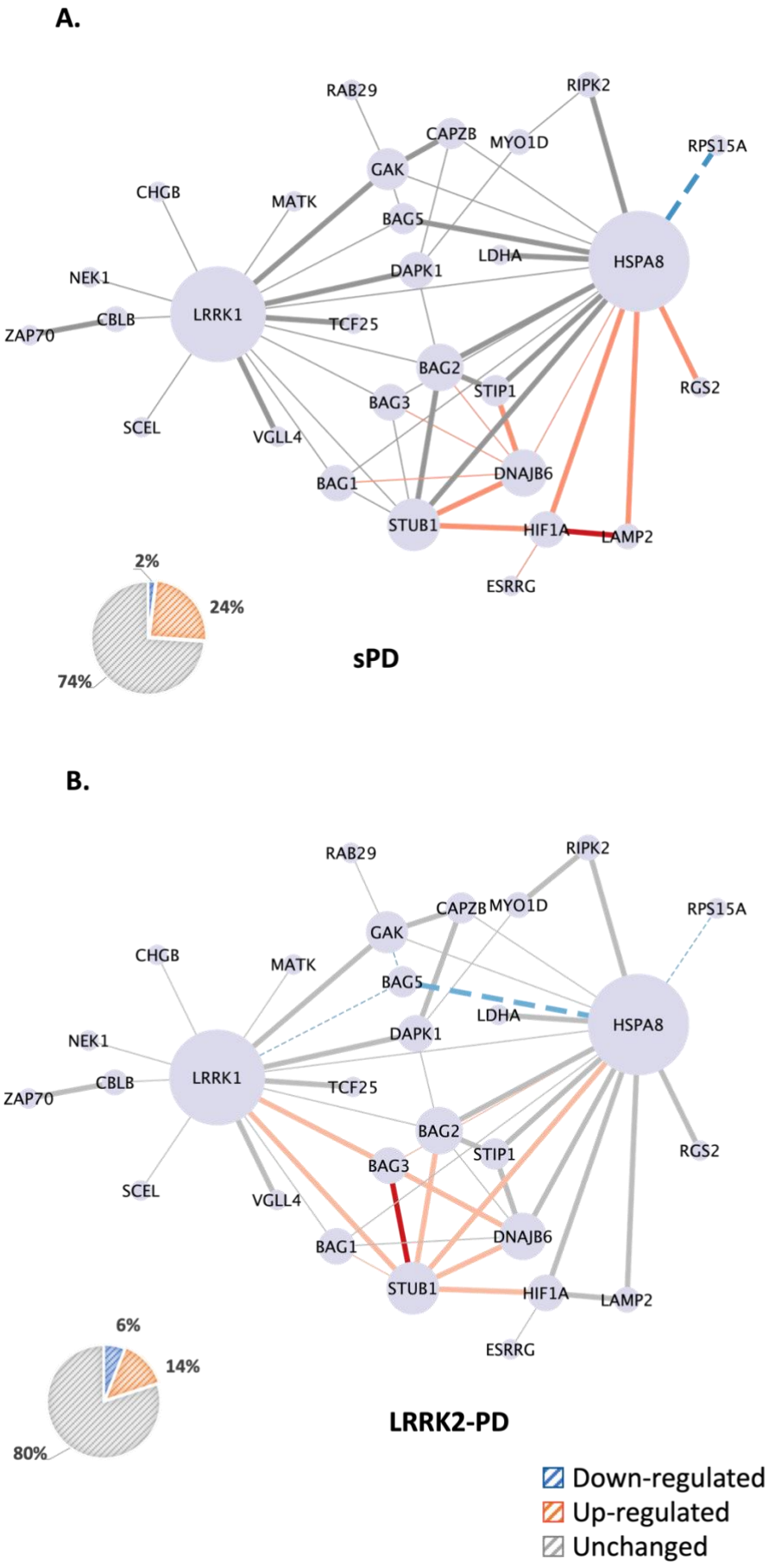

A.
RAB29
RIPK2
CAPZB
MYO1D
RPS15A
GAK
CHGB
MATK
BAG5
HSPA8
LDHA
DAPK1
NEK1
LRRK1
CBLB
ZAP70
TCF25
BAG2
STIP1
RGS2
BAG3
SCEL
VGLL4
DNAJB6
BAG1
STUB1
HIF1A
LAMP2
ESRRG
2%
24%
74%
sPD
B.
RAB29
RIPK2
CAPZB
MYO1D
RPS15A
GAK
CHGB
MATK
BAG5
LDHA
HSPA8
DAPK1
NEK1
LRRK1
CBLB
ZAP70
TCF25
BAG2
STIP1
RGS2
BAG3
SCEL
VGLL4
DNAJB6
BAG1
STUB1
HIF1A
LAMP2
ESRRG
6%
14%
80%
LRRK2-PD
Down-regulated
Up-regulated
Unchanged

Figure 44. PD-associated edge weight analysis on Topological Cluster G

The network graphs show how Cluster G was affected by the sPD and LRRK2-PD. Edge width represents the co-expression between 2 proteins. Thicker edges represent significant co-expression (Pearson's correlation coefficient > 0.5). Edge colour represents alteration in expression levels of the connected proteins, in which red edges represent that both of the 2 proteins were significantly up-regulated in PD condition as compared to the controls; orange edges represent that 1 of the 2 proteins were significantly up-regulated in PD conditions as compared to the controls; light blue edges represent that 1 of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls, while dark blue edges represent that both of the 2 proteins were significantly down-regulated in PD conditions as compared to the controls. Grey edges means that none of the 2 connected proteins presented differential expression in the PD cases vs. controls. The pie graphs show the percentages of edges that were up-regulated and down-regulated in the 2 PD conditions.

- Single SNP association analysis

After QC, a total of 136, 3887 and 4047 high-quality, protein-coding SNPs of LRRK2 interactors remained in the SNP sets of "Immuno", "NeuroX1" and "NeuroX2", respectively. Of note, "Immuno" only contains SNPs for 42 out of 418 (10.3%) LRRK2 interactors and was thereby discarded, while "NeuroX1" and "NeuroX2" contain SNPs for 323 and 322 LRRK2 interactors, respectively. "NeuroX1" was used to evaluate the association between SNPs of LRRK2 interactors and sPD (involving 339 sPD patients and 154 controls), while a merged list of "NeuroX1" and "NeuroX2" (N = 2091 SNPs, involving 49 controls and 41 LRRK2 pathogenic variant carriers) was used to assess the association between SNPs and LRRK2-PD. After adjusted by gender and 7 PCs (λ = 1.000), the logistic regression on "NeuroX1" set returned 23 SNPs with MAF higher ($p < 0.05$) in sPD patients as compared to controls, namely rs151264467, rs8178046, rs3218772, rs6133278, rs79181168, rs2230801, rs116938571, rs200046311, rs78436829, rs35625617, rs55945045, rs117843818, rs4841, rs35303786, rs41286651, rs35589976, rs1056719, rs5030752, rs1049951, rs8178017, rs79308175, rs148283548, rs3194151 (**Figure 38, Table 9**). Among these SNPs, only rs151264467, which is located on the *TTC27* passed Bonferroni's correction (adjusted-p = 0.006). Burden analysis showed a total of 15 genes associated sPD, including TUBB6, TRAF2, PRKDC, TK1, DIDO1,

KIF2A, RPL11, MFN2, CDKL3, RAB29, DYNC1H1, CYREN, DVL3, TAOK3, STK40. Only 3 GO-BPs were associated with these genes:

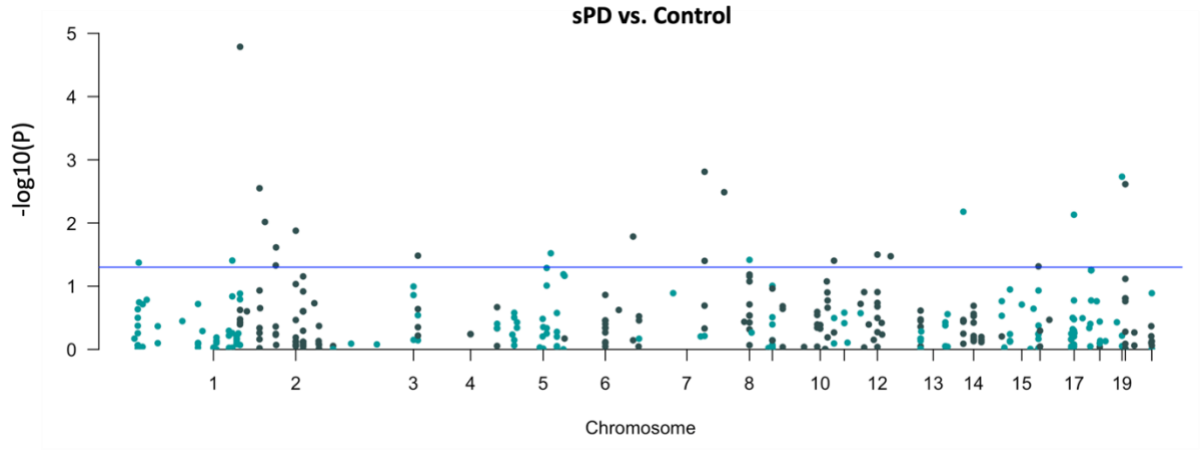


Figure 45. Manhattan plot for single SNP association analysis on LRRK2 interactors (sPD vs. controls)

Line denotes statistical significance ($p < 0.05$). The results have been adjusted for a genomic control inflation factor $\lambda = 1.000$ (sample size = 493).

Table 9. SNPs of LRRK2 interactors associated with sPD

Chr	SNP	p	Adjusted-p
2	rs151264467	<0.001	0.006
8	rs8178046	0.002	0.544
19	rs3218772	0.002	0.651
20	rs6133278	0.002	0.855
2	rs79181168	0.003	0.992
8	rs2230801	0.003	1.000
13	rs116938571	0.007	1.000
17	rs200046311	0.007	1.000
2	rs78436829	0.010	1.000
2	rs35625617	0.013	1.000
6	rs55945045	0.016	1.000
2	rs117843818	0.024	1.000
5	rs4841	0.030	1.000

12	rs35303786	0.032	1.000
4	rs41286651	0.033	1.000
12	rs35589976	0.034	1.000
9	rs1056719	0.038	1.000
1	rs5030752	0.039	1.000
10	rs1049951	0.040	1.000
8	rs8178017	0.040	1.000
1	rs79308175	0.042	1.000
2	rs148283548	0.047	1.000
16	rs3194151	0.049	1.000

In comparison, the logistic regression on "Neuro_merge" set was adjusted by gender and 13 PCs (λ = 1.029) and returned 19 SNPs with MAF higher ($p < 0.05$) in LRRK2 variant carriers as compared to controls, namely rs5030752, rs8178046, rs7170637, rs142831657, rs4842, rs62317770, rs79383654, rs16853333, rs299295, rs118103955, rs1049951, rs143662421, rs2273660, rs79385421, rs9294445, rs34688574, rs34143723, rs2924835, rs114147582 (**Figure 39, Table 10).** None of these SNPs passed Bonferroni's correction. Burden analysis showed a total of 15 genes associated with LRRK2 variants, including ITCH, ANKS4B, RIPK2, MAP2K3, PRKCZ, NFATC2, CHGB, RPS14, ARHGEF7, TRAF2, HSP90AB1, LRRK2, MRPL28, MAP2K6, AP2M1, CAPZB, GAK. A total of 41 GO-BPs were enriched for these genes, which were mainly associated with regulation of protein phosphorylation, vesicle transport, MAKP cascade and cytokine production (**Table 11**).

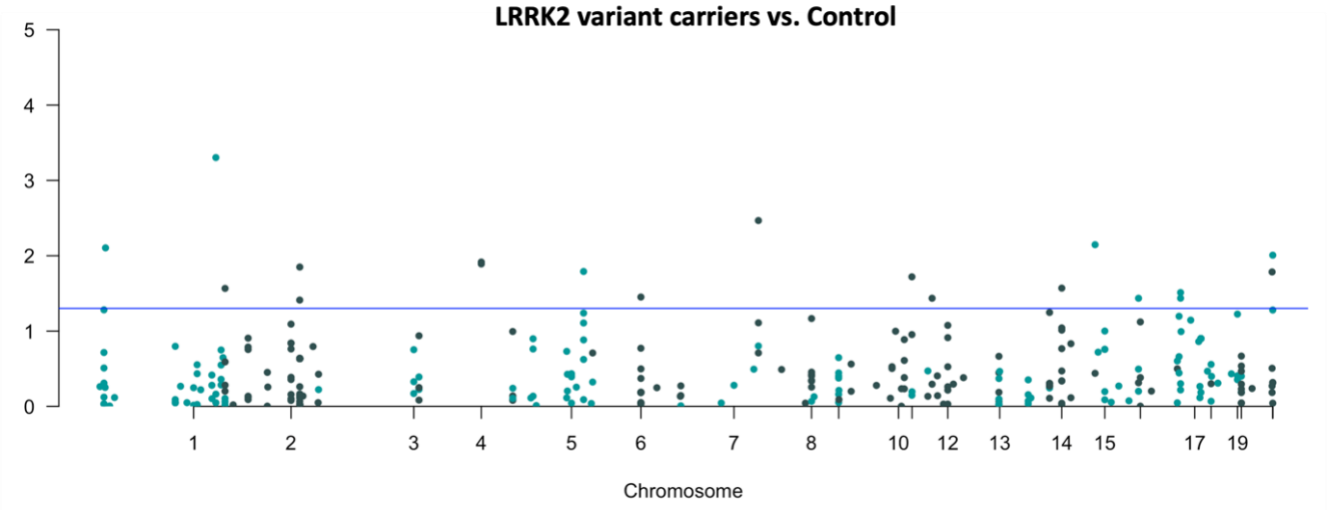


Figure 46. Manhattan plot for single SNP association analysis on LRRK2 interactors (LRRK2 variant carriers vs. controls)

Line denotes statistical significance ($p < 0.05$). The results have been adjusted for a genomic control inflation factor $\lambda = 1.029$ (sample size = 90).

Table 10. SNPs of LRRK2 interactors associated with LRRK2 variants

Chr	SNP	p	Adjusted-p
1	rs5030752	<0.001	0.152
8	rs8178046	0.003	1.000
15	rs7170637	0.007	1.000
1	rs142831657	0.008	1.000
21	rs4842	0.010	1.000
4	rs62317770	0.012	1.000
4	rs79383654	0.013	1.000
2	rs16853333	0.014	1.000
5	rs299295	0.016	1.000
20	rs118103955	0.016	1.000
10	rs1049951	0.019	1.000
14	rs143662421	0.027	1.000
2	rs2273660	0.027	1.000
17	rs79385421	0.031	1.000
6	rs9294445	0.035	1.000
17	rs34688574	0.037	1.000
12	rs34143723	0.037	1.000
15	rs2924835	0.037	1.000
2	rs114147582	0.039	1.000

Table 11. GO-BPs enriched for LRRK2 interactors with significant burden on the presence of LRRK2 pathogenic variants

Term ID	Term name	Term size	Adjusted p-value
GO:0051770	positive regulation of nitric-oxide synthase biosynthetic process	16	<0.001

GO:0043408	regulation of MAPK cascade	643	0.001
GO:0051767	nitric-oxide synthase biosynthetic process	19	0.001
GO:0051769	regulation of nitric-oxide synthase biosynthetic process	19	0.001
GO:0000165	MAPK cascade	745	0.001
GO:0043410	positive regulation of MAPK cascade	461	0.002
GO:0001817	regulation of cytokine production	773	0.002
GO:0001816	cytokine production	779	0.002
GO:0031399	regulation of protein modification process	1229	0.003
GO:0072709	cellular response to sorbitol	4	0.005
GO:0001932	regulation of protein phosphorylation	910	0.005
GO:0006468	protein phosphorylation	1346	0.006
GO:0141124	intracellular signaling cassette	1861	0.006
GO:0001934	positive regulation of protein phosphorylation	603	0.008
GO:0042325	regulation of phosphorylation	966	0.008
GO:0042327	positive regulation of phosphorylation	630	0.010
GO:0042981	regulation of apoptotic process	1463	0.011
GO:0043067	regulation of programmed cell death	1506	0.014
GO:1900244	positive regulation of synaptic vesicle endocytosis	6	0.014
GO:0072708	response to sorbitol	6	0.014
GO:0007254	JNK cascade	173	0.014
GO:0051246	regulation of protein metabolic process	2102	0.017
GO:0045937	positive regulation of phosphate metabolic process	705	0.018
GO:0010562	positive regulation of phosphorus metabolic process	705	0.018
GO:0044093	positive regulation of molecular function	1097	0.019
GO:0016310	phosphorylation	1579	0.019
GO:0019220	regulation of phosphate metabolic process	1132	0.023
GO:0051174	regulation of phosphorus metabolic process	1133	0.023
GO:0031098	stress-activated protein kinase signaling cascade	62	0.024

GO:1903423	positive regulation of synaptic vesicle recycling	8	0.025
GO:0070534	protein K63-linked ubiquitination	64	0.026
GO:0043085	positive regulation of catalytic activity	773	0.031
GO:0031401	positive regulation of protein modification process	787	0.034
GO:0048488	synaptic vesicle endocytosis	72	0.037
GO:0140238	presynaptic endocytosis	73	0.039
GO:2000514	regulation of CD4-positive, alpha-beta T cell activation	75	0.042
GO:0051247	positive regulation of protein metabolic process	1253	0.044
GO:0033554	cellular response to stress	1770	0.044
GO:0010647	positive regulation of cell communication	1777	0.046
GO:0023056	positive regulation of signaling	1778	0.046
GO:0065009	regulation of molecular function	1799	0.050

- Transcriptomic features of LRRK2 interactors differentiated the sPD and LRRK2-PD cases

Model 1 was constructed on the mRNA levels (whole blood) of the 100 LRRK2 interactors with significant differential expression in the sPD and/or LRRK2-PD cohorts (N = 371 and 116, respectively) as compared to the controls. Univariate logistic analyses found a total of 14 interactors for model construction, including STUB1, DVL1, ACTA2, CDK2, MMS19, PRKN, TUBB6, TUBG1, BAG3, HSPA1A, LMNB1, SNCA, RPS2 and SLC25A6 (**Table 12, Table D1**). Model 1 was then trained on a randomly-picked cohort of 296 sPD cases and 92 LRRK2-PD cases using the read counts of the 14 interactors listed above. A λ value of 0.008 (log(λ) = -4.869 was chosen to reach the minimum Mean-Squared Error (MSE), leaving a total of 8 interactors in the model, including STUB1 (coefficient = -1.61), ACTA2 (coefficient = -0.85), PRKN (coefficient = = -0.38), TUBB6 (coefficient = -1.21), HSPA1A (coefficient = 3.12), LMNB1 (coefficient = 0.55), SNCA (coefficient = -0.46) and SLC25A6 (coefficient = -0.68) (**Figure 48A,B**). The cut-off value on the predicted value was optimised as 0.94 to reach the maximum accuracy in the training set (AC_train = 72.9%, **Figure 48C**). The refined model was then

validated on the test set, containing 75 sPD cases and 24 LRRK2-PD cases. ROC curve showed an AUC = 0.68 (95% CI: 0.56-0.80) (**Figure 48D**).

Table 12. Significant LRRK2 interactors selected by univariate analyses for Model 1

Interactor	OR	95% CI		p-value
		Upper Limit	Lower Limit	
STUB1	0.24	0.06	0.91	0.036
DVL1	0.17	0.04	0.68	0.012
ACTA2	0.31	0.13	0.77	0.012
CDK2	0.22	0.06	0.86	0.029
MMS19	0.23	0.06	0.83	0.025
PRKN	0.18	0.06	0.56	0.003
TUBB6	0.18	0.06	0.49	0.001
TUBG1	0.25	0.07	0.90	0.035
BAG3	0.18	0.05	0.61	0.006
HSPA1A	3.23	1.25	8.34	0.016
LMNB1	2.92	1.14	7.44	0.025
SNCA	0.45	0.21	0.96	0.040
RPS2	0.29	0.10	0.86	0.026
SLC25A6	0.24	0.07	0.84	0.025

*Univariate Logistic regression was performed on the expression profile (whole blood mRNA level) of each LRRK2 interactor in the sPD and LRRK2-PD cases to identify those with significant association with phenotype (sPD or LRRK2-PD). A total of 12 LRRK2 interactors were selected by this procedure and were thereby included in further analysis. The complete result for univariate analyses see **Table D1**. Abbreviations: OR: Odd Ratio; CI: Confidence Interval*

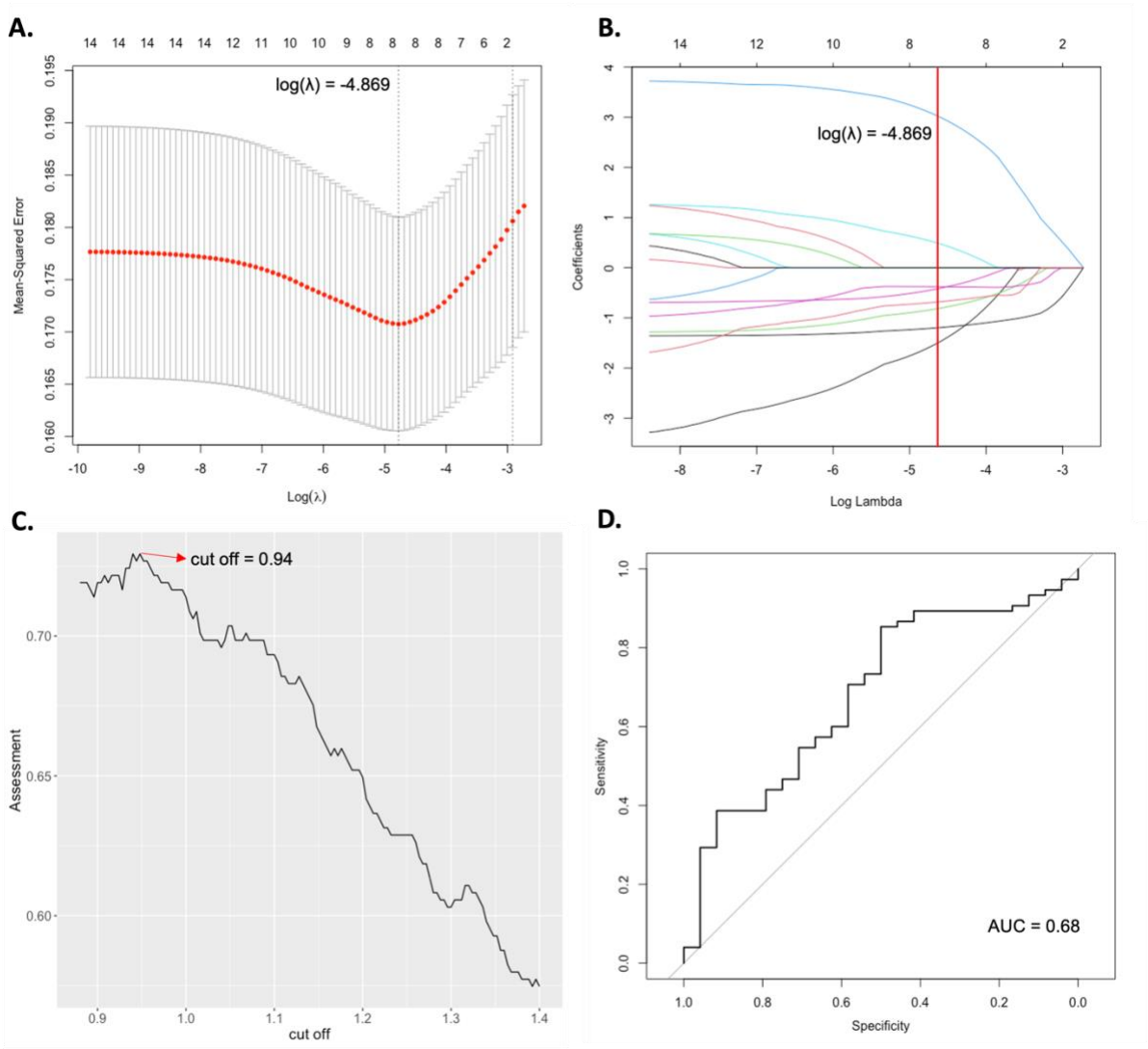


Figure 47. Classification Model 1 of sPD and LRRK2-PD based on the transcriptomic levels of LRRK2 interactors

A) *The Logistic regression model with LASSO (Least Absolute Shrinkage and Selection Operator) was adopted to further reduce the dimensionalities and to select the most significant expression profiles of LRRK2 interactors to differentiate sPD and LRRK2-PD. A λ value of 0.008, with log(λ) = -4.869 was selected according to 10-fold cross-validation;* ***B)*** *LASSO coefficient profiles of 14 LRRK2 interactors are plotted. The optimal coefficient profile was produced against the selected λ (marked as the vertical red line), comprised of 8 LRRK2 interactors, namely STUB1 (coefficient = -1.61), ACTA2 (coefficient = -0.85), PRKN (coefficient = = -0.38), TUBB6 (coefficient = -1.21), HSPA1A (coefficient = 3.12), LMNB1 (coefficient = 0.55), SNCA (coefficient = -0.46) and SLC25A6 (coefficient = -0.68).* ***C)*** *A cut-off of 0.94 was selected on the predicted values to reach the highest accuracy of 72.9%.* ***D)*** *ROC of the refined model returned an AUC of 0.68.*

Model 2 was constructed using the SNP rs151264467 (with adjusted p-value < 0.05 in univariate logistic regression in SNP association analysis for sPD vs. control), as well as the SNPs burden of 31 LRRK2 interactors, including ANKS4B, AP2M1, ARHGEF7, CAPZB, CHGB, GAK, HSP90AB1, ITCH, LRRK2, MAP2K3, MAP2K6, MRPL28, NFATC2, PRKCZ, RIPK2, RPS14, TRAF2, CDKL3, CYREN, DIDO1, DVL3, DYNC1H1, KIF2A, MFN2, PRKDC, RAB29, RPL11, STK40, TAOK3, TK1 and TUBB6 (with p-values < 0.05 in genetic burden analysis in the sPD or LRRK2-PDvs. controls). These data involved a total of 611 subjects, including 566 sPD cases and only 45 LRRK2-PD cases, which resulted in a high Imbalance Ratio (IR) of 12.5. Therefore, a random under-sampling process was performed on the sPD cohort to keep the IR ≤ 1:4 (Krawczyk, 2016), in which a total of 180 sPD cases were randomly selected to construct Model 2 together with the 45 LRRK2-PD cases. The model was firstly trained on a randomly-picked set of 144 sPD cases and 36 LRRK2-PD cases. The result showed that LASSO logistic regression failed to classify sPD and LRRK2-PD cases as all predictors were dropped from the model (**Figure 48**).

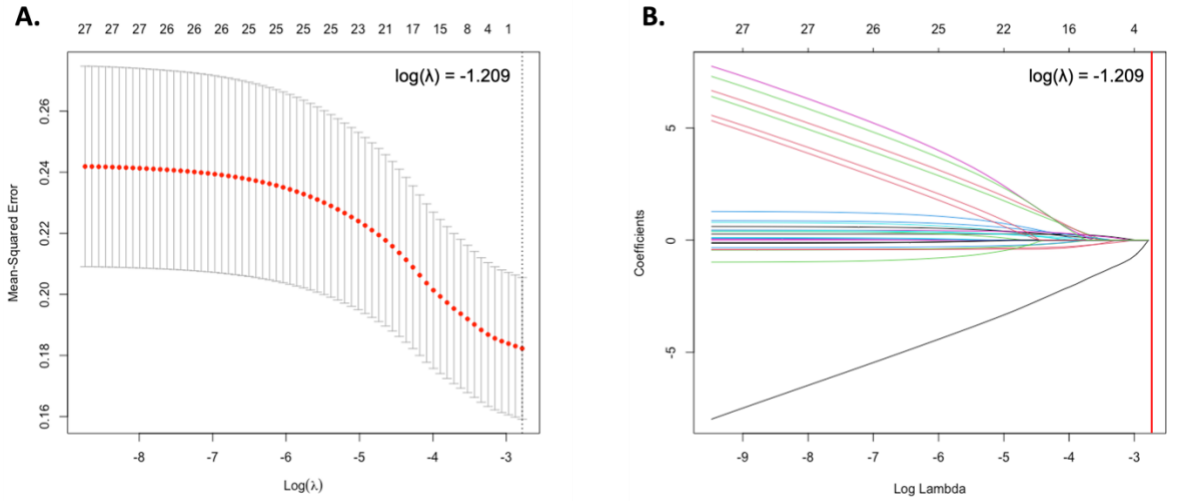


Figure 48. Classification Model 1 of sPD and LRRK2-PD based on the transcriptomic levels of LRRK2 interactors

***A)** LASSO logistic regression was adopted to evaluate the classifying performance of the combination of 1 SNP (rs151264467) and 31 LRRK2 interactor genes ANKS4B, AP2M1, ARHGEF7, CAPZB, CHGB, GAK, HSP90AB1, ITCH, LRRK2, MAP2K3, MAP2K6, MRPL28, NFATC2, PRKCZ, RIPK2, RPS14, TRAF2, CDKL3, CYREN, DIDO1, DVL3, DYNC1H1, KIF2A, MFN2, PRKDC, RAB29, RPL11, STK40, TAOK3, TK1 and TUBB6 () on the sPD and LRRK2-PD cohort. A λ value of 0.062, with log(λ) = -1.209 achieved the least MSE according to 10-fold cross-validation but left no predictors in the model. **B)** LASSO coefficient profiles of the 32 predictors are plotted. No predict remained at the λ value selected by cross-validation process (marked as red line).*

Main findings

1. Early-stage sPD patients and LRRK2-PD patients showed similar motor symptoms analysis and non-motor symptoms, such as sleeping disorders, depression, cognitive impairment and anxiety.
2. A total of 53 LRRK2 interactors presented significant differential expression in the sPD cases as compared to controls. The 29 up-regulated interactors were enriched for autophagy, negative regulation of apoptotic signalling and positive regulation of protein metabolism, while the 24 down-regulated interactors were associated with ribosomal biogenesis.
3. A total of 64 LRRK2 interactors exhibited significant differential expression in the LRRK2-PD cases as compared to Controls. The 28 up-regulated interactors were enriched for microtubule cytoskeleton organisation and positive regulation of protein metabolism, while the 36 down-regulated ones were enriched for apoptotic signalling, protein ubiquitination and peptise biosynthesis.
4. LRRK2:interactor co-expression levels remained stable across the sPD, LRRK2-PD and control cohorts.
5. WGCNA identified 3 co-expression modules of LRRK2 interactors, which were associated with intracellular protein transport, protein biosynthesis and protein metabolism, respectively. Among these 3 modules, the module that associated with protein biosynthesis were significantly down-regulated in both sPD and LRRK2-PD cases as compared to controls, while the module that related to intracellular protein transport was only down-regulated in the LRRK2-PD condition
6. In accordance with Point 5, Weighted network analysis showed that Topological Cluster A, which is responsible for ribosomal function was down-regulated in both sPD and LRRK2-PD cohort vs. controls, while Cluster F, which was responsible for mitophagy, was up-regulated only in the LRRK2-PD cohort.
7. A total of 23 and 19 SNPs in LRRK2 interactors were associated with sPD and LRRK2-PD respectively, though only 1 of them pass the multiple testing correction.

8. The sPD cases and LRRK2-PD cases can be differentiated based on the expression profiles of 8 LRRK2 interactors, including STUB1, ACTA2, PRKN, TUBB6, HSPA1A, LMNB1, SNCA and SLC25A6.
9. A total of 15 LRRK2 interactors showed a significant burden of variants (in the sPD scenario) while a different set of 15 LRRK2 interactors showed a significant burden of variants (in the LRRK2-PD scenario)

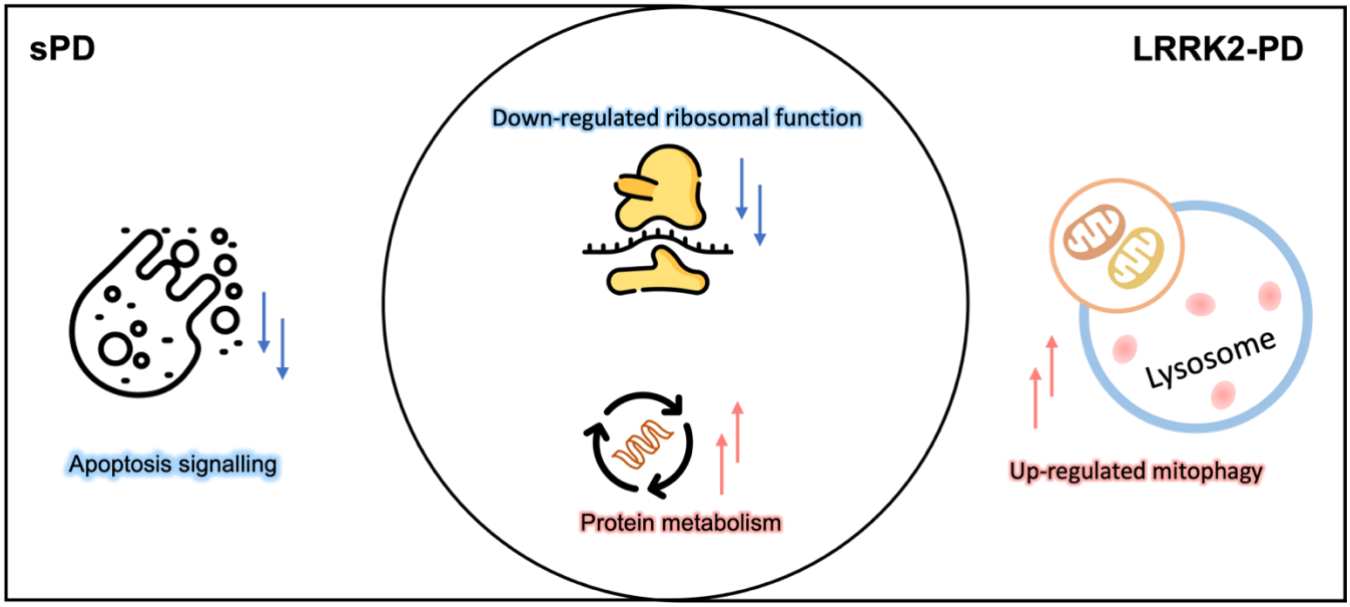


Figure 49. Summary of PD association of the LRRK2$_{int}$

Discussion

In this section, the whole blood mRNA levels of LRRK2 interactors, the LRRK2:interactor co-expression and co-expression modules withing the LRRK2net were compared between healthy controls and patients with sPD or LRRK2-PD. DEA identified a total of 29 LRRK2 up-regulated interactors in the sPD cohort as compared to controls. These interactors were related to autophagy, MAPK cascade, negative regulation of apoptotic signalling, regulation of protein metabolism and regulation of response to stress, which might suggest the early-stage sPD to be characterized by: 1) increase in cellular stress and activation of apoptosis; 2) self-adjustment mechanisms to down-regulate the apoptotic signalling to avoid excessive cell death; 3) waste disposal processes such as autophagy activated to digest or recycle damaged proteins or organelles to avoid protein aggregation and maintain cellular homeostasis.

In comparison, DEA identified a total of 28 LRRK2 up-regulated interactors in the LRRK2-PD cohort as compared to controls. LRRK2 interactors up-regulated in LRRK2-PD showed enrichment for cellular morphogenesis, unfolded protein response (UPR), microtubule cytoskeleton organisation, response to stimulus and vesicle transport suggesting that in the presence of pathogenic *LRRK2* variants, although stress-initiated cellular response is probably still present, recovering microtubule-mediated intracellular transport might become a cell priority. The implication might be that the main impact of mutated LRRK2 could be in perturbing vesicle trafficking and cytoskeleton dynamics (at least in while blood cells since the analyses were run with whole blood mRNA data). Moreover, the down-regulated interactors in the LRRK2-PD cases were also enriched for negative regulation of protein ubiquitination, suggesting that in the LRRK2-PD cases, as the pathological stress accumulates, cells might increase misfolded protein clearance by removing negative control over protein ubiquitination.

The difference in the identity and the functions of the LRRK2 interactors differentially expressed in sPD vs controls and those in LRRK2-PD vs controls suggested a potentially different molecular mechanism between these 2 conditions. Albeit the disease being the same from a clinical perspective, the disease might be able to induce changes in a different set of LRRK2 interactors depending on whether a mutation in LRRK2 was present (LRRK2-PD) or not (sPD).

It is worth noticing that in both sPD and LRRK2-PD a small set of interactors responsible for protein synthesis were down-regulated is a similar fashion in comparison with controls. This was in accordance with the results from WGCNA and weighted network analysis (see Chapter 6), showing the ribosomal-function-related co-expression module was equally down-regulated in both sPD and LRRK2-PD as compared to the controls, and the “ribosomal unit” in Topological Cluster A presented a similar pattern of alterations. This suggests that despite the molecular differences discussed above, PD might be able to induce alterations in protein synthesis exerted via the LRRK2 interactomes in both the sporadic and the familial disease scenario.

The expression levels and genetic profiles of LRRK2 interactors were used to differentiate the sporadic and the LRRK2-PD scenarios (Model 1). The results showed that using whole blood mRNA levels of 8 interactors were able to differentiate the 2 types of PD, suggesting that sPD and LRRK2-PD shared distinct molecular pathologies. However, SNPs and genetic burden of LRRK2 interactors were not able to classify the 2 types of PD (Model 2). However, it is worth noting that Model 2 was trained on a smaller sample size, involving only 36 LRRK2-PD cases and 144 sPD cases due to the limited availability of genotyping data in PPMI dataset. Therefore, the differences between sPD and LRRK2-PD pathogenesis (defined via genetic analyses) requires future investigation on larger sample size cohorts.

Conclusion

This project is focused on understanding the role of LRRK2 in the physiological and disease (PD) scenarios utilizing systems biology approaches. First of all, a protein interactome consisting of 418 robust LRRK2 interactors was constructed based on peer-reviewed data derived from multiple online resources. This is by far the most comprehensive protein interactome of LRRK2, involving a range of protein families such as ribosomal proteins, Rab GTPases, 14-3-3 proteins, heat shock proteins, cytoskeleton proteins and multiple types of protein kinases, which was in accordance with the previous 2 interactome study of LRRK2 (Manzoni et al., 2015; Porras et al., 2015b). The LRRK2 interactome was enriched for an array of diverse functions which was also expected given the fact that even the LRRK2 protein in isolation has always been associated with multiple cellular activities. Functional enrichment analysis found a “functional core” of the LRRK2 interactome, which included LRRK2 interactor that contributed to all categories of enriched biological processes. These proteins may function as the basis of LRRK2 functional network. In the meanwhile, it was possible to identify specific groups of interactors that were involved in selected LRRK2 functions (the functional units).

Next, PPIs across the LRRK2 interactors were retrieved to construct a scale-free PPI network around LRRK2, in which LRRK2 interactors were represented as nodes, while the interactions among these proteins were represented as edges. A topological core of LRRK2 PPIN was

identified, which included the mostly connected LRRK2 interactors (network hubs). These proteins form the fundamentals of LRRK2 PPIN and may function as pivotal elements able to mediate the physiological functions of LRRK2 by sustaining the flux of information within the LRRK2 network. In addition, clustering analysis was performed on the whole network and a total of 7 clusters of densely connected LRRK2 interactors (on the local scale) and associated with biological significance were identified.

These 7 clusters were all associated with different biological processes such as ribosomal function, cytoskeleton organisation, protein metabolism, autophagy, cellular response to stress and apoptotic signal transduction, which represented the mostly-reported biological functions of LRRK2 in previous studies (Albanese et al., 2019; Jiang et al., 2019; Martin et al., 2014; Parisiadou et al., 2009). These clusters of LRRK2 interactors are therefore likely to represent the local units within the networks that are specialized in helping LRRK2 sustaining very specific functions. The network model and detected clusters were then merged with i) transcriptomics data of brain regions and peripheral tissues from healthy individuals to investigate tissue-specific expression and co-expression patterns of LRRK2 interactors, and ii) transcriptomics data of whole blood from patients with sporadic PD and LRRK2-related PD to investigate impact of the disease scenario on the LRRK2 PPI network. The main new findings obtained after the work in this thesis can be recapitulated as such:

1. ***LRRK2 interactors enriched for ribosomal function were highly expressed in lung, and their expression levels (whole blood) were significantly down-regulated in both sPD and LRRK2-PD patients as compared to controls***

A total of 38 ribosomal proteins presented in the LRRK2 interactome. These proteins exhibited similar expression patterns and showed the highest mRNA levels in the lung of healthy individuals. In the meantime, they were also found in the same co-expression modules whose eigengene was highly correlated with the lung. In addition, 16 of them presented high PPI connection across each other and therefore were allocated into 1 unique topological cluster (Cluster A). This cluster showed significantly decreased connectivity in both sPD and LRRK2-PD cases when compared with controls. These findings suggested a robust “ribosomal unit” is present in the LRRK2 interactome, through which LRRK2 might control ribosomal function and protein expression. This functional unit may play an important

role especially for the biology of the lung and it probably participates in the cellular response to stress induced by disease. Also, pathogenic LRRK2 mutations seems to have little to no impact on the function of this cluster as alteration of this ribosomal cluster during disease was similar in sPD and LRRK2-PD.

2. ***LRRK2 interactors related to cytoskeleton dynamics were expressed in both brain and in the peripheral tissues, and were only down-regulated in the LRRK2-PD but not in the sPD cases when compared with controls.***

A total of 36 cytoskeleton proteins presented in the LRRK2 interactome, mainly including actin proteins and tubulin proteins (the basic unit of microtubule). Actin filament and microtubule are both essential elements of the cytoskeleton, in which the former contributes to maintaining cell morphology while the latter is responsible for intracellular vesicle trafficking. Transcriptomic analysis on healthy data showed that microtubule-associated proteins exhibited high expression levels and co-expression behaviours with LRRK2 in the brain, while actin-related proteins showed these features in the periphery. This was also confirmed by the identification of a co-expression module of LRRK2 interactors enriched for actin filament organisation that was found via WGCNA run with transcriptomics data from healthy individuals, and showed significantly high expression in the whole blood. A similar co-expression module was identified when WGCNA was run on expression data of PD patients. The eigengene of this module showed significantly lower expression level among the LRRK2-PD cases as compared to the healthy controls but not when sPD cases were compared with controls. These findings may suggest that LRRK2 and its interactors play an important role in the regulation of the cytoskeletal dynamics, with a dichotomy between the brain and the periphery where the actin cytoskeleton organisation seems a more relevant function for the peripheral tissues while microtubules dynamics (and therefore cellular trafficking) might be a specific function sustained by the LRRK2 interactome in the brain regions. Additionally, these results might suggest that one specific mechanism through which LRRK2 mutations can induce alterations correlated with disease might be via alterations of cytoskeletal dynamics, while this is probably not a molecular pathway for the sporadic disease.

3. ***The mitophagy-related topological cluster was up-regulated in the LRRK2-PD but not in the sPD cases when compared with healthy controls.***

A topological cluster (Cluster F) made of 16 interactors was detected in the LRRK2 network, which was strongly associated with mitophagy and contained other PD genes related to mitophagy (PRKN, VDAC1, RAB8A, MFN1, MFN2). This cluster presented significantly higher connectivity in the LRRK2-PD cases as compared to the controls and this finding was not recapitulated in the sporadic scenario. A mitochondrial role for LRRK2 has been reported, these new findings may suggest that this role might not be altered in the disease scenario unless in the presence of LRRK2 mutations. In other words, one specific mechanism through which LRRK2 mutations could induce alterations correlated with disease might be via alterations of mitophagy, while this is probably not a molecular pathway for the sporadic disease.

4. ***There was a dichotomy between the LRRK2 interactome in the brain vs the peripheral tissues.***

From an expression behaviour point of view, it was possible to distinguish between a "central" and a "peripheral" LRRK2 interactome. This was confirmed also when functional enrichment was taken into consideration, showing overlapping but also distinct biological functions of the LRRK2 interactors between the brain and the peripheral tissues. This might suggest that LRRK2, a protein that is indeed implicated in many different functions, might not hold the same relevance in all tissues. This might have implications for the modelling of the LRRK2 activity in physiological and disease conditions with cellular models since responses might be indeed tissue specific.

5. ***There was a dichotomy between the expression behaviour of the LRRK2$_{int}$ in the sPD cases vs LRRK2-PD cases.***

A total of 53 and 64 LRRK2 interactors presented differential expression in the sPD and LRRK2-PD cases as compared to controls, respectively. Among these interactors, 100/117 (85.5%) presented distinct alterations in the sPD and LRRK2-PD cases, suggesting that sPD and LRRK2-PD shares heterogeneous molecular signature of the LRRK2$_{int}$. In addition, the topological cluster of mitophagy (Cluster F) was significantly up-regulated in the LRRK2-PD cases but not in the sPD cases, suggesting that mitophagy is vulnerable to the pathogenic mutations of *LRRK2.* On the other hand, the "ribosomal unit" of the LRRK2$_{net}$ was down-regulated in both LRRK2-PD and sPD cases, suggesting that ribosomal function is universally affected by the 2

types of PD regardless the presence of *LRRK2* variants. In addition, LASSO logistic regression model based on transcriptomic profiles of LRRK2 interactors was able to classify sPD and LRRK2-PD cases with an accuracy of 68%. This accuracy is not sufficient in a diagnostic scenario; however, it provides the proof of principle for differences in molecular mechanisms underlying the 2 types of PD (at least when using classifiers from the whole blood). Future studies can focus on exploring the expression patterns of LRRK2 interactors in other tissues especially for the brain regions. In addition, longitudinal expression alterations of LRRK2 interactors are also worth exploring considering the long-term disease progression.

Limitations and future directions

Limitations of different types of analysis have been discussed after each chapter. Here is the summary of the main points:

1. The LRRK2 interactome and PPI network model established in this study is comprehensive but not complete, due to the nature of systems biology, which is largely dependent on the available data. It is possible that some proteins exist with strong interactions with LRRK2 that were missed by this study because they are not know or recorded yet. However, this is an inevitable disadvantage of nearly all types of interactome analysis. Therefore, it is necessary to repeat the protocol established within this thesis in the future to keep the findings updated with the PPI literature.
2. The expression data included in transcriptomic analysis were obtained from the whole blood sample from PD patients and healthy controls, which may not perfectly reflect the alterations in the brain regions. In addition, it would be helpful to analyse longitudinal changes of the expression profile of LRRK2 interactors to investigate the continuity of alterations observed in this study alongside PD progression. Therefore, future work should be focused on exploring tissue-specific gene expression patterns in relevant brain regions to better capture the intricacies of Parkinson's disease pathology. Additionally, incorporating longitudinal analyses of LRRK2 interactors' expression profiles over different stages of PD could provide a more comprehensive understanding of the molecular changes associated with disease progression.
3. Considering the biased ethnicity distribution of participant cohort included in this study, in which Caucasian account for 95% of the population, alterations observed in this study may not be applicable to other populations due to genetic variation. Future studies can

address this limitation by incorporating more diverse and representative participant cohorts that encompass a broader range of ethnic backgrounds.

4. Alterations of expression and co-expression level may not reflect the true changes in the PPIs, which may affect the accuracy of weighted network analysis in this study. Future studies should employ structural analyses, such as molecular modelling and PPI simulations, to precisely elucidate the impact of pathogenic LRRK2 variants on the structural dynamics of the LRRK2 protein and its interactions within the PPI network. Additionally, integrating experimental techniques like co-immunoprecipitation and proximity-dependent labelling assays can offer experimental validation of the predicted changes in PPIs

Reference

Aasly, J. O., Vilariño-Güell, C., Dachsel, J. C., Webber, P. J., West, A. B., Haugarvoll, K., Johansen, K. K., Toft, M., Nutt, J. G., Payami, H., Kachergus, J. M., Lincoln, S. J., Felic, A., Wider, C., Soto-Ortolaza, A. I., Cobb, S. A., White, L. R., Ross, O. A., & Farrer, M. J. (2010). Novel pathogenic LRRK2 p.Asn1437His substitution in familial Parkinson's disease. *Movement Disorders*, *25*(13), 2156–2163. https://doi.org/10.1002/MDS.23265

Aguet, F., Barbeira, A. N., Bonazzola, R., Brown, A., Castel, S. E., Jo, B., Kasela, S., Kim-Hellmuth, S., Liang, Y., Oliva, M., Flynn, E. D., Parsana, P., Fresard, L., Gamazon, E. R., Hamel, A. R., He, Y., Hormozdiari, F., Mohammadi, P., Muñoz-Aguirre, M., … Volpi, S. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, *369*(6509), 1318–1330. https://doi.org/10.1126/SCIENCE.AAZ1776

Alanis-Lobato, G., Andrade-Navarro, M. A., & Schaefer, M. H. (2017). HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Research*, *45*(D1), D408–D414. https://doi.org/10.1093/nar/gkw985

Albanese, F., Novello, S., & Morari, M. (2019). Autophagy and LRRK2 in the Aging Brain. *Frontiers in Neuroscience*, *13*, 1352. https://doi.org/10.3389/FNINS.2019.01352/BIBTEX

Alegre-Abarrategui, J., Christian, H., Lufino, M. M. P., Mutihac, R., Venda, L. L., Ansorge, O., & Wade-Martins, R. (2009). LRRK2 regulates autophagic activity and localizes to specific membrane microdomains in a novel human genomic reporter cellular model. *Human Molecular Genetics*, *18*(21), 4022. https://doi.org/10.1093/HMG/DDP346

Alonso-López, Di., Campos-Laborie, F. J., Gutiérrez, M. A., Lambourne, L., Calderwood, M. A., Vidal, M., & De Las Rivas, J. (2019). APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database : The Journal of Biological Databases and Curation*, *2019*. https://doi.org/10.1093/DATABASE/BAZ005

Alpaydin, E. (n.d.). *Machine learning*. Retrieved December 19, 2023, from https://books.google.com/books/about/Machine_Learning_revised_and_updated_edi.html?id=2nQJEAAAQBAJ

Altman, N., & Krzywinski, M. (2018). The curse(s) of dimensionality this-month. *Nature Methods*, *15*(6), 399–400. https://doi.org/10.1038/S41592-018-0019-X

Altshuler, D. M., Gibbs, R. A., Peltonen, L., Schaffner, S. F., Yu, F., Dermitzakis, E., Bonnen, P. E., De Bakker, P. I. W., Deloukas, P., Gabriel, S. B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Aarno Palotie, Parkin, M., Whittaker, P., Chang, K., Hawes, A., … McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52–58. https://doi.org/10.1038/nature09298

Ampie, L., & McGavern, D. B. (2022). Immunological defense of CNS barriers against infections. *Immunity*, *55*(5), 781. https://doi.org/10.1016/J.IMMUNI.2022.04.012

Ananthakrishnan, A. N., Khalili, H., Konijeti, G. G., Higuchi, L. M., De Silva, P., Fuchs, C. S., Willett, W. C., Richter, J. M., & Chan, A. T. (2014). Long-term intake of dietary fat and risk of ulcerative colitis and Crohn's disease. *Gut*, *63*(5), 776–784. https://doi.org/10.1136/GUTJNL-2013-305304

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Nature Precedings 2010*, 1–1. https://doi.org/10.1038/npre.2010.4282.1

Anderson, C. A., Boucher, G., Lees, C. W., Franke, A., D'Amato, M., Taylor, K. D., Lee, J. C., Goyette, P., Imielinski, M., Latiano, A., Lagacé, C., Scott, R., Amininejad, L., Bumpstead, S., Baidoo, L., Baldassano, R. N., Barclay, M., Bayless, T. M., Brand, S., … Rioux, J. D. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the

number of confirmed associations to 47. *Nature Genetics*, *43*(3), 246. https://doi.org/10.1038/NG.764

Andrade-Navarro, M., Biology, J. F. C., & 2016, undefined. (2016). Gene Set to Diseases (GS2D): disease enrichment analysis on human gene sets with literature data. *Scholar.Archive.OrgMA Andrade-Navarro, JF FontaineGenomics and Computational Biology, 2016•scholar.Archive.Org*, *2*(1), 33. https://scholar.archive.org/work/j5wlx3jlunbvpid7x7dhrhrt74/access/wayback/https://genomicscomputbiol.org/ojs/index.php/GCB/article/download/33/26

Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F. S. L., Ceol, A., Chautard, E., Dana, J. M., De Las Rivas, J., Dumousseau, M., Galeota, E., Gaulton, A., Goll, J., Hancock, R. E. W., Isserlin, R., Jimenez, R. C., Kerssemakers, J., Khadake, J., Lynn, D. J., Michaut, M., … Hermjakob, H. (2011). PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nature Methods 2011 8:7*, *8*(7), 528–529. https://doi.org/10.1038/nmeth.1637

Arita, M. (2005). Scale-Freeness and Biological Networks. *The Journal of Biochemistry*, *138*(1), 1–4. https://doi.org/10.1093/JB/MVI094

Atashrazm, F., Hammond, D., Perera, G., Bolliger, M. F., Matar, E., Halliday, G. M., Schüle, B., Lewis, S. J. G., Nichols, R. J., & Dzamko, N. (2019). LRRK2-mediated Rab10 phosphorylation in immune cells from Parkinson's disease patients. *Movement Disorders*, *34*(3), 406–415. https://doi.org/10.1002/MDS.27601

Barbosa, P., Hapuarachchi, B., Djamshidian, A., Strand, K., Lees, A. J., De Silva, R., Holton, J. L., & Warner, T. T. (2019). Lower nucleus accumbens α-synuclein load and D3 receptor levels in Parkinson's disease with impulsive compulsive behaviours. *Brain : A Journal of Neurology*, *142*(11), 3580–3591. https://doi.org/10.1093/BRAIN/AWZ298

Bardien, S., Lesage, S., Brice, A., & Carr, J. (2011). Genetic characteristics of leucine-rich repeat kinase 2 (LRRK2) associated Parkinson's disease. *Parkinsonism and Related Disorders*, *17*(7), 501–508. https://doi.org/10.1016/j.parkreldis.2010.11.008

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Brant, S. R., Silverberg, M. S., Taylor, K. D., Barmada, M. M., Bitton, A., Dassopoulos, T., Datta, L. W., Green, T., Griffiths, A. M., Kistner, E. O., Murtha, M. T., Regueiro, M. D., Rotter, J. I., … Daly, M. J. (2008a). Genome-wide association defines more than thirty distinct susceptibility loci for Crohn's disease. *Nature Genetics*, *40*(8), 955. https://doi.org/10.1038/NG.175

Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Brant, S. R., Silverberg, M. S., Taylor, K. D., Barmada, M. M., Bitton, A., Dassopoulos, T., Datta, L. W., Green, T., Griffiths, A. M., Kistner, E. O., Murtha, M. T., Regueiro, M. D., Rotter, J. I., … Daly, M. J. (2008b). Genome-wide association defines more than thirty distinct susceptibility loci for Crohn's disease. *Nature Genetics*, *40*(8), 955. https://doi.org/10.1038/NG.175

Bar-Shira, A., Hutter, C. M., Giladi, N., Zabetian, C. P., & Orr-Urtreger, A. (2009). Ashkenazi Parkinson's disease patients with the LRRK2 G2019S mutation share a common founder dating from the second to fifth centuries. *Neurogenetics*, *10*(4), 355–358. https://doi.org/10.1007/S10048-009-0186-0

Bavelas, A. (1948). A Mathematical Model for Group Structures. *Human Organization*, *7*(3), 16–30. https://doi.org/10.17730/HUMO.7.3.F4033344851GL053

Bedford, C., Sears, C., Perez-Carrion, M., Piccoli, G., & Condliffe, S. B. (2016). LRRK2 regulates voltage-gated calcium channel function. *Frontiers in Molecular Neuroscience*, *9*(MAY), 187266. https://doi.org/10.3389/FNMOL.2016.00035/BIBTEX

Beilina, A., Rudenko, I. N., Kaganovich, A., Civiero, L., Chau, H., Kalia, S. K., Kalia, L. V., Lobbestael, E., Chia, R., Ndukwe, K., Ding, J., Nalls, M. A., Olszewski, M., Hauser, D. N., Kumaran, R., Lozano, A. M., Baekelandt, V., Greene, L. E., Taymans, J. M., … Cookson, M. R. (2014). Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2626–2631. https://doi.org/10.1073/PNAS.1318306111/SUPPL_FILE/SD03.TXT

Belluzzi, E., Gonnelli, A., Cirnaru, M. D., Marte, A., Plotegher, N., Russo, I., Civiero, L., Cogo, S., Carrion, M. P., Franchin, C., Arrigoni, G., Beltramini, M., Bubacco, L., Onofri, F., Piccoli, G., & Greggio, E. (2016). LRRK2 phosphorylates pre-synaptic N-ethylmaleimide sensitive fusion (NSF) protein enhancing its ATPase activity and SNARE complex disassembling rate. *Molecular Neurodegeneration*, *11*(1), 1–16. https://doi.org/10.1186/S13024-015-0066-Z/FIGURES/6

Berwick, D. C., & Harvey, K. (2012). LRRK2 functions as a Wnt signaling scaffold, bridging cytosolic proteins and membrane-localized LRP6. *Human Molecular Genetics*, *21*(22), 4966. https://doi.org/10.1093/HMG/DDS342

Berwick, D. C., Javaheri, B., Wetzel, A., Hopkinson, M., Nixon-Abell, J., Grannò, S., Pitsillides, A. A., & Harvey, K. (2017). Pathogenic LRRK2 variants are gain-of-function mutations that enhance LRRK2-mediated repression of β-catenin signaling. *Molecular Neurodegeneration*, *12*(1), 1–17. https://doi.org/10.1186/S13024-017-0153-4/FIGURES/4

Biskup, S., Moore, D. J., Celsi, F., Higashi, S., West, A. B., Andrabi, S. A., Kurkinen, K., Yu, S. W., Savitt, J. M., Waldvogel, H. J., Faull, R. L. M., Emson, P. C., Torp, R., Ottersen, O. P., Dawson, T. M., & Dawson, V. L. (2006). Localization of LRRK2 to membranous and vesicular structures in mammalian brain. *Annals of Neurology*, *60*(5), 557–569. https://doi.org/10.1002/ANA.21019

Biskup, S., Moore, D. J., Rea, A., Lorenz-Deperieux, B., Coombes, C. E., Dawson, V. L., Dawson, T. M., & West, A. B. (2007). Dynamic and redundant regulation of LRRK2 and LRRK1 expression. *BMC Neuroscience*, *8*(1), 1–11. https://doi.org/10.1186/1471-2202-8-102/FIGURES/4

Blacque, O. E., Scheidel, N., & Kuhns, S. (2018). Rab GTPases in cilium formation and function. *Small GTPases*, *9*(1–2), 76–94. https://doi.org/10.1080/21541248.2017.1353847

Boecker, C. A., Goldsmith, J., Dou, D., Cajka, G. G., & Holzbaur, E. L. F. (2021). Increased LRRK2 kinase activity alters neuronal autophagy by disrupting the axonal transport of autophagosomes. *Current Biology*, *31*(10), 2140-2154.e6. https://doi.org/10.1016/J.CUB.2021.02.061/ATTACHMENT/1B09B35F-6BAE-4709-9134-87ED757D05E6/MMC1.PDF

Bonet-Ponce, L., Beilina, A., Williamson, C. D., Lindberg, E., Kluss, J. H., Saez-Atienzar, S., Landeck, N., Kumaran, R., Mamais, A., Bleck, C. K. E., Li, Y., & Cookson, M. R. (2020). LRRK2 mediates tubulation and vesicle sorting from lysosomes. *Science Advances*, *6*(46), eabb2454. https://doi.org/10.1126/sciadv.abb2454

Boon, J. Y., Dusonchet, J., Trengrove, C., & Wolozin, B. (2014). Interaction of LRRK2 with kinase and GTPase signaling cascades. *Frontiers in Molecular Neuroscience*, *7*(JULY), 94764. https://doi.org/10.3389/FNMOL.2014.00064/BIBTEX

Booth, H. D. E., Wessely, F., Connor-Robson, N., Rinaldi, F., Vowles, J., Browne, C., Evetts, S. G., Hu, M. T., Cowley, S. A., Webber, C., & Wade-Martins, R. (2019). RNA sequencing reveals MMP2 and TGFB1 downregulation in LRRK2 G2019S Parkinson's iPSC-derived astrocytes. *Neurobiology of Disease*, *129*, 56–66. https://doi.org/10.1016/j.nbd.2019.05.006

Bras, J. M., Guerreiro, R. J., Ribeiro, M. H., Januario, C., Morgadinho, A., Oliveira, C. R., Cunha, L., Hardy, J., & Singleton, A. (2005). G2019S dardarin substitution is a common cause of Parkinson's disease in a Portuguese cohort. *Movement Disorders : Official Journal of the Movement Disorder Society*, *20*(12), 1653–1655. https://doi.org/10.1002/mds.20682

Bravo-San Pedro, J. M., Niso-Santano, M., Gómez-Sánchez, R., Pizarro-Estrella, E., Aiastui-Pujana, A., Gorostidi, A., Climent, V., López De Maturana, R., Sanchez-Pernaute, R., López De Munain, A., Fuentes, J. M., & González-Polo, R. A. (2013). The LRRK2 G2019S mutant exacerbates basal autophagy through activation of the MEK/ERK pathway. *Cellular and Molecular Life Sciences : CMLS*, *70*(1), 121–136. https://doi.org/10.1007/S00018-012-1061-Y

Brugger, F., Erro, R., Balint, B., Kägi, G., Barone, P., & Bhatia, K. P. (2015). Why is there motor deterioration in Parkinson's disease during systemic infections-a hypothetical view. *Npj Parkinson's Disease 2015 1:1*, *1*(1), 1–5. https://doi.org/10.1038/npjparkd.2015.14

Bu, X. Le, Wang, X., Xiang, Y., Shen, L. L., Wang, Q. H., Liu, Y. H., Jiao, S. S., Wang, Y. R., Cao, H. Y., Yi, X., Liu, C. H., Deng, B., Yao, X. Q., Xu, Z. Q., Zhou, H. D., & Wang, Y. J. (2015). The association between infectious burden and Parkinson's disease: Acase-control study. *Parkinsonism and Related Disorders*, *21*(8), 877–881. https://doi.org/10.1016/j.parkreldis.2015.05.015

Buljan, M., Ciuffa, R., van Drogen, A., Vichalkovski, A., Mehnert, M., Rosenberger, G., Lee, S., Varjosalo, M., Pernas, L. E., Spegg, V., Snijder, B., Aebersold, R., & Gstaiger, M. (2020). Kinase Interaction Network Expands Functional and Disease Roles of Human Kinases. *Molecular Cell*, *79*(3), 504-520.e9. https://doi.org/10.1016/J.MOLCEL.2020.07.001

Canzler, S., Fischer, M., Ulbricht, D., Ristic, N., Hildebrand, P. W., & Staritzbichler, R. (2022). ProteinPrompt: a webserver for predicting protein–protein interactions. *Bioinformatics Advances*, *2*(1). https://doi.org/10.1093/BIOADV/VBAC059

Carrion, M. D. P., Marsicano, S., Daniele, F., Marte, A., Pischedda, F., Cairano, E. Di, Piovesana, E., Von Zweydorf, F., Kremmer, E., Gloeckner, C. J., Onofri, F., Perego, C., & Piccoli, G. (2017). The LRRK2 G2385R variant is a partial loss-of-function mutation that affects synaptic vesicle trafficking through altered protein interactions. *Scientific Reports*, *7*(1). https://doi.org/10.1038/S41598-017-05760-9

CG, W., SL, B., KJ, V., KO, W., KL, P., & RO, W. (2020). LRRK2 maintains mitochondrial homeostasis and regulates innate immune responses to Mycobacterium tuberculosis. *ELife*, *9*. https://doi.org/10.7554/ELIFE.51071

Chapman-Kiddell, C. A., Davies, P. S. W., Gillen, L., & Radford-Smith, G. L. (2010). Role of diet in the development of inflammatory bowel disease. *Inflammatory Bowel Diseases*, *16*(1), 137–151. https://doi.org/10.1002/IBD.20968

Chen, X., Xie, C., Tian, W., Sun, L., Wang, Z., Hawes, S., Chang, L., Kung, J., Ding, J., Chen, S., Le, W., & Cai, H. (2020). Parkinson's disease-related Leucine-rich repeat kinase 2

modulates nuclear morphology and genomic stability in striatal projection neurons during aging. *Molecular Neurodegeneration*, *15*(1), 1–19. https://doi.org/10.1186/S13024-020-00360-0/FIGURES/9

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *70*(6), 6. https://doi.org/10.1103/PhysRevE.70.066111

Consortium, T. G. O., Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, *224*(1). https://doi.org/10.1093/GENETICS/IYAD031

Cook, D. A., Kannarkat, G. T., Cintron, A. F., Butkovich, L. M., Fraser, K. B., Chang, J., Grigoryan, N., Factor, S. A., West, A. B., Boss, J. M., & Tansey, M. G. (2017a). LRRK2 levels in immune cells are increased in Parkinson's disease. *Npj Parkinson's Disease 2017 3:1*, *3*(1), 1–12. https://doi.org/10.1038/s41531-017-0010-8

Cook, D. A., Kannarkat, G. T., Cintron, A. F., Butkovich, L. M., Fraser, K. B., Chang, J., Grigoryan, N., Factor, S. A., West, A. B., Boss, J. M., & Tansey, M. G. (2017b). LRRK2 levels in immune cells are increased in Parkinson's disease. *Npj Parkinson's Disease 2017 3:1*, *3*(1), 1–12. https://doi.org/10.1038/s41531-017-0010-8

Cook, D. A., Kannarkat, G. T., Cintron, A. F., Butkovich, L. M., Fraser, K. B., Chang, J., Grigoryan, N., Factor, S. A., West, A. B., Boss, J. M., & Tansey, M. G. (2017c). LRRK2 levels in immune cells are increased in Parkinson's disease. *NPJ Parkinson's Disease*, *3*(1). https://doi.org/10.1038/S41531-017-0010-8

Cook, D. A., Kannarkat, G. T., Cintron, A. F., Butkovich, L. M., Fraser, K. B., Chang, J., Grigoryan, N., Factor, S. A., West, A. B., Boss, J. M., & Tansey, M. G. (2017d). LRRK2 levels in immune cells are increased in Parkinson's disease. *Npj Parkinson's Disease 2017 3:1*, *3*(1), 1–12. https://doi.org/10.1038/s41531-017-0010-8

Cookson, M. R. (2010). The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease. *Nature Reviews. Neuroscience*, *11*(12), 791. https://doi.org/10.1038/NRN2935

Correia Guedes, L., Ferreira, J. J., Rosa, M. M., Coelho, M., Bonifati, V., & Sampaio, C. (2010). Worldwide frequency of G2019S LRRK2 mutation in Parkinson's disease: A systematic review. *Parkinsonism and Related Disorders*, *16*(4), 237–242. https://doi.org/10.1016/j.parkreldis.2009.11.004

Craig, D. W., Hutchins, E., Violich, I., Alsop, E., Gibbs, J. R., Levy, S., Robison, M., Prasad, N., Foroud, T., Crawford, K. L., Toga, A. W., Whitsett, T. G., Kim, S., Casey, B., Reimer, A., Hutten, S. J., Frasier, M., Kern, F., Fehlman, T., … Van Keuren-Jensen, K. (2021). RNA sequencing of whole blood reveals early alterations in immune cells and gene expression in Parkinson's disease. *Nature Aging 2021 1:8*, *1*(8), 734–747. https://doi.org/10.1038/s43587-021-00088-6

Cuervo, A. M., Stafanis, L., Fredenburg, R., Lansbury, P. T., & Sulzer, D. (2004). Impaired degradation of mutant α-synuclein by chaperone-mediated autophagy. *Science*, *305*(5688), 1292–1295. https://doi.org/10.1126/SCIENCE.1101738/SUPPL_FILE/CUERVO.SOM.PDF

Daturpalli, S., Waudby, C. A., Meehan, S., & Jackson, S. E. (2013). Hsp90 Inhibits α-Synuclein Aggregation by Interacting with Soluble Oligomers. *Journal of Molecular Biology*, *425*(22), 4614–4628. https://doi.org/10.1016/J.JMB.2013.08.006

de Boni, L., Watson, A. H., Zaccagnini, L., Wallis, A., Zhelcheska, K., Kim, N., Sanderson, J., Jiang, H., Martin, E., Cantlon, A., Rovere, M., Liu, L., Sylvester, M., Lashley, T., Dettmer, U., Jaunmuktane, Z., & Bartels, T. (2022). Brain region-specific susceptibility of Lewy body pathology in synucleinopathies is governed by α-synuclein conformations. *Acta Neuropathologica*, *143*(4), 453–469. https://doi.org/10.1007/S00401-022-02406-7/FIGURES/5

De Guilhem De Lataillade, A., Verchere, J., Oullier, T., Prigent, A., Durand, T., Pellegrini, C., Neunlist, M., Baron, T., Rolli-Derkinderen, M., & Derkinderen, P. (2021). LRRK2 is reduced in Parkinson's disease gut. *Acta Neuropathologica*, *142*(3), 601–603. https://doi.org/10.1007/s00401-021-02334-yï

Del Tredici, K., & Braak, H. (2016). Review: Sporadic Parkinson's disease: development and distribution of α-synuclein pathology. *Neuropathology and Applied Neurobiology*, *42*(1), 33–50. https://doi.org/10.1111/NAN.12298

Deshpande, P., Flinkman, D., Hong, Y., Goltseva, E., Siino, V., Sun, L., Peltonen, S., Elo, L. L., Kaasinen, V., James, P., & Coffey, E. T. (2020). Protein synthesis is suppressed in sporadic and familial Parkinson's disease by LRRK2. *The FASEB Journal*, *34*(11), 14217–14233. https://doi.org/10.1096/FJ.202001046R

Devic, I., Hwang, H., Edgar, J. S., Izutsu, K., Presland, R., Pan, C., Goodlett, D. R., Wang, Y., Armaly, J., Tumas, V., Zabetian, C. P., Leverenz, J. B., Shi, M., & Zhang, J. (2011). Salivary α-synuclein and DJ-1: potential biomarkers for Parkinson's disease. *Brain : A Journal of Neurology*, *134*(Pt 7). https://doi.org/10.1093/BRAIN/AWR015

Dhekne, H. S., Yanatori, I., Gomez, R. C., Tonelli, F., Diez, F., Schüle, B., Steger, M., Alessi, D. R., & Pfeffer, S. R. (2018). A pathway for parkinson's disease LRRK2 kinase to block primary cilia and sonic hedgehog signaling in the brain. *ELife*, *7*. https://doi.org/10.7554/ELIFE.40202

di Domenico, A., Carola, G., Calatayud, C., Pons-Espinal, M., Muñoz, J. P., Richaud-Patin, Y., Fernandez-Carasa, I., Gut, M., Faella, A., Parameswaran, J., Soriano, J., Ferrer, I., Tolosa, E., Zorzano, A., Cuervo, A. M., Raya, A., & Consiglio, A. (2019). Patient-Specific iPSC-Derived Astrocytes Contribute to Non-Cell-Autonomous Neurodegeneration in Parkinson's Disease. *Stem Cell Reports*, *12*(2), 213–229. https://doi.org/10.1016/j.stemcr.2018.12.011

Di Fonzo, A., Tassorelli, C., De Mari, M., Chien, H. F., Ferreira, J., Rohé, C. F., Riboldazzi, G., Antonini, A., Albani, G., Mauro, A., Marconi, R., Abbruzzese, G., Lopiano, L., Fincati, E., Guidi, M., Marini, P., Stocchi, F., Onofrj, M., Toni, V., … Cossu, G. (2005). Comprehensive analysis of the LRRK2 gene in sixty families with Parkinson's disease. *European Journal of Human Genetics 2006 14:3*, *14*(3), 322–331. https://doi.org/10.1038/sj.ejhg.5201539

Doggett, E. A., Zhao, J., Mork, C. N., Hu, D., & Nichols, R. J. (2012). Phosphorylation of LRRK2 serines 955 and 973 is disrupted by Parkinson's disease mutations and LRRK2 pharmacological inhibition. *Journal of Neurochemistry*, *120*(1), 37–45. https://doi.org/10.1111/J.1471-4159.2011.07537.X

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/J.1600-0587.2012.07348.X

Dou, D., Smith, E. M., Evans, C. S., Boecker, C. A., & Holzbaur, E. L. F. (2022). Regulatory imbalance between LRRK2 kinase, PPM1H phosphatase, and ARF6 GTPase disrupts the axonal transport of autophagosomes. *BioRxiv*, 2022.11.14.516471. https://doi.org/10.1101/2022.11.14.516471

Drolet, R. E., Sanders, J. M., & Kern, J. T. (2011). Leucine-rich repeat kinase 2 (LRRK2) cellular biology: A review of recent advances in identifying physiological substrates and cellular functions. In *Journal of Neurogenetics* (Vol. 25, Issue 4, pp. 140–151). J Neurogenet. https://doi.org/10.3109/01677063.2011.627072

Dursun, E., Gezen-Ak, D., Hanağasi, H., Bilgiç, B., Lohmann, E., Ertan, S., Atasoy, I. L., Alaylioğlu, M., Araz, Ö. S., Önal, B., Gündüz, A., Apaydin, H., Kiziltan, G., Ulutin, T., Gürvit, H., & Yilmazer, S. (2015). The interleukin 1 alpha, interleukin 1 beta, interleukin 6 and alpha-2-macroglobulin serum levels in patients with early or late onset Alzheimer's disease, mild cognitive impairment or Parkinson's disease. *Journal of Neuroimmunology*, *283*, 50–57. https://doi.org/10.1016/j.jneuroim.2015.04.014

Dwyer, Z., Rudyk, C., Thompson, A., Farmer, K., Fenner, B., Fortin, T., Derksen, A., Sun, H., & Hayley, S. (2020). Leucine-rich repeat kinase-2 (LRRK2) modulates microglial phenotype and dopaminergic neurodegeneration. *Neurobiology of Aging*, *91*, 45–55. https://doi.org/10.1016/j.neurobiolaging.2020.02.017

Dzamko, N., Inesta-Vaquera, F., Zhang, J., Xie, C., Cai, H., Arthur, S., Tan, L., Choi, H., Gray, N., Cohen, P., Pedrioli, P., Clark, K., & Alessi, D. R. (2012). The IkappaB Kinase Family Phosphorylates the Parkinson's Disease Kinase LRRK2 at Ser935 and Ser910 during Toll-Like Receptor Signaling. *PLOS ONE*, *7*(6), e39132. https://doi.org/10.1371/JOURNAL.PONE.0039132

Dzamko, N., Rowe, D. B., & Halliday, G. M. (2016). Increased peripheral inflammation in asymptomatic leucine-rich repeat kinase 2 mutation carriers. *Movement Disorders*, *31*(6), 889–897. https://doi.org/10.1002/MDS.26529

El Haj, R. Ben, Salmi, A., Regragui, W., Moussa, A., Bouslam, N., Tibar, H., Benomar, A., Yahyaoui, M., & Bouhouche, A. (2017). Evidence for prehistoric origins of the G2019S mutation in the North African Berber population. *PloS One*, *12*(7). https://doi.org/10.1371/JOURNAL.PONE.0181335

El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? *Machine Learning in Radiation Oncology*, 3–11. https://doi.org/10.1007/978-3-319-18305-3_1

Ennemoser, M., Pum, A., & Kungl, A. (2022a). Disease-specific glycosaminoglycan patterns in the extracellular matrix of human lung and brain. *Carbohydrate Research*, *511*, 108480. https://doi.org/10.1016/J.CARRES.2021.108480

Ennemoser, M., Pum, A., & Kungl, A. (2022b). Disease-specific glycosaminoglycan patterns in the extracellular matrix of human lung and brain. *Carbohydrate Research*, *511*, 108480. https://doi.org/10.1016/J.CARRES.2021.108480

Fan, Y., Howden, A. J. M., Sarhan, A. R., Lis, P., Ito, G., Martinez, T. N., Brockmann, K., Gasser, T., Alessi, D. R., & Sammler, E. M. (2018). Interrogating Parkinson's disease LRRK2 kinase pathway activity by assessing Rab10 phosphorylation in human neutrophils. *The Biochemical Journal*, *475*(1), 23–44. https://doi.org/10.1042/BCJ20170803

Fan, Y., Nirujogi, R. S., Garrido, A., Ruiz-Martínez, J., Bergareche-Yarza, A., Mondragón-Rezola, E., Vinagre-Aragón, A., Croitoru, I., Gorostidi Pagola, A., Paternain Markinez, L., Alcalay, R., Hickman, R. A., Düring, J., Gomes, S., Pratuseviciute, N., Padmanabhan, S., Valldeoriola, F., Pérez Sisqués, L., Malagelada, C., … Sammler, E. M. (2021). R1441G but

not G2019S mutation enhances LRRK2 mediated Rab10 phosphorylation in human peripheral blood neutrophils. *Acta Neuropathologica*, *142*(3), 475–494. https://doi.org/10.1007/S00401-021-02325-Z/FIGURES/7

Fava, V. M., Manry, J., Cobat, A., Orlova, M., Van Thuc, N., Ba, N. N., Thai, V. H., Abel, L., Alcaïs, A., Schurr, E., Brown, E., Gibbings, D., Hayley, S., Park, D., Philpott, D. C., Rioux, J. D., & Schlossmacher, M. (2016a). A Missense LRRK2 Variant Is a Risk Factor for Excessive Inflammatory Responses in Leprosy. *PLoS Neglected Tropical Diseases*, *10*(2). https://doi.org/10.1371/JOURNAL.PNTD.0004412

Fava, V. M., Manry, J., Cobat, A., Orlova, M., Van Thuc, N., Ba, N. N., Thai, V. H., Abel, L., Alcaïs, A., Schurr, E., Brown, E., Gibbings, D., Hayley, S., Park, D., Philpott, D. C., Rioux, J. D., & Schlossmacher, M. (2016b). A Missense LRRK2 Variant Is a Risk Factor for Excessive Inflammatory Responses in Leprosy. *PLoS Neglected Tropical Diseases*, *10*(2). https://doi.org/10.1371/journal.pntd.0004412

Fava, V. M., Xu, Y. Z., Lettre, G., Thuc, N. Van, Orlova, M., Thai, V. H., Tao, S., Croteau, N., Eldeeb, M. A., MacDougall, E. J., Cambri, G., Lahiri, R., Adams, L., Fon, E. A., Trempe, J. F., Cobat, A., Alcaïs, A., Abell, L., & Schurr, E. (2019). Pleiotropic effects for Parkin and LRRK2 in leprosy type-1 reactions and Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(31), 15616–15624. https://doi.org/10.1073/pnas.1901805116

Fénelon, G., & Walusinski, O. (2021). The landmark contributions of Paul Blocq, Georges Marinesco, and Édouard Brissaud in Parkinson's disease. *Revue Neurologique*, *177*(10), 1214–1220. https://doi.org/10.1016/J.NEUROL.2021.02.386

Ferreira, J. J., Guedes, L. C., Rosa, M. M., Coelho, M., Van Doeselaar, M., Schweiger, D., Di Fonzo, A., Oostra, B. A., Sampaio, C., & Bonifati, V. (2007). High prevalence of LRRK2 mutations in familial and sporadic Parkinson's disease in Portugal. *Movement Disorders : Official Journal of the Movement Disorder Society*, *22*(8), 1194–1201. https://doi.org/10.1002/MDS.21525

Filippini, A., Gennarelli, M., & Russo, I. (2021). Leucine-rich repeat kinase 2-related functions in GLIA: an update of the last years. *Biochemical Society Transactions*. https://doi.org/10.1042/bst20201092

Flinkman, D., Hong, Y., Gnjatovic, J., Deshpande, P., Ortutay, Z., Peltonen, S., Kaasinen, V., James, P., & Coffey, E. (2023). Regulators of proteostasis are translationally repressed in fibroblasts from patients with sporadic and LRRK2-G2019S Parkinson's disease. *Npj Parkinson's Disease 2023 9:1*, *9*(1), 1–13. https://doi.org/10.1038/s41531-023-00460-w

Foffani, G., & Obeso, J. A. (2018). A Cortical Pathogenic Theory of Parkinson's Disease. *Neuron*, *99*(6), 1116–1128. https://doi.org/10.1016/J.NEURON.2018.07.028

Franke, A., McGovern, D. P. B., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., Lees, C. W., Balschun, T., Lee, J., Roberts, R., Anderson, C. A., Bis, J. C., Bumpstead, S., Ellinghaus, D., Festen, E. M., Georges, M., Green, T., Haritunians, T., Jostins, L., … Parkes, M. (2010). Meta-Analysis Increases to 71 the Tally of Confirmed Crohn's Disease Susceptibility Loci. *Nature Genetics*, *42*(12), 1118. https://doi.org/10.1038/NG.717

Fraser, K. B., Moehle, M. S., Daher, J. P. L., Webber, P. J., Williams, J. Y., Stewart, C. A., Yacoubian, T. A., Cowell, R. M., Dokland, T., Ye, T., Chen, D., Siegal, G. P., Galemmo, R. A., Tsika, E., Moore, D. J., Standaert, D. G., Kojima, K., Mobley, J. A., & West, A. B. (2013). LRRK2 secretion in exosomes is regulated by 14-3-3. *Human Molecular Genetics*, *22*(24), 4988–5000. https://doi.org/10.1093/HMG/DDT346

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35. https://doi.org/10.2307/3033543

Friedman, L. G., Lachenmayer, M. L., Wang, J., He, L., Poulose, S. M., Komatsu, M., Holstein, G. R., & Yue, Z. (2012a). Disrupted Autophagy Leads to Dopaminergic Axon and Dendrite Degeneration and Promotes Presynaptic Accumulation of α-Synuclein and LRRK2 in the Brain. *Journal of Neuroscience*, *32*(22), 7585–7593. https://doi.org/10.1523/JNEUROSCI.5809-11.2012

Friedman, L. G., Lachenmayer, M. L., Wang, J., He, L., Poulose, S. M., Komatsu, M., Holstein, G. R., & Yue, Z. (2012b). *Neurobiology of Disease Disrupted Autophagy Leads to Dopaminergic Axon and Dendrite Degeneration and Promotes Presynaptic Accumulation of-Synuclein and LRRK2 in the Brain*. https://doi.org/10.1523/JNEUROSCI.5809-11.2012

Fries, C. (1936). *Metaphysik als Naturwissenschaft: Betrachtungen zu Ludwig von Bertalanffy's theoretische Biologie*. E. Ebering.

Fuji, R. N., Flagella, M., Baca, M., Baptista, M. A. S., Brodbeck, J., Chan, B. K., Fiske, B. K., Honigberg, L., Jubb, A. M., Katavolos, P., Lee, D. W., Lewin-Koh, S. C., Lin, T., Liu, X., Liu, S., Lyssikatos, J. P., O'Mahony, J., Reichelt, M., Roose-Girma, M., … Watts, R. J. (2015). Effect of selective LRRK2 kinase inhibition on nonhuman primate lung. *Science Translational Medicine*, *7*(273), 273ra15. https://doi.org/10.1126/SCITRANSLMED.AAA3634

Funayama, M., Hasegawa, K., Kowa, H., Saito, M., Tsuji, S., & Obata, F. (2002). A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. *Annals of Neurology*, *51*(3), 296–301. https://doi.org/10.1002/ANA.10113

Funayama, M., Hasegawa, K., Ohta, E., Kawashima, N., Komiyama, M., Kowa, H., Tsuji, S., & Obata, F. (2005). An LRRK2 mutation as a cause for the parkinsonism in the original PARK8 family. *Annals of Neurology*, *57*(6), 918–921. https://doi.org/10.1002/ANA.20484

Gaig, C., Martí, M. J., Ezquerra, M., Cardozo, A., Rey, M. J., & Tolosa, E. (2009). G2019S LRRK2 mutation causing Parkinson's disease without Lewy bodies. *Case Reports*, *2009*, bcr0820080632. https://doi.org/10.1136/BCR.08.2008.0632

Gaiter, D., Westerlund, M., Carmine, A., Lindqvist, E., Sydow, O., & Olson, L. (2006). LRRK2 expression linked to dopamine-innervated areas. *Annals of Neurology*, *59*(4), 714–719. https://doi.org/10.1002/ANA.20808

Gao, X., Carroni, M., Nussbaum-Krammer, C., Mogk, A., Nillegoda, N. B., Szlachcic, A., Guilbride, D. L., Saibil, H. R., Mayer, M. P., & Bukau, B. (2015). Human Hsp70 Disaggregase Reverses Parkinson's-Linked α-Synuclein Amyloid Fibrils. *Molecular Cell*, *59*(5), 781–793. https://doi.org/10.1016/J.MOLCEL.2015.07.012

Gardet, A., Benita, Y., Li, C., Sands, B. E., Ballester, I., Stevens, C., Korzenik, J. R., Rioux, J. D., Daly, M. J., Xavier, R. J., & Podolsky, D. K. (2010a). LRRK2 is involved in the IFN-gamma response and host response to pathogens. *Journal of Immunology (Baltimore, Md. : 1950)*, *185*(9), 5577–5585. https://doi.org/10.4049/JIMMUNOL.1000548

Gardet, A., Benita, Y., Li, C., Sands, B. E., Ballester, I., Stevens, C., Korzenik, J. R., Rioux, J. D., Daly, M. J., Xavier, R. J., & Podolsky, D. K. (2010b). LRRK2 is involved in the IFN-gamma response and host response to pathogens. *Journal of Immunology (Baltimore, Md. : 1950)*, *185*(9), 5577–5585. https://doi.org/10.4049/JIMMUNOL.1000548

Ghosh, S., & Dasgupta, R. (2022a). Introduction to Artificial Intelligence (AI) Methods in Biology. *Machine Learning in Biological Sciences*, 19–27. https://doi.org/10.1007/978-981-16-8881-2_2

Ghosh, S., & Dasgupta, R. (2022b). Introduction to the Machine Learning Models. *Machine Learning in Biological Sciences*, 45–50. https://doi.org/10.1007/978-981-16-8881-2_4

Ghosh, S., & Dasgupta, R. (2022c). Machine Learning Methods. *Machine Learning in Biological Sciences*, 29–43. https://doi.org/10.1007/978-981-16-8881-2_3

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., Shamovsky, V., Weiser, J., Brunson, T., Sanati, N., … D'Eustachio, P. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, *50*(D1), D687–D692. https://doi.org/10.1093/NAR/GKAB1028

Gillis, J., & Pavlidis, P. (2011). The role of indirect connections in gene networks in predicting function. *Bioinformatics*, *27*(13), 1860–1866. https://doi.org/10.1093/BIOINFORMATICS/BTR288

Giusto, E., Yacoubian, T. A., Greggio, E., & Civiero, L. (2021). Pathways to Parkinson's disease: a spotlight on 14-3-3 proteins. *Npj Parkinson's Disease 2021 7:1*, *7*(1), 1–14. https://doi.org/10.1038/s41531-021-00230-6

Gómez-Suaga, P., Luzón-Toro, B., Churamani, D., Zhang, L., Bloor-Young, D., Patel, S., Woodman, P. G., Churchill, G. C., & Hilfiker, S. (2012). Leucine-rich repeat kinase 2 regulates autophagy through a calcium-dependent pathway involving NAADP. *Human Molecular Genetics*, *21*(3), 511–525. https://doi.org/10.1093/HMG/DDR481

Gonçalves, A. M., Pereira-Santos, A. R., Esteves, A. R., Cardoso, S. M., & Empadinhas, N. (2021). The Mitochondrial Ribosome: A World of Opportunities for Mitochondrial Dysfunction Toward Parkinson's Disease. *Antioxidants and Redox Signaling*, *34*(8), 694–711. https://doi.org/10.1089/ARS.2019.7997/ASSET/IMAGES/LARGE/ARS.2019.7997_FIGURE6.JPEG

Gopalai, A. A., Lim, J. L., Li, H. H., Zhao, Y., Lim, T. T., Eow, G. B., Puvanarajah, S., Viswanathan, S., Norlinah, M. I., Abdul Aziz, Z., Lim, S. K., Tan, C. T., Tan, A. H., Lim, S. Y., Tan, E. K., & Ahmad Annuar, A. (2019). LRRK2 N551K and R1398H variants are protective in Malays and Chinese in Malaysia: A case–control association study for Parkinson's disease. *Molecular Genetics & Genomic Medicine*, *7*(11), e604. https://doi.org/10.1002/MGG3.604

Gorostidi, A., Ruiz-Martínez, J., Lopez De Munain, A., Alzualde, A., & Martí Massó, J. F. (2009). LRRK2 G2019S and R1441G mutations associated with Parkinson's disease are common in the Basque Country, but relative prevalence is determined by ethnicity. *Neurogenetics*, *10*(2), 157–159. https://doi.org/10.1007/S10048-008-0162-0

Gradel, K. O., Nielsen, H. L., Schønheyder, H. C., Ejlertsen, T., Kristensen, B., & Nielsen, H. (2009). Increased short- and long-term risk of inflammatory bowel disease after salmonella or campylobacter gastroenteritis. *Gastroenterology*, *137*(2), 495–501. https://doi.org/10.1053/J.GASTRO.2009.04.001

Gu, Q., Cuevas, E., Raymick, J., Kanungo, J., & Sarkar, S. (2020). Downregulation of 14-3-3 Proteins in Alzheimer's Disease. *Molecular Neurobiology*, *57*(1), 32–40. https://doi.org/10.1007/S12035-019-01754-Y/FIGURES/4

Guo, L., Gandhi, P. N., Wang, W., Petersen, R. B., Wilson-Delfosse, A. L., & Chen, S. G. (2007). The Parkinson's disease-associated protein, leucine-rich repeat kinase 2 (LRRK2), is an

authentic GTPase thatstimulates kinase activity. *Experimental Cell Research*, *313*(16), 3658–3670. https://doi.org/10.1016/J.YEXCR.2007.07.007

Hakimi, M., Selvanantham, T., Swinton, E., Padmore, R. F., Tong, Y., Kabbach, G., Venderova, K., Girardin, S. E., Bulman, D. E., Scherzer, C. R., Lavoie, M. J., Gris, D., Park, D. S., Angel, J. B., Shen, J., Philpott, D. J., & Schlossmacher, M. G. (2011a). Parkinson's disease-linked LRRK2 is expressed in circulating and tissue immune cells and upregulated following recognition of microbial structures. *Journal of Neural Transmission*, *118*(5), 795–808. https://doi.org/10.1007/s00702-011-0653-2

Hakimi, M., Selvanantham, T., Swinton, E., Padmore, R. F., Tong, Y., Kabbach, G., Venderova, K., Girardin, S. E., Bulman, D. E., Scherzer, C. R., Lavoie, M. J., Gris, D., Park, D. S., Angel, J. B., Shen, J., Philpott, D. J., & Schlossmacher, M. G. (2011b). Parkinson's disease-linked LRRK2 is expressed in circulating and tissue immune cells and upregulated following recognition of microbial structures. *Journal of Neural Transmission*, *118*(5), 795–808. https://doi.org/10.1007/S00702-011-0653-2/FIGURES/5

Hakimi, M., Selvanantham, T., Swinton, E., Padmore, R. F., Tong, Y., Kabbach, G., Venderova, K., Girardin, S. E., Bulman, D. E., Scherzer, C. R., Lavoie, M. J., Gris, D., Park, D. S., Angel, J. B., Shen, J., Philpott, D. J., & Schlossmacher, M. G. (2011c). Parkinson's disease-linked LRRK2 is expressed in circulating and tissue immune cells and upregulated following recognition of microbial structures. *Journal of Neural Transmission (Vienna, Austria : 1996)*, *118*(5), 795–808. https://doi.org/10.1007/S00702-011-0653-2

Hammes, J., Theis, H., Giehl, K., Hoenig, M. C., Greuel, A., Tittgemeyer, M., Timmermann, L., Fink, G. R., Drzezga, A., Eggers, C., & Van Eimeren, T. (2019a). Dopamine metabolism of the nucleus accumbens and fronto-striatal connectivity modulate impulse control. *Brain*, *142*(3), 733–743. https://doi.org/10.1093/BRAIN/AWZ007

Hammes, J., Theis, H., Giehl, K., Hoenig, M. C., Greuel, A., Tittgemeyer, M., Timmermann, L., Fink, G. R., Drzezga, A., Eggers, C., & Van Eimeren, T. (2019b). Dopamine metabolism of the nucleus accumbens and fronto-striatal connectivity modulate impulse control. *Brain : A Journal of Neurology*, *142*(3), 733–743. https://doi.org/10.1093/BRAIN/AWZ007

Härtlova, A., Herbst, S., Peltier, J., Rodgers, A., Bilkei-Gorzo, O., Fearns, A., Dill, B. D., Lee, H., Flynn, R., Cowley, S. A., Davies, P., Lewis, P. A., Ganley, I. G., Martinez, J., Alessi, D. R., Reith, A. D., Trost, M., & Gutierrez, M. G. (2018). LRRK2 is a negative regulator of Mycobacterium tuberculosis phagosome maturation in macrophages. *The EMBO Journal*, *37*(12), e98694. https://doi.org/10.15252/EMBJ.201798694

Härtlova, A., Herbst, S., Peltier, J., Rodgers, A., Bilkei-Gorzo, O., Fearns, A., Dill, B. D., Lee, H., Flynn, R., Cowley, S. A., Davies, P., Lewis, P. A., Ganley, I. G., Martinez, J., Alessi, D. R., Reith, A. D., Trost, M., Gutierrez, M. G., A, H., … MG, G. (2018). LRRK2 is a negative regulator of Mycobacterium tuberculosis phagosome maturation in macrophages. *The EMBO Journal*, *37*(12), e98694. https://doi.org/10.15252/embj.201798694

He, J., Li, C., Ye, B., & Zhong, W. (2012). Efficient and accurate Greedy Search Methods for mining functional modules in protein interaction networks. *BMC Bioinformatics*, *13 Suppl 10*(10), 1–11. https://doi.org/10.1186/1471-2105-13-S10-S19/TABLES/7

Hedman, A. C., Smith, J. M., & Sacks, D. B. (2015). The biology of IQGAP proteins: beyond the cytoskeleton. *EMBO Reports*, *16*(4), 427–446. https://doi.org/10.15252/EMBR.201439834

Helip-Wooley, A., & Thoene, J. G. (2004). Sucrose-induced vacuolation results in increased expression of cholesterol biosynthesis and lysosomal genes. *Experimental Cell Research*, *292*(1), 89–100. https://doi.org/10.1016/J.YEXCR.2003.09.003

Henry, A. G., Aghamohammadzadeh, S., Samaroo, H., Chen, Y., Mou, K., Needle, E., & Hirst, W. D. (2015). Pathogenic LRRK2 mutations, through increased kinase activity, produce enlarged lysosomes with reduced degradative capacity and increase ATP13A2 expression. *Human Molecular Genetics*, *24*(21), 6013–6028. https://doi.org/10.1093/hmg/ddv314

Herbst, S., & Gutierrez, M. G. (2019). LRRK2 in Infection: Friend or Foe? *ACS Infectious Diseases*, *5*(6), 809–815. https://doi.org/10.1021/ACSINFECDIS.9B00051/ASSET/IMAGES/LARGE/ID-2019-000517_0001.JPEG

Higashi, S., Biskup, S., West, A. B., Trinkaus, D., Dawson, V. L., Faull, R. L. M., Waldvogel, H. J., Arai, H., Dawson, T. M., Moore, D. J., & Emson, P. C. (2007). Localization of Parkinson's disease-associated LRRK2 in normal and pathological human brain. *Brain Research*, *1155*(1), 208–219. https://doi.org/10.1016/J.BRAINRES.2007.04.034

Higuchi, L. M., Khalili, H., Chan, A. T., Richter, J. M., Bousvaros, A., & Fuchs, C. S. (2012). A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. *The American Journal of Gastroenterology*, *107*(9), 1399–1406. https://doi.org/10.1038/AJG.2012.196

Ho, D. H., Jang, J., Joe, E. H., Son, I., Seo, H., & Seol, W. (2016). G2385R and I2020T Mutations Increase LRRK2 GTPase Activity. *BioMed Research International*, *2016*. https://doi.org/10.1155/2016/7917128

Ho, P. W. L., Leung, C. T., Liu, H., Pang, S. Y. Y., Lam, C. S. C., Xian, J., Li, L., Kung, M. H. W., Ramsden, D. B., & Ho, S. L. (2020). Age-dependent accumulation of oligomeric SNCA/α-synuclein from impaired degradation in mutant LRRK2 knockin mouse model of Parkinson disease: role for therapeutic activation of chaperone-mediated autophagy (CMA). *Autophagy*, *16*(2), 347–370. https://doi.org/10.1080/15548627.2019.1603545

Hou, J. K., Abraham, B., & El-Serag, H. (2011). Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *The American Journal of Gastroenterology*, *106*(4), 563–573. https://doi.org/10.1038/AJG.2011.44

Hu, Y., Vinayagam, A., Nand, A., Comjean, A., Chung, V., Hao, T., Mohr, S. E., & Perrimon, N. (2018). Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. *Nucleic Acids Research*, *46*(D1), D567–D574. https://doi.org/10.1093/NAR/GKX1116

Hui, K. Y., Fernandez-Hernandez, H., Hu, J., Schaffner, A., Pankratz, N., Hsu, N. Y., Chuang, L. S., Carmi, S., Villaverde, N., Li, X., Rivas, M., Levine, A. P., Bao, X., Labrias, P. R., Haritunians, T., Ruane, D., Gettler, K., Chen, E., Li, D., … Peter, I. (2018). Functional variants in the LRRK2 gene confer shared effects on risk for Crohn's disease and Parkinson's disease. *Science Translational Medicine*, *10*(423), 7795. https://doi.org/10.1126/SCITRANSLMED.AAI7795/SUPPL_FILE/AAI7795_TABLE_S4_AND_S6.ZIP

Hurtado-Lorenzo, A., & Anand, V. S. (2008). Heat Shock Protein 90 Modulates LRRK2 Stability: Potential Implications for Parkinson's Disease Treatment. *The Journal of Neuroscience*, *28*(27), 6757. https://doi.org/10.1523/JNEUROSCI.1870-08.2008

Iadevaia, V., Caldarola, S., Biondini, L., Gismondi, A., Karlsson, S., Dianzani, I., & Loreni, F. (2010). PIM1 kinase is destabilized by ribosomal stress causing inhibition of cell cycle

progression. *Oncogene 2010 29:40*, *29*(40), 5490–5499. https://doi.org/10.1038/onc.2010.279

Ijcai, R. K.-, & 1995, undefined. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Researchgate.NetR KohaviIjcai, 1995•researchgate.Net*. https://www.researchgate.net/profile/Ron-Kohavi/publication/2352264_A_Study_of_Cross-Validation_and_Bootstrap_for_Accuracy_Estimation_and_Model_Selection/links/02e7e51bcc14c5e91c000000/A-Study-of-Cross-Validation-and-Bootstrap-for-Accuracy-Estimation-and-Model-Selection.pdf

Ikezu, T., Koro, L., Wolozin, B., Farraye, F. A., Strongosky, A. J., & Wszolek, Z. K. (2020). Crohn's and Parkinson's Disease-Associated LRRK2 Mutations Alter Type II Interferon Responses in Human CD14+ Blood Monocytes Ex Vivo. *Journal of Neuroimmune Pharmacology*, *15*(4), 794–800. https://doi.org/10.1007/S11481-020-09909-8/FIGURES/4

Imai, Y., Kobayashi, Y., Inoshita, T., Meng, H., Arano, T., Uemura, K., Asano, T., Yoshimi, K., Zhang, C. L., Matsumoto, G., Ohtsuka, T., Kageyama, R., Kiyonari, H., Shioi, G., Nukina, N., Hattori, N., & Takahashi, R. (2015). The Parkinson's Disease-Associated Protein Kinase LRRK2 Modulates Notch Signaling through the Endosomal Pathway. *PLoS Genetics*, *11*(9). https://doi.org/10.1371/JOURNAL.PGEN.1005503

Ito, G., Katsemonova, K., Tonelli, F., Lis, P., Baptista, M. A. S., Shpiro, N., Duddy, G., Wilson, S., Ho, P. W. L., Ho, S. L., Reith, A. D., & Alessi, D. R. (2016a). Phos-tag analysis of Rab10 phosphorylation by LRRK2: a powerful assay for assessing kinase function and inhibitors. *Biochemical Journal*, *473*(17), 2671–2685. https://doi.org/10.1042/BCJ20160557

Ito, G., Katsemonova, K., Tonelli, F., Lis, P., Baptista, M. A. S., Shpiro, N., Duddy, G., Wilson, S., Ho, P. W. L., Ho, S. L., Reith, A. D., & Alessi, D. R. (2016b). Phos-tag analysis of Rab10 phosphorylation by LRRK2: a powerful assay for assessing kinase function and inhibitors. *The Biochemical Journal*, *473*(17), 2671–2685. https://doi.org/10.1042/BCJ20160557

Ito, G., & Utsunomiya-Tate, N. (2023). Overview of the Impact of Pathogenic LRRK2 Mutations in Parkinson's Disease. *Biomolecules 2023, Vol. 13, Page 845*, *13*(5), 845. https://doi.org/10.3390/BIOM13050845

Jang, J., Oh, H., Nam, D., Seol, W., Seo, M. K., Park, S. W., Kim, H. G., Seo, H., Son, I., & Ho, D. H. (2018). Increase in anti-apoptotic molecules, nucleolin, and heat shock protein 70, against upregulated LRRK2 kinase activity. *Animal Cells and Systems*, *22*(5), 273–280. https://doi.org/10.1080/19768354.2018.1518262

Jennings, D., Huntwork-Rodriguez, S., Henry, A. G., Sasaki, J. C., Meisner, R., Diaz, D., Solanoy, H., Wang, X., Negrou, E., Bondar, V. V., Ghosh, R., Maloney, M. T., Propson, N. E., Zhu, Y., Maciuca, R. D., Harris, L., Kay, A., LeWitt, P., King, T. A., … Troyer, M. D. (2022). Preclinical and clinical evaluation of the LRRK2 inhibitor DNL201 for Parkinson's disease. *Science Translational Medicine*, *14*(648), 2658. https://doi.org/10.1126/SCITRANSLMED.ABJ2658/SUPPL_FILE/SCITRANSLMED.ABJ2658_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Jeong, G. R., Jang, E. H., Bae, J. R., Jun, S., Kang, H. C., Park, C. H., Shin, J. H., Yamamoto, Y., Tanaka-Yamamoto, K., Dawson, V. L., Dawson, T. M., Hur, E. M., & Lee, B. D. (2018). Dysregulated phosphorylation of Rab GTPases by LRRK2 induces neurodegeneration.

Molecular Neurodegeneration, *13*(1), 1–17. https://doi.org/10.1186/s13024-018-0240-1

Jiang, Z. C., Chen, X. J., Zhou, Q., Gong, X. H., Chen, X., & Wu, W. J. (2019). Downregulated LRRK2 gene expression inhibits proliferation and migration while promoting the apoptosis of thyroid cancer cells by inhibiting activation of the JNK signaling pathway. *International Journal of Oncology*, *55*(1), 21–34. https://doi.org/10.3892/IJO.2019.4816/HTML

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/SCIENCE.AAA8415

Juli, G., Gismondi, A., Monteleone, V., Caldarola, S., Iadevaia, V., Aspesi, A., Dianzani, I., Proud, C. G., & Loreni, F. (2016). Depletion of ribosomal protein S19 causes a reduction of rRNA synthesis. *Scientific Reports 2016 6:1*, *6*(1), 1–10. https://doi.org/10.1038/srep35026

Junker, B. H., & Schreiber, F. (2007). Analysis of Biological Networks. *Analysis of Biological Networks*, 1–346. https://doi.org/10.1002/9780470253489

Kachergus, J., Mata, I. F., Hulihan, M., Taylor, J. P., Lincoln, S., Aasly, J., Gibson, J. M., Ross, O. A., Lynch, T., Wiley, J., Payami, H., Nutt, J., Maraganore, D. M., Czyzewski, K., Styczynska, M., Wszolek, Z. K., Farrer, M. J., & Toft, M. (2005). Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: Evidence of a common founder across European populations. *American Journal of Human Genetics*, *76*(4), 672–680. https://doi.org/10.1086/429256

Kalia, L. V., Lang, A. E., Hazrati, L. N., Fujioka, S., Wszolek, Z. K., Dickson, D. W., Ross, O. A., Van Deerlin, V. M., Trojanowski, J. Q., Hurtig, H. I., Alcalay, R. N., Marder, K. S., Clark, L. N., Gaig, C., Tolosa, E., Ruiz-Martínez, J., Marti-Masso, J. F., Ferrer, I., López De Munain, A., … Marras, C. (2015). Clinical Correlations With Lewy Body Pathology in LRRK2-Related Parkinson Disease. *JAMA Neurology*, *72*(1), 100–105. https://doi.org/10.1001/JAMANEUROL.2014.2704

Kilisch, M., Lytovchenko, O., Arakel, E. C., Bertinetti, D., & Schwappach, B. (2016). A dual phosphorylation switch controls 14-3-3-dependent cell surface expression of TASK-1. *Journal of Cell Science*, *129*(4), 831–842. https://doi.org/10.1242/JCS.180182/260180/AM/A-DUAL-PHOSPHORYLATION-SWITCH-CONTROLS-14-3-3

Kim, C., Beilina, A., Smith, N., Li, Y., Kim, M., Kumaran, R., Kaganovich, A., Mamais, A., Adame, A., Iba, M., Kwon, S., Lee, W. J., Shin, S. J., Rissman, R. A., You, S., Lee, S. J., Singleton, A. B., Cookson, M. R., & Masliah, E. (2020). LRRK2 mediates microglial neurotoxicity via NFATc2 in rodent models of synucleinopathies. *Science Translational Medicine*, *12*(565). https://doi.org/10.1126/scitranslmed.aay0399

Kim, J. S., Park, I. S., Park, H. E., Kim, S. Y., Yun, J. A., Jung, C. K., Sung, H. Y., Lee, J. K., & Kang, W. K. (2017). α-Synuclein in the colon and premotor markers of Parkinson disease in neurologically normal subjects. *Neurological Sciences*, *38*(1), 171–179. https://doi.org/10.1007/S10072-016-2745-0/FIGURES/3

Kim, K. S., Marcogliese, P. C., Yang, J., Callaghan, S. M., Resende, V., Abdel-Messih, E., Marras, C., Visanji, N. P., Huang, J., Schlossmacher, M. G., Trinkle-Mulcahy, L., Slack, R. S., Lang, A. E., & Park, D. S. (2018). Regulation of myeloid cell phagocytosis by LRRK2 via WAVE2 complex stabilization is altered in Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(22), E5164–E5173. https://doi.org/10.1073/PNAS.1718946115/SUPPL_FILE/PNAS.1718946115.SAPP.PDF

Kingwell, K. (2023). LRRK2-targeted Parkinson disease drug advances into phase III. *Nature Reviews. Drug Discovery*, *22*(1), 3–5. https://doi.org/10.1038/D41573-022-00212-0

Koh, G. C. K. W., Porras, P., Aranda, B., Hermjakob, H., & Orchard, S. E. (2012). Analyzing protein-protein interaction networks. *Journal of Proteome Research*, *11*(4), 2014–2031. https://doi.org/10.1021/PR201211W/SUPPL_FILE/PR201211W_SI_002.PDF

Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*, *51*(W1), W207–W212. https://doi.org/10.1093/NAR/GKAD347

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. https://doi.org/10.1007/S13748-016-0094-0/TABLES/1

Krishna, K., & Murty, M. N. (1999). Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *29*(3), 433–439. https://doi.org/10.1109/3477.764879

Kubo, M., Kamiya, Y., Nagashima, R., Maekawa, T., Eshima, K., Azuma, S., Ohta, E., & Obata, F. (2010). LRRK2 is expressed in B-2 but not in B-1 B cells, and downregulated by cellular activation. *Journal of Neuroimmunology*, *229*(1–2), 123–128. https://doi.org/10.1016/J.JNEUROIM.2010.07.021

Kuhlmann, N., & Milnerwood, A. J. (2020). A Critical LRRK at the Synapse? The Neurobiological Function and Pathophysiological Dysfunction of LRRK2. *Frontiers in Molecular Neuroscience*, *13*, 571518. https://doi.org/10.3389/FNMOL.2020.00153/BIBTEX

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Maayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/NAR/GKW377

Kurita, T. (2020). Principal Component Analysis (PCA). *Computer Vision*, 1–4. https://doi.org/10.1007/978-3-030-03243-2_649-1

Kuwahara, T., & Iwatsubo, T. (2020). The Emerging Functions of LRRK2 and Rab GTPases in the Endolysosomal System. *Frontiers in Neuroscience*, *14*, 528270. https://doi.org/10.3389/FNINS.2020.00227/BIBTEX

Lai, S. W., Liao, K. F., Lin, C. L., & Sung, F. C. (2014). Irritable bowel syndrome correlates with increased risk of Parkinson's disease in Taiwan. *European Journal of Epidemiology*, *29*(1), 57–62. https://doi.org/10.1007/S10654-014-9878-3

Lamonaca, G., & Volta, M. (2020). Alpha-Synuclein and LRRK2 in Synaptic Autophagy: Linking Early Dysfunction to Late-Stage Pathology in Parkinson's Disease. *Cells 2020, Vol. 9, Page 1115*, *9*(5), 1115. https://doi.org/10.3390/CELLS9051115

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*. https://doi.org/10.1186/1471-2105-9-559

Lavalley, N. J., Slone, S. R., Ding, H., West, A. B., & Yacoubian, T. A. (2016). 14-3-3 Proteins regulate mutant LRRK2 kinase activity and neurite shortening. *Human Molecular Genetics*, *25*(1), 109–122. https://doi.org/10.1093/HMG/DDV453

Lee, A. J., Wang, Y., Alcalay, R. N., Mejia-Santana, H., Saunders-Pullman, R., Bressman, S., Corvol, J. C., Brice, A., Lesage, S., Mangone, G., Tolosa, E., Pont-Sunyer, C., Vilas, D., Schüle, B., Kausar, F., Foroud, T., Berg, D., Brockmann, K., Goldwurm, S., … Marder, K.

(2017). Penetrance estimate of LRRK2 p.G2019S mutation in individuals of non-Ashkenazi Jewish ancestry. *Movement Disorders*, *32*(10), 1432–1438. https://doi.org/10.1002/MDS.27059

Lesage, S., Dürr, A., Tazir, M., Lohmann, E., Leutenegger, A.-L., Janin, S., Pollak, P., & Brice, A. (2006). LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs. In *The New England journal of medicine* (Vol. 354, Issue 4, pp. 422–423). https://doi.org/10.1056/NEJMc055540

Lesniak, R. K., Nichols, R. J., & Montine, T. J. (2022). Development of mutation-selective LRRK2 kinase inhibitors as precision medicine for Parkinson's disease and other diseases for which carriers are at increased risk. *Frontiers in Neurology*, *13*. https://doi.org/10.3389/FNEUR.2022.1016040

Lewitt, P. A., & Library, W. O. (2015). Levodopa therapy for Parkinson's disease: Pharmacokinetics and pharmacodynamics. *Movement Disorders*, *30*(1), 64–72. https://doi.org/10.1002/MDS.26082

Li, C., Wei, Z., Fan, Y., Fukuda, M., Khater, M., & Correspondence, G. W. (2017). The GTPase Rab43 Controls the Anterograde ER-Golgi Trafficking and Sorting of GPCRs. *CellReports*, *21*, 1089–1101. https://doi.org/10.1016/j.celrep.2017.10.011

Li, W. (2012). Volcano plots in analyzing differential expressions with mRNA microarrays. *Journal of Bioinformatics and Computational Biology*, *10*(6). https://doi.org/10.1142/S0219720012310038

Li, X., Wang, Q. J., Pan, N., Lee, S., Zhao, Y., Chait, B. T., & Yue, Z. (2011). Phosphorylation-Dependent 14-3-3 Binding to LRRK2 Is Impaired by Common Mutations of Familial Parkinson's Disease. *PLOS ONE*, *6*(3), e17153. https://doi.org/10.1371/JOURNAL.PONE.0017153

Li, Y. yan, Zhou, T. tian, Zhang, Y., Chen, N. H., & Yuan, Y. H. (2022). Distribution of α-Synuclein Aggregation in the Peripheral Tissues. *Neurochemical Research*, *47*(12), 3627–3634. https://doi.org/10.1007/S11064-022-03586-0/METRICS

Lin, J. C., Lin, C. S., Hsu, C. W., Lin, C. L., & Kao, C. H. (2016). Association Between Parkinson's Disease and Inflammatory Bowel Disease: a Nationwide Taiwanese Retrospective Cohort Study. *Inflammatory Bowel Diseases*, *22*(5), 1049–1055. https://doi.org/10.1097/MIB.0000000000000735

Liu, X., Liu, B., Huang, Z., Shi, T., Chen, Y., & Zhang, J. (2012). SPPS: a sequence-based method for predicting probability of protein-protein interaction partners. *PloS One*, *7*(1). https://doi.org/10.1371/JOURNAL.PONE.0030938

Liu, Z., Lee, J., Krummey, S., Lu, W., Cai, H., & Lenardo, M. J. (2011). The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease. *Nature Immunology*, *12*(11), 1063–1070. https://doi.org/10.1038/NI.2113

Liu, Z., Xu, E., Zhao, H. T., Cole, T., & West, A. B. (2020). LRRK2 and Rab10 coordinate macropinocytosis to mediate immunological responses in phagocytes. *The EMBO Journal*, *39*(20), e104862. https://doi.org/10.15252/EMBJ.2020104862

Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Lü, J., & Wang, P. (2020). Modeling and Analysis of Bio-molecular Networks. *Modeling and Analysis of Bio-Molecular Networks*, 1–464. https://doi.org/10.1007/978-981-15-9144-0/COVER

Lu, J., Wu, M., & Yue, Z. (2020). Autophagy and parkinson's disease. *Advances in Experimental Medicine and Biology*, *1207*, 21–51. https://doi.org/10.1007/978-981-15-4272-5_2/COVER

Luciano, M. S., Lipton, R. B., Wang, C., Katz, M., Zimmerman, M. E., Sanders, A. E., Ozelius, L. J., Bressman, S. B., & Saunders-Pullman, R. (2010). Clinical expression of LRRK2 G2019S mutations in the elderly. *Movement Disorders*, *25*(15), 2571–2576. https://doi.org/10.1002/MDS.23330

Ma'ayan, A. (2011). Introduction to network analysis in systems biology. *Science Signaling*, *4*(190). https://doi.org/10.1126/SCISIGNAL.2001965/SUPPL_FILE/ANSWERKEY.PDF

Maćkiewicz, A., & Ratajczak, W. (1993). Principal components analysis (PCA). *Computers & Geosciences*, *19*(3), 303–342. https://doi.org/10.1016/0098-3004(93)90090-R

MacLeod, D. A., Rhinn, H., Kuwahara, T., Zolin, A., Di Paolo, G., MacCabe, B. D., Marder, K. S., Honig, L. S., Clark, L. N., Small, S. A., & Abeliovich, A. (2013). RAB7L1 Interacts with LRRK2 to Modify Intraneuronal Protein Sorting and Parkinson's Disease Risk. *Neuron*, *77*(3), 425–439. https://doi.org/10.1016/j.neuron.2012.11.033

MacLeod, D., Dowman, J., Hammond, R., Leete, T., Inoue, K., & Abeliovich, A. (2006). The Familial Parkinsonism Gene LRRK2 Regulates Neurite Process Morphology. *Neuron*, *52*(4), 587–593. https://doi.org/10.1016/j.neuron.2006.10.008

Madero-Pérez, J., Fernández, B., Lara Ordóñez, A. J., Fdez, E., Lobbestael, E., Baekelandt, V., & Hilfiker, S. (2018a). RAB7L1-mediated relocalization of LRRK2 to the golgi complex causes centrosomal deficits via RAB8A. *Frontiers in Molecular Neuroscience*, *11*, 421081. https://doi.org/10.3389/FNMOL.2018.00417/BIBTEX

Madero-Pérez, J., Fernández, B., Lara Ordóñez, A. J., Fdez, E., Lobbestael, E., Baekelandt, V., & Hilfiker, S. (2018b). RAB7L1-mediated relocalization of LRRK2 to the golgi complex causes centrosomal deficits via RAB8A. *Frontiers in Molecular Neuroscience*, *11*, 421081. https://doi.org/10.3389/FNMOL.2018.00417/BIBTEX

Madureira, M., Connor-Robson, N., & Wade-Martins, R. (2020). "LRRK2: Autophagy and Lysosomal Activity." *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/FNINS.2020.00498

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, *2014*(1), e593. https://doi.org/10.7717/PEERJ.593/SUPP-8

Mahid, S. S., Minor, K. S., Soto, R. E., Hornung, C. A., & Galandiuk, S. (2006). Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clinic Proceedings*, *81*(11), 1462–1471. https://doi.org/10.4065/81.11.1462

Manconi, A., Vargiu, E., Armano, G., & Milanesi, L. (2012). Literature retrieval and mining in bioinformatics: State of the art and challenges. *Advances in Bioinformatics*, *2012*. https://doi.org/10.1155/2012/573846

Mandemakers, W., Snellinx, A., O'Neill, M. J., & de Strooper, B. (2012). LRRK2 expression is enriched in the striosomal compartment of mouse striatum. *Neurobiology of Disease*, *48*(3), 582–593. https://doi.org/10.1016/J.NBD.2012.07.017

Manes, J. L., Tjaden, K., Parrish, T., Simuni, T., Roberts, A., Greenlee, J. D., Corcos, D. M., & Kurani, A. S. (2018). Altered resting-state functional connectivity of the putamen and

internal globus pallidus is related to speech impairment in Parkinson's disease. *Brain and Behavior*, *8*(9). https://doi.org/10.1002/BRB3.1073

Manschwetus, J. T., Wallbott, M., Fachinger, A., Obergruber, C., Pautz, S., Bertinetti, D., Schmidt, S. H., & Herberg, F. W. (2020a). Binding of the Human 14-3-3 Isoforms to Distinct Sites in the Leucine-Rich Repeat Kinase 2. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/FNINS.2020.00302

Manschwetus, J. T., Wallbott, M., Fachinger, A., Obergruber, C., Pautz, S., Bertinetti, D., Schmidt, S. H., & Herberg, F. W. (2020b). Binding of the Human 14-3-3 Isoforms to Distinct Sites in the Leucine-Rich Repeat Kinase 2. *Frontiers in Neuroscience*, *14*, 531922. https://doi.org/10.3389/FNINS.2020.00302/BIBTEX

Manzoni, C., Denny, P., Lovering, R. C., & Lewis, P. A. (2015). Computational analysis of the LRRK2 interactome. *PeerJ*, *2015*(2). https://doi.org/10.7717/peerj.778

Manzoni, C., Mamais, A., Dihanich, S., Abeti, R., Soutar, M. P. M., Plun-Favreau, H., Giunti, P., Tooze, S. A., Bandopadhyay, R., & Lewis, P. A. (2013). Inhibition of LRRK2 kinase activity stimulates macroautophagy. *Biochimica et Biophysica Acta*, *1833*(12), 2900–2910. https://doi.org/10.1016/J.BBAMCR.2013.07.020

Manzoni, C., Mamais, A., Roosen, D. A., Dihanich, S., Soutar, M. P. M., Plun-Favreau, H., Bandopadhyay, R., Hardy, J., Tooze, S. A., Cookson, M. R., & Lewis, P. A. (2016). mTOR independent regulation of macroautophagy by Leucine Rich Repeat Kinase 2 via Beclin-1. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP35106

Marchand, A., Drouyer, M., Sarchione, A., Chartier-Harlin, M. C., & Taymans, J. M. (2020). LRRK2 Phosphorylation, More Than an Epiphenomenon. *Frontiers in Neuroscience*, *14*, 539593. https://doi.org/10.3389/FNINS.2020.00527/BIBTEX

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C. S., Caspell-Garcia, C., Simuni, T., Jennings, D., Tanner, C. M., Trojanowski, J. Q., Shaw, L. M., Seibyl, J., Schuff, N., Singleton, A., Kieburtz, K., Toga, A. W., Mollenhauer, B., Galasko, D., Chahine, L. M., … Taucher, J. (2018). The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, *5*(12), 1460–1477. https://doi.org/10.1002/acn3.644

Maries, E., Dass, B., Collier, T. J., Kordower, J. H., & Steece-Collier, K. (2003). The role of α-synuclein in Parkinson's disease: insights from animal models. *Nature Reviews Neuroscience 2003 4:9*, *4*(9), 727–738. https://doi.org/10.1038/nrn1199

Marte, A., Russo, I., Rebosio, C., Valente, P., Belluzzi, E., Pischedda, F., Montani, C., Lavarello, C., Petretto, A., Fedele, E., Baldelli, P., Benfenati, F., Piccoli, G., Greggio, E., & Onofri, F. (2019). Leucine-rich repeat kinase 2 phosphorylation on synapsin I regulates glutamate release at pre-synaptic sites. *Journal of Neurochemistry*, *150*(3), 264–281. https://doi.org/10.1111/JNC.14778

Martin, I., Kim, J. W., Lee, B. D., Kang, H. C., Xu, J. C., Jia, H., Stankowski, J., Kim, M. S., Zhong, J., Kumar, M., Andrabi, S. A., Xiong, Y., Dickson, D. W., Wszolek, Z. K., Pandey, A., Dawson, T. M., & Dawson, V. L. (2014). Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. *Cell*, *157*(2), 472. https://doi.org/10.1016/J.CELL.2014.01.064

Mata, I. F., Kachergus, J. M., Taylor, J. P., Lincoln, S., Aasly, J., Lynch, T., Hulihan, M. M., Cobb, S. A., Wu, R. M., Lu, C. S., Lahoz, C., Wszolek, Z. K., & Farrer, M. J. (2005). Lrrk2 pathogenic substitutions in Parkinson's disease. *Neurogenetics*, *6*(4), 171–177. https://doi.org/10.1007/S10048-005-0005-1/TABLES/3

Matejuk, A., Vandenbark, A. A., & Offner, H. (2021). Cross-Talk of the CNS With Immune Cells and Functions in Health and Disease. *Frontiers in Neurology*, *12*, 672455. https://doi.org/10.3389/FNEUR.2021.672455/BIBTEX

Maza, E., Frasse, P., Senin, P., Bouzayen, M., & Zouine, M. (2013). Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments. *Communicative & Integrative Biology*, *6*(6). https://doi.org/10.4161/CIB.25849

Melrose, H. L., Kent, C. B., Taylor, J. P., Dachsel, J. C., Hinkle, K. M., Lincoln, S. J., Mok, S. S., Culvenor, J. G., Masters, C. L., Tyndall, G. M., Bass, D. I., Ahmed, Z., Andorfer, C. A., Ross, O. A., Wszolek, Z. K., Delldonne, A., Dickson, D. W., & Farrer, M. J. (2007). A comparative analysis of leucine-rich repeat kinase 2 (Lrrk2) expression in mouse brain and Lewy body disease. *Neuroscience*, *147*(4), 1047–1058. https://doi.org/10.1016/j.neuroscience.2007.05.027

Mesarović, M. D. (1968). Systems Theory and Biology—View of a Theoretician. *Systems Theory and Biology*, 59–87. https://doi.org/10.1007/978-3-642-88343-9_3

Mesarovic, M. D., Sreenath, S. N., & Keene, J. D. (2004). Search for organising principles: understanding in systems biology. *Systems Biology*, *1*(1), 19–27. https://doi.org/10.1049/SB:20045010

Miklossy, J., Arai, T., Guo, J. P., Klegeris, A., Yu, S., McGeer, E. G., & McGeer, P. L. (2006). LRRK2 Expression in Normal and Pathologic Human Brain and in Human Cell Lines. *Journal of Neuropathology & Experimental Neurology*, *65*(10), 953–963. https://doi.org/10.1097/01.JNEN.0000235121.98052.54

Mills, R. D., Mulhern, T. D., Cheng, H. C., & Culvenor, J. G. (2012). Analysis of LRRK2 accessory repeat domains: prediction of repeat length, number and sites of Parkinson's disease mutations. *Biochemical Society Transactions*, *40*(5), 1086–1089. https://doi.org/10.1042/BST20120088

Mizuno-Yamasaki, E., Rivera-Molina, F., & Novick, P. (2012). GTPase Networks in Membrane Traffic. *Https://Doi.Org/10.1146/Annurev-Biochem-052810-093700*, *81*, 637–659. https://doi.org/10.1146/ANNUREV-BIOCHEM-052810-093700

Moehle, M. S., Webber, P. J., Tse, T., Sukar, N., Standaert, D. G., Desilva, T. M., Cowell, R. M., & West, A. B. (2012). LRRK2 inhibition attenuates microglial inflammatory responses. *Journal of Neuroscience*, *32*(5), 1602–1611. https://doi.org/10.1523/JNEUROSCI.5601-11.2012

Mukhopadhyay, A., Nieves, E., Che, F. Y., Wang, J., Jin, L., Murray, J. W., Gordon, K., Angeletti, R. H., & Wolkoff, A. W. (2011). Proteomic analysis of endocytic vesicles: Rab1a regulates motility of early endocytic vesicles. *Journal of Cell Science*, *124*(5), 765–775. https://doi.org/10.1242/JCS.079020

Nazish, I., Arber, C., Piers, T. M., Warner, T. T., Hardy, J. A., Lewis, P. A., Pocock, J. M., & Bandopadhyay, R. (2021). Abrogation of LRRK2 dependent Rab10 phosphorylation with TLR4 activation and alterations in evoked cytokine release in immune cells. *Neurochemistry International*, *147*, 105070. https://doi.org/10.1016/J.NEUINT.2021.105070

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *69*(6), 5. https://doi.org/10.1103/PHYSREVE.69.066133/FIGURES/5/MEDIUM

Nguyen, A. P. T., & Moore, D. J. (2017). Understanding the GTPase Activity of LRRK2: Regulation, Function, and Neurotoxicity. *Advances in Neurobiology*, *14*, 71. https://doi.org/10.1007/978-3-319-49969-7_4

Nguyen, M., Wong, Y. C., Ysselstein, D., Severino, A., & Krainc, D. (2019). Synaptic, Mitochondrial, and Lysosomal Dysfunction in Parkinson's Disease. *Trends in Neurosciences*, *42*(2), 140–149. https://doi.org/10.1016/j.tins.2018.11.001

Nicholl, D. J., Vaughan, J. R., Khan, N. L., Ho, S. L., Aldous, D. E. W., Lincoln, S., Farrer, M., Gayton, J. D., Davis, M. B., Piccini, P., Daniel, S. E., Lennox, G. G., Brooks, D. J., Williams, A. C., & Wood, N. W. (2002). Two large British kindreds with familial Parkinson's disease: a clinico-pathological and genetic study. *Brain*, *125*(1), 44–57. https://doi.org/10.1093/BRAIN/AWF013

Nichols, R. J., Dzamko, N., Morrice, N. A., Campbell, D. G., Deak, M., Ordureau, A., Macartney, T., Tong, Y., Shen, J., Prescott, A. R., & Alessi, D. R. (2010a). 14-3-3 binding to LRRK2 is disrupted by multiple Parkinson's disease-associated mutations and regulates cytoplasmic localization. *Biochemical Journal*, *430*(3), 393–404. https://doi.org/10.1042/BJ20100483

Nichols, R. J., Dzamko, N., Morrice, N. A., Campbell, D. G., Deak, M., Ordureau, A., Macartney, T., Tong, Y., Shen, J., Prescott, A. R., & Alessi, D. R. (2010b). 14-3-3 binding to LRRK2 is disrupted by multiple Parkinson's disease-associated mutations and regulates cytoplasmic localization. *The Biochemical Journal*, *430*(3), 393–404. https://doi.org/10.1042/BJ20100483

Nichols, R. J., Dzamko, N., Morrice, N. A., Campbell, D. G., Deak, M., Ordureau, A., Macartney, T., Tong, Y., Shen, J., Prescott, A. R., & Alessi, D. R. (2010c). 14-3-3 binding to LRRK2 is disrupted by multiple Parkinson's disease-associated mutations and regulates cytoplasmic localization. *Biochemical Journal*, *430*(Pt 3), 393. https://doi.org/10.1042/BJ20100483

Nielsen, F. (2016). *Hierarchical Clustering*. 195–211. https://doi.org/10.1007/978-3-319-21903-5_8

Niemann, N., Billnitzer, A., & Jankovic, J. (2021). Parkinson's disease and skin. *Parkinsonism & Related Disorders*, *82*, 61–76. https://doi.org/10.1016/J.PARKRELDIS.2020.11.017

Nixon-Abell, J., Berwick, D. C., Grannó, S., Spain, V. A., Blackstone, C., & Harvey, K. (2016). Protective LRRK2 R1398H variant enhances GTPase and Wnt signaling activity. *Frontiers in Molecular Neuroscience*, *9*(MAR), 179717. https://doi.org/10.3389/FNMOL.2016.00018/BIBTEX

Noyce, A. J., Lees, A. J., & Schrag, A. E. (2016). The prediagnostic phase of Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *87*(8), 871–878. https://doi.org/10.1136/JNNP-2015-311890/-/DC1

Nuytemans, K., Rademakers, R., Theuns, J., Pals, P., Engelborghs, S., Pickut, B., de Pooter, T., Peeters, K., Mattheijssens, M., Van den Broeck, M., Cras, P., De Deyn, P. P., & van Broeckhoven, C. (2008). Founder mutation p.R1441C in the leucine-rich repeat kinase 2 gene in Belgian Parkinson's disease patients. *European Journal of Human Genetics : EJHG*, *16*(4), 471–479. https://doi.org/10.1038/SJ.EJHG.5201986

Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Baker, J., Malangone, C., Lopez, I., Miranda, A., Cruz-Castillo, C., Fumis, L., Bernal-Llinares, M., Tsukanov, K., Cornu, H., Tsirigos, K., Razuvayevskaya, O., Buniello, A., Schwartzentruber, J., Karim, M., Ariano, B., … McDonagh, E. M. (2023). The next-generation Open Targets Platform: reimagined,

redesigned, rebuilt. *Nucleic Acids Research*, *51*(D1), D1353–D1359. https://doi.org/10.1093/NAR/GKAC1046

Ohta, E., Kawakami, F., Kubo, M., & Obata, F. (2011). LRRK2 directly phosphorylates Akt1 as a possible physiological substrate: impairment of the kinase activity by Parkinson's disease-associated mutations. *FEBS Letters*, *585*(14), 2165–2170. https://doi.org/10.1016/J.FEBSLET.2011.05.044

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., Del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., … Hermjakob, H. (2014). The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1115

Ousman, S. S., & Kubes, P. (2012). Immune surveillance in the central nervous system. In *Nature Neuroscience* (Vol. 15, Issue 8, pp. 1096–1101). Nature Publishing Group. https://doi.org/10.1038/nn.3161

Ozelius, L. J., Senthil, G., Saunders-Pullman, R., Ohmann, E., Deligtisch, A., Tagliati, M., Hunt, A. L., Klein, C., Henick, B., Hailpern, S. M., Lipton, R. B., Soto-Valencia, J., Risch, N., & Bressman, S. B. (2006). LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. In *The New England journal of medicine* (Vol. 354, Issue 4, pp. 424–425). https://doi.org/10.1056/NEJMc055509

Paisán-Ruíz, C., Jain, S., Evans, E. W., Gilks, W. P., Simón, J., Van Der Brug, M., De Munain, A. L., Aparicio, S., Gil, A. M., Khan, N., Johnson, J., Martinez, J. R., Nicholl, D., Carrera, I. M., Peňa, A. S., De Silva, R., Lees, A., Martí-Massó, J. F., Pérez-Tur, J., … Singleton, A. B. (2004). Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron*, *44*(4), 595–600. https://doi.org/10.1016/j.neuron.2004.10.023

Pan, P. Y., Li, X., Wang, J., Powell, J., Wang, Q., Zhang, Y., Chen, Z., Wicinski, B., Hof, P., Ryan, T. A., & Yue, Z. (2017). Parkinson's Disease-Associated LRRK2 Hyperactive Kinase Mutant Disrupts Synaptic Vesicle Trafficking in Ventral Midbrain Neurons. *Journal of Neuroscience*, *37*(47), 11366–11376. https://doi.org/10.1523/JNEUROSCI.0964-17.2017

Panagiotakopoulou, V., Ivanyuk, D., De Cicco, S., Haq, W., Arsić, A., Yu, C., Messelodi, D., Oldrati, M., Schöndorf, D. C., Perez, M. J., Cassatella, R. P., Jakobi, M., Schneiderhan-Marra, N., Gasser, T., Nikić-Spiegel, I., & Deleidi, M. (2020). Interferon-γ signaling synergizes with LRRK2 in neurons and microglia derived from human induced pluripotent stem cells. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18755-4

Parisiadou, L., Xie, C., Hyun, J. C., Lin, X., Gu, X. L., Long, C. X., Lobbestael, E., Baekelandt, V., Taymans, J. M., Sun, L., & Cai, H. (2009). Phosphorylation of Ezrin/Radixin/Moesin Proteins by LRRK2 Promotes the Rearrangement of Actin Cytoskeleton in Neuronal Morphogenesis. *Journal of Neuroscience*, *29*(44), 13971–13980. https://doi.org/10.1523/JNEUROSCI.3799-09.2009

Parkinson, J. (2002). An essay on the shaking palsy. 1817. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *14*(2). https://doi.org/10.1176/JNP.14.2.223

Peter, I., Dubinsky, M., Bressman, S., Park, A., Lu, C., Chen, N., & Wang, A. (2018). Anti-Tumor Necrosis Factor Therapy and Incidence of Parkinson Disease Among Patients With Inflammatory Bowel Disease. *JAMA Neurology*, *75*(8), 939–946. https://doi.org/10.1001/JAMANEUROL.2018.0605

Pfeffer, S. R. (2017). Rab GTPases: Master regulators that establish the secretory and endocytic pathways. *Molecular Biology of the Cell*, *28*(6), 712–715. https://doi.org/10.1091/MBC.E16-10-0737/ASSET/IMAGES/LARGE/712FIG1.JPEG

Picca, A., Guerra, F., Calvani, R., Romano, R., Josécoelho-Júnior, H., Bucci, C., & Marzetti, E. (2021). Mitochondrial Dysfunction, Protein Misfolding and Neuroinflammation in Parkinson's Disease: Roads to Biomarker Discovery. *Biomolecules 2021, Vol. 11, Page 1508*, *11*(10), 1508. https://doi.org/10.3390/BIOM11101508

Playford, E. D., Jenkins, I. H., Passingham, R. E., Nutt, J., Frackowiak, R. S. J., & Brooks, D. J. (1992). Impaired mesial frontal and putamen activation in Parkinson's disease: a positron emission tomography study. *Annals of Neurology*, *32*(2), 151–161. https://doi.org/10.1002/ANA.410320206

Plotegher, N., & Civiero, L. (2012). Neuronal Autophagy, α-Synuclein Clearance, and LRRK2 Regulation: A Lost Equilibrium in Parkinsonian Brain. *Journal of Neuroscience*, *32*(43), 14851–14853. https://doi.org/10.1523/JNEUROSCI.3588-12.2012

Poewe, W. (2008). Non-motor symptoms in Parkinson's disease. In *European Journal of Neurology* (Vol. 15, Issue SUPPL. 1, pp. 14–20). Eur J Neurol. https://doi.org/10.1111/j.1468-1331.2008.02056.x

Porras, P., Duesbury, M., Fabregat, A., Ueffing, M., Orchard, S., Gloeckner, C. J., & Hermjakob, H. (2015a). A visual review of the interactome of LRRK2: Using deep-curated molecular interaction data to represent biology. In *Proteomics* (Vol. 15, Issue 8, pp. 1390–1404). Wiley-VCH Verlag. https://doi.org/10.1002/pmic.201400390

Porras, P., Duesbury, M., Fabregat, A., Ueffing, M., Orchard, S., Gloeckner, C. J., & Hermjakob, H. (2015b). A visual review of the interactome of LRRK2: Using deep-curated molecular interaction data to represent biology. *Proteomics*, *15*(8), 1390. https://doi.org/10.1002/PMIC.201400390

Proulx, S. T., & Engelhardt, B. (2022). Central nervous system zoning: How brain barriers establish subdivisions for CNS immune privilege and immune surveillance. *Journal of Internal Medicine*, *292*(1), 47–67. https://doi.org/10.1111/JOIM.13469

Purlyte, E., Dhekne, H. S., Sarhan, A. R., Gomez, R., Lis, P., Wightman, M., Martinez, T. N., Tonelli, F., Pfeffer, S. R., & Alessi, D. R. (2018a). Rab29 activation of the Parkinson's disease-associated LRRK2 kinase. *The EMBO Journal*, *37*(1), 1–18. https://doi.org/10.15252/EMBJ.201798099

Purlyte, E., Dhekne, H. S., Sarhan, A. R., Gomez, R., Lis, P., Wightman, M., Martinez, T. N., Tonelli, F., Pfeffer, S. R., & Alessi, D. R. (2018b). Rab29 activation of the Parkinson's disease-associated LRRK2 kinase. *The EMBO Journal*, *37*(1), 1–18. https://doi.org/10.15252/EMBJ.201798099

Qi, W., Schlapbach, R., & Rehrauer, H. (2017). RNA-Seq Data Analysis: From Raw Data Quality Control to Differential Expression Analysis. *Methods in Molecular Biology*, *1669*, 295–307. https://doi.org/10.1007/978-1-4939-7286-9_23

Rahiminejad, S., Maurya, M. R., & Subramaniam, S. (2019). Topological and functional comparison of community detection algorithms in biological networks. *BMC Bioinformatics*, *20*(1), 1–25. https://doi.org/10.1186/S12859-019-2746-0/FIGURES/6

Rivero-Ríos, P., Romo-Lozano, M., Madero-Pérez, J., Thomas, A. P., Biosa, A., Greggio, E., & Hilfiker, S. (2019). The G2019S variant of leucine-rich repeat kinase 2 (LRRK2) alters endolysosomal trafficking by impairing the function of the GTPase RAB8A. *Journal of Biological Chemistry*, *294*(13), 4738–4758.

https://doi.org/10.1074/JBC.RA118.005008/ATTACHMENT/AB87E7C2-DB39-4183-8369-D2A3B28C83E0/MMC1.ZIP

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/BIOINFORMATICS/BTP616

Ross, G. W., Petrovitch, H., Abbott, R. D., Nelson, J., Markesbery, W., Davis, D., Hardman, J., Launer, L., Masaki, K., Tanner, C. M., & White, L. R. (2004). Parkinsonian signs and substantia nigra neuron density in decendents elders without PD. *Annals of Neurology*, *56*(4), 532–539. https://doi.org/10.1002/ANA.20226

Ross, O. A., Wu, Y. R., Lee, M. C., Funayama, M., Chen, M. L., Soto, A. I., Mata, I. F., Lee-Chen, G. J., Chiung, M. C., Tang, M., Zhao, Y., Hattori, N., Farrer, M. J., Tan, E. K., & Wu, R. M. (2008). Analysis of Lrrk2 R1628P as a risk factor for Parkinson's disease. *Annals of Neurology*, *64*(1), 88–92. https://doi.org/10.1002/ANA.21405

Rota, L., Pellegrini, C., Benvenuti, L., Antonioli, L., Fornai, M., Blandizzi, C., Cattaneo, A., & Colla, E. (2019). Constipation, deficit in colon contractions and alpha-synuclein inclusions within the colon precede motor abnormalities and neurodegeneration in the central nervous system in a mouse model of alpha-synucleinopathy. *Translational Neurodegeneration*, *8*(1), 1–15. https://doi.org/10.1186/S40035-019-0146-Z/FIGURES/6

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Russo, I., Berti, G., Plotegher, N., Bernardo, G., Filograna, R., Bubacco, L., & Greggio, E. (2015). Leucine-rich repeat kinase 2 positively regulates inflammation and down-regulates NF-κB p50 signaling in cultured microglia cells. *Journal of Neuroinflammation*, *12*(1), 1–13. https://doi.org/10.1186/s12974-015-0449-7

Ryan, K. J., White, C. C., Patel, K., Xu, J., Olah, M., Replogle, J. M., Frangieh, M., Cimpean, M., Winn, P., McHenry, A., Kaskow, B. J., Chan, G., Cuerdon, N., Bennett, D. A., Boyd, J. D., Imitola, J., Elyaman, W., De Jager, P. L., & Bradshaw, E. M. (2017). A human microglia-like cellular model for assessing the effects of neurodegenerative disease gene variants. *Science Translational Medicine*, *9*(421), 7635. https://doi.org/10.1126/SCITRANSLMED.AAI7635/SUPPL_FILE/AAI7635_TABLES_S1_TO_S3.XLSX

Saez-Atienzar, S., Bonet-Ponce, L., Blesa, J. R., Romero, F. J., Murphy, M. P., Jordan, J., & Galindo, M. F. (2014). The LRRK2 inhibitor GSK2578215A induces protective autophagy in SH-SY5Y cells: involvement of Drp-1-mediated mitochondrial fission and mitochondrial-derived ROS signaling. *Cell Death & Disease*, *5*(8). https://doi.org/10.1038/CDDIS.2014.320

Sánchez-Danés, A., Richaud-Patin, Y., Carballo-Carbajal, I., Jiménez-Delgado, S., Caig, C., Mora, S., Di Guglielmo, C., Ezquerra, M., Patel, B., Giralt, A., Canals, J. M., Memo, M., Alberch, J., López-Barneo, J., Vila, M., Cuervo, A. M., Tolosa, E., Consiglio, A., & Raya, A. (2012). Disease-specific phenotypes in dopamine neurons from human iPS-based models of genetic and sporadic Parkinson's disease. *EMBO Molecular Medicine*, *4*(5), 380–395. https://doi.org/10.1002/EMMM.201200215

Schapansky, J., Nardozzi, J. D., Felizia, F., & LaVoie, M. J. (2014). Membrane recruitment of endogenous LRRK2 precedes its potent regulation of autophagy. *Human Molecular Genetics*, *23*(16), 4201–4214. https://doi.org/10.1093/HMG/DDU138

Schlüter, O. M., Schmitz, F., Jahn, R., Rosenmund, C., & Südhof, T. C. (2004). A Complete Genetic Analysis of Neuronal Rab3 Function. *Journal of Neuroscience*, *24*(29), 6629–6637. https://doi.org/10.1523/JNEUROSCI.1610-04.2004

Sen, S., Webber, P. J., & West, A. B. (2009). Dependence of Leucine-rich Repeat Kinase 2 (LRRK2) Kinase Activity on Dimerization. *Journal of Biological Chemistry*, *284*(52), 36346–36356. https://doi.org/10.1074/JBC.M109.025437

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, *50*(W1), W216–W221. https://doi.org/10.1093/NAR/GKAC194

Shin, N., Jeong, H., Kwon, J., Heo, H. Y., Kwon, J. J., Yun, H. J., Kim, C. H., Han, B. S., Tong, Y., Shen, J., Hatano, T., Hattori, N., Kim, K. S., Chang, S., & Seol, W. (2008). LRRK2 regulates synaptic vesicle endocytosis. *Experimental Cell Research*, *314*(10), 2055–2065. https://doi.org/10.1016/J.YEXCR.2008.02.015

Shu, Y., Ming, J., Zhang, P., Wang, Q., Jiao, F., & Tian, B. (2016). Parkinson-Related LRRK2 Mutation R1628P Enables Cdk5 Phosphorylation of LRRK2 and Upregulates Its Kinase Activity. *PLOS ONE*, *11*(3), e0149739. https://doi.org/10.1371/JOURNAL.PONE.0149739

Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., Huang, J., Du, Y., Lin, L., Dong, Z., Yang, L., Liu, X., Jiang, H., Xu, X., Wang, J., … Mulder, J. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, *367*(6482). https://doi.org/10.1126/SCIENCE.AAY4106

Sonninen, T. M., Hämäläinen, R. H., Koskuvi, M., Oksanen, M., Shakirzyanova, A., Wojciechowski, S., Puttonen, K., Naumenko, N., Goldsteins, G., Laham-Karam, N., Lehtonen, M., Tavi, P., Koistinaho, J., & Lehtonen, Š. (2020). Metabolic alterations in Parkinson's disease astrocytes. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-71329-8

Soudy, M., Anwar, A. M., Ahmed, E. A., Osama, A., Ezzeldin, S., Mahgoub, S., & Magdeldin, S. (2020). UniprotR: Retrieving and visualizing protein sequence and functional information from Universal Protein Resource (UniProt knowledgebase). *Journal of Proteomics*, *213*, 103613. https://doi.org/10.1016/j.jprot.2019.103613

Spanaki, C., Latsoudis, H., & Plaitakis, A. (2006). LRRK2 mutations on Crete: R1441H associated with PD evolving to PSP. *Neurology*, *67*(8), 1518–1519. https://doi.org/10.1212/01.WNL.0000239829.33936.73

Steger, M., Diez, F., Dhekne, H. S., Lis, P., Nirujogi, R. S., Karayel, O., Tonelli, F., Martinez, T. N., Lorentzen, E., Pfeffer, S. R., Alessi, D. R., & Mann, M. (2017a). Systematic proteomic analysis of LRRK2-mediated rab GTPase phosphorylation establishes a connection to ciliogenesis. *ELife*, *6*. https://doi.org/10.7554/ELIFE.31012

Steger, M., Diez, F., Dhekne, H. S., Lis, P., Nirujogi, R. S., Karayel, O., Tonelli, F., Martinez, T. N., Lorentzen, E., Pfeffer, S. R., Alessi, D. R., & Mann, M. (2017b). Systematic proteomic analysis of LRRK2-mediated Rab GTPase phosphorylation establishes a connection to ciliogenesis. *ELife*, *6*. https://doi.org/10.7554/ELIFE.31012

Steger, M., Tonelli, F., Ito, G., Davies, P., Trost, M., Vetter, M., Wachter, S., Lorentzen, E., Duddy, G., Wilson, S., Baptista, M. A. S., Fiske, B. K., Fell, M. J., Morrow, J. A., Reith, A. D., Alessi, D. R., & Mann, M. (2016a). Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *ELife*, *5*(JANUARY2016). https://doi.org/10.7554/eLife.12813.001

Steger, M., Tonelli, F., Ito, G., Davies, P., Trost, M., Vetter, M., Wachter, S., Lorentzen, E., Duddy, G., Wilson, S., Baptista, M. A. S., Fiske, B. K., Fell, M. J., Morrow, J. A., Reith, A. D., Alessi, D. R., & Mann, M. (2016b). Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *ELife*, *5*(JANUARY2016). https://doi.org/10.7554/ELIFE.12813.001

Stevenson, R., Samokhina, E., Rossetti, I., Morley, J. W., & Buskila, Y. (2020). Neuromodulation of Glial Function During Neurodegeneration. In *Frontiers in Cellular Neuroscience* (Vol. 14, p. 278). Frontiers Media S.A. https://doi.org/10.3389/fncel.2020.00278

Stevers, L. M., De Vries, R. M. J. M., Doveston, R. G., Milroy, L. G., Brunsveld, L., & Ottmann, C. (2017). Structural interface between LRRK2 and 14-3-3 protein. *Biochemical Journal*, *474*(7), 1273–1287. https://doi.org/10.1042/BCJ20161078

Stokholm, M. G., Danielsen, E. H., Hamilton-Dutoit, S. J., & Borghammer, P. (2016). Pathological α-synuclein in gastrointestinal tissues from prodromal Parkinson disease patients. *Annals of Neurology*, *79*(6), 940–949. https://doi.org/10.1002/ANA.24648

Stolzenberg, E., Berry, D., Yang, D., Lee, E. Y., Kroemer, A., Kaufman, S., Wong, G. C. L., Oppenheim, J. J., Sen, S., Fishbein, T., Bax, A., Harris, B., Barbut, D., & Zasloff, M. A. (2017). A Role for Neuronal Alpha-Synuclein in Gastrointestinal Immunity. *Journal of Innate Immunity*, *9*(5), 456–463. https://doi.org/10.1159/000477990

Suzuki, K., Fujita, H., Matsubara, T., & Hirata, K. (2019). Non-motor symptoms in postural instability/gait difficulty subtype in the early stage of Parkinson's disease. *European Journal of Neurology*, *26*(3), e37–e37. https://doi.org/10.1111/ENE.13828

Tan, E. K., Peng, R., Teo, Y. Y., Tan, L. C., Angeles, D., Ho, P., Chen, M. L., Lin, C. H., Mao, X. Y., Chang, X. L., Prakash, K. M., Liu, J. J., Au, W. L., Le, W. D., Jankovic, J., Burgunder, J. M., Zhao, Y., & Wu, R. M. (2010). Multiple LRRK2 variants modulate risk of Parkinson disease: a chinese multicenter study. *Human Mutation*, *31*(5), 561–568. https://doi.org/10.1002/HUMU.21225

Tang, M., Sun, J., Shimizu, K., & Kadota, K. (2015). Evaluation of methods for differential expression analysis on multi-group RNA-seq count data. *BMC Bioinformatics*, *16*(1), 1–14. https://doi.org/10.1186/S12859-015-0794-7/FIGURES/2

Tarassov, K., Messier, V., Landry, C. R., Radinovic, S., Serna Molina, M. M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., & Michnick, S. W. (2008). An in vivo map of the yeast protein interactome. *Science*, *320*(5882), 1465–1470. https://doi.org/10.1126/SCIENCE.1153878/SUPPL_FILE/IVICI_MANUAL.PDF

Tezuka, T., Taniguchi, D., Sano, M., Shimada, T., Oji, Y., Tsunemi, T., Ikeda, A., Li, Y., Yoshino, H., Ogata, J., Shiba-Fukushima, K., Funayama, M., Nishioka, K., Imai, Y., & Hattori, N. (2022). Pathophysiological evaluation of the LRRK2 G2385R risk variant for Parkinson's disease. *Npj Parkinson's Disease 2022 8:1*, *8*(1), 1–7. https://doi.org/10.1038/s41531-022-00367-y

Thévenet, J., Gobert, R., van Huijsduijnen, R. H., Wiessner, C., & Sagot, Y. J. (2011a). Regulation of LRRK2 Expression Points to a Functional Role in Human Monocyte Maturation. *PLOS ONE*, *6*(6), e21519. https://doi.org/10.1371/JOURNAL.PONE.0021519

Thévenet, J., Gobert, R., van Huijsduijnen, R. H., Wiessner, C., & Sagot, Y. J. (2011b). Regulation of LRRK2 expression points to a functional role in human monocyte maturation. *PloS One*, *6*(6). https://doi.org/10.1371/JOURNAL.PONE.0021519

Thirstrup, K., Dächsel, J. C., Oppermann, F. S., Williamson, D. S., Smith, G. P., Fog, K., & Christensen, K. V. (2017a). Selective LRRK2 kinase inhibition reduces phosphorylation of endogenous Rab10 and Rab12 in human peripheral mononuclear blood cells. *Scientific Reports*, *7*(1). https://doi.org/10.1038/S41598-017-10501-Z

Thirstrup, K., Dächsel, J. C., Oppermann, F. S., Williamson, D. S., Smith, G. P., Fog, K., & Christensen, K. V. (2017b). Selective LRRK2 kinase inhibition reduces phosphorylation of endogenous Rab10 and Rab12 in human peripheral mononuclear blood cells. *Scientific Reports*, *7*(1). https://doi.org/10.1038/S41598-017-10501-Z

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. https://doi.org/10.1007/BF02289263/METRICS

Tian, Y., Lv, J., Su, Z., Wu, T., Li, X., Hu, X., Zhang, J., & Wu, L. (2021). LRRK2 plays essential roles in maintaining lung homeostasis and preventing the development of pulmonary fibrosis. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(35), e2106685118. https://doi.org/10.1073/PNAS.2106685118/SUPPL_FILE/PNAS.2106685118.SAPP.PDF

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

Tombesi, G., Chen, C., Favetta, G., Plotegher, N., Sevegnani, M., Marte, A., Battisti, I., Civiero, L., Onofri, F., Piccoli, G., Arrigoni, G., Manzoni, C., Parisiadou, L., & Greggio, E. (2022). LRRK2 regulates dendritic spine dynamics through interaction with post-synaptic actin cytoskeleton. *BioRxiv*, 2022.10.31.514622. https://doi.org/10.1101/2022.10.31.514622

Tombesi, G., Kompella, S., Favetta, G., Chen, C., Zhao, Y., Sevegnani, M., Marte, A., Battisti, I., Morosin, E., Ornaghi, M., Iannotta, L., Plotegher, N., Civiero, L., Onofri, F., Eickholt, B. J., Piccoli, G., Arrigoni, G., Beccano-Kelly, D., Manzoni, C., … Greggio, E. (2023). LRRK2 regulates synaptic function through BDNF signaling and actin cytoskeleton. *BioRxiv*, 2022.10.31.514622. https://doi.org/10.1101/2022.10.31.514622

Tomkins, J. E., Ferrari, R., Vavouraki, N., Hardy, J., Hardy, J., Hardy, J., Hardy, J., Hardy, J., Lovering, R. C., Lewis, P. A., Lewis, P. A., Lewis, P. A., McGuffin, L. J., Manzoni, C., & Manzoni, C. (2020). PINOT: An intuitive resource for integrating protein-protein interactions. *Cell Communication and Signaling*, *18*(1). https://doi.org/10.1186/s12964-020-00554-5

Tomkins, J. E., & Manzoni, C. (2021). Advances in protein-protein interaction network analysis for Parkinson's disease. *Neurobiology of Disease*, *155*, 105395. https://doi.org/10.1016/J.NBD.2021.105395

Trinh, J., Guella, I., James Farrer, M., Affiliations, A., & Author, C. (2014). Disease Penetrance of Late-Onset Parkinsonism: A Meta-analysis. *JAMA Neurology*, *71*(12), 1535–1539. https://doi.org/10.1001/JAMANEUROL.2014.1909

Turski, P., Chaberska, I., Szukało, P., Pyska, P., Milanowski, Ł., Szlufik, S., Figura, M., Hoffman-Zacharska, D., Siuda, J., & Koziorowski, D. (2022). Review of the epidemiology and variability of LRRK2 non-p.Gly2019Ser pathogenic mutations in Parkinson's disease. *Frontiers in Neuroscience*, *16*, 971270. https://doi.org/10.3389/FNINS.2022.971270/BIBTEX

Valente, T. W. (2010). Centrality. *Social Networks and Health*, 81–99. https://doi.org/10.1093/ACPROF:OSO/9780195301014.003.0005

Van Dongen, S. (2008). Graph Clustering Via a Discrete Uncoupling Process. *Https://Doi.Org/10.1137/040608635*, *30*(1), 121–141. https://doi.org/10.1137/040608635

Verma, M., Callio, J., Anthony Otero, P., Sekler, I., Wills, Z. P., & Chu, C. T. (2017). Mitochondrial Calcium Dysregulation Contributes to Dendrite Degeneration Mediated by PD/LBD-Associated LRRK2 Mutants. *Journal of Neuroscience*, *37*(46), 11151–11165. https://doi.org/10.1523/JNEUROSCI.3791-16.2017

Verma, M., Lizama, B. N., & Chu, C. T. (2022). Excitotoxicity, calcium and mitochondria: a triad in synaptic neurodegeneration. *Translational Neurodegeneration 2022 11:1*, *11*(1), 1–14. https://doi.org/10.1186/S40035-021-00278-7

Voit, R., & Grummt, I. (2001). Phosphorylation of UBF at serine 388 is required for interaction with RNA polymerase I and activation of rDNA transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(24), 13631–13636. https://doi.org/10.1073/PNAS.231071698/ASSET/16418696-0C49-4EEF-885C-6A4C007F44BA/ASSETS/GRAPHIC/PQ2310716005.JPEG

von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., & Bork, P. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Research*, *33*(Database issue). https://doi.org/10.1093/NAR/GKI005

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., & Bork, P. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. *Nature 2002 417:6887*, *417*(6887), 399–403. https://doi.org/10.1038/nature750

Von Spiczak, S., Helbig, K. L., Shinde, D. N., Huether, R., Pendziwiat, M., Lourenço, C., Nunes, M. E., Sarco, D. P., Kaplan, R. A., Dlugos, D. J., Kirsch, H., Slavotinek, A., Cilio, M. R., Cervenka, M. C., Cohen, J. S., McClellan, R., Fatemi, A., Yuen, A., Sagawa, Y., … Helbig, I. (2017). DNM1 encephalopathy. *Neurology*, *89*(4), 385–394. https://doi.org/10.1212/WNL.0000000000004152

Wallings, R., Connor-Robson, N., & Wade-Martins, R. (2019). LRRK2 interacts with the vacuolar-type H+-ATPase pump a1 subunit to regulate lysosomal function. *Human Molecular Genetics*, *28*(16), 2696–2710. https://doi.org/10.1093/hmg/ddz088

Wallings, R. L., Herrick, M. K., & Tansey, M. G. (2020). LRRK2 at the Interface Between Peripheral and Central Immune Function in Parkinson's. *Frontiers in Neuroscience*, *14*, 531890. https://doi.org/10.3389/FNINS.2020.00443/BIBTEX

Wallings, R. L., & Tansey, M. G. (2019a). LRRK2 regulation of immune-pathways and inflammatory disease. *Biochemical Society Transactions*, *47*(6), 1581. https://doi.org/10.1042/BST20180463

Wallings, R. L., & Tansey, M. G. (2019b). LRRK2 regulation of immune-pathways and inflammatory disease. *Biochemical Society Transactions*, *47*(6), 1581–1595. https://doi.org/10.1042/BST20180463

Waltl, I., & Kalinke, U. (2022). Beneficial and detrimental functions of microglia during viral encephalitis. *Trends in Neurosciences*, *45*(2), 158–170. https://doi.org/10.1016/J.TINS.2021.11.004

Wang, D., Lou, J., Ouyang, C., Chen, W., Liu, Y., Liu, X., Cao, X., Wang, J., & Lu, L. (2010). Ras-related protein Rab10 facilitates TLR4 signaling by promoting replenishment of TLR4 onto the plasma membrane. *Proceedings of the National Academy of Sciences of the*

United States of America, *107*(31), 13806–13811. https://doi.org/10.1073/PNAS.1009428107/-/DCSUPPLEMENTAL

Wang, J., Zhang, J. R., Zang, Y. F., & Wu, T. (2018). Consistent decreased activity in the putamen in Parkinson's disease: a meta-analysis and an independent validation of resting-state fMRI. *GigaScience*, *7*(6). https://doi.org/10.1093/GIGASCIENCE/GIY071

Wang, L., Xie, C., Greggio, E., Parisiadou, L., Shim, H., Sun, L., Chandran, J., Lin, X., Lai, C., Yang, W. J., Moore, D. J., Dawson, T. M., Dawson, V. L., Chiosis, G., Cookson, M. R., & Cai, H. (2008). The Chaperone Activity of Heat Shock Protein 90 Is Critical for Maintaining the Stability of Leucine-Rich Repeat Kinase 2. *The Journal of Neuroscience*, *28*(13), 3384. https://doi.org/10.1523/JNEUROSCI.0185-08.2008

Wang, R., Wang, C., & Liu, G. (2020). A novel graph clustering method with a greedy heuristic search algorithm for mining protein complexes from dynamic and static PPI networks. *Information Sciences*, *522*, 275–298. https://doi.org/10.1016/J.INS.2020.02.063

Wang, Y., Hicks, S. C., & Hansen, K. D. (2020). Co-expression analysis is biased by a mean-correlation relationship. *BioRxiv*, 2020.02.13.944777. https://doi.org/10.1101/2020.02.13.944777

Weimers, P., Halfvarson, J., Sachs, M. C., Saunders-Pullman, R., Ludvigsson, J. F., Peter, I., Burisch, J., & Olén, O. (2019). Inflammatory Bowel Disease and Parkinson's Disease: A Nationwide Swedish Cohort Study. *Inflammatory Bowel Diseases*, *25*(1), 111–123. https://doi.org/10.1093/IBD/IZY190

Weindel, C. G., Bell, S. L., Huntington, T. E., Vail, K. J., Patrick, K. L., & Watson, R. O. (2019). LRRK2 regulates innate immune responses and neuroinflammation during Mycobacterium tuberculosis infection. *BioRxiv*, 699066. https://doi.org/10.1101/699066

Westerlund, M., Belin, A. C., Anvret, A., Bickford, P., Olson, L., & Galter, D. (2008). Developmental regulation of leucine-rich repeat kinase 1 and 2 expression in the brain and other rodent and human organs: Implications for Parkinson's disease. *Neuroscience*, *152*(2), 429–436. https://doi.org/10.1016/j.neuroscience.2007.10.062

Williams-Gray, C. H., Wijeyekoon, R., Yarnall, A. J., Lawson, R. A., Breen, D. P., Evans, J. R., Cummins, G. A., Duncan, G. W., Khoo, T. K., Burn, D. J., & Barker, R. A. (2016). Serum immune markers and disease progression in an incident Parkinson's disease cohort (ICICLE-PD). *Movement Disorders*, *31*(7), 995–1003. https://doi.org/10.1002/MDS.26563

Wirdefeldt, K., Adami, H.-O., Cole, P., Trichopoulos, D., Mandel, J., Wirdefeldt, K., Adami, H.-O., Trichopoulos, Á. D., Cole, P., & Mandel, J. (2011). Epidemiology and etiology of Parkinson's disease: a review of the evidence. *European Journal of Epidemiology 2011 26:1*, *26*(1), 1–58. https://doi.org/10.1007/S10654-011-9581-6

Wszolek, Z. K., Vieregge, P., Uitti, R. J., Gasser, T., Yasuhara, O., McGeer, P., Berry, K., Calne, D. B., Vingerhoets, F. J. G., Klein, C., & Pfeiffer, R. F. (1997). German-Canadian family (Family A) with Parkinsonism, Amyotrophy, and Dementia-longitudinal observations. *Parkinsonism and Related Disorders*, *3*(3), 125–139. https://doi.org/10.1016/S1353-8020(97)00013-8

Wu, C. X., Liao, J., Park, Y., Reed, X., Engel, V. A., Hoang, N. C., Takagi, Y., Johnson, S. M., Wang, M., Federici, M., Jeremy Nichols, R., Sanishvili, R., Cookson, M. R., & Hoang, Q. Q. (2019). Parkinson's disease-associated mutations in the GTPase domain of LRRK2

impair its nucleotide-dependent conformational dynamics. *Journal of Biological Chemistry*, *294*(15), 5907–5913. https://doi.org/10.1074/jbc.RA119.007631

Wu, J., Yang, S., Wu, H., Huang, Y., & Miao, Y. (2023). Knockdown of LRRK2 inhibits the progression of lung cancer by regulating TLR4/NF-κB pathways and NLRP3 inflammasome. *Journal of Clinical Biochemistry and Nutrition*, *73*(3), 178. https://doi.org/10.3164/JCBN.22-122

Xiao, N., Zhou, A., Kempher, M. L., Zhou, B. Y., Shi, Z. J., Yuan, M., Guo, X., Wu, L., Ning, D., van Nostrand, J., Firestone, M. K., & Zhou, J. (2022). Disentangling direct from indirect relationships in association networks. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(2), e2109995119. https://doi.org/10.1073/PNAS.2109995119/SUPPL_FILE/PNAS.2109995119.SAPP.PDF

Xie, H., Hu, H., Chang, M., Huang, D., Gu, X., Xiong, X., Xiong, R., Hu, L., & Li, G. (2016). Identification of chaperones in a MPP+-induced and ATRA/TPA-differentiated SH-SY5Y cell PD model. *American Journal of Translational Research*, *8*(12), 5659. /pmc/articles/PMC5209517/

Xiong, Y., Coombes, C. E., Kilaru, A., Li, X., Gitler, A. D., Bowers, W. J., Dawson, V. L., Dawson, T. M., & Moore, D. J. (2010). GTPase Activity Plays a Key Role in the Pathobiology of LRRK2. *PLOS Genetics*, *6*(4), e1000902. https://doi.org/10.1371/JOURNAL.PGEN.1000902

Yoon, J. H., Mo, J. S., Kim, M. Y., Ann, E. J., Ahn, J. S., Jo, E. H., Lee, H. J., Lee, Y. C., Seol, W., Yarmoluk, S. M., Gasser, T., Kahle, P. J., Liu, G. H., Belmonte, J. C. I., & Park, H. S. (2017). LRRK2 functions as a scaffolding kinase of ASK1-mediated neuronal cell death. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1864*(12), 2356–2368. https://doi.org/10.1016/J.BBAMCR.2017.09.001

Yun, H. J., Kim, H., Ga, I., Oh, H., Ho, D. H., Kim, J., Seo, H., Son, I., & Seol, W. (2015). An early endosome regulator, Rab5b, is an LRRK2 kinase substrate. *Journal of Biochemistry*, *157*(6), 485–495. https://doi.org/10.1093/JB/MVV005

Zabetian, C. P., Samii, A., Mosley, A. D., Roberts, J. W., Leis, B. C., Yearout, D., Raskind, W. H., & Griffith, A. (2005). A clinic-based study of the LRRK2 gene in Parkinson disease yields new mutations. *Neurology*, *65*(5), 741–744. https://doi.org/10.1212/01.WNL.0000172630.22804.73

Zerial, M., & McBride, H. (2001). Rab proteins as membrane organizers. *Nature Reviews Molecular Cell Biology 2001 2:2*, *2*(2), 107–117. https://doi.org/10.1038/35052055

Zhao, Y., Li, M. C., Konaté, M. M., Chen, L., Das, B., Karlovich, C., Williams, P. M., Evrard, Y. A., Doroshow, J. H., & McShane, L. M. (2021). TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *Journal of Translational Medicine*, *19*(1), 1–15. https://doi.org/10.1186/S12967-021-02936-W/FIGURES/5

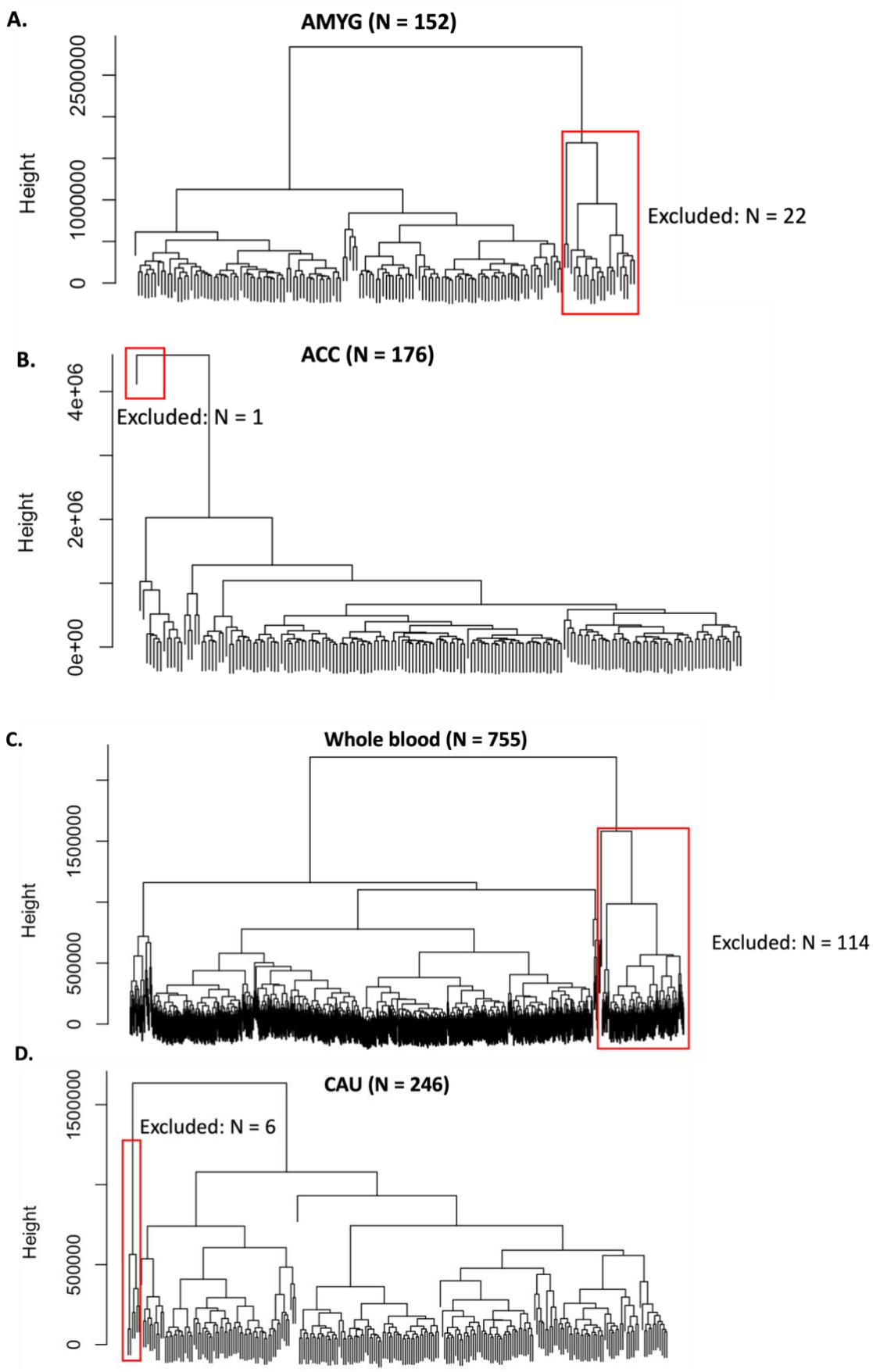
Zhu, B., Park, J.-M., Coffey, S., Hsu, I.-U., Lam, T. T., Gopal, P. P., Ginsberg, S. D., Wang, J., Su, C., Zhao, H., Hafler, D. A., Chandra, S. S., & Zhang, L. (2022). Single-cell transcriptomic and proteomic analysis of Parkinson's disease Brains. *BioRxiv*, 2022.02.14.480397. https://doi.org/10.1101/2022.02.14.480397

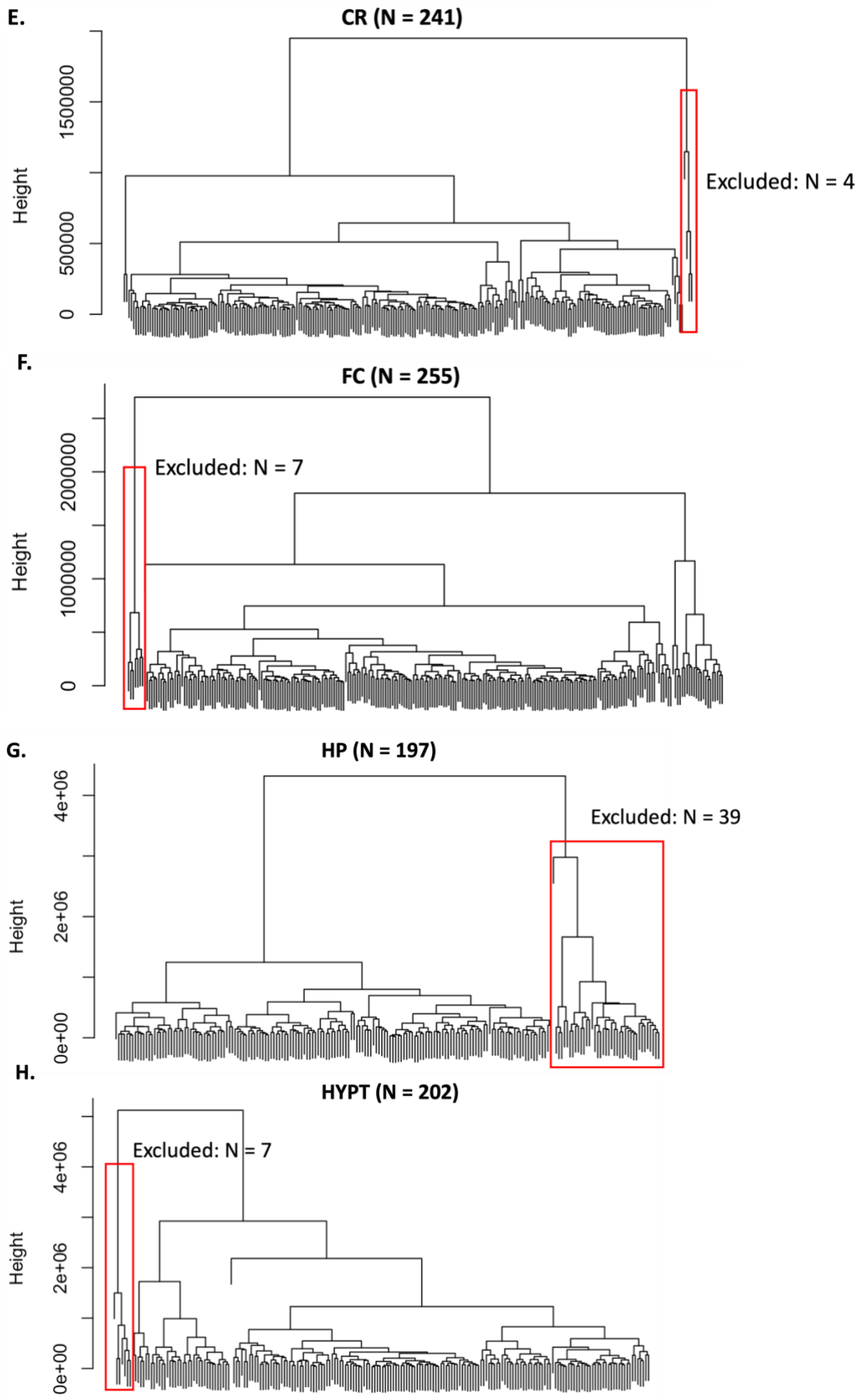
Zhuang, X., Adipietro, K. A., Datta, S., Northup, J. K., & Ray, K. (2010). Rab1 Small GTP-Binding Protein Regulates Cell Surface Trafficking of the Human Calcium-Sensing Receptor. *Endocrinology*, *151*(11), 5114–5123. https://doi.org/10.1210/EN.2010-0422

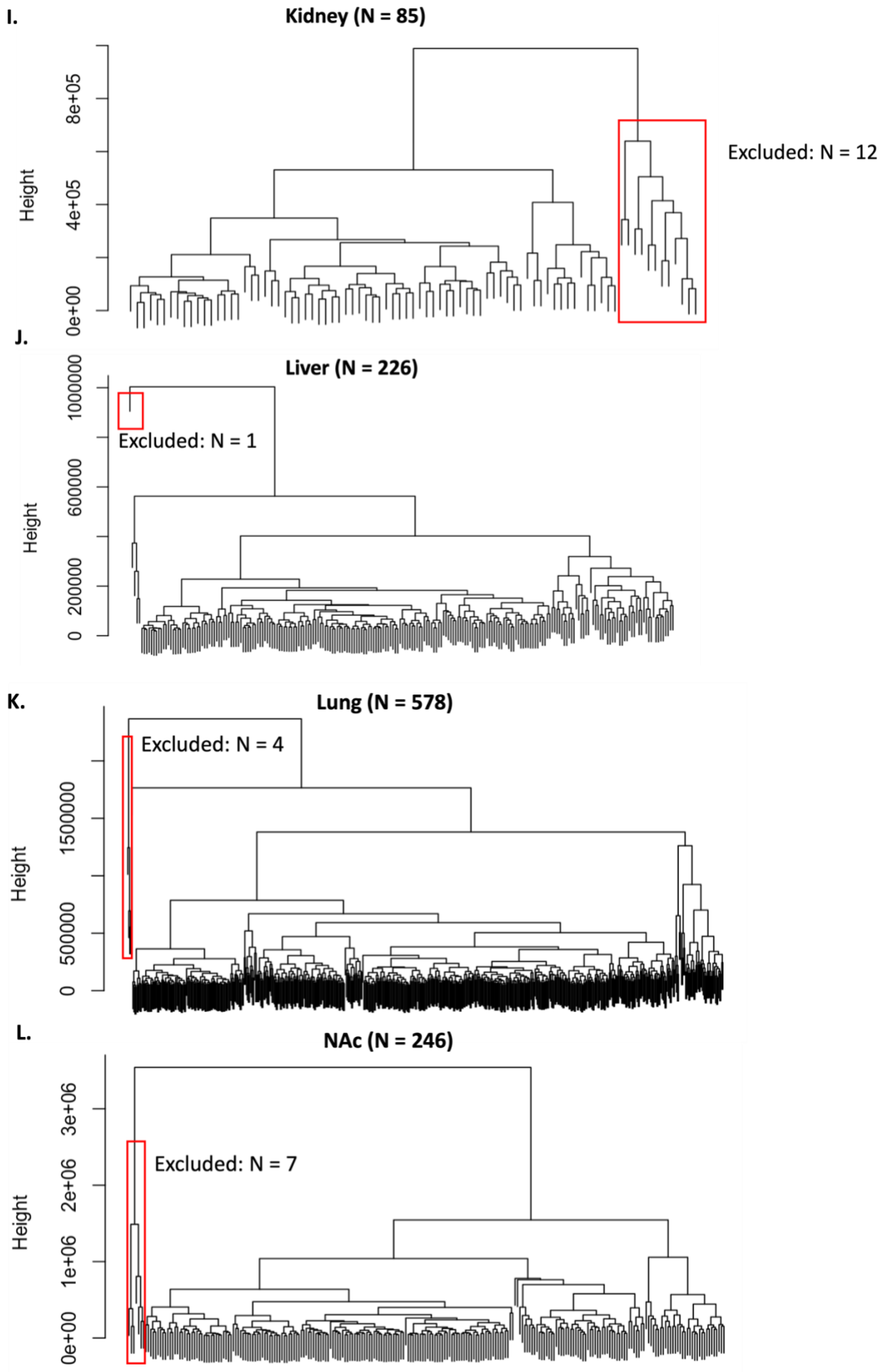
Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R. J., Calne, D. B., Stoessl, A. J., Pfeiffer, R. F., Patenge, N., Carbajal, I. C.,

Vieregge, P., Asmus, F., Müller-Myhsok, B., Dickson, D. W., Meitinger, T., … Gasser, T. (2004). Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*, *44*(4), 601–607. https://doi.org/10.1016/j.neuron.2004.11.005

Appendix A: Supplementary Figures: QC on GTEx samples




E.
CR (N = 241)
Height
0
500000
1500000
Excluded: N = 4
F.
FC (N = 255)
Height
0
1000000
2000000
Excluded: N = 7
G.
HP (N = 197)
Height
0e+00
2e+06
4e+06
Excluded: N = 39
H.
HYPT (N = 202)
Height
0e+00
2e+06
4e+06
Excluded: N = 7


I.
Kidney (N = 85)
Height
0e+00
4e+05
8e+05
Excluded: N = 12
J.
Liver (N = 226)
Height
0
200000
600000
1000000
Excluded: N = 1
K.
Lung (N = 578)
Height
0
500000
1500000
Excluded: N = 4
L.
NAc (N = 246)
Height
0e+00
1e+06
2e+06
3e+06
Excluded: N = 7

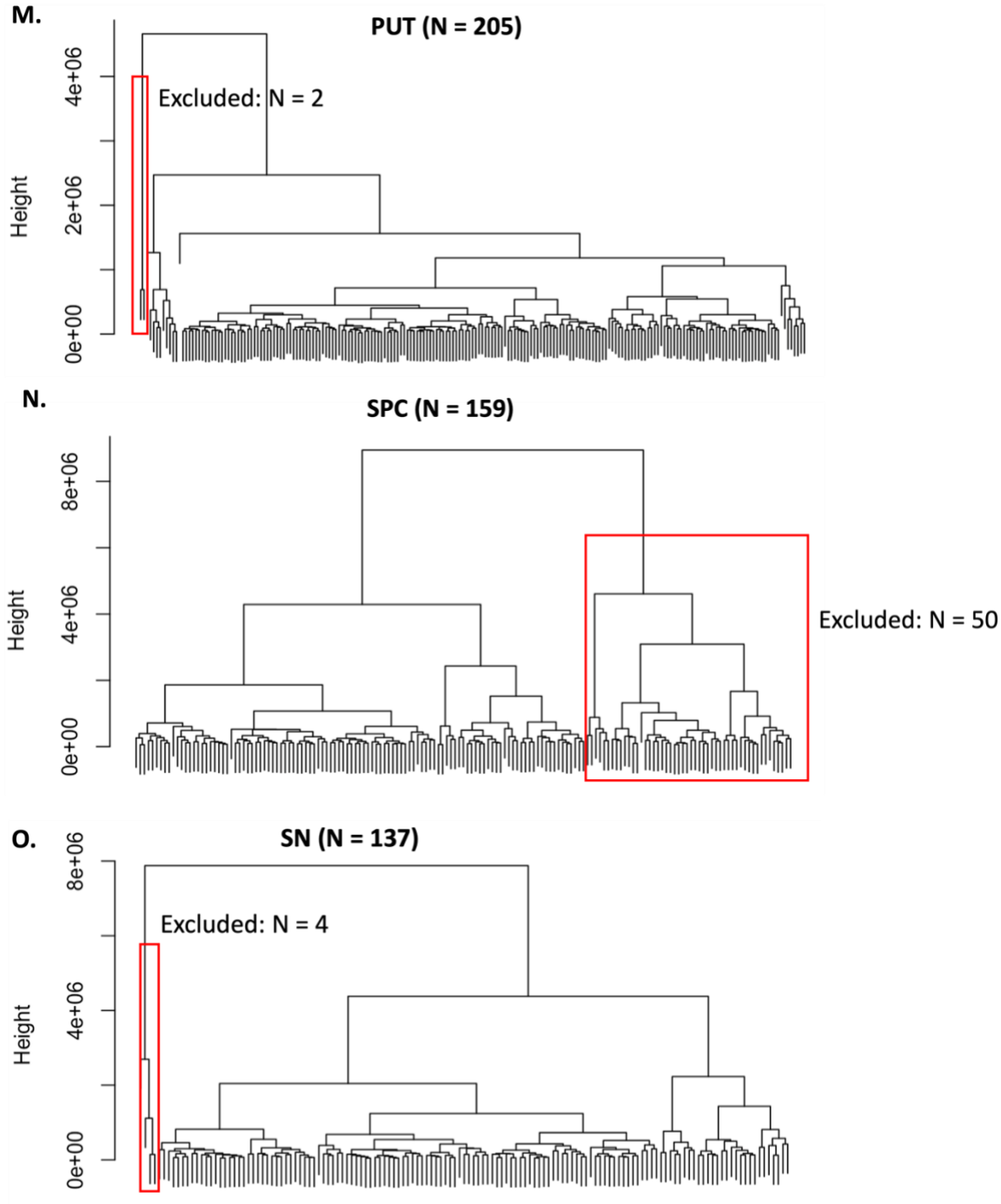


Figure A1. Subject QC on GTEx mRNA dataset

Hierarchical Clustering was performed to identify outlier samples in expression dataset of each of the 15 tissue included in Chapter 2. Outliers are marked in red rectangular and were discarded in the following expression analysis. Abbreviations: AMYG: amygdala; ACC: anterior cingulate cortex; CAU: caudate; CR: cerebellum; FC: frontal cortex; HP: hippocampus; HYPT: hypothalamus; NAc: Nucleus Accumbens; PUT: putamen; SN: substantia nigra; SPC: spinal cord c-1.

Appendix B: GO terms enriched for LRRK2 interactors

Table B1. GO-BP terms enriched for Functional Unit 1 of the $LRRK2_{int}$

Term ID	Term Name	Term Size	Adjusted p-value
GO:0006259	DNA metabolic process	1054	<0.001
GO:0044265	cellular macromolecule catabolic process	1030	<0.001
GO:0009266	response to temperature stimulus	184	<0.001
GO:0071214	cellular response to abiotic stimulus	336	0.011
GO:0009314	response to radiation	449	0.031
GO:0104004	cellular response to environmental stimulus	336	0.011
GO:0030163	protein catabolic process	1031	<0.001
GO:0034599	cellular response to oxidative stress	275	<0.001
GO:1901565	organonitrogen compound catabolic process	1386	<0.001
GO:0006281	DNA repair	591	0.006
GO:0007006	mitochondrial membrane organization	116	<0.001
GO:0000723	telomere maintenance	158	0.011
GO:0061726	mitochondrion disassembly	94	0.009
GO:0097193	intrinsic apoptotic signalling pathway	299	<0.001
GO:0009057	macromolecule catabolic process	1391	<0.001
GO:0071310	cellular response to organic substance	1993	<0.001
GO:0035966	response to topologically incorrect protein	161	<0.001
GO:0034097	response to cytokine	914	<0.001
GO:0010243	response to organonitrogen compound	1034	0.001
GO:0009725	response to hormone	861	0.003
GO:0014070	response to organic cyclic compound	933	0.003
GO:0034614	cellular response to reactive oxygen species	142	0.003
GO:1901653	cellular response to peptide	363	0.010
GO:0071495	cellular response to endogenous stimulus	1445	<0.001
GO:0072331	signal transduction by p53 class mediator	175	<0.001
GO:0009628	response to abiotic stimulus	1141	<0.001
GO:0009719	response to endogenous stimulus	1692	<0.001
GO:0033554	cellular response to stress	1941	<0.001
GO:0034605	cellular response to heat	63	<0.001
GO:0062197	cellular response to chemical stress	332	<0.001
GO:0006974	cellular response to DNA damage stimulus	881	0.001
GO:0034976	response to endoplasmic reticulum stress	259	0.001
GO:0031098	stress-activated protein kinase signalling cascade	246	0.001
GO:0035967	cellular response to topologically incorrect protein	119	0.002
GO:0098780	response to mitochondrial depolarisation	21	0.010

GO:0043632	modification-dependent macromolecule catabolic process	713	0.005
GO:0006914	autophagy	562	<0.001
GO:0016236	macroautophagy	321	<0.001
GO:0061684	chaperone-mediated autophagy	16	<0.001
GO:0061912	selective autophagy	89	<0.001
GO:0000422	autophagy of mitochondrion	94	0.009
GO:0061919	process utilizing autophagic mechanism	562	<0.001
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	455	0.001
GO:0061024	membrane organization	886	<0.001
GO:0009411	response to UV	148	0.001
GO:0034644	cellular response to UV	88	0.036
GO:0070266	necroptotic process	39	0.042
GO:0051603	proteolysis involved in protein catabolic process	799	0.001
GO:0019941	modification-dependent protein catabolic process	701	0.003
GO:0051403	stress-activated MAPK cascade	239	0.004
GO:0000423	mitophagy	37	<0.001
GO:0006511	ubiquitin-dependent protein catabolic process	691	0.002
GO:0098779	positive regulation of mitophagy in response to mitochondrial depolarization	10	0.003
GO:0061734	parkin-mediated stimulation of mitophagy in response to mitochondrial depolarization	6	0.010
GO:0044248	cellular catabolic process	2138	<0.001
GO:0012501	programmed cell death	1948	<0.001
GO:1901575	organic substance catabolic process	2148	<0.001
GO:0007254	JNK cascade	176	0.039
GO:0006915	apoptotic process	1894	<0.001
GO:0097300	programmed necrotic cell death	46	0.001
GO:1901700	response to oxygen-containing compound	1665	<0.001
GO:1901698	response to nitrogen compound	1123	<0.001
GO:0009408	response to heat	110	<0.001
GO:0006979	response to oxidative stress	420	0.003
GO:0071345	cellular response to cytokine stimulus	823	<0.001
GO:1904923	regulation of autophagy of mitochondrion in response to mitochondrial depolarization	15	0.034
GO:0097190	apoptotic signalling pathway	597	<0.001
GO:0000302	response to reactive oxygen species	194	0.005
GO:0008637	apoptotic mitochondrial changes	105	0.001
GO:0032870	cellular response to hormone stimulus	598	0.001

GO:0071407	cellular response to organic cyclic compound	579	0.004
GO:0071417	cellular response to organonitrogen compound	642	0.032
GO:0009416	response to light stimulus	319	0.018
GO:0006986	response to unfolded protein	139	<0.001
GO:1904925	positive regulation of autophagy of mitochondrion in response to mitochondrial depolarization	14	0.023
GO:1901701	cellular response to oxygen-containing compound	1191	<0.001
GO:1901699	cellular response to nitrogen compound	698	<0.001
GO:0010498	proteasomal protein catabolic process	526	0.001
GO:0008219	cell death	2113	<0.001
GO:0070997	neuron death	367	<0.001
GO:0036473	cell death in response to oxidative stress	91	0.001
GO:0070265	necrotic cell death	62	0.013
GO:0008631	intrinsic apoptotic signalling pathway in response to oxidative stress	45	0.010
GO:0070059	intrinsic apoptotic signalling pathway in response to endoplasmic reticulum stress	64	0.017
GO:0072332	intrinsic apoptotic signalling pathway by p53 class mediator	82	0.019
GO:0034620	cellular response to unfolded protein	99	0.015
GO:0007005	mitochondrion organization	539	<0.001

Table B2. GO-CC terms enriched for Functional Unit 1 of the $LRRK2_{int}$

Term ID	Term Name	Term size	Adjusted p-value
GO:0070062	extracellular exosome	2109	<0.001
GO:1903561	extracellular vesicle	2133	<0.001
GO:0043230	extracellular organelle	2134	<0.001
GO:0065010	extracellular membrane-bounded organelle	2134	<0.001
GO:0005856	cytoskeleton	2407	<0.001
GO:0005925	focal adhesion	421	<0.001
GO:0030055	cell-substrate junction	431	<0.001
GO:0030054	cell junction	2214	<0.001
GO:0015630	microtubule cytoskeleton	1388	<0.001
GO:0070161	anchoring junction	903	<0.001
GO:0045202	synapse	1451	<0.001
GO:1902494	catalytic complex	1781	<0.001
GO:0005739	mitochondrion	1672	<0.001
GO:0048471	perinuclear region of cytoplasm	733	<0.001
GO:0042995	cell projection	2329	<0.001
GO:0031410	cytoplasmic vesicle	2476	<0.001

GO:0097708	intracellular vesicle	2481	<0.001
GO:0099080	supramolecular complex	1404	<0.001
GO:0044297	cell body	567	<0.001
GO:0099512	supramolecular fibre	1034	<0.001
GO:0016234	inclusion body	74	<0.001
GO:0099081	supramolecular polymer	1042	<0.001
GO:0120025	plasma membrane bounded cell projection	2217	<0.001
GO:0005819	spindle	431	<0.001
GO:0101031	protein folding chaperone complex	41	<0.001
GO:0048770	pigment granule	111	<0.001
GO:0042470	melanosome	111	<0.001
GO:0099513	polymeric cytoskeletal fibre	793	<0.001
GO:0043005	neuron projection	1299	<0.001
GO:0036477	somatodendritic compartment	849	<0.001
GO:0140535	intracellular protein-containing complex	951	<0.001
GO:0030424	axon	641	<0.001
GO:0098687	chromosomal region	399	<0.001
GO:0031967	organelle envelope	1293	<0.001
GO:0031975	envelope	1293	<0.001
GO:0030425	dendrite	621	<0.001
GO:0097447	dendritic tree	623	<0.001
GO:0043025	neuronal cell body	500	<0.001
GO:0098978	glutamatergic synapse	426	<0.001
GO:0098562	cytoplasmic side of membrane	213	<0.001
GO:0005874	microtubule	474	<0.001
GO:0005815	microtubule organizing center	855	<0.001
GO:0000781	chromosome, telomeric region	170	<0.001
GO:0098794	postsynapse	644	<0.001
GO:0098793	presynapse	561	<0.001
GO:0005938	cell cortex	314	<0.001
GO:0031968	organelle outer membrane	249	<0.001
GO:0101002	ficolin-1-rich granule	184	<0.001
GO:0019867	outer membrane	251	<0.001
GO:0031966	mitochondrial membrane	771	<0.001
GO:0005741	mitochondrial outer membrane	220	<0.001
GO:0005768	endosome	1047	<0.001
GO:0022626	cytosolic ribosome	110	<0.001
GO:0035861	site of double-strand break	86	<0.001
GO:0098588	bounding membrane of organelle	2160	<0.001
GO:0031252	cell leading edge	424	<0.001
GO:0045335	phagocytic vesicle	141	<0.001
GO:0150034	distal axon	277	<0.001
GO:0005740	mitochondrial envelope	820	<0.001
GO:0090734	site of DNA damage	117	<0.001

GO:0099503	secretory vesicle	1039	<0.001
GO:0098857	membrane microdomain	286	<0.001
GO:0005813	centrosome	647	<0.001
GO:0030139	endocytic vesicle	342	<0.001
GO:0022627	cytosolic small ribosomal subunit	44	<0.001
GO:0045121	membrane raft	285	<0.001
GO:0005757	mitochondrial permeability transition pore complex	7	<0.001
GO:0120111	neuron projection cytoplasm	90	<0.001
GO:0030141	secretory granule	871	<0.001
GO:0034399	nuclear periphery	151	<0.001
GO:1904813	ficolin-1-rich granule lumen	124	<0.001
GO:0031264	death-inducing signalling complex	9	0.001
GO:0030670	phagocytic vesicle membrane	78	0.001
GO:0015935	small ribosomal subunit	79	0.001
GO:0016235	aggresome	36	0.001
GO:0030426	growth cone	170	0.001
GO:0043197	dendritic spine	171	0.001
GO:0098796	membrane protein complex	1357	0.001
GO:0044309	neuron spine	175	0.001
GO:0030427	site of polarized growth	175	0.001
GO:1990234	transferase complex	892	0.001
GO:0005838	proteasome regulatory particle	22	0.001
GO:0001726	ruffle	181	0.001
GO:1904115	axon cytoplasm	63	0.002
GO:0005853	eukaryotic translation elongation factor 1 complex	4	0.002
GO:0005694	chromosome	1942	0.002
GO:0044391	ribosomal subunit	189	0.002
GO:0022624	proteasome accessory complex	25	0.002
GO:0012506	vesicle membrane	1230	0.002
GO:0015629	actin cytoskeleton	501	0.002
GO:0030666	endocytic vesicle membrane	194	0.003
GO:0043209	myelin sheath	45	0.003
GO:1990909	Wnt signalosome	13	0.003
GO:0019866	organelle inner membrane	566	0.004
GO:1902554	serine/threonine protein kinase complex	136	0.005
GO:0009898	cytoplasmic side of plasma membrane	175	0.007
GO:0097342	ripoptosome	6	0.009
GO:0005903	brush border	111	0.009
GO:0043679	axon terminus	112	0.009
GO:0030659	cytoplasmic vesicle membrane	1213	0.012
GO:0043204	perikaryon	155	0.015
GO:0051233	spindle midzone	37	0.016
GO:0000793	condensed chromosome	278	0.016

GO:1902911	protein kinase complex	156	0.016
GO:0061695	transferase complex, transferring phosphorus-containing groups	325	0.017
GO:0005769	early endosome	428	0.020
GO:1905368	peptidase complex	125	0.021
GO:0044306	neuron projection terminus	127	0.023
GO:0030496	midbody	206	0.026
GO:0000502	proteasome complex	67	0.028
GO:0000922	spindle pole	172	0.034
GO:0032587	ruffle membrane	100	0.035
GO:0044294	dendritic growth cone	9	0.035
GO:0000775	chromosome, centromeric region	258	0.037
GO:0043195	terminal bouton	45	0.041
GO:1990565	HSP90-CDC37 chaperone complex	2	0.042
GO:0030135	coated vesicle	310	0.044
GO:0031256	leading edge membrane	179	0.046

Table B3. GO-BP enriched for Functional Unit 2 of the $LRRK2_{int}$

Term ID	**Term name**	**Term size**	**Adjusted p-value**
GO:0045184	establishment of protein localization	1898	<0.001
GO:0030029	actin filament-based process	813	<0.001
GO:0015031	protein transport	1788	<0.001
GO:0072594	establishment of protein localization to organelle	664	<0.001
GO:0140694	non-membrane-bounded organelle assembly	386	<0.001
GO:0097479	synaptic vesicle localization	53	0.036
GO:0051649	establishment of localization in cell	2340	<0.001
GO:0046907	intracellular transport	1806	<0.001
GO:0036465	synaptic vesicle recycling	80	<0.001
GO:0051650	establishment of vesicle localization	194	<0.001
GO:0099504	synaptic vesicle cycle	196	<0.001
GO:0140238	presynaptic endocytosis	67	<0.001
GO:0051261	protein depolymerization	123	<0.001
GO:0098813	nuclear chromosome segregation	340	0.003
GO:0006886	intracellular protein transport	1095	<0.001
GO:0051169	nuclear transport	330	<0.001
GO:0030705	cytoskeleton-dependent intracellular transport	205	<0.001
GO:0072384	organelle transport along microtubule	88	0.001
GO:0000226	microtubule cytoskeleton organization	652	<0.001
GO:0030036	actin cytoskeleton organization	727	<0.001
GO:0007059	chromosome segregation	434	<0.001
GO:0033365	protein localization to organelle	1144	<0.001
GO:0034504	protein localization to nucleus	309	<0.001

GO:0070585	protein localization to mitochondrion	128	<0.001
GO:0007052	mitotic spindle organization	135	0.010
GO:0051640	organelle localization	577	<0.001
GO:0048488	synaptic vesicle endocytosis	67	<0.001
GO:0006913	nucleocytoplasmic transport	330	<0.001
GO:0051168	nuclear export	165	<0.001
GO:0051170	import into nucleus	170	<0.001
GO:1903047	mitotic cell cycle process	773	<0.001
GO:0097435	supramolecular fibre organization	814	<0.001
GO:0008088	axo-dendritic transport	76	0.009
GO:0042254	ribosome biogenesis	311	0.001
GO:0042274	ribosomal small subunit biogenesis	77	0.001
GO:0000819	sister chromatid segregation	252	0.038
GO:0007017	microtubule-based process	939	<0.001
GO:0007018	microtubule-based movement	419	0.001
GO:0042255	ribosome assembly	59	<0.001
GO:0051225	spindle assembly	129	0.006
GO:0010970	transport along microtubule	167	0.001
GO:0099518	vesicle cytoskeletal trafficking	70	0.037
GO:0051656	establishment of organelle localization	443	<0.001
GO:0007051	spindle organization	199	<0.001
GO:0031109	microtubule polymerization or depolymerization	126	<0.001
GO:0051648	vesicle localization	211	<0.001
GO:0071705	nitrogen compound transport	2252	<0.001
GO:0051646	mitochondrion localization	50	<0.001
GO:0015931	nucleobase-containing compound transport	227	0.009
GO:0072319	vesicle uncoating	9	<0.001
GO:0072318	clathrin coat disassembly	8	<0.001
GO:0008154	actin polymerization or depolymerization	202	0.002
GO:0099111	microtubule-based transport	208	0.001
GO:0000028	ribosomal small subunit assembly	17	<0.001
GO:0048489	synaptic vesicle transport	40	0.004
GO:0016192	vesicle-mediated transport	1638	0.015
GO:0007015	actin filament organization	452	<0.001
GO:0072655	establishment of protein localization to mitochondrion	122	<0.001
GO:0046785	microtubule polymerization	85	<0.001
GO:0051258	protein polymerization	284	<0.001
GO:0030041	actin filament polymerization	172	0.001
GO:0006611	protein export from nucleus	59	<0.001
GO:0006606	protein import into nucleus	165	<0.001
GO:0007049	cell cycle	1818	<0.001
GO:0099003	vesicle-mediated transport in synapse	217	<0.001

GO:0007010	cytoskeleton organization	1507	<0.001
GO:0070925	organelle assembly	944	<0.001
GO:0000278	mitotic cell cycle	929	<0.001
GO:0022402	cell cycle process	1270	<0.001
GO:0051276	chromosome organization	635	<0.001
GO:0016191	synaptic vesicle uncoating	7	<0.001

Table B4. GO-CC enriched for Functional Unit 2 of the $LRRK2_{int}$

Term ID	Term Name	Term Size	Adjusted p-value
GO:0005856	cytoskeleton	2407	<0.001
GO:0070062	extracellular exosome	2109	<0.001
GO:1903561	extracellular vesicle	2133	<0.001
GO:0043230	extracellular organelle	2134	<0.001
GO:0065010	extracellular membrane-bounded organelle	2134	<0.001
GO:0015630	microtubule cytoskeleton	1388	<0.001
GO:0099080	supramolecular complex	1404	<0.001
GO:0005925	focal adhesion	421	<0.001
GO:0030054	cell junction	2214	<0.001
GO:0030055	cell-substrate junction	431	<0.001
GO:0099512	supramolecular fibre	1034	<0.001
GO:0099081	supramolecular polymer	1042	<0.001
GO:0099513	polymeric cytoskeletal fibre	793	<0.001
GO:0070161	anchoring junction	903	<0.001
GO:0045202	synapse	1451	<0.001
GO:0015629	actin cytoskeleton	501	<0.001
GO:0005874	microtubule	474	<0.001
GO:0022626	cytosolic ribosome	110	<0.001
GO:0042995	cell projection	2329	<0.001
GO:0005819	spindle	431	<0.001
GO:0120025	plasma membrane bounded cell projection	2217	<0.001
GO:0005815	microtubule organizing centre	855	<0.001
GO:0022627	cytosolic small ribosomal subunit	44	<0.001
GO:0044391	ribosomal subunit	189	<0.001
GO:0031410	cytoplasmic vesicle	2476	<0.001
GO:0097708	intracellular vesicle	2481	<0.001
GO:0015935	small ribosomal subunit	79	<0.001
GO:0031252	cell leading edge	424	<0.001
GO:0098794	postsynapse	644	<0.001
GO:0005813	centrosome	647	<0.001
GO:0048471	perinuclear region of cytoplasm	733	<0.001
GO:0048770	pigment granule	111	<0.001
GO:0042470	melanosome	111	<0.001
GO:0043005	neuron projection	1299	<0.001
GO:0098687	chromosomal region	399	<0.001

GO:0030424	axon	641	<0.001
GO:0044297	cell body	567	<0.001
GO:0098978	glutamatergic synapse	426	<0.001
GO:0072686	mitotic spindle	184	<0.001
GO:0005768	endosome	1047	<0.001
GO:0000793	condensed chromosome	278	<0.001
GO:0031975	envelope	1293	<0.001
GO:0031967	organelle envelope	1293	<0.001
GO:0036477	somatodendritic compartment	849	<0.001
GO:0098793	presynapse	561	<0.001
GO:0150034	distal axon	277	<0.001
GO:0030425	dendrite	621	<0.001
GO:0097447	dendritic tree	623	<0.001
GO:0005885	Arp2/3 protein complex	11	<0.001
GO:0043025	neuronal cell body	500	<0.001
GO:0005938	cell cortex	314	<0.001
GO:0030684	preribosome	108	<0.001
GO:0005903	brush border	111	<0.001
GO:0045335	phagocytic vesicle	141	<0.001
GO:0030426	growth cone	170	<0.001
GO:0032040	small-subunit processome	73	<0.001
GO:0005739	mitochondrion	1672	<0.001
GO:0016234	inclusion body	74	<0.001
GO:0030427	site of polarized growth	175	<0.001
GO:0000779	condensed chromosome, centromeric region	178	<0.001
GO:0042641	actomyosin	77	<0.001
GO:0034399	nuclear periphery	151	<0.001
GO:0043292	contractile fiber	245	<0.001
GO:0099503	secretory vesicle	1039	<0.001
GO:0005840	ribosome	524	<0.001
GO:0000775	chromosome, centromeric region	258	<0.001
GO:0035861	site of double-strand break	86	<0.001
GO:0000776	kinetochore	167	<0.001
GO:0030139	endocytic vesicle	342	<0.001
GO:0030027	lamellipodium	201	<0.001
GO:0030016	myofibril	236	<0.001
GO:0098588	bounding membrane of organelle	2160	<0.001
GO:0000922	spindle pole	172	<0.001
GO:0001725	stress fibre	70	<0.001
GO:0097517	contractile actin filament bundle	70	<0.001
GO:0030017	sarcomere	215	<0.001
GO:0030141	secretory granule	871	<0.001
GO:0098984	neuron to neuron synapse	371	<0.001
GO:0032432	actin filament bundle	78	<0.001
GO:0005635	nuclear envelope	502	<0.001

GO:0032279	asymmetric synapse	339	<0.001
GO:0101031	protein folding chaperone complex	41	<0.001
GO:0012506	vesicle membrane	1230	<0.001
GO:0005876	spindle microtubule	83	<0.001
GO:0098862	cluster of actin-based cell projections	167	<0.001
GO:1904115	axon cytoplasm	63	<0.001
GO:0099572	postsynaptic specialization	352	<0.001
GO:0005884	actin filament	113	<0.001
GO:0043209	myelin sheath	45	<0.001
GO:0030659	cytoplasmic vesicle membrane	1213	<0.001
GO:0120111	neuron projection cytoplasm	90	<0.001
GO:0014069	postsynaptic density	323	<0.001
GO:0090734	site of DNA damage	117	<0.001
GO:0005929	cilium	747	<0.001
GO:0005844	polysome	73	0.001
GO:0005643	nuclear pore	99	0.001
GO:0005694	chromosome	1942	0.001
GO:0022625	cytosolic large ribosomal subunit	54	0.001
GO:0030670	phagocytic vesicle membrane	78	0.001
GO:0000781	chromosome, telomeric region	170	0.001
GO:0043197	dendritic spine	171	0.001
GO:0098857	membrane microdomain	286	0.002
GO:0044309	neuron spine	175	0.002
GO:0031968	organelle outer membrane	249	0.002
GO:0099568	cytoplasmic region	290	0.002
GO:0019867	outer membrane	251	0.002
GO:0005862	muscle thin filament tropomyosin	4	0.002
GO:0032838	plasma membrane bounded cell projection cytoplasm	252	0.002
GO:0036464	cytoplasmic ribonucleoprotein granule	255	0.002
GO:0001726	ruffle	181	0.003
GO:0030133	transport vesicle	427	0.003
GO:0005769	early endosome	428	0.003
GO:0005741	mitochondrial outer membrane	220	0.003
GO:0101002	ficolin-1-rich granule	184	0.003
GO:1990909	Wnt signalosome	13	0.004
GO:0030666	endocytic vesicle membrane	194	0.005
GO:0035770	ribonucleoprotein granule	272	0.005
GO:0055037	recycling endosome	201	0.007
GO:0031966	mitochondrial membrane	771	0.008
GO:0045121	membrane raft	285	0.008
GO:0030175	filopodium	107	0.011
GO:0005802	trans-Golgi network	257	0.014
GO:0098858	actin-based cell projection	219	0.015
GO:0043679	axon terminus	112	0.015

GO:0016459	myosin complex	57	0.016
GO:0015934	large ribosomal subunit	114	0.017
GO:0005757	mitochondrial permeability transition pore complex	7	0.019
GO:0016235	aggresome	36	0.019
GO:0030135	coated vesicle	310	0.019
GO:0005740	mitochondrial envelope	820	0.020
GO:0010008	endosome membrane	551	0.020
GO:0000940	outer kinetochore	19	0.021
GO:0070382	exocytic vesicle	228	0.021
GO:0051233	spindle midzone	37	0.022
GO:0098590	plasma membrane region	1304	0.023
GO:0043204	perikaryon	155	0.025
GO:0005865	striated muscle thin filament	20	0.025
GO:0140535	intracellular protein-containing complex	951	0.026
GO:0005814	centriole	156	0.026
GO:1902911	protein kinase complex	156	0.026
GO:0008290	F-actin capping protein complex	8	0.030
GO:0044306	neuron projection terminus	127	0.037
GO:0098791	Golgi apparatus subcompartment	377	0.037
GO:0016363	nuclear matrix	128	0.039
GO:0044294	dendritic growth cone	9	0.044
GO:0016461	unconventional myosin complex	9	0.044
GO:0030496	midbody	206	0.045
GO:1990565	HSP90-CDC37 chaperone complex	2	0.050

Table B5. GO-BP enriched for Functional Unit 3 of the LRRK2$_{int}$

Term ID	Term name	Term size	Adjusted p-value
GO:1903321	negative regulation of protein modification by small protein conjugation or removal	96	<0.001
GO:0043085	positive regulation of catalytic activity	1181	<0.001
GO:0043086	negative regulation of catalytic activity	768	<0.001
GO:0051338	regulation of transferase activity	908	<0.001
GO:0051336	regulation of hydrolase activity	1007	0.003
GO:0051348	negative regulation of transferase activity	271	<0.001
GO:0043549	regulation of kinase activity	775	<0.001
GO:0051438	regulation of ubiquitin-protein transferase activity	59	<0.001
GO:1903052	positive regulation of proteolysis involved in protein catabolic process	138	0.013
GO:0051347	positive regulation of transferase activity	584	<0.001
GO:0010952	positive regulation of peptidase activity	189	0.019
GO:2000058	regulation of ubiquitin-dependent protein catabolic process	173	<0.001

GO:0061136	regulation of proteasomal protein catabolic process	196	<0.001
GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	142	<0.001
GO:0042326	negative regulation of phosphorylation	363	<0.001
GO:0051054	positive regulation of DNA metabolic process	299	<0.001
GO:0042327	positive regulation of phosphorylation	822	<0.001
GO:0031396	regulation of protein ubiquitination	210	<0.001
GO:0031397	negative regulation of protein ubiquitination	83	<0.001
GO:0030162	regulation of proteolysis	737	<0.001
GO:0071900	regulation of protein serine/threonine kinase activity	386	0.002
GO:1901800	positive regulation of proteasomal protein catabolic process	119	0.002
GO:0051052	regulation of DNA metabolic process	535	<0.001
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	2051	<0.001
GO:1904666	regulation of ubiquitin protein ligase activity	27	<0.001
GO:0032069	regulation of nuclease activity	22	0.001
GO:0032436	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	95	0.010
GO:0032075	positive regulation of nuclease activity	7	0.023
GO:1904667	negative regulation of ubiquitin protein ligase activity	13	<0.001
GO:0051444	negative regulation of ubiquitin-protein transferase activity	23	0.001
GO:0033673	negative regulation of kinase activity	228	0.002
GO:0043393	regulation of protein binding	202	<0.001
GO:0032091	negative regulation of protein binding	95	<0.001
GO:0032092	positive regulation of protein binding	89	0.039
GO:0031324	negative regulation of cellular metabolic process	2244	<0.001
GO:0019220	regulation of phosphate metabolic process	1400	<0.001
GO:0042325	regulation of phosphorylation	1244	<0.001
GO:0045936	negative regulation of phosphate metabolic process	419	<0.001
GO:0045937	positive regulation of phosphate metabolic process	907	<0.001
GO:0051174	regulation of phosphorus metabolic process	1401	<0.001

GO:0010563	negative regulation of phosphorus metabolic process	420	<0.001
GO:0010562	positive regulation of phosphorus metabolic process	907	<0.001
GO:0031399	regulation of protein modification process	1540	<0.001
GO:0031400	negative regulation of protein modification process	493	<0.001
GO:1903320	regulation of protein modification by small protein conjugation or removal	252	<0.001
GO:0001932	regulation of protein phosphorylation	1101	<0.001
GO:0031401	positive regulation of protein modification process	1005	<0.001
GO:0071902	positive regulation of protein serine/threonine kinase activity	227	0.041
GO:0045859	regulation of protein kinase activity	661	<0.001
GO:0001934	positive regulation of protein phosphorylation	745	<0.001
GO:0045862	positive regulation of proteolysis	369	<0.001
GO:1903050	regulation of proteolysis involved in protein catabolic process	231	<0.001
GO:0033674	positive regulation of kinase activity	494	0.017
GO:0051100	negative regulation of binding	162	0.015
GO:0010950	positive regulation of endopeptidase activity	169	0.024
GO:0050790	regulation of catalytic activity	2356	<0.001
GO:0051098	regulation of binding	372	<0.001
GO:0044093	positive regulation of molecular function	1580	<0.001
GO:0032070	regulation of deoxyribonuclease activity	10	0.003
GO:0044092	negative regulation of molecular function	1115	<0.001
GO:0051345	positive regulation of hydrolase activity	582	<0.001
GO:0001933	negative regulation of protein phosphorylation	320	<0.001

Table B6. GO-CC enriched for Functional Unit 3 of the LRRK2int

Term ID	Term Name	Term size	Adjusted p-value
GO:0005925	focal adhesion	421	<0.001
GO:0005856	cytoskeleton	2407	<0.001
GO:0030055	cell-substrate junction	431	<0.001
GO:0016234	inclusion body	74	<0.001
GO:0030054	cell junction	2214	<0.001
GO:0015630	microtubule cytoskeleton	1388	<0.001
GO:1902494	catalytic complex	1781	<0.001
GO:0070062	extracellular exosome	2109	<0.001
GO:1903561	extracellular vesicle	2133	<0.001

GO:0065010	extracellular membrane-bounded organelle	2134	<0.001
GO:0043230	extracellular organelle	2134	<0.001
GO:0070161	anchoring junction	903	<0.001
GO:0045202	synapse	1451	<0.001
GO:0044297	cell body	567	<0.001
GO:0101031	protein folding chaperone complex	41	<0.001
GO:0098687	chromosomal region	399	<0.001
GO:0022626	cytosolic ribosome	110	<0.001
GO:0098562	cytoplasmic side of membrane	213	<0.001
GO:0140535	intracellular protein-containing complex	951	<0.001
GO:0042995	cell projection	2329	<0.001
GO:0022627	cytosolic small ribosomal subunit	44	<0.001
GO:0043005	neuron projection	1299	<0.001
GO:0036477	somatodendritic compartment	849	<0.001
GO:0005815	microtubule organizing centre	855	<0.001
GO:0048471	perinuclear region of cytoplasm	733	<0.001
GO:0005694	chromosome	1942	<0.001
GO:0120025	plasma membrane bounded cell projection	2217	<0.001
GO:0030424	axon	641	<0.001
GO:0005739	mitochondrion	1672	<0.001
GO:0000781	chromosome, telomeric region	170	<0.001
GO:0015935	small ribosomal subunit	79	<0.001
GO:0005813	centrosome	647	<0.001
GO:0098978	glutamatergic synapse	426	<0.001
GO:0043025	neuronal cell body	500	<0.001
GO:0030425	dendrite	621	<0.001
GO:0097447	dendritic tree	623	<0.001
GO:0099080	supramolecular complex	1404	<0.001
GO:0098794	postsynapse	644	<0.001
GO:0044391	ribosomal subunit	189	<0.001
GO:0005819	spindle	431	<0.001
GO:0016235	aggresome	36	<0.001
GO:0031410	cytoplasmic vesicle	2476	<0.001
GO:0097708	intracellular vesicle	2481	<0.001
GO:0031264	death-inducing signalling complex	9	<0.001
GO:0120111	neuron projection cytoplasm	90	<0.001
GO:0099512	supramolecular fibre	1034	<0.001
GO:0099081	supramolecular polymer	1042	<0.001
GO:0009898	cytoplasmic side of plasma membrane	175	<0.001
GO:1902554	serine/threonine protein kinase complex	136	<0.001
GO:0005853	eukaryotic translation elongation factor 1 complex	4	0.001
GO:0000793	condensed chromosome	278	0.001
GO:1990909	Wnt signalosome	13	0.001

GO:0005903	brush border	111	0.001
GO:0048770	pigment granule	111	0.001
GO:0042470	melanosome	111	0.001
GO:0090734	site of DNA damage	117	0.001
GO:1902911	protein kinase complex	156	0.001
GO:0035861	site of double-strand break	86	0.002
GO:1904813	ficolin-1-rich granule lumen	124	0.002
GO:0031252	cell leading edge	424	0.002
GO:0005938	cell cortex	314	0.002
GO:0030426	growth cone	170	0.002
GO:0097342	ripoptosome	6	0.003
GO:1904115	axon cytoplasm	63	0.003
GO:0030427	site of polarized growth	175	0.003
GO:0150034	distal axon	277	0.004
GO:0101002	ficolin-1-rich granule	184	0.005
GO:0045121	membrane raft	285	0.005
GO:0098857	membrane microdomain	286	0.005
GO:0032993	protein-DNA complex	1435	0.006
GO:0016328	lateral plasma membrane	72	0.006
GO:0005838	proteasome regulatory particle	22	0.007
GO:0005874	microtubule	474	0.007
GO:1990234	transferase complex	892	0.007
GO:0034399	nuclear periphery	151	0.008
GO:0031975	envelope	1293	0.008
GO:0031967	organelle envelope	1293	0.008
GO:0000775	chromosome, centromeric region	258	0.011
GO:0022624	proteasome accessory complex	25	0.012
GO:0098862	cluster of actin-based cell projections	167	0.016
GO:0099513	polymeric cytoskeletal fibre	793	0.016
GO:0030141	secretory granule	871	0.017
GO:1990565	HSP90-CDC37 chaperone complex	2	0.018
GO:0098552	side of membrane	656	0.019
GO:0000785	chromatin	1366	0.020
GO:0099503	secretory vesicle	1039	0.021
GO:0016604	nuclear body	888	0.022
GO:0000779	condensed chromosome, centromeric region	178	0.025
GO:0042405	nuclear inclusion body	12	0.028
GO:0099568	cytoplasmic region	290	0.030
GO:0070603	SWI/SNF superfamily-type complex	96	0.032
GO:0016607	nuclear speck	418	0.035
GO:0098793	presynapse	561	0.043
GO:0098574	cytoplasmic side of lysosomal membrane	14	0.047

Table B7. GO-BP terms enriched for Functional Unit 4 of the LRRK2int

Term ID	Term name	Term size	Adjusted p-value
GO:0070507	regulation of microtubule cytoskeleton organization	158	<0.001
GO:0032956	regulation of actin cytoskeleton organization	364	0.003
GO:0051495	positive regulation of cytoskeleton organization	194	0.026
GO:0051493	regulation of cytoskeleton organization	541	<0.001
GO:0010638	positive regulation of organelle organization	507	<0.001
GO:0010639	negative regulation of organelle organization	367	<0.001
GO:0010821	regulation of mitochondrion organization	151	<0.001
GO:0051247	positive regulation of protein metabolic process	1499	<0.001
GO:0042176	regulation of protein catabolic process	366	<0.001
GO:0030833	regulation of actin filament polymerization	142	0.003
GO:0031113	regulation of microtubule polymerization	54	0.004
GO:1903146	regulation of autophagy of mitochondrion	37	0.002
GO:0032273	positive regulation of protein polymerization	92	0.007
GO:0044087	regulation of cellular component biogenesis	955	<0.001
GO:0016241	regulation of macroautophagy	160	0.002
GO:0051128	regulation of cellular component organization	2365	<0.001
GO:0033043	regulation of organelle organization	1189	<0.001
GO:0043254	regulation of protein-containing complex assembly	409	<0.001
GO:0051130	positive regulation of cellular component organization	1056	<0.001
GO:0051129	negative regulation of cellular component organization	702	<0.001
GO:1902903	regulation of supramolecular fibre organization	381	<0.001
GO:0031110	regulation of microtubule polymerization or depolymerization	89	0.005
GO:0031344	regulation of cell projection organization	656	<0.001
GO:0001558	regulation of cell growth	429	0.047
GO:0120035	regulation of plasma membrane bounded cell projection organization	640	<0.001
GO:0110053	regulation of actin filament organization	276	<0.001
GO:0010506	regulation of autophagy	348	<0.001

GO:0031331	positive regulation of cellular catabolic process	420	<0.001
GO:1902743	regulation of lamellipodium organization	53	0.003
GO:0010975	regulation of neuron projection development	447	0.009
GO:0008064	regulation of actin polymerization or depolymerization	161	0.002
GO:0032535	regulation of cellular component size	373	<0.001
GO:0030832	regulation of actin filament length	164	0.003
GO:0032886	regulation of microtubule-based process	259	0.001
GO:0032970	regulation of actin filament-based process	404	0.001
GO:0032271	regulation of protein polymerization	203	<0.001
GO:1903599	positive regulation of autophagy of mitochondrion	13	0.015
GO:0031334	positive regulation of protein-containing complex assembly	193	<0.001
GO:1902745	positive regulation of lamellipodium organization	37	0.002
GO:0010592	positive regulation of lamellipodium assembly	29	0.005
GO:0009896	positive regulation of catabolic process	521	<0.001
GO:0010591	regulation of lamellipodium assembly	41	0.005
GO:0009894	regulation of catabolic process	992	<0.001
GO:0031329	regulation of cellular catabolic process	790	<0.001
GO:0090066	regulation of anatomical structure size	511	<0.001
GO:0044089	positive regulation of cellular component biogenesis	502	<0.001
GO:0010823	negative regulation of mitochondrion organization	53	0.003
GO:0031346	positive regulation of cell projection organization	353	0.007
GO:1902905	positive regulation of supramolecular fibre organization	175	0.036
GO:0045732	positive regulation of protein catabolic process	212	<0.001

Table B8. GO-CC enriched for Functional Unit 4 of the $LRRK2_{int}$

Term ID	Term name	Term size	Adjusted p-value
GO:0005856	cytoskeleton	2407	<0.001
GO:0070062	extracellular exosome	2109	<0.001
GO:1903561	extracellular vesicle	2133	<0.001
GO:0043230	extracellular organelle	2134	<0.001
GO:0065010	extracellular membrane-bounded organelle	2134	<0.001
GO:0005925	focal adhesion	421	<0.001
GO:0030055	cell-substrate junction	431	<0.001

GO:0030054	cell junction	2214	<0.001
GO:0015630	microtubule cytoskeleton	1388	<0.001
GO:0005739	mitochondrion	1672	<0.001
GO:0070161	anchoring junction	903	<0.001
GO:0016234	inclusion body	74	<0.001
GO:0045202	synapse	1451	<0.001
GO:0044297	cell body	567	<0.001
GO:0042995	cell projection	2329	<0.001
GO:0120025	plasma membrane bounded cell projection	2217	<0.001
GO:1902494	catalytic complex	1781	<0.001
GO:0005819	spindle	431	<0.001
GO:0043005	neuron projection	1299	<0.001
GO:0098794	postsynapse	644	<0.001
GO:0036477	somatodendritic compartment	849	<0.001
GO:0043025	neuronal cell body	500	<0.001
GO:0098687	chromosomal region	399	<0.001
GO:0098562	cytoplasmic side of membrane	213	<0.001
GO:0005815	microtubule organizing centre	855	<0.001
GO:0005938	cell cortex	314	<0.001
GO:0048471	perinuclear region of cytoplasm	733	<0.001
GO:0030425	dendrite	621	<0.001
GO:0097447	dendritic tree	623	<0.001
GO:0030424	axon	641	<0.001
GO:0005813	centrosome	647	<0.001
GO:0022627	cytosolic small ribosomal subunit	44	<0.001
GO:0022626	cytosolic ribosome	110	<0.001
GO:0015935	small ribosomal subunit	79	<0.001
GO:0005757	mitochondrial permeability transition pore complex	7	<0.001
GO:0031967	organelle envelope	1293	<0.001
GO:0031975	envelope	1293	<0.001
GO:0016235	aggresome	36	<0.001
GO:0120111	neuron projection cytoplasm	90	<0.001
GO:0031252	cell leading edge	424	<0.001
GO:0098978	glutamatergic synapse	426	<0.001
GO:0031410	cytoplasmic vesicle	2476	<0.001
GO:0097708	intracellular vesicle	2481	<0.001
GO:0031966	mitochondrial membrane	771	<0.001
GO:0030426	growth cone	170	<0.001
GO:0031264	death-inducing signaling complex	9	<0.001
GO:0098796	membrane protein complex	1357	<0.001
GO:0009898	cytoplasmic side of plasma membrane	175	<0.001
GO:0030427	site of polarized growth	175	<0.001
GO:0150034	distal axon	277	<0.001
GO:0099512	supramolecular fibre	1034	<0.001
GO:0099081	supramolecular polymer	1042	<0.001
GO:0045121	membrane raft	285	<0.001
GO:0098857	membrane microdomain	286	<0.001
GO:0099080	supramolecular complex	1404	<0.001

GO:0005740	mitochondrial envelope	820	<0.001
GO:0044391	ribosomal subunit	189	<0.001
GO:0048770	pigment granule	111	<0.001
GO:0042470	melanosome	111	<0.001
GO:1990909	Wnt signalosome	13	<0.001
GO:0035861	site of double-strand break	86	0.001
GO:1990234	transferase complex	892	0.001
GO:0000781	chromosome, telomeric region	170	0.001
GO:0043197	dendritic spine	171	0.001
GO:0044309	neuron spine	175	0.001
GO:0005694	chromosome	1942	0.001
GO:1904115	axon cytoplasm	63	0.001
GO:0051233	spindle midzone	37	0.002
GO:0140535	intracellular protein-containing complex	951	0.002
GO:0097342	ripoptosome	6	0.002
GO:0031968	organelle outer membrane	249	0.003
GO:0019867	outer membrane	251	0.003
GO:0101031	protein folding chaperone complex	41	0.003
GO:0016328	lateral plasma membrane	72	0.003
GO:0030496	midbody	206	0.004
GO:0090734	site of DNA damage	117	0.005
GO:0005741	mitochondrial outer membrane	220	0.007
GO:0000922	spindle pole	172	0.008
GO:0044294	dendritic growth cone	9	0.009
GO:0046930	pore complex	26	0.010
GO:0099568	cytoplasmic region	290	0.010
GO:0097225	sperm midpiece	54	0.011
GO:1902554	serine/threonine protein kinase complex	136	0.014
GO:1990565	HSP90-CDC37 chaperone complex	2	0.017
GO:0032838	plasma membrane bounded cell projection cytoplasm	252	0.019
GO:0030027	lamellipodium	201	0.023
GO:0005768	endosome	1047	0.026
GO:0034399	nuclear periphery	151	0.026
GO:0099513	polymeric cytoskeletal fibre	793	0.030
GO:0005874	microtubule	474	0.031
GO:0043204	perikaryon	155	0.031
GO:1902911	protein kinase complex	156	0.032
GO:0044292	dendrite terminus	14	0.038
GO:0005903	brush border	111	0.038
GO:0019866	organelle inner membrane	566	0.041
GO:0031616	spindle pole centrosome	15	0.047
GO:0099073	mitochondrion-derived vesicle	3	0.050
GO:0000322	storage vacuole	3	0.050

Table B9. GO-BP enriched for Functional Unit 5 of the LRRK2$_{int}$

Term ID	Term name	Term size	Adjusted p-value
GO:1902531	regulation of intracellular signal transduction	1725	<0.001

GO:0010648	negative regulation of cell communication	1353	0.005
GO:2001242	regulation of intrinsic apoptotic signalling pathway	168	<0.001
GO:1902533	positive regulation of intracellular signal transduction	995	<0.001
GO:0048584	positive regulation of response to stimulus	2267	<0.001
GO:0023056	positive regulation of signalling	1727	<0.001
GO:1901796	regulation of signal transduction by p53 class mediator	105	0.001
GO:2001243	negative regulation of intrinsic apoptotic signalling pathway	100	0.002
GO:1902532	negative regulation of intracellular signal transduction	514	0.011
GO:0009967	positive regulation of signal transduction	1528	0.002
GO:0009968	negative regulation of signal transduction	1248	0.002
GO:1902176	negative regulation of oxidative stress-induced intrinsic apoptotic signalling pathway	19	0.005
GO:0080135	regulation of cellular response to stress	711	<0.001
GO:1900407	regulation of cellular response to oxidative stress	82	0.019
GO:0080134	regulation of response to stress	1441	<0.001
GO:0048585	negative regulation of response to stimulus	1611	0.040
GO:1903202	negative regulation of oxidative stress-induced cell death	47	0.014
GO:2001234	negative regulation of apoptotic signalling pathway	231	0.001
GO:0010942	positive regulation of cell death	588	<0.001
GO:0010647	positive regulation of cell communication	1729	<0.001
GO:0010941	regulation of cell death	1638	<0.001
GO:0043067	regulation of programmed cell death	1483	<0.001
GO:0060548	negative regulation of cell death	1001	<0.001
GO:1901214	regulation of neuron death	324	<0.001
GO:1903201	regulation of oxidative stress-induced cell death	71	0.005
GO:1902882	regulation of response to oxidative stress	91	0.048
GO:0042981	regulation of apoptotic process	1454	<0.001

GO:0043068	positive regulation of programmed cell death	522	<0.001
GO:0043069	negative regulation of programmed cell death	900	<0.001
GO:0023057	negative regulation of signalling	1353	0.005
GO:0043066	negative regulation of apoptotic process	881	<0.001
GO:0043065	positive regulation of apoptotic process	509	<0.001
GO:2001233	regulation of apoptotic signalling pathway	366	<0.001
GO:2001235	positive regulation of apoptotic signalling pathway	130	0.006
GO:1901216	positive regulation of neuron death	93	<0.001

Table B10. GO-CC enriched for Functional Unit 5 of the $LRRK2_{int}$

Term iD	**Term name**	**Term size**	Adjusted p-value
GO:0005856	cytoskeleton	2407	<0.001
GO:0070062	extracellular exosome	2109	<0.001
GO:1903561	extracellular vesicle	2133	<0.001
GO:0043230	extracellular organelle	2134	<0.001
GO:0065010	extracellular membrane-bounded organelle	2134	<0.001
GO:0005925	focal adhesion	421	<0.001
GO:0030055	cell-substrate junction	431	<0.001
GO:0030054	cell junction	2214	<0.001
GO:0015630	microtubule cytoskeleton	1388	<0.001
GO:0005739	mitochondrion	1672	<0.001
GO:0070161	anchoring junction	903	<0.001
GO:0016234	inclusion body	74	<0.001
GO:0045202	synapse	1451	<0.001
GO:0044297	cell body	567	<0.001
GO:0042995	cell projection	2329	<0.001
GO:0120025	plasma membrane bounded cell projection	2217	<0.001
GO:1902494	catalytic complex	1781	<0.001
GO:0005819	spindle	431	<0.001
GO:0043005	neuron projection	1299	<0.001
GO:0098794	postsynapse	644	<0.001
GO:0036477	somatodendritic compartment	849	<0.001
GO:0043025	neuronal cell body	500	<0.001
GO:0098687	chromosomal region	399	<0.001
GO:0098562	cytoplasmic side of membrane	213	<0.001
GO:0005815	microtubule organizing center	855	<0.001
GO:0005938	cell cortex	314	<0.001
GO:0048471	perinuclear region of cytoplasm	733	<0.001
GO:0030425	dendrite	621	<0.001
GO:0097447	dendritic tree	623	<0.001

GO:0030424	axon	641	<0.001
GO:0005813	centrosome	647	<0.001
GO:0022627	cytosolic small ribosomal subunit	44	<0.001
GO:0022626	cytosolic ribosome	110	<0.001
GO:0015935	small ribosomal subunit	79	<0.001
GO:0005757	mitochondrial permeability transition pore complex	7	<0.001
GO:0031967	organelle envelope	1293	<0.001
GO:0031975	envelope	1293	<0.001
GO:0016235	aggresome	36	<0.001
GO:0120111	neuron projection cytoplasm	90	<0.001
GO:0031252	cell leading edge	424	<0.001
GO:0098978	glutamatergic synapse	426	<0.001
GO:0031410	cytoplasmic vesicle	2476	<0.001
GO:0097708	intracellular vesicle	2481	<0.001
GO:0031966	mitochondrial membrane	771	<0.001
GO:0030426	growth cone	170	<0.001
GO:0031264	death-inducing signalling complex	9	<0.001
GO:0098796	membrane protein complex	1357	<0.001
GO:0009898	cytoplasmic side of plasma membrane	175	<0.001
GO:0030427	site of polarized growth	175	<0.001
GO:0150034	distal axon	277	<0.001
GO:0099512	supramolecular fibre	1034	<0.001
GO:0099081	supramolecular polymer	1042	<0.001
GO:0045121	membrane raft	285	<0.001
GO:0098857	membrane microdomain	286	<0.001
GO:0099080	supramolecular complex	1404	<0.001
GO:0005740	mitochondrial envelope	820	<0.001
GO:0044391	ribosomal subunit	189	<0.001
GO:0048770	pigment granule	111	<0.001
GO:0042470	melanosome	111	<0.001
GO:1990909	Wnt signalosome	13	<0.001
GO:0035861	site of double-strand break	86	0.001
GO:1990234	transferase complex	892	0.001
GO:0000781	chromosome, telomeric region	170	0.001
GO:0043197	dendritic spine	171	0.001
GO:0044309	neuron spine	175	0.001
GO:0005694	chromosome	1942	0.001
GO:1904115	axon cytoplasm	63	0.001
GO:0051233	spindle midzone	37	0.002
GO:0140535	intracellular protein-containing complex	951	0.002
GO:0097342	ripoptosome	6	0.002
GO:0031968	organelle outer membrane	249	0.003
GO:0019867	outer membrane	251	0.003
GO:0101031	protein folding chaperone complex	41	0.003
GO:0016328	lateral plasma membrane	72	0.003
GO:0030496	midbody	206	0.004
GO:0090734	site of DNA damage	117	0.005
GO:0005741	mitochondrial outer membrane	220	0.007
GO:0000922	spindle pole	172	0.008

GO:0044294	dendritic growth cone	9	0.009
GO:0046930	pore complex	26	0.010
GO:0099568	cytoplasmic region	290	0.010
GO:0097225	sperm midpiece	54	0.011
GO:1902554	serine/threonine protein kinase complex	136	0.014
GO:0032838	plasma membrane bounded cell projection cytoplasm	252	0.019
GO:0030027	lamellipodium	201	0.023
GO:0005768	endosome	1047	0.026
GO:0034399	nuclear periphery	151	0.026
GO:0099513	polymeric cytoskeletal fibre	793	0.030
GO:0005874	microtubule	474	0.031
GO:0043204	perikaryon	155	0.031
GO:1902911	protein kinase complex	156	0.032
GO:0044292	dendrite terminus	14	0.038
GO:0005903	brush border	111	0.038
GO:0019866	organelle inner membrane	566	0.041
GO:0031616	spindle pole centrosome	15	0.047

Table B11. GO-BP enriched for Functional Unit 6 of the LRRK2$_{int}$

Term ID	Term name	Term size	Adjusted p-value
GO:0060560	developmental growth involved in morphogenesis	238	0.018
GO:0030030	cell projection organization	1556	<0.001
GO:0032989	cellular component morphogenesis	778	<0.001
GO:0097581	lamellipodium organization	91	0.048
GO:0040007	growth	933	0.008
GO:0120039	plasma membrane bounded cell projection morphogenesis	651	<0.001
GO:0031175	neuron projection development	976	<0.001
GO:0120036	plasma membrane bounded cell projection organization	1518	<0.001
GO:0048588	developmental cell growth	224	<0.001
GO:0016049	cell growth	497	<0.001
GO:0048666	neuron development	1130	<0.001
GO:0032990	cell part morphogenesis	675	<0.001
GO:0048812	neuron projection morphogenesis	635	<0.001
GO:0030182	neuron differentiation	1401	<0.001
GO:0022008	neurogenesis	1688	<0.001
GO:0030032	lamellipodium assembly	73	0.006
GO:0048858	cell projection morphogenesis	656	<0.001
GO:0000902	cell morphogenesis	965	<0.001
GO:0048699	generation of neurons	1473	<0.001
GO:1990138	neuron projection extension	175	0.001

Table B12. GO-CC enriched for Functional Unit 6 of the LRRK2$_{int}$

Term ID	Term name	Term size	Adjusted p-value
GO:0005856	cytoskeleton	2407	<0.001
GO:0042995	cell projection	2329	<0.001
GO:0030054	cell junction	2214	<0.001
GO:0120025	plasma membrane bounded cell projection	2217	<0.001
GO:0005925	focal adhesion	421	<0.001
GO:0030055	cell-substrate junction	431	<0.001
GO:0045202	synapse	1451	<0.001
GO:0070161	anchoring junction	903	<0.001
GO:0043005	neuron projection	1299	<0.001
GO:0070062	extracellular exosome	2109	<0.001
GO:1903561	extracellular vesicle	2133	<0.001
GO:0065010	extracellular membrane-bounded organelle	2134	<0.001
GO:0043230	extracellular organelle	2134	<0.001
GO:0030424	axon	641	<0.001
GO:0031252	cell leading edge	424	<0.001
GO:0036477	somatodendritic compartment	849	<0.001
GO:0043025	neuronal cell body	500	<0.001
GO:0015630	microtubule cytoskeleton	1388	<0.001
GO:0098978	glutamatergic synapse	426	<0.001
GO:0030425	dendrite	621	<0.001
GO:0097447	dendritic tree	623	<0.001
GO:0044297	cell body	567	<0.001
GO:0098794	postsynapse	644	<0.001
GO:0048471	perinuclear region of cytoplasm	733	<0.001
GO:0150034	distal axon	277	<0.001
GO:0031410	cytoplasmic vesicle	2476	<0.001
GO:0097708	intracellular vesicle	2481	<0.001
GO:0015629	actin cytoskeleton	501	<0.001
GO:0099080	supramolecular complex	1404	<0.001
GO:0099512	supramolecular fiber	1034	<0.001
GO:0099081	supramolecular polymer	1042	<0.001
GO:0030426	growth cone	170	<0.001
GO:0030427	site of polarized growth	175	<0.001
GO:0005903	brush border	111	<0.001
GO:0099513	polymeric cytoskeletal fiber	793	<0.001
GO:0098793	presynapse	561	<0.001
GO:0030027	lamellipodium	201	<0.001
GO:0005938	cell cortex	314	<0.001
GO:0048770	pigment granule	111	<0.001
GO:0042470	melanosome	111	<0.001
GO:0098862	cluster of actin-based cell projections	167	<0.001
GO:0005874	microtubule	474	<0.001

GO:0005815	microtubule organizing center	855	<0.001
GO:0005911	cell-cell junction	517	<0.001
GO:0016234	inclusion body	74	<0.001
GO:0098590	plasma membrane region	1304	<0.001
GO:0042641	actomyosin	77	<0.001
GO:0043197	dendritic spine	171	<0.001
GO:0031253	cell projection membrane	354	<0.001
GO:0044309	neuron spine	175	<0.001
GO:0001726	ruffle	181	<0.001
GO:0120111	neuron projection cytoplasm	90	<0.001
GO:1990909	Wnt signalosome	13	<0.001
GO:0099503	secretory vesicle	1039	<0.001
GO:0005768	endosome	1047	<0.001
GO:0032279	asymmetric synapse	339	<0.001
GO:0097517	contractile actin filament bundle	70	<0.001
GO:0001725	stress fiber	70	<0.001
GO:0005739	mitochondrion	1672	0.001
GO:0098984	neuron to neuron synapse	371	0.001
GO:0032432	actin filament bundle	78	0.001
GO:0043209	myelin sheath	45	0.001
GO:0014069	postsynaptic density	323	0.002
GO:0045335	phagocytic vesicle	141	0.002
GO:0030139	endocytic vesicle	342	0.003
GO:0044294	dendritic growth cone	9	0.003
GO:0005819	spindle	431	0.004
GO:0099572	postsynaptic specialization	352	0.004
GO:0098562	cytoplasmic side of membrane	213	0.004
GO:0043204	perikaryon	155	0.004
GO:0099568	cytoplasmic region	290	0.005
GO:1904115	axon cytoplasm	63	0.005
GO:0005885	Arp2/3 protein complex	11	0.007
GO:0005929	cilium	747	0.007
GO:0043679	axon terminus	112	0.007
GO:0030141	secretory granule	871	0.009
GO:0016328	lateral plasma membrane	72	0.010
GO:0016235	aggresome	36	0.010
GO:0101002	ficolin-1-rich granule	184	0.013
GO:0032838	plasma membrane bounded cell projection cytoplasm	252	0.013
GO:0005794	Golgi apparatus	1613	0.014
GO:0044292	dendrite terminus	14	0.014
GO:0098845	postsynaptic endosome	14	0.014
GO:0044306	neuron projection terminus	127	0.014
GO:0005769	early endosome	428	0.019
GO:0000322	storage vacuole	3	0.025

GO:0099073	mitochondrion-derived vesicle	3	0.025
GO:0043195	terminal bouton	45	0.025
GO:0005813	centrosome	647	0.031
GO:0031143	pseudopodium	18	0.032
GO:0098858	actin-based cell projection	219	0.039
GO:0012506	vesicle membrane	1230	0.041
GO:0098685	Schaffer collateral - CA1 synapse	97	0.043
GO:0098588	bounding membrane of organelle	2160	0.045

Table B13. GO-BP enriched for Functional Unit 7 of the $LRRK2_{int}$

Term ID	Term name	Term size	Adjusted p-value
GO:0046827	positive regulation of protein export from nucleus	18	<0.001
GO:0046825	regulation of protein export from nucleus	32	0.001
GO:0032386	regulation of intracellular transport	335	<0.001
GO:0017157	regulation of exocytosis	215	0.020
GO:1903421	regulation of synaptic vesicle recycling	20	0.008
GO:0046822	regulation of nucleocytoplasmic transport	110	0.007
GO:0070201	regulation of establishment of protein localization	523	0.002
GO:0046824	positive regulation of nucleocytoplasmic transport	61	0.011
GO:0060627	regulation of vesicle-mediated transport	538	<0.001
GO:1904951	positive regulation of establishment of protein localization	315	0.015
GO:0051050	positive regulation of transport	905	0.004
GO:0032388	positive regulation of intracellular transport	199	0.001
GO:0032880	regulation of protein localization	854	<0.001
GO:0051049	regulation of transport	1754	<0.001
GO:0060341	regulation of cellular localization	963	<0.001
GO:0032879	regulation of localization	2101	<0.001

Table B14. GO-CC enriched for Functional Unit 7 of the $LRRK2_{int}$

Term ID	Term name	Term size	Adjusted p-value
GO:0005856	cytoskeleton	2407	<0.001
GO:0030054	cell junction	2214	<0.001
GO:0045202	synapse	1451	<0.001
GO:0015630	microtubule cytoskeleton	1388	<0.001
GO:0070062	extracellular exosome	2109	<0.001
GO:1903561	extracellular vesicle	2133	<0.001
GO:0065010	extracellular membrane-bounded organelle	2134	<0.001
GO:0043230	extracellular organelle	2134	<0.001
GO:0098794	postsynapse	644	<0.001

GO:0043005	neuron projection	1299	<0.001
GO:0120025	plasma membrane bounded cell projection	2217	<0.001
GO:0031410	cytoplasmic vesicle	2476	<0.001
GO:0097708	intracellular vesicle	2481	<0.001
GO:0048471	perinuclear region of cytoplasm	733	<0.001
GO:0042995	cell projection	2329	<0.001
GO:0098978	glutamatergic synapse	426	<0.001
GO:0098793	presynapse	561	<0.001
GO:0044297	cell body	567	<0.001
GO:0030424	axon	641	<0.001
GO:0150034	distal axon	277	<0.001
GO:0036477	somatodendritic compartment	849	<0.001
GO:0099512	supramolecular fiber	1034	<0.001
GO:0099081	supramolecular polymer	1042	<0.001
GO:0030425	dendrite	621	<0.001
GO:0097447	dendritic tree	623	<0.001
GO:0099080	supramolecular complex	1404	<0.001
GO:0005815	microtubule organizing centre	855	<0.001
GO:0042470	melanosome	111	<0.001
GO:0048770	pigment granule	111	<0.001
GO:0005925	focal adhesion	421	<0.001
GO:0070161	anchoring junction	903	<0.001
GO:0030055	cell-substrate junction	431	<0.001
GO:0099503	secretory vesicle	1039	<0.001
GO:0005813	centrosome	647	<0.001
GO:0016234	inclusion body	74	<0.001
GO:0043025	neuronal cell body	500	<0.001
GO:0005819	spindle	431	<0.001
GO:0005768	endosome	1047	<0.001
GO:0099513	polymeric cytoskeletal fibre	793	<0.001
GO:0005874	microtubule	474	<0.001
GO:0030426	growth cone	170	<0.001
GO:0030139	endocytic vesicle	342	<0.001
GO:0030427	site of polarized growth	175	<0.001
GO:0045335	phagocytic vesicle	141	<0.001
GO:0005739	mitochondrion	1672	<0.001
GO:0012506	vesicle membrane	1230	<0.001
GO:0098588	bounding membrane of organelle	2160	<0.001
GO:0031252	cell leading edge	424	<0.001
GO:0043197	dendritic spine	171	<0.001
GO:0030141	secretory granule	871	<0.001
GO:0070382	exocytic vesicle	228	<0.001
GO:0044309	neuron spine	175	<0.001
GO:0030659	cytoplasmic vesicle membrane	1213	<0.001
GO:0005938	cell cortex	314	<0.001

GO:0005757	mitochondrial permeability transition pore complex	7	<0.001
GO:0098590	plasma membrane region	1304	<0.001
GO:0101031	protein folding chaperone complex	41	<0.001
GO:0008021	synaptic vesicle	212	<0.001
GO:0043679	axon terminus	112	<0.001
GO:0030133	transport vesicle	427	<0.001
GO:0000922	spindle pole	172	<0.001
GO:0044306	neuron projection terminus	127	<0.001
GO:0015629	actin cytoskeleton	501	<0.001
GO:0030496	midbody	206	<0.001
GO:0005911	cell-cell junction	517	<0.001
GO:0098562	cytoplasmic side of membrane	213	<0.001
GO:0045121	membrane raft	285	<0.001
GO:0098857	membrane microdomain	286	<0.001
GO:0005773	vacuole	838	<0.001
GO:0016235	aggresome	36	<0.001
GO:0098984	neuron to neuron synapse	371	<0.001
GO:0010008	endosome membrane	551	<0.001
GO:0030670	phagocytic vesicle membrane	78	<0.001
GO:0030672	synaptic vesicle membrane	127	<0.001
GO:0099501	exocytic vesicle membrane	127	<0.001
GO:0043195	terminal bouton	45	<0.001
GO:0005802	trans-Golgi network	257	<0.001
GO:0032279	asymmetric synapse	339	<0.001
GO:0030666	endocytic vesicle membrane	194	<0.001
GO:0120111	neuron projection cytoplasm	90	0.001
GO:0099572	postsynaptic specialization	352	0.001
GO:0098791	Golgi apparatus subcompartment	377	0.001
GO:0046930	pore complex	26	0.001
GO:0030658	transport vesicle membrane	232	0.002
GO:0044294	dendritic growth cone	9	0.002
GO:1904115	axon cytoplasm	63	0.002
GO:0014069	postsynaptic density	323	0.002
GO:0098796	membrane protein complex	1357	0.003
GO:0001726	ruffle	181	0.003
GO:0031984	organelle subcompartment	1505	0.003
GO:0072686	mitotic spindle	184	0.004
GO:0101002	ficolin-1-rich granule	184	0.004
GO:0005764	lysosome	745	0.004
GO:0000323	lytic vacuole	745	0.004
GO:0005905	clathrin-coated pit	75	0.005
GO:0051233	spindle midzone	37	0.005
GO:1990909	Wnt signalosome	13	0.006
GO:0030027	lamellipodium	201	0.006
GO:0055037	recycling endosome	201	0.006

GO:0031975	envelope	1293	0.006
GO:0031967	organelle envelope	1293	0.006
GO:0044292	dendrite terminus	14	0.007
GO:0098845	postsynaptic endosome	14	0.007
GO:0005876	spindle microtubule	83	0.007
GO:0099568	cytoplasmic region	290	0.008
GO:0005794	Golgi apparatus	1613	0.009
GO:0031616	spindle pole centrosome	15	0.009
GO:1902494	catalytic complex	1781	0.010
GO:0098858	actin-based cell projection	219	0.011
GO:0005770	late endosome	305	0.012
GO:0043204	perikaryon	155	0.014
GO:0005798	Golgi-associated vesicle	96	0.015
GO:0030016	myofibril	236	0.017
GO:0030132	clathrin coat of coated pit	19	0.019
GO:0055038	recycling endosome membrane	103	0.021
GO:0043292	contractile fibre	245	0.022
GO:0099522	cytosolic region	20	0.022
GO:0097225	sperm midpiece	54	0.022
GO:0030175	filopodium	107	0.025
GO:0032838	plasma membrane bounded cell projection cytoplasm	252	0.026
GO:0031256	leading edge membrane	179	0.030
GO:0031253	cell projection membrane	354	0.034
GO:0140535	intracellular protein-containing complex	951	0.045

Appendix C: GO-BPs enriched for topological clusters of the LRRK2$_{net}$

Table C1. GO-BP enriched for topological cluster A of the LRRK2$_{net}$

Term ID	Term Name	Term size	Adjusted p-value	Term Group
GO:0006412	translation	743	<0.001	translation
GO:0002181	cytoplasmic translation	159	<0.001	translation
GO:0043043	peptide biosynthetic process	772	<0.001	translation
GO:0043604	amide biosynthetic process	905	<0.001	translation
GO:0006518	peptide metabolic process	926	<0.001	translation
GO:0034645	cellular macromolecule biosynthetic process	1208	<0.001	translation
GO:0006260	DNA replication	285	<0.001	translation
GO:0043603	amide metabolic process	1204	<0.001	translation
GO:1901566	organonitrogen compound biosynthetic process	1805	<0.001	translation
GO:0140694	non-membrane-bounded organelle assembly	386	<0.001	translation
GO:0042255	ribosome assembly	59	<0.001	translation
GO:0042274	ribosomal small subunit biogenesis	77	<0.001	translation
GO:0000028	ribosomal small subunit assembly	17	<0.001	translation
GO:0030490	maturation of SSU-rRNA	54	<0.001	translation
GO:0070925	organelle assembly	944	<0.001	translation
GO:0042254	ribosome biogenesis	311	0.001	translation
GO:0006364	rRNA processing	229	0.002	translation
GO:0045005	DNA-templated DNA replication maintenance of fidelity	56	0.004	translation
GO:0045004	DNA replication proofreading	2	0.006	translation
GO:0006261	DNA-templated DNA replication	162	0.007	translation
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	38	0.009	translation
GO:0016072	rRNA metabolic process	268	0.010	translation
GO:0006281	DNA repair	591	0.013	translation

Table C2. GO-BP enriched for Topological Cluster B of the LRRK2$_{net}$

Term ID	Term Name	Term size	Adjusted p-value	Term Group
GO:0043065	positive regulation of apoptotic process	509	<0.001	cell death

GO:2001235	positive regulation of apoptotic signalling pathway	130	0.020	cell death
GO:0043068	positive regulation of programmed cell death	522	<0.001	cell death
GO:0062100	positive regulation of programmed necrotic cell death	7	0.042	cell death
GO:0010942	positive regulation of cell death	588	<0.001	cell death
GO:0097190	apoptotic signalling pathway	597	<0.001	cell death
GO:0097191	extrinsic apoptotic signalling pathway	224	<0.001	cell death
GO:0097193	intrinsic apoptotic signalling pathway	299	0.027	cell death
GO:0042981	regulation of apoptotic process	1454	<0.001	cell death
GO:2001233	regulation of apoptotic signalling pathway	366	<0.001	cell death
GO:0043067	regulation of programmed cell death	1483	<0.001	cell death
GO:0062098	regulation of programmed necrotic cell death	28	<0.001	cell death
GO:0043069	negative regulation of programmed cell death	900	0.005	cell death
GO:0097300	programmed necrotic cell death	46	<0.001	cell death
GO:0070266	necroptotic process	39	<0.001	cell death
GO:0010941	regulation of cell death	1638	<0.001	cell death
GO:0010939	regulation of necrotic cell death	42	<0.001	cell death
GO:0060548	negative regulation of cell death	1001	0.011	cell death
GO:0070265	necrotic cell death	62	<0.001	cell death
GO:0006915	apoptotic process	1894	<0.001	cell death
GO:0060544	regulation of necroptotic process	25	<0.001	cell death
GO:0060546	negative regulation of necroptotic process	16	0.001	cell death
GO:0060545	positive regulation of necroptotic process	6	0.030	cell death
GO:0012501	programmed cell death	1948	<0.001	cell death
GO:0062099	negative regulation of programmed necrotic cell death	17	0.001	cell death

GO:0008219	cell death	2113	<0.001	cell death
GO:2001236	regulation of extrinsic apoptotic signalling pathway	153	0.037	cell death
GO:0060547	negative regulation of necrotic cell death	22	0.003	cell death
GO:2001238	positive regulation of extrinsic apoptotic signalling pathway	50	0.038	cell death
GO:0043122	regulation of I-kappaB kinase/NF-kappaB signalling	254	<0.001	cell death
GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB signalling	191	<0.001	cell death
GO:0043124	negative regulation of I-kappaB kinase/NF-kappaB signalling	46	0.029	cell death
GO:0001959	regulation of cytokine-mediated signalling pathway	152	0.001	cell death
GO:1902531	regulation of intracellular signal transduction	1725	0.001	cell death
GO:0009967	positive regulation of signal transduction	1528	0.003	cell death
GO:0009968	negative regulation of signal transduction	1248	0.006	cell death
GO:1902533	positive regulation of intracellular signal transduction	995	0.011	cell death
GO:1902532	negative regulation of intracellular signal transduction	514	0.025	cell death
GO:0023056	positive regulation of signalling	1727	0.010	cell death
GO:0023057	negative regulation of signalling	1353	0.011	cell death
GO:0010647	positive regulation of cell communication	1729	0.010	cell death
GO:0010648	negative regulation of cell communication	1353	0.011	cell death
GO:0010803	regulation of tumor necrosis factor-mediated signalling pathway	54	0.001	cell death
GO:1903265	positive regulation of tumor necrosis factor-mediated signalling pathway	12	<0.001	cell death

GO:0001961	positive regulation of cytokine-mediated signalling pathway	57	0.001	cell death
GO:0060760	positive regulation of response to cytokine stimulus	64	0.001	cell death
GO:0060759	regulation of response to cytokine stimulus	163	0.001	cell death
GO:0048585	negative regulation of response to stimulus	1611	0.005	cell death
GO:0044093	positive regulation of molecular function	1580	<0.001	protein metabolism
GO:0043085	positive regulation of catalytic activity	1181	<0.001	protein metabolism
GO:0051091	positive regulation of DNA-binding transcription factor activity	268	0.016	protein metabolism
GO:0050790	regulation of catalytic activity	2356	0.022	protein metabolism
GO:0045862	positive regulation of proteolysis	369	<0.001	protein metabolism
GO:0010952	positive regulation of peptidase activity	189	0.003	protein metabolism
GO:0051345	positive regulation of hydrolase activity	582	0.003	protein metabolism
GO:2001056	positive regulation of cysteine-type endopeptidase activity	143	0.001	protein metabolism
GO:0043280	positive regulation of cysteine-type endopeptidase activity involved in apoptotic process	124	0.016	protein metabolism
GO:0030162	regulation of proteolysis	737	0.016	protein metabolism
GO:0010950	positive regulation of endopeptidase activity	169	0.002	protein metabolism
GO:0032075	positive regulation of nuclease activity	7	0.042	protein metabolism
GO:2000116	regulation of cysteine-type endopeptidase activity	234	0.008	protein metabolism
GO:0051092	positive regulation of NF-kappaB transcription factor activity	157	0.001	protein metabolism
GO:0034976	response to endoplasmic reticulum stress	259	0.013	response to stress

GO:0071345	cellular response to cytokine stimulus	823	<0.001	response to stress
GO:0071356	cellular response to tumor necrosis factor	225	<0.001	response to stress
GO:0019221	cytokine-mediated signalling pathway	480	0.017	response to stress
GO:0033209	tumor necrosis factor-mediated signalling pathway	105	<0.001	response to stress
GO:0034612	response to tumor necrosis factor	245	<0.001	response to stress
GO:0034097	response to cytokine	914	<0.001	response to stress
GO:0031098	stress-activated protein kinase signalling cascade	246	0.010	response to stress
GO:0062197	cellular response to chemical stress	332	0.002	response to stress
GO:0034599	cellular response to oxidative stress	275	0.018	response to stress
GO:0033554	cellular response to stress	1941	0.003	response to stress
GO:0071310	cellular response to organic substance	1993	0.004	response to stress

Table C3. GO-BP enriched for Topological Cluster C of the $LRRK2_{net}$

Term ID	Term Name	Term size	Adjusted p-value	Term Group
GO:0006897	endocytosis	660	0.045	Vesicular transport
GO:2000781	positive regulation of double-strand break repair	86	0.034	Translation
GO:0010569	regulation of double-strand break repair via homologous recombination	73	0.018	Translation
GO:1905168	positive regulation of double-strand break repair via homologous recombination	39	0.001	Translation
GO:0045911	positive regulation of DNA recombination	71	0.016	Translation
GO:0099003	vesicle-mediated transport in synapse	217	0.005	Vesicular transport
GO:0099504	synaptic vesicle cycle	196	0.003	Vesicular transport
GO:0140238	presynaptic endocytosis	67	<0.001	Vesicular transport

GO:0048488	synaptic vesicle endocytosis	67	<0.001	Vesicular transport
GO:0036465	synaptic vesicle recycling	80	<0.001	Vesicular transport
GO:0120039	plasma membrane bounded cell projection morphogenesis	651	0.041	Cytoskeleton organisation
GO:1990138	neuron projection extension	175	0.032	Cytoskeleton organisation
GO:0048812	neuron projection morphogenesis	635	0.034	Cytoskeleton organisation
GO:0000902	cell morphogenesis	965	0.015	Cytoskeleton organisation
GO:0048858	cell projection morphogenesis	656	0.044	Cytoskeleton organisation
GO:0031529	ruffle organization	58	0.007	Cytoskeleton organisation
GO:0120036	plasma membrane bounded cell projection organization	1518	0.004	Cytoskeleton organisation
GO:0030030	cell projection organization	1556	0.005	Cytoskeleton organisation
GO:0120034	positive regulation of plasma membrane bounded cell projection assembly	112	0.004	Cytoskeleton organisation
GO:1902743	regulation of lamellipodium organization	53	<0.001	Cytoskeleton organisation
GO:0090066	regulation of anatomical structure size	511	<0.001	Cytoskeleton organisation
GO:0030032	lamellipodium assembly	73	<0.001	Cytoskeleton organisation
GO:0097581	lamellipodium organization	91	<0.001	Cytoskeleton organisation
GO:0010591	regulation of lamellipodium assembly	41	<0.001	Cytoskeleton organisation
GO:0044087	regulation of cellular component biogenesis	955	<0.001	Cytoskeleton organisation
GO:0043254	regulation of protein-containing complex assembly	409	0.001	Cytoskeleton organisation
GO:0051128	regulation of cellular component organization	2365	<0.001	Cytoskeleton organisation
GO:0010592	positive regulation of lamellipodium assembly	29	<0.001	Cytoskeleton organisation
GO:1902745	positive regulation of lamellipodium organization	37	<0.001	Cytoskeleton organisation

GO:0032271	regulation of protein polymerization	203	<0.001	Cytoskeleton organisation
GO:0051258	protein polymerization	284	<0.001	Cytoskeleton organisation
GO:0034314	Arp2/3 complex-mediated actin nucleation	47	<0.001	Cytoskeleton organisation
GO:0032535	regulation of cellular component size	373	<0.001	Cytoskeleton organisation
GO:0030833	regulation of actin filament polymerization	142	<0.001	Cytoskeleton organisation
GO:0033043	regulation of organelle organization	1189	<0.001	Cytoskeleton organisation
GO:1902903	regulation of supramolecular fiber organization	381	<0.001	Cytoskeleton organisation
GO:0032970	regulation of actin filament-based process	404	<0.001	Cytoskeleton organisation
GO:0030041	actin filament polymerization	172	<0.001	Cytoskeleton organisation
GO:0008064	regulation of actin polymerization or depolymerization	161	<0.001	Cytoskeleton organisation
GO:0030832	regulation of actin filament length	164	<0.001	Cytoskeleton organisation
GO:0110053	regulation of actin filament organization	276	<0.001	Cytoskeleton organisation
GO:0032956	regulation of actin cytoskeleton organization	364	<0.001	Cytoskeleton organisation
GO:0051493	regulation of cytoskeleton organization	541	<0.001	Cytoskeleton organisation
GO:0099188	postsynaptic cytoskeleton organization	12	0.002	Cytoskeleton organisation
GO:0007010	cytoskeleton organization	1507	<0.001	Cytoskeleton organisation
GO:0097435	supramolecular fiber organization	814	<0.001	Cytoskeleton organisation
GO:0098974	postsynaptic actin cytoskeleton organization	10	0.001	Cytoskeleton organisation
GO:0030036	actin cytoskeleton organization	727	<0.001	Cytoskeleton organisation
GO:0030029	actin filament-based process	813	<0.001	Cytoskeleton organisation
GO:0045010	actin nucleation	62	<0.001	Cytoskeleton organisation
GO:0008154	actin polymerization or depolymerization	202	<0.001	Cytoskeleton organisation

GO:0007015	actin filament organization	452	<0.001	Cytoskeleton organisation

Table C4. GO-BP enriched for Topological Cluster D of the LRRK2$_{net}$

Term ID	Term Name	Term size	Adjusted p-value	Term Group
GO:0008219	cell death	2113	0.003	Cell death
GO:0006915	apoptotic process	1894	0.005	Cell death
GO:0012501	programmed cell death	1948	0.001	Cell death
GO:0010942	positive regulation of cell death	588	0.002	Cell death
GO:0010941	regulation of cell death	1638	0.001	Cell death
GO:0043065	positive regulation of apoptotic process	509	0.006	Cell death
GO:0042981	regulation of apoptotic process	1454	0.002	Cell death
GO:0043068	positive regulation of programmed cell death	522	0.001	Cell death
GO:0043067	regulation of programmed cell death	1483	<0.001	Cell death
GO:0071705	nitrogen compound transport	2252	0.040	Protein localisation
GO:1903829	positive regulation of protein localization	459	0.033	Protein localisation
GO:0015031	protein transport	1788	0.017	Protein localisation
GO:0045184	establishment of protein localization	1898	0.032	Protein localisation
GO:1903078	positive regulation of protein localization to plasma membrane	60	0.009	Protein localisation
GO:1904377	positive regulation of protein localization to cell periphery	68	0.014	Protein localisation
GO:1904375	regulation of protein localization to cell periphery	125	0.006	Protein localisation
GO:0006886	intracellular protein transport	1095	0.005	Protein localisation
GO:0046907	intracellular transport	1806	0.003	Protein localisation
GO:0051649	establishment of localization in cell	2340	0.002	Protein localisation
GO:0051170	import into nucleus	170	0.027	Protein localisation
GO:0006913	nucleocytoplasmic transport	330	<0.001	Protein localisation

GO:0051169	nuclear transport	330	<0.001	Protein localisation
GO:0006606	protein import into nucleus	165	0.024	Protein localisation
GO:0034504	protein localization to nucleus	309	<0.001	Protein localisation
GO:0043603	amide metabolic process	1204	<0.001	Protein Metabolism
GO:0050790	regulation of catalytic activity	2356	0.013	Protein Metabolism
GO:0043086	negative regulation of catalytic activity	768	0.014	Protein Metabolism
GO:0044092	negative regulation of molecular function	1115	0.006	Protein Metabolism
GO:0045859	regulation of protein kinase activity	661	0.004	Protein Metabolism
GO:0001934	positive regulation of protein phosphorylation	745	0.001	Protein Metabolism
GO:0042327	positive regulation of phosphorylation	822	0.003	Protein Metabolism
GO:0010563	negative regulation of phosphorus metabolic process	420	0.018	Protein Metabolism
GO:0010562	positive regulation of phosphorus metabolic process	907	0.007	Protein Metabolism
GO:0051174	regulation of phosphorus metabolic process	1401	<0.001	Protein Metabolism
GO:0045936	negative regulation of phosphate metabolic process	419	0.018	Protein Metabolism
GO:0045937	positive regulation of phosphate metabolic process	907	0.007	Protein Metabolism
GO:0042325	regulation of phosphorylation	1244	<0.001	Protein Metabolism
GO:0019220	regulation of phosphate metabolic process	1400	<0.001	Protein Metabolism
GO:0043549	regulation of kinase activity	775	0.002	Protein Metabolism
GO:0051348	negative regulation of transferase activity	271	0.001	Protein Metabolism
GO:0051338	regulation of transferase activity	908	<0.001	Protein Metabolism
GO:0031401	positive regulation of protein modification process	1005	0.017	Protein Metabolism

GO:0001932	regulation of protein phosphorylation	1101	<0.001	Protein Metabolism
GO:0031399	regulation of protein modification process	1540	<0.001	Protein Metabolism
GO:0001933	negative regulation of protein phosphorylation	320	0.045	Protein Metabolism
GO:0031400	negative regulation of protein modification process	493	<0.001	Protein Metabolism
GO:0006518	peptide metabolic process	926	<0.001	Translation
GO:0042273	ribosomal large subunit biogenesis	76	0.022	Translation
GO:0042255	ribosome assembly	59	0.008	Translation
GO:0042254	ribosome biogenesis	311	0.038	Translation
GO:1901566	organonitrogen compound biosynthetic process	1805	<0.001	Translation
GO:0034645	cellular macromolecule biosynthetic process	1208	<0.001	Translation
GO:0043043	peptide biosynthetic process	772	<0.001	Translation
GO:0043604	amide biosynthetic process	905	<0.001	Translation
GO:0002181	cytoplasmic translation	159	<0.001	Translation
GO:0006412	translation	743	<0.001	Translation

Table C5. GO-BP enriched for Topological Cluster E of the $LRRK2_{net}$

Term ID	Term Name	Term size	Adjusted p-value	Term Group
GO:0010564	regulation of cell cycle process	720	0.020	Cell cycle
GO:0051726	regulation of cell cycle	1124	0.033	Cell cycle
GO:0000226	microtubule cytoskeleton organization	652	0.009	Cell cycle
GO:0007017	microtubule-based process	939	0.005	Cell cycle
GO:1903047	mitotic cell cycle process	773	0.038	Cell cycle
GO:0007051	spindle organization	199	0.022	Cell cycle
GO:0000278	mitotic cell cycle	929	0.031	Cell cycle
GO:0022402	cell cycle process	1270	0.004	Cell cycle
GO:0007049	cell cycle	1818	<0.001	Cell cycle
GO:0015031	protein transport	1788	0.040	Protein localisation
GO:0051170	import into nucleus	170	0.009	Protein localisation
GO:0006913	nucleocytoplasmic transport	330	<0.001	Protein localisation
GO:0034502	protein localization to chromosome	117	0.024	Protein localisation

GO:0033365	protein localization to organelle	1144	<0.001	Protein localisation
GO:0006886	intracellular protein transport	1095	0.001	Protein localisation
GO:0051169	nuclear transport	330	<0.001	Protein localisation
GO:0051649	establishment of localization in cell	2340	<0.001	Protein localisation
GO:0046907	intracellular transport	1806	<0.001	Protein localisation
GO:0045184	establishment of protein localization	1898	0.001	Protein localisation
GO:0006606	protein import into nucleus	165	0.007	Protein localisation
GO:0034504	protein localization to nucleus	309	<0.001	Protein localisation
GO:0051338	regulation of transferase activity	908	0.025	Protein metabolism
GO:0051603	proteolysis involved in protein catabolic process	799	0.007	Protein metabolism
GO:0043632	modification-dependent macromolecule catabolic process	713	0.002	Protein metabolism
GO:0019941	modification-dependent protein catabolic process	701	0.002	Protein metabolism
GO:0006511	ubiquitin-dependent protein catabolic process	691	0.002	Protein metabolism
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	455	<0.001	Protein metabolism
GO:0010498	proteasomal protein catabolic process	526	0.001	Protein metabolism
GO:0006508	proteolysis	1787	0.040	Protein metabolism
GO:0044265	cellular macromolecule catabolic process	1030	0.002	Protein metabolism
GO:0051438	regulation of ubiquitin-protein transferase activity	59	0.032	Protein metabolism
GO:0090175	regulation of establishment of planar polarity	57	0.028	Response to stress
GO:0033554	cellular response to stress	1941	0.024	Response to stress
GO:0008630	intrinsic apoptotic signalling pathway in response to DNA damage	108	0.016	Response to stress

GO:0006974	cellular response to DNA damage stimulus	881	0.019	Response to stress
GO:0090179	planar cell polarity pathway involved in neural tube closure	12	0.007	Response to stress
GO:0090178	regulation of establishment of planar polarity involved in neural tube closure	13	0.009	Response to stress
GO:0060071	Wnt signalling pathway, planar cell polarity pathway	53	0.021	Response to stress
GO:0035567	non-canonical Wnt signalling pathway	67	0.002	Response to stress
GO:0072332	intrinsic apoptotic signalling pathway by p53 class mediator	82	0.004	Response to stress
GO:0072331	signal transduction by p53 class mediator	175	0.001	Response to stress

Table C6. GO-BP enriched for Topological Cluster F of the $LRRK2_{net}$

Term ID	Term Name	Term size	Adjusted p-value	Term Group
GO:0000423	Autophagy	37	0.003	Autophagy
GO:0061912	selective autophagy	89	0.044	Autophagy
GO:0006914	autophagy	562	0.043	Autophagy
GO:0061919	process utilizing autophagic mechanism	562	0.043	Autophagy
GO:0007005	mitochondrion organization	539	0.035	Autophagy
GO:0098780	response to mitochondrial depolarisation	21	0.001	Autophagy
GO:1904923	regulation of autophagy of mitochondrion in response to mitochondrial depolarization	15	<0.001	Autophagy
GO:1904925	positive regulation of autophagy of mitochondrion in response to mitochondrial depolarization	14	<0.001	Autophagy
GO:0008053	mitochondrial fusion	29	0.001	Autophagy
GO:0048284	organelle fusion	154	<0.001	Autophagy
GO:0061734	parkin-mediated stimulation of Autophagy in response to mitochondrial depolarization	6	<0.001	Autophagy

GO:0098779	positive regulation of Autophagy in response to mitochondrial depolarization	10	<0.001	Autophagy

Table C7. GO-BP enriched for Topological Cluster G

Term ID	Term Name	Term size	Adjusted p-value	Term Group
GO:0061919	process utilizing autophagic mechanism	562	0.011	Autophagy
GO:0071211	protein targeting to vacuole involved in autophagy	4	0.017	Autophagy
GO:0006914	autophagy	562	0.011	Autophagy
GO:0061740	protein targeting to lysosome involved in chaperone-mediated autophagy	3	0.008	Autophagy
GO:0061684	chaperone-mediated autophagy	16	<0.001	Autophagy
GO:0051084	'de novo' post-translational protein folding	36	0.024	Protein metabolism
GO:0018108	peptidyl-tyrosine phosphorylation	379	0.014	Protein metabolism
GO:0018212	peptidyl-tyrosine modification	381	0.015	Protein metabolism
GO:0009896	positive regulation of catabolic process	521	<0.001	Protein metabolism
GO:0009894	regulation of catabolic process	992	0.005	Protein metabolism
GO:0051085	chaperone cofactor-dependent protein refolding	31	0.015	Protein metabolism
GO:0031396	regulation of protein ubiquitination	210	0.013	Protein metabolism
GO:1903320	regulation of protein modification by small protein conjugation or removal	252	0.001	Protein metabolism
GO:0006458	'de novo' protein folding	40	0.033	Protein metabolism
GO:0061077	chaperone-mediated protein folding	70	0.004	Protein metabolism
GO:0006457	protein folding	219	<0.001	Protein metabolism

Appendix D: Univariate Logistic analyses on the association between mRNA levels of LRRK2 interactors and cohort phenotype (sPD/LRRK2-PD)

Table D1. Univariate logistic analyses on LRRK2 interactor expression levels and cohort phenotype

Interactor	OR	95% CI		p-value
		Upper Limit	Lower Limit	
TUBB6	0.18	0.06	0.49	0.001
PRKN	0.18	0.06	0.56	0.003
BAG3	0.18	0.05	0.61	0.006
ACTA2	0.31	0.13	0.77	0.012
DVL1	0.17	0.04	0.68	0.012
HSPA1A	3.23	1.25	8.34	0.016
SLC25A6	0.24	0.07	0.84	0.025
MMS19	0.23	0.06	0.83	0.025
LMNB1	2.92	1.14	7.44	0.025
RPS2	0.29	0.10	0.86	0.026
CDK2	0.22	0.06	0.86	0.029
TUBG1	0.25	0.07	0.90	0.035
STUB1	0.24	0.06	0.91	0.036
SNCA	0.45	0.21	0.96	0.040
SPATS2	0.28	0.07	1.11	0.071
LRRK2	2.24	0.91	5.55	0.080
PSMC6	2.36	0.87	6.44	0.093
TRADD	0.34	0.10	1.21	0.097
RAB5B	0.36	0.11	1.24	0.106
RPS27	1.59	0.88	2.89	0.125
RACK1	0.42	0.13	1.29	0.127
RPLP0	0.47	0.18	1.25	0.131
BAX	0.40	0.11	1.43	0.158
RPL11	1.70	0.79	3.65	0.172
TUBA1C	0.42	0.12	1.46	0.172
RPL23	1.54	0.82	2.87	0.178
RPL13	0.51	0.19	1.40	0.193
RPL34	1.50	0.80	2.83	0.207
EEF2	0.48	0.15	1.52	0.210
CYREN	0.43	0.11	1.64	0.218
ACTG1	0.50	0.16	1.55	0.231
CFAP20	0.47	0.14	1.65	0.239

RPS15A	1.52	0.75	3.07	0.243
HSPA4	1.87	0.64	5.43	0.251
DNM1	0.68	0.34	1.33	0.258
KPNB1	1.94	0.60	6.30	0.272
USP39	0.50	0.14	1.84	0.297
YWHAB	1.78	0.60	5.30	0.302
EPRS1	1.69	0.62	4.61	0.304
RAB8A	0.52	0.14	1.89	0.320
MAPK3	0.57	0.18	1.74	0.320
RBIS	1.57	0.62	3.95	0.337
RIF1	1.63	0.59	4.52	0.344
RPL9	1.33	0.72	2.48	0.364
RPS3A	1.30	0.73	2.31	0.377
HS71B	0.63	0.23	1.76	0.381
BAG5	1.62	0.53	4.95	0.398
ZFAND5	1.58	0.54	4.58	0.403
RPL30	1.39	0.64	3.03	0.407
HSP90AB1	1.62	0.51	5.12	0.409
OPA1	1.52	0.56	4.10	0.410
ITGB3BP	1.49	0.57	3.93	0.417
PRKCZ	0.56	0.13	2.33	0.421
RPS13	1.43	0.59	3.48	0.426
TUBA1A	0.63	0.20	1.98	0.434
RAC1	0.62	0.19	2.03	0.434
RPL23A	1.42	0.56	3.62	0.458
FADD	0.63	0.19	2.15	0.464
HIF1A	1.48	0.52	4.27	0.465
MAP2K4	1.47	0.51	4.23	0.470
RAB5A	1.48	0.48	4.54	0.493
ITCH	1.46	0.48	4.46	0.504
CCDC43	1.43	0.50	4.05	0.505
RGS2	1.33	0.57	3.10	0.506
FANCM	1.40	0.49	3.98	0.527
MBP	0.69	0.21	2.27	0.544
LARP4	1.40	0.46	4.26	0.550
EEF1A1	1.28	0.55	2.95	0.567
PADI4	0.77	0.32	1.88	0.569
CMAS	1.35	0.48	3.83	0.571
RPS5	0.76	0.29	1.99	0.576
RAB1A	1.34	0.47	3.84	0.583
RPS7	1.20	0.62	2.35	0.587
PRKDC	0.74	0.24	2.31	0.600

NUP107	1.30	0.46	3.73	0.620
SKA3	0.82	0.36	1.84	0.625
LAMP2	1.27	0.46	3.53	0.640
FAM192A	0.78	0.23	2.67	0.694
IQGAP1	1.23	0.43	3.54	0.700
ARPC2	1.26	0.39	4.03	0.702
RPL19	0.83	0.30	2.32	0.727
AFG3L2	1.25	0.35	4.42	0.728
RPL10A	0.85	0.34	2.14	0.729
RPS11	1.18	0.46	2.99	0.731
TMPO	1.19	0.43	3.24	0.738
SRPK1	1.20	0.40	3.55	0.745
RPS18	1.12	0.53	2.38	0.769
RHBDD1	0.84	0.24	2.90	0.779
RPS14	1.12	0.47	2.66	0.791
DNAJB6	1.17	0.35	3.88	0.803
TAOK3	1.16	0.34	3.90	0.815
DNM1L	1.12	0.37	3.38	0.841
RPSA	1.08	0.44	2.66	0.873
SSR4	1.07	0.34	3.42	0.906
RPS8	1.05	0.43	2.53	0.921
TMOD3	1.06	0.32	3.50	0.930
RPL36A	1.04	0.31	3.48	0.949
TOR1AIP1	0.97	0.31	3.00	0.954
ATP5MG	0.99	0.36	2.75	0.992
GSK3B	1.00	0.30	3.38	1.000

Abbreviations: OR: Odd Ratio; CI: Confidence Interval