Advances in protein-protein interaction network analysis for Parkinson's disease

James E. Tomkins\textsuperscript{a}, Claudia Manzoni\textsuperscript{b, *}

\textsuperscript{a} School of Pharmacy, University of Reading, UK
\textsuperscript{b} School of Pharmacy, University College London, UK

\section*{ARTICLE INFO}

\textbf{Keywords:}
Protein interactions
Network analysis
Parkinson's disease
Data integration
Disease modelling
Interactome

\section*{ABSTRACT}

Protein-protein interactions (PPIs) are a key component of the subcellular molecular networks which enable cells to function. Due to their importance in homeostasis, alterations to the networks can be detrimental, leading to cellular dysfunction and ultimately disease states. Parkinson's disease (PD) is a progressive neurodegenerative condition with multifactorial aetiology, spanning genetic variation and environmental modifiers. At a molecular and systems level, the characterisation of PD is the focus of extensive research, largely due to an unmet need for disease modifying therapies.

PPI network analysis approaches are a valuable strategy to accelerate our understanding of the molecular crosstalk and biological processes underlying PD pathogenesis, especially due to the complex nature of this disease. In this review, we describe the utility of PPI network approaches in modelling complex systems, focusing on previous work in PD research. We discuss four principal strategies for using PPI network approaches: to infer PD related cellular functions, pathways and novel genes; to support genomics studies; to study the interactome of single PD related genes; and to compare the molecular basis of PD to other neurodegenerative disorders.

This is an evolving area of research which is likely to further expand as omics data generation and availability increase. These approaches complement and bridge-the-gap between genetics and functional research to inform future investigations. In this review we outline several limitations that require consideration, acknowledging that ongoing challenges in this field continue to be addressed and the refinement of these approaches will facilitate further advances using PPI network analysis for understanding complex diseases.

\section{1. Protein-protein interactions}

Molecular interactions are critical for coordinating subcellular mechanisms. These interaction events are tightly regulated within a dynamic interconnected landscape of molecular pathways which enable cells to execute intricate processes, such as the recruitment of molecular assemblies required for the completion of cellular functions and the transduction of molecular messages between cellular compartments, in support of homeostasis. Impairment of molecular pathways can contribute towards aberrant signaling and ultimately cellular dysfunc-

...
Institute (EMBL-EBI), provides query capability across many molecular interaction repositories from a single submission. This maximises the coverage of captured data and streamlines the process to a single user interface. To support the utility of these data, open-access resources, which collate data from numerous primary databases and process these data for downstream applications, have been developed. Examples include the search tool for recurring instances of neighbouring genes (STRING), which captures interactions between interactors of a protein of interest and encompasses prediction data in addition to reported experimentally-derived findings (Szklarczyk et al., 2019), the protein interaction network online tool (PINOT), whereby a data processing focus is on wide coverage, data traceability to source publications and a transparent confidence scoring procedure (Tomkins et al., 2020); the human integrated PPI reference (HIPPIE), which incorporates tissue-level expression and functional annotation filters to PPI data processing (Alamis-Lobato et al., 2017); and the molecular interaction search tool (MIST), which supports PPI queries across many model organisms and PPI ortholog mapping across species (Hu et al., 2018).

Protein interaction query tools such as STRING, PINOT, HIPPIE and MIST, together with network visualisation software such as Cytoscape (Shannon et al., 2003), facilitate PPI network analysis.

2. Protein-protein interaction network analysis

We can think of PPIs within the subcellular landscape as a web of “nodes” (i.e. proteins) connected by “edges” that represent interactions between proteins. Proteins of interest which form the basis of a PPI network investigation are often termed “seeds”. The full extent of molecular interactions within a cell is known as the “interactome” or “cell interactome” (Vidal et al., 2011), however, this term is also sometimes used in the context of more localised interaction profiles, for particular proteins of interest, for example (“protein interactome”). In this way, different subcellular PPI landscapes can be mapped adapting the mathematical representation of “networks” (also referred to as “graphs”), which are wiring diagrams able to describe a complex system whereby analysis is defined by graph theory (Pavlopoulos et al., 2011). This mathematical approach was developed to assess the topological features and “describe the detailed behaviour of a system consisting of hundreds to billions of interacting components” (Barabasi, 2016).

For PPI networks, graph theory assists in assessing the interconnectivity of proteins which underpin subcellular functionality whilst retaining the wider molecular context of the (inherently complex) system. Using topological analysis, features can be identified which may be important or relevant to the investigation, such as hub proteins (hubs) and protein clusters. Hubs are defined as the most connected nodes within the network (Fig. 1), thus they are responsible to sustain network connectivity (Pavlopoulos et al., 2011). Protein clusters are communities of nodes with an increased degree of interconnection relative to the whole network and may represent particular functional modules and/or protein complexes.

Functional insight into protein interaction networks is often obtained via functional enrichment analyses and pathway annotation. These complementary approaches tend to utilise data from large-scale functional annotation projects, such as Gene Ontology (GO) (Gene Ontology Consortium, 2018), Reactome (Jassal et al., 2020), KEGG (Kanehisa et al., 2021) and WikiPathways (Slenter et al., 2018), to identify biological processes and pathways which are particularly relevant to specific PPI networks and/or sub-networks. Moreover, based on the guilt-by-association principle (Oliver, 2000), functional annotation data are useful to infer functional associations of proteins which are poorly characterised or have unknown functions. In this way, data integration strategies, whereby PPI data are considered in conjunction with functional annotation data, enhance the utility of in silico models of biological systems, in support of our developing understanding of health and disease.

3. Protein interaction networks and complex neurodegenerative diseases

PPI data can be used to generate in silico networks to model disease. These models can be interrogated to formulate predictions regarding global disease dynamics. Another type of analysis can be implemented to provide prioritization of new molecular players in disease, both in terms of potential causal genes or hypothetical drug targets and biomarkers (Manzoni et al., 2018). For example, protein interaction network analysis performed around the genes known to cause ~70% of Noonan syndrome cases led to the prioritization of a novel potential gene (SHOC2). This gene was then confirmed to be mutated in a cohort of Noonan-like syndrome patients that did not present with alterations in the other known disease genes (Cordero et al., 2009).

The network approach is therefore very promising, especially for the study of complex diseases, where modelling the interaction profiles of multiple molecular players synchronously is required to define the disease landscape.

Recently, an increasing number of research groups have focussed on developing and applying different pipelines and approaches to integrate PPI data in the context of modelling neurodegenerative diseases. These efforts have mainly been focused on the integration of multiple genetic information to identify molecular pathways underlying disease and guide functional work on target validation and biomarkers discovery. Disease classification is another area of current investigation whereby PPI network analysis is applied to subdivide diseases into endophenotypes with defined molecular profiles. Studies have focussed on Alzheimer’s disease (AD) (Hu et al., 2017; Santiago et al., 2019), frontotemporal dementia (FTD) (Bonham et al., 2019, 2018; Ferrari et al., 2017) amyotrophic lateral sclerosis (ALS) (Beltran et al., 2019; Dervishi et al., 2018) and hereditary spastic paraplegia (HSP) (Bis-Brewer et al., 2019; Vavouraki et al., 2021), just to name a few. Albeit the absence of a gold standard pipeline for PPI network analysis in neurodegeneration and the uncertainty about the questions that can be confidently asked of such analysis, there are indeed numerous approaches that we can review within Parkinson’s research, to get a
glimpse of the utility this technique has to offer, in respect to progressing our understanding of complex neurodegenerative diseases.

4. The case of Parkinson's disease

Parkinson's disease (PD) is a complex neurodegenerative condition where multiple molecular players have a role in triggering a progressive brain degeneration that heavily affects the substantia nigra, thus creating a dopamine deficit within the brain circuits that control voluntary movements. The neurodegeneration is not solely confined to the mid-brain and with the progression of the disease it spreads to other brain regions such as the cortices, thus inducing additional, non-motor symptoms such as psychosis and cognitive decline (Lewis and Spillane, 2019).

Complex neurodegenerative disorders are multifactorial; in the case of PD, mutations in several genes are associated with monogenic familial cases (Hernandez et al., 2016; Reed et al., 2019). These genes are often referred to as PD Mendelian or familial genes. One of the difficulties, when working in silico with PD Mendelian genes, is to univocally define which are the PD genes to be considered and where to draw the line between classical PD and other PD-spectrum diseases.

This problem, in fact, aligns with the existence of atypical PD and hereditary parkinsonisms, which are syndromes that present with a PD motor component but are not considered as classical PD by some authors (McFarland and Hess, 2017). For example, LRRK2 can be considered a PD Mendelian gene as mutations in LRRK2 are generally associated with cases of familial, classical PD. Whereas ATP13A2 is more difficult to classify, mutations in this gene are associated with a range of phenotypes which include juvenile parkinsonism, young onset PD, Kufor-Rakeb syndrome (Di Fonzo et al., 2007; Ramirez et al., 2006) and hereditary spastic paraplegia (Odake et al., 2020), where the motor component presents with other, non-classical PD clinical manifestations.

More information on Mendelian inheritance in PD and related syndromes can be found in the Online Mendelian Inheritance in Man (OMIM) repository (Amberger et al., 2015).

In addition to PD Mendelian genes, where we can isolate coding mutations potentially associated with an alteration of protein functionality which contributes to pathogenesis, we need to consider the existence of genetic risk factors. These are single nucleotide polymorphisms (SNPs) associated with an increase in risk of PD, identified via genome wide association studies (GWAS). The latest PD GWAS meta-analysis reported 90 variants in 78 independent risk loci (Nalls et al., 2019). The challenge with these risk signals is that SNPs are most often non-coding; they map to intergenic, intronic or regulatory regions of the genome. As a consequence, it is not clear which gene(s) they influence and how the SNPs functionally impact on the molecular mechanisms of PD (Manzoni et al., 2020).

The influence of the environment is another consideration in the multifactorial-nature of PD pathogenesis. There are several factors that have been linked with increased risk of PD (Kieburtz and Wunderle, 2013); aging is a major risk factor for PD and some chemical exposures, such as paraquat and rotenone (constituents of pesticides), are associated with increased PD prevalence (Tanner et al., 2011). Furthermore, MPTP intoxication results in selective degeneration of the substantia nigra and triggers a PD phenocopy (Langston, 2017). Although there are links between PD and the environment, it is very difficult to discern the molecular basis of specific environmental risk factors in relation to PD onset and progression in a robustly controlled manner. Some environmental factors acutely interfere with relevant cellular functions (such as mitochondrial toxicity upon MPTP intoxication (Langston, 2017)), while others exert a more subtle, progressive effect (such as in the case of aging-related brain changes, associated with progressive epigenetic modifications, reduced efficiency of cell waste disposal and genomic instability (Hou et al., 2019)).

The current clinical approach to PD is focused on symptomatic relief in an attempt to improve quality of life for patients, however disease modifying therapies that can stop or slow the progression of neurodegeneration are yet to be developed. This is partially due to a lack of understanding of the molecular mechanism(s) underlying the disease and absence of knowledge regarding the mechanisms behind cell-specific vulnerability. This scenario is additionally complicated by an intrinsic difficulty in recapitulating the many, different molecular players and their reciprocal connections within the same disease landscape.

This situation provides a well-suited opportunity to use network analysis with the aim of generating a holistic model which accounts for the multiple PD-associated molecular players, to predict disease mechanisms and dynamics. Different PPI network approaches have been used and if we focus on the target goals of these studies, we can divide them into 4 groups (Fig. 2): those that use the PPI network model to infer cellular functions and pathways involved in PD (George et al., 2019; Keane et al., 2015; Rakshit et al., 2014; Shen et al., 2020); those that use the PPI network model to support prioritization of novel molecular players identified via genomics techniques (Ferrari et al., 2018; Kia et al., 2021; Siitonen et al., 2019); those that study the functional space around a single PD Mendelian gene (Hernandez et al., 2020; Manzoni et al., 2015; Forras et al., 2015); and finally those that use a PPI network approach to contrast and compare PD with other neurodegenerative diseases (Ferrari et al., 2018; Haenig et al., 2020; Nguyen et al., 2014) (Table 1).

If we examine the technical construction of network models, all these studies can be sub-divided into 2 groups based on the source of seeds which form the foundation for generating the PPI network model: studies which use a list of (literature derived) PD associated genes and those which initially use gene expression profiling (in PD vs healthy relevant tissues, for example) to obtain seed lists.

5. Approach 1: use of PPI networks to infer PD cellular functions, pathways and novel genes

One of the major challenges in PD research is that of recapitulating, within one unifying model, the plethora of genetic and environmental
factors contributing to disease. It is assumed that such a model would assist in defining the functional/dysfunctional landscape underlying the disease, thus allowing drug target and biomarker prioritization.

Generating this unifying model is a complex task and, even if we only focus on the genetic components of PD (without considering environmental influences), the effort needed to combine and co-evaluate the many different genetic contributors represents a massive endeavour for classical approaches in biomedical research. However, network modelling approaches are well suited for attempting to shed light on the complex nature of PD pathogenesis. Networks are, in fact, the mathematical tools for evaluating multiple objects as a collective dynamic structure rather than single isolated entities.

Applying PPI network analysis approaches to investigate the molecular mechanisms underlying PD is an active research area. A study published in 2015 utilised PPI network analysis to prioritize proteins responsible for the toxicity observed in a MPP\(^+\) in vitro model (Keane et al., 2015). The PPI network was built around seeds defined as proteins that have been reported in literature to contribute to mitochondrial dysfunction or autophagy alterations in the MPP\(^+\) PD model. PPI data were sourced from iRefIndex (Razick et al., 2008). Four proteins (p62, GABARAP, GBRL1 and GBRL2) were predicted via network analysis to be responsible for the observed toxicity in the MPP\(^+\) model. The authors functionally validated this prediction showing that indeed, combined (but not individual) knockdown or over-expression of the prioritized genes modulated the toxic phenotype.

Other studies have taken advantage of gene expression analysis and constructed PPI networks using differentially expressed genes, in the context of PD, as seeds. This approach has been used in a publication (Rakshit et al., 2014) where the authors used differentially expressed genes from post-mortem substantia nigra and frontal cortex of PD patients as seeds to build PPI networks via the Genes2FANs and POiNeT tools (Dannenfelser et al., 2012; Lee et al., 2009). Topological analysis of the PPI networks led to the prioritization of 37 proteins thought to be particularly relevant to PD. Moving forward with these results, these genes require functional verification to evaluate whether they could serve as PD biomarkers or drug targets.

A more recent study, in 2019 (George et al., 2019), used a combination of 5 selected PD markers and differentially expressed genes resulting from the study of multiple developmental stages of PD-IPSC derived dopaminergic neurons as a starting point to build PPI networks using the Bisogenet Cytoscape plugin (Martin et al., 2010). The networks underwent pathway annotation leading to the identification of functions related to the ubiquitin proteasomal system (UPS), immune response and apoptosis. Further topological investigation led to the identification of 51 (new) genes that require further functional verification to establish their status as potential novel PD genes.

Finally, in 2020 (Shen et al., 2020), protein interactions across differentially expressed genes identified from PD blood transcriptomics were mapped using STRING (Szklarczyk et al., 2015). This approach led to the creation of a PPI network which upon topological analysis resulted in the identification of 10 hub proteins. Expression levels of these hubs were used to predict the disease status in a pilot cohort of PD patients and healthy controls. Among the 10 proteins, GP1BA, GP6, ITGB5, and P2RY12 were suggested as particularly relevant to PD. Moving forward with these results, these genes require functional verification to establish whether they could serve as PD biomarkers or drug targets.

6. Approach 2: use of PPI networks to support genomics studies

Genetic screening of PD cohorts is performed via whole exome or whole genome sequencing (WES or WGS) to identify (usually rare) pathogenic variants associated with disease. Another genetic approach utilizes array genotyping to perform GWAS analysis and identify common risk variants associated with PD. Many groups are trying to incorporate PPI network analysis as part of these genetic approaches.

In the case of GWAS, the vast majority of isolated risk variants are...
non-coding thus posing the challenge of identifying which gene(s) they influence (Cannon and Mohlke, 2018). The initial strategy was to nominate the open reading frame (ORF) closest to the SNP of interest, therefore utilizing proximity as a criterion for candidate gene prioritization. However, this approach has proven unsuccessful and highly misleading, with the consequent demand for an efficient post-GWAS pipeline able to confidently nominate candidate genes for functional validation. Multiple approaches are now under investigation, with expression quantitative trait loci (eQTL) as the technique of choice for many.

eQTL analyses make use of tissue-specific transcriptomics databases and evaluate gene expression changes in the presence of a specific SNP of interest (Schaid et al., 2018). If a GWAS marker is associated with a statistically significant increase or decrease in expression level of a particular gene, we can hypothesise that the risk SNP acts via modulation of that specific gene expression. Depending on the genomic distance that is investigated around the SNP of interest, eQTL analyses are divided in cis (where only genes in proximity of the SNP are investigated) and trans (where the analysis is performed on a wider scale) (Nica and Dermitzakis, 2013).

However, eQTL does not explain all the non-coding risk variants, opening to the possibility of additional strategies, such as PPI network analysis, to be integrated into post-GWAS pipelines (Chen et al., 2011; Wang et al., 2018).

A paper published in 2018 by our group explored the possibility of using PPI network analysis to prioritize genes at the PD-GWAS loci (Ferrari et al., 2018). We selected PD Mendelian genes as seeds to build the PD-PPI network using the weighted protein-protein interaction network analysis (WPPINA) pipeline (Ferrari et al., 2017) applied to PPI data from IMEx (member or observer) databases (Orchard et al., 2012). This pipeline was further developed and automated to create the open-access online tool named PINOT (Tomkins et al., 2020). We overapped the relevant nodes obtained via topological and functional analyses of the PD-PPI network with the ORFs located in different linkage disequilibrium blocks around the PD-GWAS markers. With this procedure we confirmed 10 candidate genes (previously nominated by proximity) and prioritized 17 novel ORFs within 13 of the 32 PD-GWAS markers known at the time of the analysis (Nalls et al., 2014). One of the prioritized novel genes (KAT78) was further supported by functional validation (Soutar et al., 2020).

In 2021 (Kia et al., 2021), the prioritization of candidate genes for the updated version of the PD-GWAS (Chang et al., 2017) was attempted using eQTLs and splicing analysis with a combination of Coloc and transcriptome-wide association study (TWAS) analyses applied to expression data from Brainexac (Ramasamy et al., 2014; Trabzuni et al., 2011), the Genotype-Tissue Expression (GTEx) portal (GTEx Consortium, 2013), and CommonMind (Dobyn et al., 2018). In particular, in this paper (Kia et al., 2021), PPI network analysis was used as a follow-up investigation of the 11 candidate genes prioritized via eQTL and splicing analysis to retrospectively verify whether they were part of the PPI network built around the PD Mendelian genes using the WPPINA pipeline. The prioritized genes were, indeed, reported to be more connected to PD Mendelian genes as opposed to randomly sampled genes, thus suggesting a possible functional connection between the PD Mendelian genes and the risk candidate genes at the PD-GWAS loci.

In the case of WES and WGS, one of the main challenges is that of reaching discovery power for association of rare variants with the disease phenotype. One of the strategies that has been implemented within the analytical pipelines is that of focusing on a selection of variants within an array of genes of interest. PPI network analysis has been proposed as a tool to prioritize genes (and therefore variants) to be used for the case/control association.

In 2019, a study was published proposing a PPI-based variant reduction method applied to WES from a Finnish cohort of PD patients (Siitonen et al., 2019). The authors selected 36 genes with suggested association to PD as per text mining of the UniProt database and used these seeds to build the PD-PPI network with PPI data from the integrated interactions database (IID) (Kotlyar et al., 2016). The proteins/nodes of this network were then used to focus the single-variant and polygenic risk score association tests of the WES results. While the isolated single variants failed in the replication cohort, the polygenic risk score led to a significant association with PD albeit with low predictive power.

7. Approach 3: use of PPI networks to study single PD gene interactomes

Another interesting use of protein interaction network analysis is applying PPI data to model the functional landscape proximal to PD Mendelian genes, to decipher their function in health and disease. Many studies have focused on LRRK2 as this is the major genetic player in familial PD. Although the precise role of LRRK2 remains to be elucidated, LRRK2 kinase inhibitors are currently under clinical investigation as potential PD therapeutic agents (Atashzam and Dzamko, 2016). Two papers modelling the LRRK2 protein interactome were independently published in 2015. The first publication (Porras et al., 2015), explored the utility of PPI information curated into the IntAct database (Orchard et al., 2014) and LRRK2 was used as a working example. The authors showed that the LRRK2 interactome is vast and likely associated with different pathways such as regulation of the cell cycle, intrinsic apoptosis, axon development, membrane trafficking, response to stress, and EGFR signaling (analysis obtained via Reactome pathway analysis (Jassal et al., 2020)).

In the second publication (Manzoni et al., 2015) we used PPI data from BioGRID (Oughtred et al., 2021) and IntAct (Orchard et al., 2014) to describe the extent of the LRRK2 interactome as principally associated with transport, vesicle trafficking and possible enzymatic regulation of the cytoskeleton (data obtained via Gene Ontology enrichment analysis (Gene Ontology Consortium, 2015)). LRRK2 was suggested to be a hub protein, whose interactome is likely to be cell type and developmental stage specific, thus accounting for the plethora of different functions that have been highlighted for LRRK2 in different experimental models. More recent updated insights into the LRRK2 protein interactome highlight the scale of PPI data being generated and available for this protein (Tomkins et al., 2018; Gloeckner and Porras, 2020).

Another interesting PD protein is alpha synuclein (αSyn), not only point mutations (Polymeropoulos et al., 1997) and gene copy number variants (Singleton et al., 2003) in the SNCA gene (coding for αSyn) are associated with different forms of PD, but deposition of misfolded αSyn is one of the pathological hallmarks of PD (Braak et al., 2003; Spillantini et al., 1997). Little is known about the role of αSyn in health and disease, although multiple functions have been proposed (Stefanis, 2012). Due to the genetic and pathological links of αSyn to PD, αSyn is also a potential target for therapeutic intervention, whereby research is focussed on mechanisms to reduce production, cell-to-cell transmission, aggregation and/or increase clearance of αSyn (Fields et al., 2019).

In a recent review on αSyn (Hernandez et al., 2020), the authors constructed and analysed the αSyn protein interactome using the Ingenuity Pathway Analysis software (marketed by Qiagen) which suggested roles for this protein in transcription, translation, protein folding, trafficking, secretion, mitochondrial functionality, and degradative pathways.

8. Approach 4: use of PPI networks to contrast and compare PD against other neurodegenerative disorders

Another use of PPI network analysis that has been explored utilises the network approach to compare the interaction and functional profiles of PD with other neurodegenerative disorders, to identify similarities and differences both in terms of functions or pathways associated with disease and specific nodes or network clusters unique to or shared across different conditions.
In 2014, a publication described a PPI network comparing 10 neurodegenerative diseases, including PD (Nguyen et al., 2014). Seeds were selected as disease associated genes as per the Online Mendelian Inheritance in Man (OMIM) database (Amberger et al., 2015), while PPI data were collected using the Interologous Interaction Database (12d) (Brown and Jurisica, 2005). Topological analysis of the networks led to the evaluation of the positions of the nodes in respect to the different diseases, to identify shared proteins most likely responsible for disease linkages. This first multi-disease analysis showed pleiotropy across the different diseases both in terms of molecular players and common pathological mechanisms, with a focus on the commonalities between PD vs FTD and PD vs ALS. The Toll-like receptor signaling pathway was suggested as a potential shared point of intervention across these neurodegenerative diseases.

In 2018, we published a comparison between PD and FTD performed via network analysis (Ferrari et al., 2018). We started by selecting Mendelian genes for PD and FTD as seeds for PPI network construction, using PPI data from IMEx associated databases, as per the WPPiNA pipeline. Topological and functional comparison between the PD and FTD networks showed that, even though the same general functions were associated with the 2 disorders, the granular details of those processes were quite different. Waste disposal was a communal process; however, the PD network was more associated with the autophagy-lysosomal pathway while the FTD network was more enriched in GO terms associated with the unfolded protein response (UPR) involving ubiquitin and the proteasome. Signaling functions were associated with development of the nervous system for FTD and the immune system in the case of PD. Finally, cell death related terms specifically related to DNA damage associated responses in FTD, while pointing towards mitochondria related cell death in PD. In 2020, a large project was published comparing the PPI network landscape of multiple neurodegenerative disorders: AD, PD, Huntington’s disease, ALS and spinocerebellar ataxia type 1 (Haenig et al., 2020). The PPI network was built based on yeast two-hybrid screening with data integration from literature mining and showed that predicted disease modules, contained both disease-specific interactions and PPIs commonly associated to multiple disorders, as well as an enrichment in aggregation-prone proteins.

9. Limitations and future directions

Protein interaction network analysis is one of the new frontiers of research providing the biomedical field with an in silico tool to study the molecular pathways of disease from a global viewpoint. A dynamic research scenario is forging ideas and testing new pipelines, and these efforts are starting to provide us with a sense of what we can achieve through the pairing of network modelling with genetics and functional research. These initial studies are also providing us with opportunities to critique these approaches and highlight elements of the pipelines for further development to aid future research.

If we consider the work done so far within the PD field, we can identify 3 major problems that need to be addressed within the development of the pipelines. The first problem is that of using classical graph theory to analyse PPI networks that are incomplete by definition as they rely on experimentally-derived PPI data and manual curation, thus some proteins might be understudied and some fields of research less curated than others. This is not a trivial issue, in that it can lead to misinterpretations of the results (de Silva et al., 2006). An example is the search for hubs, defined within classical graph theory as the nodes within the network with the highest number of connections. This measure is intrinsically biased in incomplete PPI networks and this can affect the interpretation of the results. We can consider 2 proteins, protein A with 10 known interactions and protein B with 1 known interaction; classical graph theory will highlight protein A as a hub. Since PPI networks are incomplete, this mathematically correct consideration might not have biological support. Maybe, in the real scenario, protein B has many more interactors, being more of a hub than protein A, however since protein B is understudied or protein B PPI data are not yet curated into databases used for these analyses, this information is not available in the network. This comment does not intend to reduce the relevance and importance of what can be done with topological analysis of PPI networks. Instead, it is aimed at raising awareness of required considerations for interpretation of classical topological analysis of PPI networks and at stressing the need for the development of an adjusted network theory to be more suitable in the PPI scenario. It was already suggested that “overly simple measures of network topology” might lead to bias and discrepancies in the results (Prinzl, 2011) and that such measures could benefit from the development of more complex graph theory methods where topology measures are associated with local properties and additional information to weight the protein interactions (Kanazoz et al., 2004; Nabieva et al., 2005). For example, in an alternative approach (i.e. local clustering hypothesis (Barabasi et al., 2011)) it has been observed that disease relevant nodes are closer to each other in comparison with other components of the graph (they have the tendency to cluster). Even if the PPI network is incomplete, these disease relevant clusters represent an informative structure that can be isolated. As a second example, in a study conducted by our group, we have introduced the concept of inter interactome hubs (I IHs), where instead of searching for nodes with the highest degree (or absolute number of connections) we search for nodes able to connect a de ned percentage of seeds (Ferrari et al., 2017).

The second problem is the definition of “PD genes”. Many PPI networks are, in fact, built based on a selection of PD genes used as seeds. However, the classification of a “PD gene” is not clear and currently it is often a personal, yet informed, decision by the authors. We clearly see, in the examples reported here, that different authors have made very different choices regarding the selection of “PD genes” to be used as seeds. Some authors have selected only 5 PD associated genes (George et al., 2019), while others broadened this de nition to include up to 36 PD genes (Siitonen et al., 2019). This clearly introduces a strong element of difference across all the PD-PPI networks that have been generated so far.

Finally, there is the compelling issue of cell and tissue speci city, a type of information that is not usually associated with PPIs as they are detected ex vivo or in generic model systems. There are tools that allow for filtering for gene expression in tissues and cell lines, however, these are post-analysis filters; the raw PPI data do not (usually) come with this information. Moreover, caution needs to be exercised when inferring protein abundance measures from gene expression data. The relationships between gene and protein expression are highly variable and dynamic.

A new frontier of PPI network analysis is that of combining PPIs with other –omics data. Thanks to recent technical advances, a valuable resource that is becoming available at large-scale is that of quantitative proteomics applied to large cohorts of genotyped patients and controls; for specific tissues and even at the single-cell level (Dyring-Andersen et al., 2020). Availability of large cohorts of high-quality proteomics and genomics data opens the possibility of pQTL analysis, a type of quantitative trait loci that is used to evaluate correlation between the presence of certain genetic markers (linked to disease risk) and changes in protein expression levels (Sun et al., 2018). Albeit the novelty of this approach, there are already examples in the literature describing how pQTL could be used both as a tool to fine map GWAS results thus leading to seed prioritization prior to PPI network analysis, as well as to filter PPI networks during analysis thus isolating the more genetically relevant clusters to infer disease-function connections (Johnson et al., 2021; Kibinge et al., 2020).

10. Conclusive remarks

With the increase in -omics data generation and availability, network
Neurobiology of Disease 155 (2021) 105395

7

J.E. Tomkins and C. Manzoni

Bonham, L.W., Steele, N.Z.R., Karch, C.M., Broce, I., Geier, E.G., Wen, N.L., Momeni, P., Chang, D., Nalls, M.A., Hallgrímsdóttir, B., Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., Hamosh, A., 2015. OMIM. studies in the years to come. are living in a thriving time for this type of bioinformatics research, and
pipelines, evaluating potential biases and testing new applications. We
there is a clear and substantial amount of work ongoing in this area, in
and functional research in neurodegeneration. Although the pipelines
Barabási, A.-L., Gulbahce, N., Loscalzo, J., 2011. Network medicine: a network-based
resource for mining gene and protein interaction data. Nucleic Acids Res. 46, D1049–D1056. https://doi.org/10.1093/nar/gku1179.

Gene Ontology Consortium, 2015. Gene ontology consortium: going forward. Nucleic Acids Res. 43, D1049–D1056. https://doi.org/10.1093/nar/gku1179.

George, G., Singh, S., Lokappa, S.B., Varkey, J., 2019. Gene co-expression network analysis for identifying genetic markers in Parkinson’s disease - a three-way comparative approach. Genomics 111, 819–830. https://doi.org/10.1016/j.genomics.2018.08.049.

Fields, C.R., Bengoa-Vergniory, N., Wade-Martins, R., 2019. Targeting alpha-synuclein as a approach to unraveling inherited axonopathies. Sci. Rep. 9, 1692. https://doi.org/10.1038/s41598-019-32902-4.

Di Fonzo, A., Chien, H.F., Socal, M., Giraudo, S., Tassorelli, C., Iliceto, G., Fabbriini, G., Marconi, R., Fincati, E., Abbruzzese, G., Marino, P., Squitieri, F., Horstwin, M.N., Montagna, P., Libera, A.D., Stochchi, F., Goldwurm, S., Ferreira, J.R., Meco, G., Martignoni, E., Lopiano, L., Hardjim, L.B.O., Osawa, B.A., Barbas, E.R., Bonam, V., 2007. ATP13A2 missense mutations in juvenileparkinsonism and young onset Parkinson disease. Neurology 65, 1567–1568. https://doi.org/10.1212/01.wnl.0000260903.07111.08.

Dobyn, A., Huckins, L.M., Boocock, J., Slomefman, L.G., Glucksberg, B.S., Giambartolomei, C., Hoffman, G.E., Perumal, T.M., Girdhar, K., Jiang, Y., Raj, T., Ruderfer, D.M., Kramer, R.S., Pinto, D., Abbruzzese, S., Rouns, P., Donini, R., Devlin, B., Sklar, P., Stahl, E.A., Sieberts, S.K., 2018. Landscape of conditional eQTL
in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. Am. J. Hum. Genet. 102, 1169–1184. https://doi.org/10.1016/j.ajhg.2018.04.011.

Dyring, G., Anderson, S., Vartdal, F., Severinsen, T., Hoff, M.J., et al., 2017. Weighted protein-protein interaction network analysis for identifying genetic markers in Parkinson’s disease using weighted protein-protein interaction network analysis. BMC Genomics 19, 452. https://doi.org/10.1186/s12862-018-4543-1.

Bennett, D.A., Corral, D.A., Kirou, T., Kuroki, M., Bocchini, C.A., Scott, A.F., Hamosh, A., 2012. Genes2FANs: connecting genes through protein-protein interactions and molecular pathways, upstream regulators, interactome domains, and novel targets in ALS. Sci. Rep. 2, 14732. https://doi.org/10.1038/srep02129.

Bennett, D.A., Corral, D.A., Kirou, T., Kuroki, M., Bocchini, C.A., Scott, A.F., Hamosh, A., 2012. Genes2FANs: connecting genes through protein-protein interactions and molecular pathways, upstream regulators, interactome domains, and novel targets in ALS. Sci. Rep. 2, 14732. https://doi.org/10.1038/srep02129.

Bennett, D.A., Corral, D.A., Kirou, T., Kuroki, M., Bocchini, C.A., Scott, A.F., Hamosh, A., 2012. Genes2FANs: connecting genes through protein-protein interactions and molecular pathways, upstream regulators, interactome domains, and novel targets in ALS. Sci. Rep. 2, 14732. https://doi.org/10.1038/srep02129.
Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., Stagljar, I., 2015. Fundamentals of
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N.,
de Silva, E., Thorne, T., Ingram, P., Agrafioti, I., Swire, J., Wiuf, C., Stumpf, M.P.H.,
Santiago, J.A., Bottero, V., Potashkin, J.A., 2019. Transcriptomic and network analysis
Shen, J., Chen, X.-C., Li, W.-J., Han, Q., Chen, C., Lu, J.-M., Zheng, J.-Y., Xue, S.-R., 2020.
Schaid, D.J., Chen, W., Larson, N.B., 2018. From genome-wide associations to candidate
J.E. Tomkins and C. Manzoni
Whiting, P.J., Blauwendraat, C., Whitworth, A.J., Manzoni, C., Ryten, M., Lewis, P.
Gandhi, S., Houlden, H., Cookson, M.R., Wood, N.W., Singleton, A.B., Hardy, J.,
Guelfi, S., Zhang, D., Pittman, A., Trabzuni, D., Pan, K.S., Kia, D.A., Bictash, M.,
Cirillo, E., Coort, S.L., Digles, D., Ehrhart, F., Giebertz, P., Kalafati, M., Martens, M.,
Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Evelo, C.T.,
Pico, A.R., Willighagen, E.L., 2018. WikiPathways: a multifaceted pathway database
Miller, R., Nishida, K., Rieswijk, L., Waagmeester, A., Eijssen, L.M.T., Evelo, C.T.,
Priestley, B., Fernandez, H.H., Cambi, F., Umbach, D.M., Blair, A., Sandler, D.P., Langston, J.W., 2011. Rotenone, parakariton, and Parkinson disease. Environ. Health Perspect. 119, 866–872. https://doi.org/10.1289/ehp.1002839.
Tomkins, J.E., Dinh, M., Bella, A., Ferrari, R., Iacqua, N., Cookson, M.R., Lewis, P. A., Manzoni, C. 2018. Comparative protein interaction network analysis identifies shared and distinct functions for the human ROCO proteins. Proteomics 18, e1700444. https://doi.org/10.1002/pmc.201700444.
Tomkins, J.E., Ferrari, R., Vavouzaki, N., Hardy, J., Lovering, R.C., Lewis, P.A., McGuffin, L.J., Manzoni, C., 2020. PINOT: an intuitive resource for integrating protein–protein interactions. Cell Commun. Signal. 18, 92. https://doi.org/10.1186/s12964-020-00554-5.
Trabzuni, D., Ryten, M., Walker, R., Smith, C., Imran, S., Ramasamy, A., Weale, M.E., Hardy, J., 2011. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. J. Neurochem. 119, 275–282. https://doi.org/10.1111/j.1471-4159.2011.07432.x.
Vavouzaki, N., Tomkins, J.E., Kara, E., Houlden, H., Hardy, J., Tindall, M.J., Lewis, P.A., Manzoni, C. 2021. Integrating protein networks and machine learning for disease stratification in the Hereditary Spastic Paraplegias. iScience, 102484. https://doi.org/10.1016/j.isci.2021.102484.
Vidal, M., Cestic, M.E., Barabási, A.-L., 2011. Interactome networks and human disease. Cell 144, 986–998. https://doi.org/10.1016/j.cell.2011.02.016.
Wang, J., Zheng, J., Wang, Z., Li, H., Deng, M., 2018. Inferring gene-disease association by an integrative analysis of eQTL genome-wide association study and protein–protein interaction data. Hum. Hered. 83, 117–129. https://doi.org/10.1159/000489761.