A Systems Biological Approach to Parkinson’s Disease

KATHARINA FRIEDLINDE HEIL
Abstract

Parkinson’s Disease (PD) is the second most common neurodegenerative disease in the Western world. It shows a high degree of genetic and phenotypic complexity with many implicated factors, various disease manifestations but few clear causal links. Ongoing research has identified a growing number of molecular alterations linked to the disease.

Dopaminergic neurons in the substantia nigra, specifically their synapses, are the key-affected region in PD. Therefore, this work focuses on understanding the disease effects on the synapse, aiming to identify potential genetic triggers and synaptic PD associated mechanisms. Currently, one of the main challenges in this area is data quality and accessibility.

In order to study PD, publicly available data were systematically retrieved and analysed. 418 PD associated genes could be identified, based on mutations and curated annotations. I curated an up-to-date and complete synaptic proteome map containing a total of 6,706 proteins. Region specific datasets describing the presynapse, postsynapse and synaptosome were also delimited. These datasets were analysed, investigating similarities and differences, including reproducibility and functional interpretations.

The use of Protein-Protein-Interaction Network (PPIN) analysis was chosen to gain deeper knowledge regarding specific effects of PD on the synapse. Thus I generated a customised, filtered, human specific Protein-Protein Interaction (PPI) dataset, containing 211,824 direct interactions, from four public databases. Proteomics data and PPI information allowed the construction of PPINs. These were analysed and a set of low level statistics, including modularity, clustering coefficient and node degree, explaining the network’s topology from a mathematical point of view were obtained.

Apart from low-level network statistics, high-level topology of the PPINs was studied. To identify functional network subgroups, different clustering algorithms were investigated. In the context of biological networks, the underlying hypothesis is that proteins in a structural community are more likely to share common functions. Therefore I attempted to identify PD enriched communities of synaptic proteins. Once identified, they were compared amongst each other. Three community clusters could be identified as containing largely overlapping gene sets. These contain 24 PD associated genes. Apart from the known disease associated genes in these communities, a total of 322 genes was identified. Each of the three clusters is specifically enriched for specific biological processes and cellular components, which include neurotransmitter secretion, positive regulation of synapse assembly, pre- and post-synaptic membrane, scaffolding proteins, neuromuscular junction development and complement activation (classical pathway) amongst others.

The presented approach combined a curated set of PD associated genes, filtered PPI information and synaptic proteomes. Various small- and large-scale analytical approaches, including PPIN topology analysis, clustering algorithms and enrichment studies identified highly PD affected synaptic proteins and subregions. Specific disease associated functions confirmed known research insights and allowed me to propose a new list of so far unknown potential disease associated genes. Due to the open design, this approach can be used to answer similar research questions regarding other complex diseases amongst others.