Parkinson’s Disease Gene Networks

Murray O. Robinson
Molquant, Inc.
Email address: murray@molquant.com

Abstract: Using recently developed analytical tools, we identified correlation-based networks among a set of 16 Parkinson’s Disease (PD) genes and linked these PD gene networks to associated biological networks. Two major correlated groups were identified, one containing PARK2, LRRK2, PINK1 and NOVA2, another containing SNCA, SYNJ1, RPE65 and MAPT. The PARK2 containing network exhibited correlation linkage to mitochondrial biogenesis through a PGC1a network, as well as to vesicle transport, lysosomal biogenesis and hypoxia networks. The SCNA containing network exhibited correlation linkage to axonal transport through axon motor networks anchored by KIF1A and MYO5A. A PABPC1L network exhibited linkage to an ATG4B seeded autophagy network; PARK7 linked directly to mitochondria through a network comprising mitochondrial respiratory complex genes. Core genes common to the member networks of the PARK2 containing (494 genes) the SNCA containing network (313 genes) resulted in candidate PD gene lists suitable for prioritization or polygenic analysis of PD GWAS. These results confirm and extend current hypotheses on biological processes affected by PD, identify distinct biologically related subsets of PD genes, link candidate PD genes to the identified networks, and provide proof of concept for the application of these correlation analysis tools to biological interpretation of genomics data.

1. Introduction

Parkinson's Disease (PD), a progressive neurodegenerative disease that affects several neuronal systems, typically presents as motor defects (rigidity/tremors/dyskinesias) resulting from the death of dopaminergic neurons in the substantia nigra (SN) of the midbrain. With a US incidence of approximately 60,000, roughly 1 million Americans are afflicted by the disorder. Approximately 10 million people are affected by PD worldwide [1]. Historically, PD was thought to be caused largely by environmental toxins, supported by the increased incidence of PD linked to certain pesticide exposure, as well as the rapid inducement of PD in individuals exposed to MPTP, a neurotoxic contaminant of synthetic opiate preparations. However, approximately 15% of PD exhibits familial association, pointing to genetic underpinnings.

In the last decade more than a dozen PD associated genes have been uncovered through a number of approaches: involvement in pathology (e.g. SNCA), positional cloning (e.g. PARK2, PINK1, PARK7), linkage to other genetic conditions (e.g. GBA) or next generation sequencing in family cohorts (SYNJ1, SMPD1 and others). These approaches to gene identification typically lack information regarding biological function or pathway involvement. Intensive effort has made great progress towards the elucidation of the biological underpinnings of PD, however, targeted therapeutics for PD.

To further uncover PD genes, particularly among the 85% of PD patients with sporadic disease, several Genome Wide Association Studies (GWAS) have been conducted to identify PD associated gene variants. A recent meta-analysis [2] and additional associated data has been compiled in the web resource PDGene.org [3]. The meta-analysis identified 11 loci with SNPs that met the p<1x10^{-8} significance criteria for association, five of which (GBA, MAPT, LRRK2, PARK16, SNCA) had been previously linked to familial PD. PDgene maintains a database of over 900 loci reported to be associated with PD in one or more of the GWA studies.

We sought to explore the biology and gene associations of PD using gene network analysis tools that we recently developed. We first used correlation analysis algorithms and a large database of expression data compiled from a number of sources to build PD gene correlation networks, and then assessed the networks for inter-network relationships. We next assessed the relationships between the PD gene correlation networks and a selected set of biological networks, which were also produced using correlation algorithms. Last, we conducted a search for additional candidate PD associated genes by examining the intersection of two PD gene networks and genes associated with loci identified in PD GWA studies.
2. Results and Discussion

2.1. Biological Networks

Using correlation algorithms applied to a large set of gene expression data, we developed a set of networks linked to biological processes. Three classes of biology are included in the networks: biochemical and signaling pathways (e.g. glycolysis), cellular structures (e.g. mitochondria) and cell type/tissue specific networks (e.g. skeletal muscle). To produce a biological network, we start with a set of relevant biologically characterized genes, and conduct multiple correlations to identify one or a few genes, which act as robust “seeds” for network generation. Correlation to the seed gene then defines each network. Except for the choice of the seed gene, networks are produced mathematically. Networks comprise different numbers of genes depending on correlation significance (typically several hundred genes), however, for convenience plots typically comprise only 10 genes. Figure 1 shows network and matrix “bathymetry” plots for 19 biologically defined networks. In the network plot (1A), 10 genes for each network are plotted as nodes, correlation represent the length of the edges (edges not rendered here). The bathymetry plot (1B) (a 2D matrix, so data are duplicated below the diagonal) shows correlation relationships among the 190 genes rendered as color and “height”. Networks include: vesicle transport, autophagy, axon motors, TFEB anchored lysosomal biogenesis, skeletal muscle, adipose tissue, ribosomes, PGC1α anchored mitochondrial biogenesis, mitochondria, glycolysis, mitosis, bone/cartilage, hypoxia/blood vessels, T cells, NK cells, CD68 macrophages, CD14 macrophages, neutrophils. Most of the plotted biological networks exhibit limited correlation to other networks, except the T cell, NK cell and macrophage networks, whose cells are known to share many genes in common.

2.2. Parkinson’s Disease Gene Networks

A review of literature and PD databases [4] resulted in the selection of 12 published genes in which genetic alterations were associated with risk of PD: ATP13A2, GBA, FBXO7, LRRK2, MAPT, PARK2, PARK7, PINK1, SMPD1, SNCA, SYNJ1, UCHL1. In addition, genes from an unpublished report identifying four additional genes: NOVA2, OR56B4, PABPC1L, RPE65 were included [5].

Each of the 16 genes served as the “seed” gene for network formation and the resulting network and bathymetry plots are shown in figure 2. Because the seed genes were chosen by their role in PD and not for robust biological correlation, the genes in each PD network do not always exhibit as high network correlation coefficients as the biological networks (note SNCA, SMPD1, FBXO7, OR56B4, GBA in figure 2B).

The networks are manually ordered along the diagonal to group networks with the highest degree of correlation. Correlations were observed among most of the networks (figure 2B), however, two distinct network clusters emerged from the analysis. The set of networks with the highest inter-network correlation included PARK2, LRRK2, PINK1 and the newly reported gene NOVA2. For convenience this will be referred to as the PARK2 meta-network. A second set of networks exhibiting robust inter-network correlation comprised SNCA, RPE65, SYNJ1 and MAPT (SNCA meta-network). Three other genes also exhibited correlation to these two clusters (FBXO7, SMPD1, and UCHL1), whereas PARK7, PABPC1L, OR56B4, GBA and ATP13A2 networks exhibited limited correlation linkage. Also, note that the two meta-networks exhibited correlation linkage to one another (figure 2B). The data used to generate these networks are drawn from many sources; we interpret these networks to...
represent the normal biology of these genes, not the pathologic state associated with PD.

2.3 Links between Parkinson’s Gene Networks and Biological Networks

To provide biological annotation of the Parkinson’s gene networks, biological networks and Parkinson’s gene networks were plotted together. Of more than 20 biological networks assessed, nine relevant networks were plotted in Figure 3. Consistent with the identification of meta-networks among the Parkinson’s Genes, each meta-network exhibited correlation to distinct biological networks.

As shown in Figure 3, the PARK2 meta-network exhibited correlation to four of the queried biological networks, a vesicle network seeded by VAMP2, a PGC1a (PPARGC1A) seeded network, a lysosomal biogenesis network seeded by TFEB, and a hypoxia network seeded by VEGFA. These correlations enable a hypothesis that the non-pathogenic versions of LRRK2, PARK2, PINK1 and NOVA2 normally function in intracellular processes that control the inte-
Integration of nutrient and oxygen sensing with the cellular machinery that controls mitochondrial homeostasis. The fact that the pathogenic forms of these genes are directly linked to disease suggests that the linked biological pathways may be altered in PD.

The observations and hypothesis are consistent with several previous findings in the field. Links have recently been reported connecting oxygen sensing machinery (represented as hypoxia here) and mitochondrial biogenesis [6, 7]. PARK2 and PINK1 have been shown previously to function in the control of mitochondrial homeostasis, including motility [8, 9], as has FBXO7 [10], which here exhibits modest correlation linkage to the PARK2 meta-network. The transcriptional co-activator PGC1α/PPARGC1A is thought to be a central controller of this process [11, 12]. PGC1α variants have been linked to pathological retinal angiogenesis [13]. Neuropathy phenotypes in the PGC1α knockout mouse have led to interest in PGC1α modulation in neurodegenerative diseases [14]. SNPs near PGC1α have been associated with age of onset in PD [15]. In addition, meta-analysis of expression data from Parkinson’s affected tissues identified a PGC1α signature as one of the top altered pathways [16]. Another network identified here is one seeded by Transcription Factor EB (TFEB), a master regulator of lysosomal biogenesis [17]. TFEB has been linked to PGC1α in Huntington’s disease [18] in work suggesting that up-regulation of these linked pathways was protective in Huntington’s.

This analysis further links PGC1α and these PD genes to biological processes that are not directly related to the mitochondrion per se (note the limited correlation linkage to integral mitochondrial genes), rather to components and regulators of intracellular organelle trafficking. LRRK2 in particular, exhibits tight correlations to several cytoskeletal, vesicle/intracellular transport genes (COL4A3, COL4A4, SYNE1, ARHGAP24,31, GPRASP1, LMBRD1, SLC6A13, STX12, TRAK2/Milton), suggesting a role in modulation of intracellular transport. A recent paper identified RPS11, RPS15 and other ribosomal subunits as LRRK2 substrates [19]. Here, a tightly correlated ribo-

![Figure 3. “Bathymetry plot” showing relationships among sixteen Parkinson’s Disease networks and identified biological networks. The ten top genes from each network are plotted, labels are superimposed on the ten core genes of each network. Correlation between each gene is indicated by color and height.

Figure 3. “Bathymetry plot” showing relationships among sixteen Parkinson’s Disease networks and identified biological networks. The ten top genes from each network are plotted, labels are superimposed on the ten core genes of each network. Correlation between each gene is indicated by color and height.
omosomal network anchored by RPS8 that includes RPS11 and RPS15 exhibited no correlation link to LRRK2 or other PD gene networks.

Also consistent with the hypothesis, the Parkinson’s associated lysosomal gene SMAD1 also exhibits linkage to the PARK2 meta-networks. Last, this analysis links the RNA binding protein NOVA2 to the network. NOVA2 has previously been linked to the Disabled/Reelin neuronal development pathway through control of Disabled RNA splicing [20]. Recent reports have linked the reelin pathway to control of neuronal angiogenesis and lymphatic vessel biology [21-23]. Consistent with these reports, note that here NOVA2 exhibits the strongest linkage to the hypoxia network (figure 3).

The SNCA meta-network also exhibited correlation to PPARGC1A (PGC1a) network, but the strongest linkages are to cytoskeletal motor protein networks: either the kinesin motor KIF1A, or the myosin motor gene myosin 5A (MYO5A). The RPE65 network exhibited tightest linkage to the MYO5A motor network, whereas SYN1, a known synaptic vesicle protein exhibited tightest linkage to the KIF1A network; SNCA shared linkages with both motor networks, whereas the microtubule associated protein MAPT exhibited best linkage to RPE65 rather than the motor networks. Although not included in the SNCA meta-network, ubiquitin hydrolase UCHL1 also exhibits linkage with both RPE65 and KIF1A networks. Together, these observations further support a hypothesis that these Parkinson’s genes normally function as components of and regulators of axonal transport, which is intimately linked to mitochondrial homeostasis [24].

The other queried Parkinson’s genes exhibit distinct linkage correlations. PABPC1L, an RNA binding protein of unknown function, and to a lesser extent, ATP13A2 exhibit linkage to an autophagy network seeded by ATG4B. PARK7 exhibits tight linkage to the integral mitochondrial protein network anchored by COX6A1 (as does the PD gene candidate HTRA2 (data not shown)). Olfactory Receptor OR56B4 doesn’t exhibit significant linkage to any networks analyzed. Although not shown here, PD associated gene VPS35 exhibits weak linkage to SYN1J, and PD gene PLA2G6 doesn’t exhibit significant linkage to any networks analyzed (data not shown). Further analysis exploring a broader set of biological networks may reveal additional linkages.

2.4 PGC1a, a potential protective network in some genetically defined forms of PD

Consistent with the previously reported observation that a PGC1a associated signatures are altered in PD [16], this analysis hypothesizes a role for PGC1a in biological processes associated with some of the identified PD genes. The PGC1a (PPARGC1A) network exhibits linkage to the both the PARK2 and SNCA meta-networks (figure 3). Although several lines of evidence indicate that PGC1a is associated with mitochondrial biogenesis, networks here proliferatory networks (those comprising mitochondrial genes) and PGC1a associated networks. PGC1a exhibits linkage to components of vesicle transport, including lysosomal structures, cytoskeletal motors as well as sensors of oxygen concentration (hypoxia). Previous reports have suggested a direct link between PGC1a overexpression and up-regulation of intrinsic mitochondrial genes, however, others have noted the inherent difficulty in distinguishing between direct and indirect effects of PGC1a on gene expression [7]. The network analysis shown here postulates that PGC1a’s impact on mitochondrial biogenesis occurs through its modulation of these associated transport and sensor pathways. While enforced over-expression of PGC1a in neurons by genetic means has had varied results in PD models [25, 26], PGC1a can be up-regulated through endogenous mechanisms by the thiazolidinedione class of diabetes drugs, which appear to activate PGC1a through AMP Kinase (AMPK) [12, 27]. Development of an AMP Kinase network was not feasible using Molquant tools. The seven identified AMP Kinase genes (two alpha, two beta, three gamma subunits) exhibit little correlated expression, nor did key biological pathway genes, precluding selection of a seed gene. Based on previous links between the thiazolidinedione drugs, PGC1a and PD, a phase 2 trial was initiated in 2011[28]. Given the association of PGC1a pathways with specific subgroups of PD genes, assessment of the genotype of patients in such a study may provide additional information regarding genotype/efficacy relationships.

The recently reported drug candidate R118, a potent AMPK activator (paper on related molecule [29]) just entered the clinic, intended for treatment of peripheral artery disease. If this molecule achieves adequate exposure and tolerability in human studies, it may represent an intriguing candidate for PD.

2.5 Parkinson’s network gene lists

One challenge to establishing PD genotype/efficacy relationships for any study is the limited number of known genes contributing to PD. It is estimated that the known genes contribute to approximately 30% of familial and 5% of sporadic PD [30]. More comprehensive cataloguing of PD associated genes could greatly advance both the management of, and potentially the treatment of PD.

In addition to providing tools for biological interpretation, the Parkinson’s gene networks identified here comprise candidate PD gene lists suitable for exploration of current and future GWAS analysis, or for targeted genotyping studies.

Here, only examining genes of the two meta networks, figure 4 shows the overlap of the top 1000 network genes for each PD gene from either the PARK2 or SNCA meta-network. The identified genes are those that correlate with at least three of the four PD genes in each network, and therefore hypothesized to participate in the same biological processes.
An excel file of the genes is available for download (http://media.virbcdn.com/files/9b/876744869e90d886-PA
RK2SNCAmetaoverlap.pdf).

The PD Gene database is a curated and broadly inclusive list of genes and loci that have been linked to PD. Intersection of the PARK2 meta-network (494 genes) or the SNCA meta-network (313 genes) with 723 genes identified in the online database resulted in 35 genes that are present both in the PARK2 and/or SNCA meta-networks and the PDgene database. (Of 915 genes or loci in the PDgene.org database, 723 genes were also found in the Molquant datasets). The 35 genes are indicated in the supplementary table. Eleven are discussed further below.

2.6 Parkinson’s gene candidates from network/GWAS overlap

Of 35 genes overlapping with the PARK2 or SNCA meta-networks, four have been highlighted elsewhere as significant or nearly significant in PD GWAS association: RIT2 [31], SYT11 [2], ELAVL4 [32], RFX4 [33]. Table 1 highlights seven additional genes identified in the network overlaps, chosen based on p value of SNP correlation (<1X10^-5) in GWAS studies listed in GWAS central (gwascentral.org). Given the robust network linkages of these 11 genes, combined with their GWAS associations, these may represent strong candidates for further exploration into genetic susceptibility to Parkinson’s Disease.

2.7 Chromosome 1 PD locus gene S1PR1 present in PARK2 meta-network

A stratified analysis of PD GWAS data identified a previously unrecognized PD locus on Chr 1 that included candidate genes DPH5, OLFM3, S1PR1, SLC30A7, VCAM [34]. We note that the sphingosine phosphate receptor S1PR1 is included in the PARK2 meta-network. S1PR1 expression also exhibits correlation to MAPT and SYNJ1 networks as well, providing a strong network support to hypothesize that S1PR1 may represent the relevant affected gene for the locus. S1PR1 is widely expressed, exhibiting multifunctional roles including lymphocyte trafficking (the target of the anti-MS drug fingolimod) and angiogenesis. The network linkages observed here support a role for S1PR1 in PD gene associated function, including the previously mentioned hypoxia network.

3. Limitations

3.1 Only transcript level regulation is captured

Although transcript level regulation is widespread, it represents only one of many ways to regulate biological processes. Genes that are widely expressed and exhibit a high degree of post-translational regulation may not be captured in these networks. Nevertheless, our observations have been that, given a large enough sample size with sufficient biological complexity, i.e. many diverse samples, we can obtain networks exhibiting tight correlation of genes comprising well known biological processes (glycolysis, mitochondria, ribosomes, skeletal muscle, adipose tissue, taste receptor cells and others). However, several other attempts to generate correlation based biological networks were unsuccessful (circadian rhythm genes, Notch pathway genes, retina specific genes, nonsense mediated RNA decay and others). Certainly these networks will not comprehensively capture all of the component genes of a particularly biology, but the observations presented here argue that this approach represents a useful tool to

Table 1

<table>
<thead>
<tr>
<th>gene</th>
<th>SNPs &lt;1x10^-5</th>
<th>P value of best SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLG2</td>
<td>9</td>
<td>7x10^-9</td>
</tr>
<tr>
<td>C10ORF32</td>
<td>3</td>
<td>4 x10^-3</td>
</tr>
<tr>
<td>FYN</td>
<td>6</td>
<td>5 x 10^-4</td>
</tr>
<tr>
<td>NEGR1</td>
<td>4</td>
<td>2 x 10^-5</td>
</tr>
<tr>
<td>MRAS</td>
<td>4</td>
<td>9 x 10^-5</td>
</tr>
<tr>
<td>NFIA</td>
<td>3</td>
<td>8 x 10^-4</td>
</tr>
<tr>
<td>SNTG1</td>
<td>7</td>
<td>1 x 10^-3</td>
</tr>
</tbody>
</table>

*Table 1: PD SNP associations of selected PARK2/SNCA meta-network genes*
analyze and interpret complex biology.

3.2 Untangling co-occurring biological processes

Many biological processes, although distinct, occur in the same tissues or conditions. If an uncharacterized gene correlates with two (or more) biological networks, it is difficult to assign it to a specific process. For example, although we recognize proteasomal degradation, splicing and DNA replication to be distinct biological processes, networks comprising core genes from each of these functions exhibit a high degree of cross correlation in our datasets (data not shown). Uncharacterized gene C20ORF24 correlates highly with networks seeded by proteasomal gene PSMA5, pre-replication complex gene MCM10, and splicing factor SNRPB, precluding assignment to any of those processes. The co-regulation of these three networks probably tells an interesting story itself, but this analysis merely identifies the phenomenon, not its resolution.

3.3 Data source bias

Although more than 7000 individual profiles were explored to derive the correlations shown here, the networks identified are subject to the composition of the samples included. Many genes are known to exhibit distinct networks in different tissues, so the correlations observed here likely reflect what is occurring in the dominantly represented tissue(s) or cells. For example, PGC1α may exhibit different correlations in brown adipose tissue or cardiac cells or neurons. Furthermore, our examination of network correlations in large subsets of the overall dataset (e.g. 1600 human tissue RNAseq samples (GTEXportal.org)) demonstrates that while many networks are well replicated in the subsets, other networks are not well recapitulated. We look forward to the rapidly increasing number of profiled samples to improve the networks.

References


