Repositioning drugs by targeting network modules: a Parkinson’s disease case study

Zongliang Yue1,2, Itika Arora2, Eric Y. Zhang2, Vincent Laufer3, S. Louis Bridges3 and Jake Y. Chen1,2,4*

Abstract

Background: Much effort has been devoted to the discovery of specific mechanisms between drugs and single targets to date. However, as biological systems maintain homeostasis at the level of functional networks robustly controlling the internal environment, such networks commonly contain multiple redundant mechanisms designed to counteract loss or perturbation of a single member of the network. As such, investigation of therapeutics that target dysregulated pathways or processes, rather than single targets, may identify agents that function at a level of the biological organization more relevant to the pathology of complex diseases such as Parkinson’s Disease (PD). Genome-wide association studies (GWAS) in PD have identified common variants underlying disease susceptibility, while gene expression microarray data provide genome-wide transcriptional profiles. These genomic studies can illustrate upstream perturbations causing the dysfunction in signaling pathways and downstream biochemical mechanisms leading to the PD phenotype. We hypothesize that drugs acting at the level of a gene expression module specific to PD can overcome the lack of efficacy associated with targeting a single gene in polygenic diseases. Thus, this approach represents a promising new direction for module-based drug discovery in human diseases such as PD.

Results: We built a framework that integrates GWAS data with gene co-expression modules from tissues representing three brain regions—the frontal gyrus, the lateral substantia, and the medial substantia in PD patients. Using weighted gene correlation network analysis (WGCNA) software package in R, we conducted enrichment analysis of data from a GWAS of PD. This led to the identification of two over-represented PD-specific gene co-expression network modules: the Brown Module (Br) containing 449 genes and the Turquoise module (T) containing 905 genes. Further enrichment analysis identified four functional pathways within the Br module (cellular respiration, intracellular transport, energy coupled proton transport against the electrochemical gradient, and microtubule-based movement), and one functional pathway within the T module (M-phase). Next, we utilized drug-protein regulatory relationship databases (DMAP) and developed a Drug Effect Sum Score (DESS) to evaluate all candidate drugs that might restore gene expression to normal level across the Br and T modules. Among the drugs with the 12 highest DESS scores, 5 had been reported as potential treatments for PD and 6 hold potential repositioning applications.

(Continued on next page)
Conclusion: In this study, we present a systems pharmacology framework which draws on genetic data from GWAS and gene expression microarray data to reposition drugs for PD. Our innovative approach integrates gene co-expression modules with biomolecular interaction network analysis to identify network modules critical to the PD pathway and disease mechanism. We quantify the positive effects of drugs in a DESS score that is based on known drug-target activity profiles. Our results illustrate that this modular approach is promising for repositioning drugs for use in polygenic diseases such as PD, and is capable of addressing challenges of the hindered gene target in drug repositioning approaches to date.

Background
Parkinson’s Disease (PD) is a disorder characterized by depletion of dopamine in the basal ganglia, including the substantia nigra. While the exact etiology of PD is unknown, major advances have been made in understanding underlying disease mechanisms through technologies in genetics, transcriptomics, epigenetics, proteomics and imaging [1]. These advances have increased recognition of the heterogeneity and etiological complexity of PD as a disease. Nevertheless, there is hope for broad-spectrum therapeutic intervention, as even distinct disease subtypes implicate genes intersecting in common pathways [2]. Recently described “Network Medicine” [3] approaches offer a platform to study the molecular complexity of a particular disease systematically. These approaches are well-suited to the identification of disease modules and pathways as well as the molecular relationships between apparently distinct phenotypes [4]. Despite progress towards understanding of genetic factors that contribute to the etiology of PD, current treatments are aimed at clinically apparent PD — after patients are suffering from the onset of neurodegeneration. While, preventative drugs aim at treatment before or during the pre-clinical stage of PD are lacking, as are curative drugs aimed at the underlying molecular mechanisms have had limited success [5].

The associations discovered in GWAS of PD allow for the identification of disease-specific modules playing a role in triggering the disease. Similarly, gene expression microarray data provides a gross overview of gene expression changes that are associated with diseases like PD. However, future studies of complex diseases will need to move beyond the analysis of single genes and include analysis of interactions between genes or proteins, in order to better understand how functional pathways and networks become dysfunctional [6]. For instance, network-based approaches have already been used to examine various disease molecular mechanisms, e.g., type-2 diabetes [7], cancer [8], and neuronal degeneration specifically [9]. Bioinformatics techniques to characterize network topology and functional modules have been developed recently for functional genomics [10]. The identification of disease modules involving specific mutated genes and the molecular pathways to which they belong will provide new targets for drug development. GWAS and whole exome profiling data are combined in systems biology to illustrate upstream perturbations causing dysfunction in pathways and mechanisms leading to the disease phenotype. Therefore, we introduce the approach of discovering disease-specific modules to reveal the etiology of PD.

In this study, we hypothesize that study of PD GWAS [11] and co-expression data [12] will enable identification of disease-specific modules caused by a variation in multiple components of a functional pathway or network. Thus, we propose using a network-based approach called Weighted Gene Co-expression Network Analysis (WGCNA) [13] to detect modules of co-expressed gene networks associated with PD. We then integrate these co-expression clusters with gene regulatory network information and perform enrichment analysis to find PD-specific modules. This method, in combination with functional enrichment and network topology measures, will be used to identify potential targets. This is done by selecting drugs that reverse the altered gene expression signatures found within the PD modules.

PD modules which show significant perturbation is identified by comparing global co-expression networks in PD to regulatory networks identified using GWAS ‘hits’. After selecting the PD-specific modules for further analysis, we find significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology terms associated with PD modules. Afterward, we use knowledge of these functional pathways as the basis for modular drug discovery—the discovery of drugs that act on many nodes within the disease-specific module. This is accomplished through our innovative Drug Effect Sum Score (DESS) system and then cross-validated through rigorous analysis of published literature.

Methods
An overview of the framework
The pipeline is divided into two color-coded sections as shown in Fig. 1. The first section (colored red) contains steps for construction of PD modules, and the second section (colored green) contains steps to perform modular drug repositioning. The construction of PD modules was carried out in 6 steps: 1) We filtered genes with significant expression changes between the case vs. control
samples (with the False Discovery Rate set to 0.05), using a Bayesian inference technique available in the limma package in R [14]. 2) We performed WGCNA, which yields clusters (modules) of highly correlated genes having significant changes across three tissues. 3) We compiled PD-specific GWAS candidate genes and performed one layer extension to generate a gene regulatory network by retrieving the gene-gene regulatory relationship from the PAGER database [15]. 4) We performed enrichment analysis by finding overlapping genes shared between co-expression clusters and GWAS candidate genes, extracting these enriched clusters as PD-specific modules. 5) We constructed PD-specific network module by retrieving the gene-gene interactions for the genes in PD-specific modules from the HAPPI-2 database [16]. 6) Finally, we annotated PD-specific modules with functional groups using ClueGO [17].

The Drug repositioning section (green) was comprised of four steps. First, we calculated a $P$-score, which is an intuitive pharmacology score that combines the probability for each interaction and the weight of the drug-target interaction using data from the DMAP database (see details in Methods). Second, we calculated the $RP$-score, which is a measure of Relevant Protein importance in the PD modules network (see details in Methods). Third, we calculated the Drug Effect Score ($DES$) of each module. Finally, the $DESS$ was calculated across all modules. Using these steps, we obtained a ranked “modular drug list” consisting of candidate treatments based on PD-specific modules.

**Preparation of PD-specific omics, gene-gene interaction, and drug-protein regulation data**

Datasets from whole genome expression transcriptional profiling (on the GSE8397-GPL-96 array) were retrieved from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GS E8397). In the gene expression profile, 47 samples from PD patients and controls were used in three brain regions: the Frontal Gyrus (FG: 8 tissue samples), Lateral Substantia (LS: 16 tissue samples) and Medial Substantia

![Diagram](image-url)
MS: 23 tissue samples) [12]. SNP data was obtained from a PD paper [11], in which a GWAS was carried out. We mapped probe IDs to gene symbols using the NCBI microarray toolkit and assigned gene expression scores by the averaging probe expression values after adjustment and trimming of background noises by using the standard deviation of the mean values from all samples. Since the standard deviation of the mean values were small enough (0.02 in this study), no samples had been trimmed. After performing probe transformation and synonymous gene merging on data from the Affymetrix Human Genome U133A Array [HG-U133A] and Affymetrix Human Genome U133B Array [HG-U133B], 12,995 genes were mapped by 22,283 probes in the merged matrix from the two arrays. In the prior study, 54 genes were reported as having had significant enrichment [11] in GWAS. The PAGER database [15] was used to obtain gene-gene regulatory relationships (22,127 pairs curated from 645,385 in total). The HAPPI-2 database [16] was used to obtain protein-protein interaction (PPI) data. This integrated protein interaction database comprehensively integrates weighted human protein-protein interaction data from a wide variety of protein-protein database sources. After mapping the proteins to genes using UniProt IDs, we obtained 2,658,799 gene-gene interactions. The drug-target regulatory relationships data was from the DMAP database [18], which consisted of curated 438,004 drug-protein regulatory relationships.

**PD-specific network module identifications**

Whole-genome expression data on 12,995 genes was filtered down to 2895 candidate genes, based on a multigroup empirical Bayesian (eBayes) moderated t-test with p-value ≤ 0.05. Next, we performed WGCNA to cluster these genes based on their co-expression. To do this, we first performed our pipeline steps to identify excessive missing values and outlier microarray samples. The detection of the outlier was performed by trimming the hierarchy tree of average Euclidean distance method using cutoff tree height of 100. Second, we chose an exponent for soft thresholding based on analysis of network topology, to further reduce noise and amplify stronger connections in the scale-free topological model. Third, we performed one-step network construction and module detection using hierarchy tree of unsigned TOM-based dissimilarity distance. Fourth, we visualized the genes in modules in a hierarchy tree based on average linkage clustering [13]. Fifth, we analyzed the cluster (Principal Components) and sample (expression data) correlation using Pearson correlation and asymptotic p-value.

An initial regulatory relational network was seeded using the 54 candidate genes identified by Moran et al. and expanded using the gene-gene regulatory relationship data. The resulting expanded regulatory relational network consists of 288 genes and 1983 gene-gene regulatory relationships. Subsequently, we performed the

![Fig. 2 An example of calculating the DESS for PD-specific gene expression modules. Green indicates increased gene expression, red indicates a decrease. Note that the drug action acts to reverse the direction of gene expression found in the pathological state. Exp. stands for expression value, RP-score stands for the protein relevant score, p. stands for the priority score, P-score stands for the intuitive pharmacology score, DES stands for the Drug Effect Score, module f(ptx) stands for the enrichment score, and the DESS stands for Drug Effect Sum Score](image)
enrichment testing of the genes in the expanded regulatory relational network to measure enrichment in the co-expression clusters using the hypergeometric test and assigned the $f(pts)$ score using the formula:

$$f(pts) = \text{sign} \left( \frac{K - k}{N - n} \right) \times \log \left( \frac{\binom{K}{k}}{\binom{N}{n}} \right)$$

where $N$ is the total number of the genes in co-expression clusters, $K$ is the number of overlapping genes between co-expression clusters and genetic candidate genes, $n$ is the number of the genes in the co-expression cluster selected, and $k$ is the overlap genes between selected co-expression cluster and the genetic candidate genes. A positive value for $f(pts)$ indicates the over-representation in the expanded regulatory relational network. PD-specific modules were defined as the over-represented co-expression clusters. We then generated the network of PD-specific modules by applying high-confidence gene-gene interactions (as indicated by 3-star or above in the HAPPI-2 database).

In the final step, we performed ClueGO analysis to elucidate mechanisms involved in the PD-specific modules. We applied Bonferroni correction and selected those with post-correction $p$-value $\leq 0.05$ and Kappa score $\geq 0.5$ (moderate network strength or stronger) [17].

**Modular drug repositioning**

DESS was calculated using the $P$-score from the DMAP database, the $RP$-score from the PD modules, and the

---

**Fig. 3** The expanded regulatory relational network generated. The color of the nodes indicates the direction of change of expression; red nodes indicate the up-regulated genes, while green nodes stand for the down-regulated genes. Nodes in gray were not assayed by our whole-genome transcriptional profiling. The color scale measures the expression changes accumulated from the three brain regions.
module enrichment score \( f(p|t) \) from the PD modules. We calculated a \( P\)-score, an intuitive pharmacology score in the DMAP database, via a probability-weighted summary of all the evidence mined from literature or other drug target databases to determine an overall mechanism of “edge action” for each specific chemical-protein interaction using \( \text{conf}(d,p) \):

\[
\text{conf}(d,p) = \sum_{i=1}^{N} (\text{prob}_i(d,p) \times \text{sign}_i)
\]

(2)

where \( d \) and \( p \) are specific drugs and proteins, respectively. \( N \) is the number of types of evidence for the interaction between \( d \) and \( p \). \( \text{prob}_i(d,p) \) is confidence in each type of evidence \( i \) with a value within the range of \([0,1]\). \( \text{sign}_i \) has a value of 1 if the evidence \( i \) suggests activation and a value of -1 if the evidence \( i \) suggests inhibition. Afterwards, to rank each interaction, we used the algorithm in HAPPI [19] by assigning a \( \text{weight}(p) \) for the proteins interacting with each drug using the following formula adapted from [20].

\[
\text{weight}(p) = k \times \ln\left(\frac{\sum_{p,q \in \text{NET} \ merely}\text{conf}(p,q)}{\sum_{p,q \in \text{NET} \ merely} N(p,q)}\right)^{-1}
\]

(3)

Here, \( p \) and \( q \) are proteins in the protein interaction network, \( k \) is an empirical constant (\( k = 2 \) in this study), \( \text{conf}(p,q) \) is the confidence score of interaction between protein \( p \) and \( q \) assigned by HAPPI-2, and \( N(p,q) \) holds the value of 1 if protein \( p \) interacts with \( q \) or the value of 0 if protein \( p \) does not interact with \( q \). Thus, the foregoing probabilities and weights for each interaction were combined into \( P\)-score\((d,p) \), which includes both information on each drug’s effects on interacting proteins and the importance of the protein in the protein-protein interaction network:

\[
P\text{-score}(d,p) = \text{conf}(p,q) \times \text{weight}(p)
\]

(4)

We applied each gene’s \( RP\text{-score} \) calculation in a manner similar to formula (3) in PD-specific modules using the formula:

\[
RP\text{-score} = e^k \cdot \ln\left(\frac{\sum_{p,q \in \text{NET} \ merely} \text{conf}(p,q)}{\sum_{p,q \in \text{NET} \ merely} N(p,q)}\right)^{-1}
\]

(5)

where \( p \) and \( q \) are the indexes of proteins from the selected module, \( k \) is a constant (\( k = 2 \) in this study). The term \( \text{conf}(p,q) \) is the interaction confidence score assigned by HAPPI-2, where \( \text{conf}(p,q) \in [0,1] \).

Further, we calculated a \( DES(d,m) \) by using the drug weight score and the module gene \( RP\text{-score} \) according to the formula:

\[
\text{DES}(d,m) = \sum_{i=\text{inTarget}}^n \left[ \text{sign}_i \times \log_2 (p\text{-score}(d,i)) \right] \times \log_2 (\text{RP}\text{-score}(d,i)) \times p_i
\]

(6)

where \( m \) is the module, \( i \) is the index of the proteins in the PD-specific module, \( \text{sign}_i \) is the direction of the effect drug \( d \) on protein expression, and \( P\text{-score}(d,i) \) is the pharmacology score of the drug \( d \) to target \( i \), \( p_i \) is the priority score which indicates the source of the candidate. We assigned a value of \( p_i=1/2^0 \) when the candidates were from GWAS, \( p_i=1/2^1 \) when the candidates were from regulatory one-layer extension of GWAS, and \( p_i=1/2^2 \) when the other candidates were from the same module.

---

**Fig. 4** The WGCNA analysis of the five co-expression modules - Brown (Br), Yellow (Y), Blue (Bl), Green (G), and Turquoise (T). The dendrogram illustrates the degree similarity using hierarchy tree of TOM-based dissimilarity distance in each module cluster, which forms the basis for subsequent functional pathway identification.
The $DESS(d,M)$ was calculated by integrating all PD-specific modules according to the expression:

$$DESS(d,M) = \sum_{m \in M} DES(d,m) \times f_m(pts)$$  \hspace{1cm} (7)$$

where $M$ is the module set of module $m$, $f_m(pts)$ is probability mass function (pmf) transform score of the PD-specific module $m$. An example of how the $DESS$ score is calculated for a drug is shown in Fig. 2. Based on the total $DESS$, modular drugs (drugs selected based on their predicted effect at a module level) and their targets in the modules were collected. We pulled out modular drugs or drugs selected based on their predicted effect at a module level alongside their associated targets. Finally, we applied a single regulatory layer expansion and retrieved drug-target regulatory relationships (DMAP database) and protein-protein interactions (HAPPI-2 database) to generate the "extended modular drug-target network".

**Results**

**Construction of PD genetic association networks**

The PD genetic association network was constructed using the neighborhood extension method. Starting from the original 54 genes identified using GWAS (described in Materials and Methods, above), we obtained PD genetic association networks consisting of a total of 288 genes and 1983 regulatory relationships. The candidates of significant expression change (eBayes moderates t-test $p$-value $\leq 0.05$) are colored in the PD genetic association networks provided in Fig. 3.

**PD-specific network modules identified**

The details of the gene co-expression network construction with WGCNA have been previously described [13]. By applying the steps described above in Materials and Methods, 5 co-expression modules were identified. We color-coded these as the Brown (Br) module, the Yellow (Y) module, the Blue (Bl) module, the Green (G) module and the Turquoise (T) module, all of which are shown in Fig. 4. The number of genes in each module is as follows: the Br module containing 544 genes, the Y module containing 199 genes, the Bl module containing 821 genes, the G module containing 116 genes, and the Turquoise containing 1190 genes.

**Enrichment analysis results of two PD-specific network modules**

Based on the enrichment analysis, we identified two PD-specific modules (the T module and the Br module) shown in Table 1 and Fig. 4. The genes in these modules as displayed in the dendrogram are grouped tightly enough to be susceptible to a modular drug (a drug that acts on many members of the PD-specific module rather than on one target). 2895 genes are included in the gene co-expression modules. 50 of these genes in the co-expression modules overlapped with genes identified from our analysis of genetic data. These 50 genes are distributed among the modules as follows: 21 in the T module, 10 in

| Module | Genes in all | Candidate | Module | Module candidate | f(pts) | Rank |
|--------|-------------|-----------|--------|------------------|--------|------|
| Turquoise | 2895 | 50 | 1190 | 21 | 0.94 | 1 |
| Brown | 2895 | 50 | 544 | 10 | 0.86 | 2 |
| Green | 2895 | 50 | 116 | 2 | -0.55 | 3 |
| Yellow | 2895 | 50 | 199 | 3 | -0.65 | 4 |
| Blue | 2895 | 50 | 821 | 13 | -0.92 | 5 |

**Table 1** The 5 co-expression module enrichment based on GWAS results
Fig. 6 Global view of the protein-protein interaction network of the 2 modules. a. The Br module consists of 449 genes and 2373 gene-gene interactions. b. The T module consists of 905 genes and 5156 gene-gene interactions. The nodes in red color are up-regulated and the nodes in green color are down-regulated.
the Br module, 2 in the G module, 3 in the Y module, and 13 in the Bl module. Using the hypergeometric test, we identified two PD-specific modules (modules having positive $f(pts)$, see Methods) the T module, which had $f(pts) = 0.94$ and the Br module, which had $f(pts) = 0.86$. Figure 5 illustrates the correlation of gene expression to case-control status. Specifically, the Pearson correlation coefficient for the expression level of the genes belonging to each module was reported for each sample. Overall, cases and controls are well discriminated by the gene expression signature of the genes in the module. For instance, in the Br module, control samples have a positive correlation with modular gene expression, while disease samples are negatively correlated with gene expression of genes found in each module. The remaining relationships

![Diagram A: Gene Ontology - biological processes (GO-BP) relating to each PD-specific module as identified by ClueGO analysis.](image)

![Diagram B: Gene Ontology - biological processes (GO-BP) relating to the two PD-specific modules](image)

**Table 2** Gene Ontology - biological processes (GO-BP) relating to the two PD-specific modules

| Module         | Function                                                                 | Groups | Gene numbers |
|----------------|---------------------------------------------------------------------------|--------|--------------|
| Br module KEGG | Oxidative phosphorylation pathway                                         | Group1 | 25           |
|                | Synaptic vesicle cycle                                                   | Group0 | 23           |
| Br module GO-BP| cellular respiration                                                      | Group3 | 35           |
|                | energy coupled proton transport, against electrochemical gradient         | Group0 | 10           |
|                | intracellular transport                                                  | Group2 | 104          |
|                | microtubule-based movement                                               | Group1 | 17           |
| T module GO-BP | M phase                                                                   | Group0 | 94           |
are illustrated in Fig. 5. There are 2 comparisons (lateral control VS lateral case in both T-module and Br module) significantly different with \( p \)-value \( \leq 0.001 \) in the student t-test.

The Br network module (Fig. 6) contains 449 genes and 2373 gene-gene interactions, of which 94 genes are up-regulated and 355 genes are down-regulated. The T network module contains 905 genes and 5156 gene-gene interactions, of which 221 genes are down-regulated and 684 genes are up-regulated.

ClueGO analysis of PD-specific modules
The ClueGO analysis of the Br modules identified 4 GO biological processes, which are shown in Fig. 7 and Table 2. These are cellular respiration, intracellular transport, energy-coupled proton transport, and microtubule-based movement. Furthermore, we identified two KEGG [21] pathways, “synaptic vesicle cycling” and “oxidative phosphorylation”. The ClueGO analysis of the T module identified one GO biological process “M phase”.

Identifying drugs with predicted therapeutic effects on the Br and T modules
We generated a ranked list of the drugs based on their DESS scores. While there were 1246 (1201 unique drug-bankID) candidate drugs for drug repositioning that targeted one or more genes in the gene co-expression module in Additional file 1: Table S1, we selected only 12 (the top 1% according to DESS) candidate drugs as potential treatments (Fig. 8). The components of DESS and number of the drug targets for each drug in the T module and Br module are shown in Fig. 9.

Furthermore, the drugs are listed in Table 3 and are discussed below. The Br and T modules’ network diagrams for the extended network illustrating which disordered genes are stimulated and inhibited by these 12 drugs is provided in Fig. 10, Additional file 2: Table S2 and Additional file 3: Table S3.

Conclusions and discussion
In this work, we present a framework that identified candidate drugs for repositioning based on analysis of GWAS and gene expression microarray data. Starting with genes identified through a standard GWAS, we extended the analysis to one-layer extension by gene-gene regulatory relationship and built an extended regulatory network. Significant results based on an enrichment analysis were then used to generate PD modules. We improved gene co-expression module cohesion by
removing isolated or weakly connected genes. PD network modules were then further informed by the integration of data from Protein-Protein interaction databases.

Using this approach, we initially identified over 1201 candidates for drug repurposing. We trimmed this to 12 modular drug candidates based on their DESS. There were three important characteristics of finding within these 12 modular drugs. First, they are noteworthy in that they target PD at the level of the gene co-expression module as opposed to a specific target. Second, most of the genes on the list belong to drug families, which should be expected if data relating to drug target efficacy are accurate and internally consistent. Third, there are general 4 drug families found (steroids and steroid derivatives, lipids and lipid-like molecules, phenylpropanoids and polyketides, and other small molecules), and each family of drugs identified has been previously studied in relation to neurodegenerative disease, suggesting the external validity of our findings as well.

The top candidate drug was estradiol, a steroidal estrogen critical in the regulation of the menstrual cycle. It is currently used pharmaceutically in hormone replacement therapies for menopause and hypogonadism. Several studies support a role for the use of estradiol in several studies support a role for the use of estradiol in menopause and hypogonadism. Several studies support a role for the use of estradiol in

| Compound ID | Compound Name | Score | Drugbank | DrugName | Class |
|-------------|---------------|-------|----------|----------|-------|
| CID5757 | CID5757 | 290.3 | DB00783 | Estradiol | Steroids and steroid derivatives |
| CID445154 | SAM001246888 | 253.9 | DB02709 | Resveratrol | stilbenes |
| CID444795 | Retinoic Acid | 248.9 | DB00523 | Alitretinoin | Prenol lipids |
| CID5538 | Accutane Roche | 246.3 | DB00755 | Tretinoin | Lipids and lipid-like molecules |
| CID5282379 | Isotretinoin (USP) | 246.3 | DB00982 | Isotretinoin | Prenol lipids |
| CID5280961 | NCCG00025005-02 | 244.3 | DB01645 | Genistein | Isoflavonoids |
| CID5460439 | Rapamune | 224.9 | DB00877 | Sirolimus | Macrolide lactams |
| CID6436030 | Perceiva | 224.9 | DB00877 | Sirolimus | Macrolide lactams |
| CID667476 | follicidiene | 199 | DB00890 | Diestrol | Phenypropanoids and polyketides |
| CID448537 | oekolp | 199 | DB00255 | Diethylstilbestrol | Phenypropanoids and polyketides |
| CID6010 | component of Tylosterone | 199 | DB00510 | Methyltestosterone | Lipids and lipid-like molecules |
| CID2756 | C4522_SIGMA | 198.2 | DB00501 | Cimetidine | Organo-heterocyclic compounds |

The third drug alitretinoin, fourth drug tretinoin, and fifth drug isotretinoin are most highly ranked candidates also belonging to a single family of compounds, retinoids. The first is retinoic acid, a retinoid morphogen crucial to the embryonic development of the anterior-
Fig. 10 The extended modular drug-target network for the Br module (a) and the T module (b). The diamond nodes are drugs; circles are genes. The color of the nodes varies from green to red and indicates down-regulation or up-regulation situation of disordered genes, respectively. The color of the edges stands for the type of action. Red edges mean stimulation, green edges mean inhibition and gray edges means Protein-Protein Interactions. The node size represents the RP-score which indicates the relevance of the gene in the module calculated in Method.
posterior axis in vertebrates, as well as differentiation and maintenance of neural cell lineage. Currently, in-vivo animal studies suggest the possibility of therapeutic applications of retinoic acid for PD through nanoparticle delivery [31]. Isotretinoin, trademarked under the name Accutane, is prescribed as a treatment for severe acne vulgaris. Although isotretinoin is a known teratogen [32], it might be well-suited to treatment of PD given its typical later age of onset.

Our seventh and eighth hits, Sirolimus (Rapamune) and Sirolimus (Perceiva), are again related. Perceiva is an ocular formulation of the macrolide compound sirolimus (commonly known as rapamycin) and was developed to treat neovascular age-related macular degeneration. Sirolimus is used for the treatment of Lymphangioleiomyomatosis, as well as in prevention of organ transplant rejection. Interestingly, sirolimus has been shown to improve cognitive deficits in mouse model of Alzheimer’s Diseases through inhibition of the mTOR signaling pathway, a pathway which is thought to protect against neuronal death in mouse models of PD [33].

In addition to these twelve candidates, our ClueGO analysis suggests that investigation of two additional biological processes may be profitable. Our analysis of KEGG pathways in relation to the T module implicated mitochondrial respiration as a potential disease mechanism [34]. Interestingly, it has previously been reported that defects in mitochondrial respiration are etiologically related to the pathogenesis of PD. Thus, preservation and restoration of mitochondrial function may hold promise as a potential therapeutic intervention to halt the progression of dopaminergic neurodegeneration in PD. Secondly, in PD, neuronal cells undergo mitotic catastrophe and endoreduplication prior to cell death. It has previously been shown [35] that overexpression of DNA poly β was involved in the rotenone-mediated pathology of cellular and animal models of PD. In a cell culture model, increased levels of DNA poly β promoted rotenone-mediated endoreduplication. Selective injury to dopaminergic neurons by rotenone resulted in the upregulation of DNA poly β as the neuronal cell cycle was reactivated.

In summary, we perform drug repositioning by integrating weighted drug-protein regulations on all genes, using our novel DESS to quantitate drug effects on entire co-expression networks. As biological systems use functional pathways and networks to maintain homeostasis, by selecting drugs that act at the level of a gene module we were able to target a level of the biological organization more relevant to the disease pathologies of complex disorders such as PD. Although this approach is still in its infancy, our results suggest that it may circumvent issues associated with single-gene targeting in polygenic diseases like PD. Our analysis has identified several families of related drug candidates, all of which have previously been investigated in relation to PD and other neurodegenerative diseases. As such, we believe our framework gives internally and externally valid results and may be extended to provide complementary insights to other disease-module findings and drug-repositioning projects.

The significance of our work should be considered in light of its limitations. First, several of the classes of drugs mentioned have already studied in relation to PD and related phenotypes, as described above. However, members of the families of drugs identified have not resulted in a clinically efficacious treatment for PD to date. As such, a future direction for this line of research is to include a mechanism to account for both additive and potentially non-additive interaction effects between drugs on a disease-specific module. In addition, many of the most highly ranked modular drugs we identified show much promise, but have known adverse effects. Future research will include a method of incorporation of drug side effects into the final priority score.

Additional files

- **Additional file 1:** 1246 (1201 unique drugbankID) candidate drugs for drug repositioning that targeted to PD modules. (XLSX 136 kb)
- **Additional file 2:** The Protein-Protein Interactions and drug-target regulations in Br module’s network. (XLSX 26 kb)
- **Additional file 3:** The Protein-Protein Interactions and drug-target regulations in T module’s network. (XLSX 98 kb)

Acknowledgements

We acknowledge the partial support of this research by the University of Alabama at Birmingham (UAB) and the UAB Informatics Institute during the implementation of the project. Our database servers and web applications are maintained by the UAB high-performance computing group.

Funding

The publication cost of the paper is from Dr.Chen’s lab start-up funding in the University of Alabama at Birmingham in informatics institute and UL1 TR001417 “UAB Center for Clinical and Translational Science” grant award.

Availability of data and materials

PAGER database is available online http://discovery.informatics.uab.edu/PAGER. DMAP database is available online http://discovery.informatics.uab.edu/cmaps. The supplemental material are available for download.

About this supplement

This article has been published as part of BMC Bioinformatics Volume 18 Supplement 14, 2017: Proceedings of the 14th Annual MCBiOS conference. The full contents of the supplement are available online at https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-14.

Authors’ contributions

ZLY performed the construction of the framework, data analysis and led the writing of the manuscript under the guidance of JYC. IA, EZ participated in drug repositioning performance evaluations. VL, SLB and JYC participated in the revision of the manuscript. JYC conceived the project, supervised the entire research team with frequent feedback in the design, implementation, and evaluation of the project. All authors contributed to the completion of the manuscripts and approved final manuscript.
Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors consent for publication.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1. Center for Biomedical Big Data, Wenzhou Medical University First Affiliated Hospital, Wenzhou, Zhejiang Province, China. 2. Informatics Institute in School of Medicine, University of Alabama at Birmingham, Birmingham 35233, AL, USA. 3. Division of Clinical Immunology and Rheumatology in School of Medicine, University of Alabama at Birmingham, Birmingham 35233, AL, USA. 4. Wenzhou Yuekang InfoTech, Ltd., Wenzhou, Zhejiang Province, China.

Published: 28 December 2017

References
1. Jankovic J, Sherer T. The future of research in Parkinson disease. JAMA Neurol. 2014;71(11):1351–2.
2. Dawson TM, Dawson VL. Molecular pathways of neurodegeneration in Parkinson’s disease. Science. 2003;302(5646):819–22.
3. Chen JY, Piquette-Miller M, Smith BP. Network medicine: finding the links to personalized therapy. Clin Pharmacol Ther. 2013;94(6):613–6.
4. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(11):1566–68.
5. Dexter DT, Jenner P. Parkinson disease: from pathology to molecular disease mechanisms. Free Radic Biol Med. 2013;62:132–44.
6. Fujita KA, Otsawaziki M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, Crespo I, Perumal TM, Jurkowski W, Antony PM, et al. Integrating pathways of Parkinson’s disease in a molecular interaction network map. Mol Neurobiol 2014;56:1389–92.
7. Mlok SJ, Lopez-Yunez AM, Chen JY. Genome-wide meta-analysis of genetic regulation of DNA polymerase beta in rotenone-based Parkinson’s disease. J Theor Biol. 2014;357:138–51.
8. Langefeld P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
9. Ritchie ME, Pirson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-seq and microarray studies. Nucleic Acids Res. 2013;41(7):e47.
10. Yue Z, Khishigsaar NM, Nguyen T, Suphavilai C, Eyton MT, Zhu L, Ratliff T, Chen JY. PAGER: constructing PAGs and new PAG-PAG relationships for network biology. Bioinformatics. 2015;31(12):2507–7.
11. Chen JY, Pandey R, Nguyen TM, HAPPI-2: a comprehensive and high-quality map of human annotated and predicted protein interactions. BMC Genomics. 2017;18(1):182.
12. Bindas G, Miklenik B, Hadi H, Charoenpong P, Tossolini M, Kirilovsky A, Fridman WH, Pages F, Tojiasniak Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091–3.
13. Huang H, Nguyen T, Ibrahim S, Shantharam S, Yue Z, Chen JY. DMAP: a connectivity map database to enable identification of novel drug repositioning candidates. BMC bioinformatics. 2015;16 Suppl 13:54.
14. Chen JY, Mamidipalli S, Huan T. HAPPI: an online database of comprehensive human annotated and predicted protein interactions. BMC Genomics. 2009;10 Suppl 1:156.
15. Chen JY, Shen C, Sivachenko AY. Mining Alzheimer disease relevant proteins from integrated protein interactome data. Pac Symp Biocomput. 2006:367–78.
16. Kamehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457–62.
17. Sawada H, Ib I, Kihara T, Honda K, Nakamoto T, Kani K, Nakanishi M, Sakka N, Akaike A, Shimohama S. Estradiol protects dopaminergic neurons in a MPP+-Parkinson’s disease model. Neuropsychopharmacology. 2002;24(2):1056–64.
18. Currie LJ, Harrison MB, Trugman JM, Bennett JP, Wooten GF. Postmenopausal estrogen use affects risk for Parkinson disease. Arch Neurol. 2004;61(6):886–8.
19. O’Reilly EJ, Mireazi F, Forman MR, Ascherio A. Diethylstilbestrol exposure in utero and depression in women. Am J Epidemiol. 2010;171(8):876–82.
20. Harman SM, Tistouzas PD. Reproductive hormones in aging men. I. Measurement of sex steroids, basal luteinizing hormone, and Leydig cell response to human chorionic gonadotropin. J Clin Endocrinol Metab. 1980;51(1):35–40.
21. Okun MS, Fernandez HH, Rodriguez RL, Romrell J, Suelter M, Munson S, Louis ED, Mulligan T, Foster PS, Shenal BV, et al. Testosterone therapy in men with Parkinson disease: results of the TET-PT study. Arch Neurol. 2006;63(5):729–33.
22. Liu LX, Chen WP, Xie XJ, Wong MS. Neuroprotective effects of geranin on dopaminergic neurons in the mice model of Parkinson’s disease. Neurosci Res. 2008;60(2):156–61.
23. Turner RS, Thomas RG, Craft S, van Dyck CH, Minter J, Reynolds BA, Brewer JB, Rissman RA, Raman R, Aisen PS, et al. A randomized, double-blind, placebo-controlled trial of resveratrol for Alzheimer disease. Neurology. 2015;85(16):1383–91.
24. Derfert A, Gaballo A, Tanzarella P, Piccoli C, Capitanio N, Rico B, Annese T, Di Paola M, Dell’Aquila C, De Mari M et al: Effect of resveratrol on mitochondrial function: implications in parkin-associated familiar Parkinson’s disease. Biochim Biophys Acta. 2014;1842(7):902–915.
25. Jin F, Wu Q, YL G, Gong QH, Shi JS. Neuroprotective effect of resveratrol on 6-OHDA-induced Parkinson’s disease in rats. Eur J Pharmacol. 2008;600(1–3):78–82.
26. Estes M, Cristavo AO, Saraiva T, Rocha SM, Baltazar G, Ferreira L, Bernardino L. Retinoic acid-loaded polymeric nanoparticles induce neuroprotection in a mouse model for Parkinson’s disease. Front Aging Neurosci. 2013;5:20.
27. Kontakakis VP, Soukides D, Ferentinou P, Havaki-Kontakakis BJ, Papadimitriou GN. Isotretinoin and psychopathology: a review. Ann General Psychiatry. 2009;9:22.
28. Malagelada C, Jn ZH, Jackson LW, Povedborski S, Greene LA. Rapamycin protects against neuron death in vitro and in vivo models of Parkinson’s disease. J Neurosci. 2010;30(3):1166–75.
29. Peiter C, Vila M. Mitochondrial biology and Parkinson’s disease. Cold Spring Harbor Perspectives Med. 2012;2(2):a009332.
30. Wang H, Chen Y, Chen J, Zhang Z, Lao W, Li X, Huang J, Wang T. Cell cycle regulation of DNA polymerase beta in rotenone-based Parkinson’s disease models. PLoS One. 2014;9(10):e109697.