# An Effective Method to Identify Shared Pathways and Common Factors among Neurodegenerative Diseases

Ping Li[1,2], Yaling Nie[1,2], Jingkai Yu[1]*

1 National Key Laboratory of Biochemical Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing, 100190, China, 2 University of Chinese Academy of Sciences, Beijing, 100049, China

* jkyu@ipe.ac.cn

## Abstract

Groups of distinct but related diseases often share common symptoms, which suggest likely overlaps in underlying pathogenic mechanisms. Identifying the shared pathways and common factors among those disorders can be expected to deepen our understanding for them and help designing new treatment strategies effected on those diseases. Neurodegeneration diseases, including Alzheimer's disease (AD), Parkinson's disease (PD) and Huntington's disease (HD), were taken as a case study in this research. Reported susceptibility genes for AD, PD and HD were collected and human protein-protein interaction network (hPPIN) was used to identify biological pathways related to neurodegeneration. 81 KEGG pathways were found to be correlated with neurodegenerative disorders. 36 out of the 81 are human disease pathways, and the remaining ones are involved in miscellaneous human functional pathways. Cancers and infectious diseases are two major subclasses within the disease group. Apoptosis is one of the most significant functional pathways. Most of those pathways found here are actually consistent with prior knowledge of neurodegenerative diseases except two cell communication pathways: adherens and tight junctions. Gene expression analysis showed a high probability that the two pathways were related to neurodegenerative diseases. A combination of common susceptibility genes and hPPIN is an effective method to study shared pathways involved in a group of closely related disorders. Common modules, which might play a bridging role in linking neurodegenerative disorders and the enriched pathways, were identified by clustering analysis. The identified shared pathways and common modules can be expected to yield clues for effective target discovery efforts on neurodegeneration.

## Introduction

Healthcare improvements coupled with low fertility are expected to cause an increasingly larger proportion of old population, which leads to more chronic illnesses [1]. A representative type of chronic disease is neurodegenerative disorders, such as Alzheimer's disease (AD),

Parkinson's disease (PD) and Huntington's disease (HD). Neurodegenerative diseases bring enormous suffering in terms of economical cost and emotional trauma. Unfortunately, the etiologies and pathogeneses of these disorders remain not well understood. Current therapies for these diseases are palliative rather than curative and their effectiveness is still far from satisfactory [2]. It is thus critical to elucidate factors underlying these disorders for better design of intervention strategies. However, the traditional strategy of "one disease-one target-one drug" is no longer effective and challenged in many cases, especially with regard to multi-factorial diseases [3, 4], which is the case for neurodegenerative disorders. Physiological redundancies in biological networks could also limit efficacy of administered drugs [5]. For complex diseases, multiple targets or pathways have to be affected for successful treatment outcomes.

AD, PD and HD share at least two common symptoms: motor and cognitive impairment [6–8]. Similar phenotypic traits suggest that there are likely overlaps in the pathogenic mechanisms underlying distinct neurodegenerative disorders. Compared to studying individual diseases separately, identification and analysis of the common dysfunctional proteins or dysregulated modules/pathways of the three diseases can be expected to provide deeper insights into their pathogenic processes. Understanding the common pathogenic processes could facilitate efforts to design treatment strategies utilizing optimal drug combinations that could work effectively for the diseases.

Differentially expression genes (DEG) and genome-wide association studies (GWAS) are usually applied to study related biological pathways of a specific disease. For multiple diseases, however, there is lack of effected method to study their shared pathways and common factors. In this paper, we proposed a simple and effective approach which integrated common susceptibility genes of multiple disorders and the human protein-protein interaction data (Fig 1). AD and PD susceptibility genes were acquired from public online databases. HD susceptibility genes were acquired through literature mining and the random walk algorithm [9]. Common genes of the three susceptibility gene sets and their first neighbors in the human protein-protein interaction network (hPPIN), called as CFNN, were extracted to perform pathway enrichment analysis, which identified pathways related with neurodegenerative diseases. Gene expression data sets from NCBI GEO database [10] were applied to evaluate the computed pathways. Meanwhile, pathway clustering analysis obtained the common modules in CFNN shared by distinct pathways. Those modules might play a bridging role in linking enriched pathways and neurodegeneration.

## Materials and Methods

### Data source

Human protein-protein interaction network (hPPIN) was constructed by integrating four existing databases, i.e., BioGrid [11], HPRD [12], IntAct [13], and HomoMINT [14]. Protein identifiers were mapped to the genes coding for the proteins, and redundant interactions were removed. The comprehensive protein-protein interaction network covers 15,710 human genes and 143,237 interactions.

AD and PD susceptibility genes were acquired from the GAD [15], CTD [16] and OMIM [17] database. These public data sources store associations between genes and diseases, but focus on different aspects of the phenotype-genotype relationship. After integrating all the records in the databases, 433 and 188 distinct susceptibility genes were collected for AD and PD, respectively. The three databases does not have sufficient data for HD, whose susceptibility genes were collected by text-mining of biomedical literatures from PubMed (http://www.ncbi.nlm.nih.gov/pubmed/). It produced 20 HD susceptibility genes. Compared with AD and PD, the number of collected HD susceptibility genes is still rather low, which might be due to the
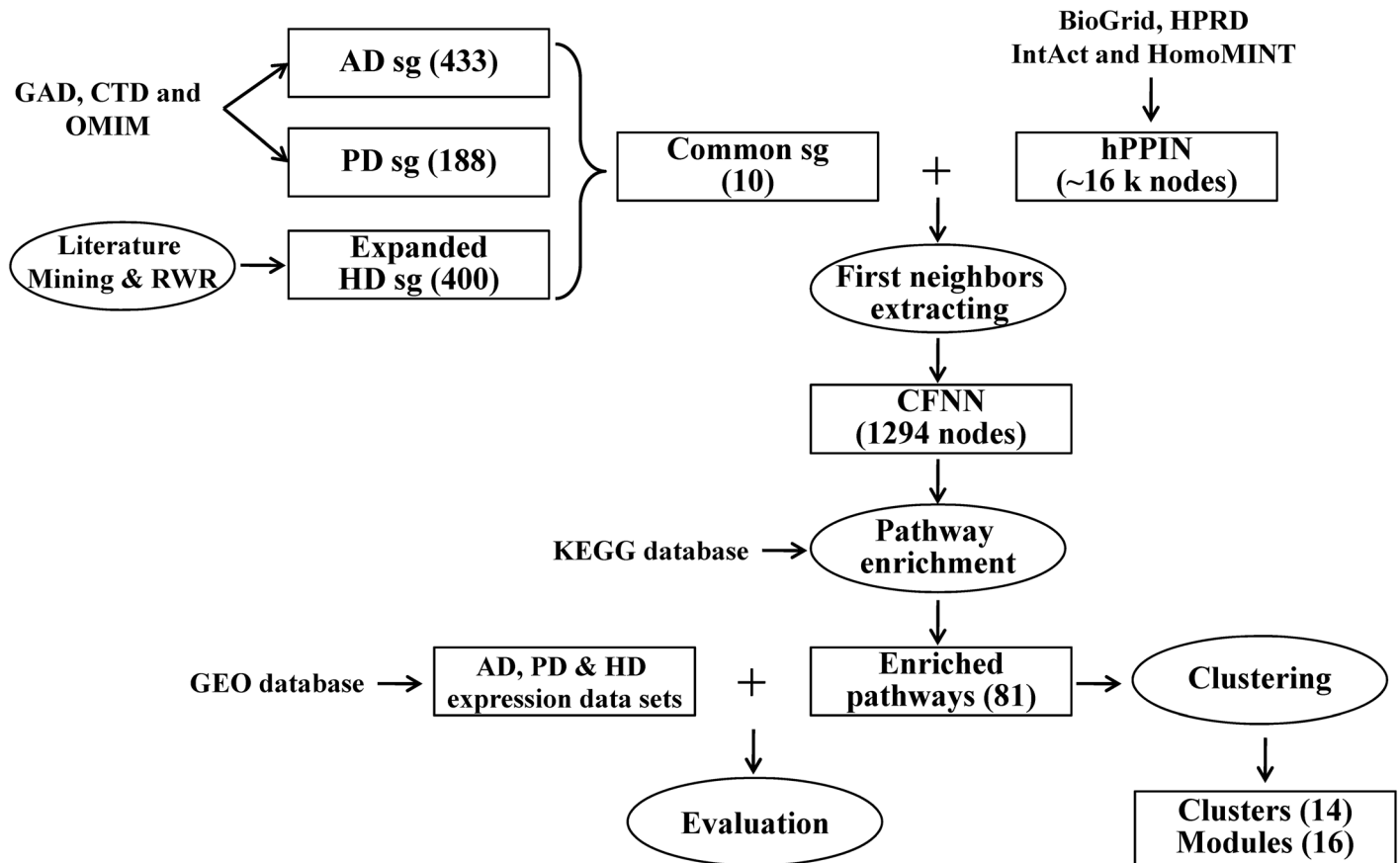
**Fig 1. Workflow for identification of shared pathways and common modules among AD, PD and HD.** In the first step, AD, PD and HD susceptibility gene (sg) sets were collected and their intersection were defined as common sg. Meanwhile, common sg's first neighbors in the human protein-protein interaction network (hPPIN) was extracted to construct common gene first neighbor network (CFNN). Then, KEGG pathway enrichment analysis was applied to the nodes in CFNN to get shared pathways of AD, PD and HD, following by gene expression analysis to evaluate the found pathways. Finally, hierarchal clustering was applied to cluster the enriched pathways and indentify common modules in CFNN. RWR: random walk with restart.

much lower prevalence of HD than AD and PD [18–20]. To bring the number of HD's susceptibility genes to the same level as those of AD and PD, a random walk algorithm [9] was applied to expand the number of HD susceptibility genes through the hPPIN, using manually collected HD susceptibility genes as seed nodes. The top 400 genes ranked by random walk (including the seed genes) were selected as the expanded set of HD susceptibility genes.

## Random walk with restart (RWR)

RWR is a variant of random walk. It mimics an iterative walker that moves from a current node to a randomly selected adjacent node, and allows the restart of the walk in every time step at source nodes with predefined probability $\gamma$ [9]. RWR is formally defined as follows:

$$p_{t+1} = (1 - \gamma)Wp_t + \gamma p_0$$

Where $W$ is the column-normalized adjacency matrix of the graph and $p_t$ is a vector in which the $i$th element holds the probability of being at node $i$ at time step $t$. $p_0$ is the initial probability vector where equal probabilities were assign to the source nodes, with the sum of the probabilities equal to 1.

In this study, RWR was used to prioritize susceptibility genes from among genes that have not been associated with HD. The set of source nodes consists of genes known to be associated with HD. The predefined probability $\gamma$ was set to 0.75, as was done by Kohler et al [21]. All genes in the network are eventually ranked according to their steady-state probabilities and the top 400 genes were selected.

## Common susceptibility genes and their first neighbor network construction

We took the intersection of AD, PD and expanded HD susceptibility genes and called it the set of common susceptibility genes of the three disorders. To check the significance of those common genes, we randomly generated three gene sets of the same size as that of AD, PD and expanded HD susceptibility genes from hPPIN and computed the number of common genes among them. The process was repeated $10^4$ times. A p-value was then computed for the observed number of common genes.

Nearest neighbors of the common genes were extracted from the hPPIN to construct a network consisting of the common genes and their first neighbors, which was called the Common gene First Neighbor Network (CFNN).

## Pathway enrichment and clustering analysis

CFNN consists of the common susceptibility genes and their direct interaction partner in hPPIN. Pathways enriched with genes in CFNN are very likely shared pathways of AD, PD and HD. ClueGO v2.0.7 [22] was used to perform KEGG [23] pathway enrichment for all nodes in CFNN. ClueGO, an Cytoscape [24] plug-in, can identify biological pathways enriched with a list of genes. Two-sided (enrichment/depletion) method based on the hypergeometric distribution was used for statistical test with a multiple testing p-value correction using the Benjamini-Hochberg method [25]. Pathways with adjusted p-value $< 0.05$ were regarded as related biological pathways to CFNN genes and were selected for further analysis.

Hierarchical clustering approach was use for clustering analysis. Genes appearing in both the CFNN and enriched KEGG pathways were named as **associated genes** (Fig 2(A)). A binary associated gene-pathway matrix was created (0: absent, 1: present). Based on this matrix, a cosine similarity matrix of pathways was built and used to group the pathways into clusters. To getting meaningful clusters, we manually checked the dendrogram plot of results and chose clustering distance $d = 1.1$ as the final cutting point. For each cluster, each member pathway's associated genes were intersected to obtain their common associated genes. Those common associated genes were then mapped to CFNN to get their interaction subnetwork, called common module (Fig 2(B)). The average clustering coefficients of the acquired modules were computed.

## Gene expression analysis

Twenty AD, PD and HD gene expression data sets (March 16, 2014), attached raw data, were collected from the NCBI GEO database (see S1 File). Among those extracted expression sets, only GSE7621 [26], GSE8397 [27], GSE20168 and GSE20292 [28, 29] on PD patients and GSE45596 [30] on AD patients (see Table 1), have significantly differentially expressed genes (methods were explained below). 4 of the 5 expression sets were on PD vs. Normal. The combine of differentially expressed genes acquired in the 4 expression sets were defined as the finally differentially expressed gene set on PD.

For Affymetrix HG_U133 (including A chip and B chip) and HG-U133_Plus_2 platform, the CEL source files were preprocessed by the RMA algorithm with default parameters in the R
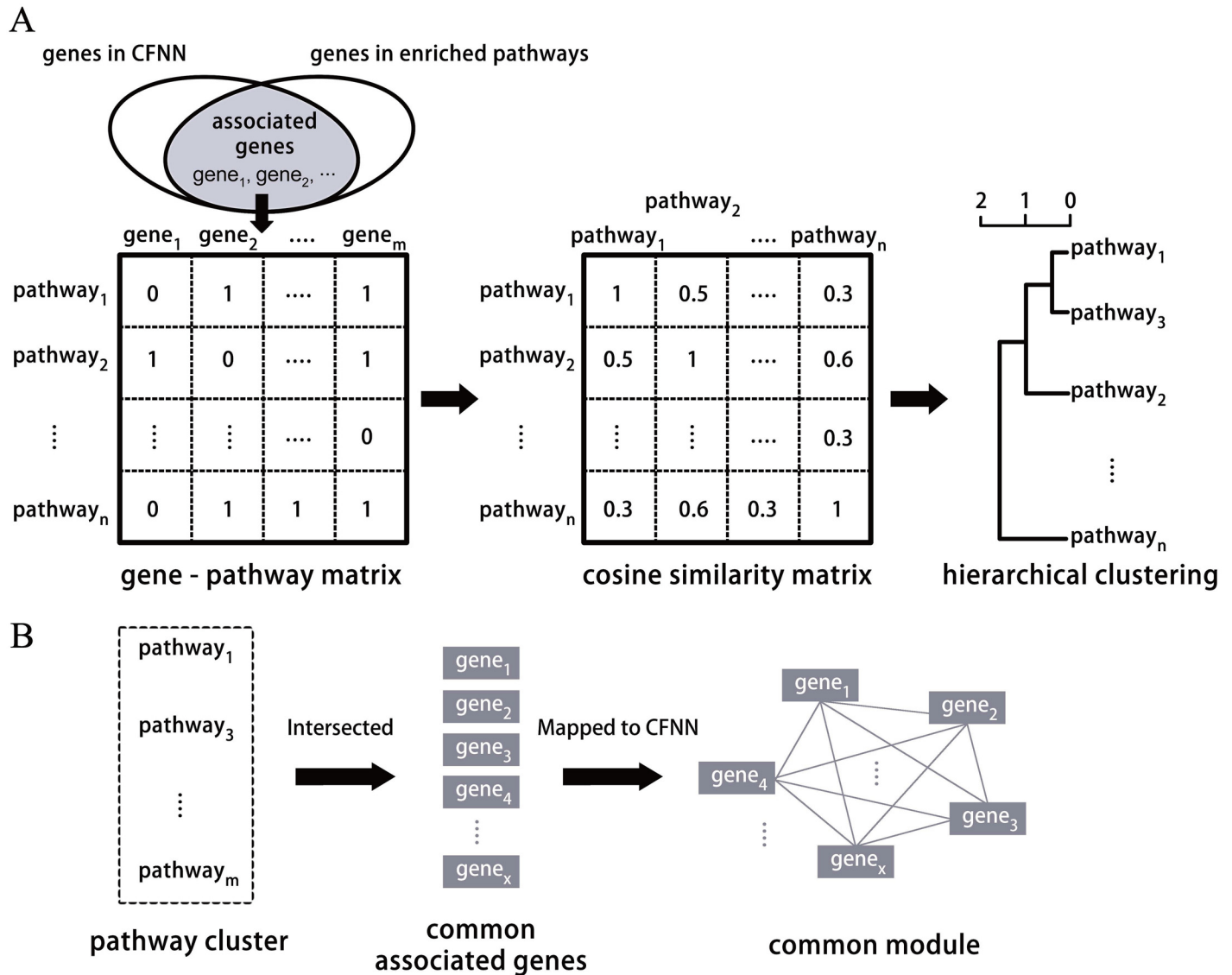
**Fig 2. Pathway clustering analysis and common module extracting.** (A) Diagram of clustering analysis. Hierarchical clustering was applied and only associated genes in a specific pathways were considered, not the full genes. (B) Extracting common module from a pathway cluster. The common module is a part of CFNN, and might playing a bridging role in linking enriched pathways and neurodegeneration.

Bioconductor package[31]. For Agilent-014850 platform, preprocessing steps of the TXT source files included background correction with the "normexp" method to subtract the background intensity from the foreground intensity for each spot [32], within-array normalization with the "loess" method to normalize the M-values for each array separately, and between-array normalization with the "quantile" method to normalize intensities or log-ratios for them to be comparable across arrays [33]. The package limma [34] in Bioconductor was then used to perform differential expression analysis for the preprocessed microarray data. Probe sets were mapped to NCBI entrez genes using R package GEOquery [35]. In cases where there were multiple probe sets that correspond to the same gene, expression values of those probe sets were averaged. Genes that were significantly differentially expressed with a Benjamini and Hochberg adjusted p-value less than 0.05 [25] were picked for later analysis.

**Table 1. List of selected gene expression data sets.**

| GEO accession | Sample tissue | Platform | Nr. of Sig. Diff. [a] |
|---|---|---|---|
| PD *vs.* Normal | | | |
| GSE7621 | Substantia nigra | HG-U133_Plus_2 | 143 |
| GSE8397 | Substantia nigra | HG-U133A/B | 655 |
| GSE20168 | Prefrontal area | HG-U133A | 169 |
| GSE20292 | Substantia nigra | HG-U133A | 24 |
| AD *vs.* Normal | | | |
| GSE45596 | Brain microvessel | Agilent-014850 | 2063 |

[a]: The number of significantly differentially expressed genes.

doi:10.1371/journal.pone.0143045.t001

To evaluate the enriched KEGG pathways, each node of the pathway was considered as a component. Those components were a mixture of one protein node and multi-protein node. Multi-protein component, which contains more than one protein, was also regarded as a single component. That is to say, if any individual protein of the multi-protein component was found to be significantly differentially expressed in gene expression analysis, the corresponding multi-protein component was taken as significantly differentially expressed. For example, α-Catenin, a multi-protein component in adherens junction, is composed of catenin alpha-1, catenin alpha-2 and catenin alpha-3. If one of the three proteins was shown to be significantly differentially expressed, α-Catenin was defined as a significantly differentially expressed component. Gene symbols of proteins involved in all components were extracted from KEGG. To check the significance of obtaining those differentially expressed components in an enriched pathway, we randomly generated gene set of the same size as that of computed differentially expressed genes from human gene set, and computed the number of components involved in the enriched pathways. The process was repeated $10^4$ times. A p-value was then computed for the observed number of differentially expressed components.

## Results and Discussion

### Common susceptibility genes of AD, PD and HD show high statistical significance

AD, PD and HD share 10 common susceptibility genes, which were obtained by taking intersection of susceptibility gene sets of the three disorders. P-value for finding same or larger size of common gene set was found to be $1.17 \times 10^{-6}$ (Fig 3), showing that the acquired 10 common genes was statistically significant.

Table 2 showed clinical indications for 5 of the 10 common genes. Interestingly, three of them had been used to treat cancers, i.e., PARP1, GSK3B and UCHL1. It suggests that cancers and neurodegenerative disorders could be correlated. GSK3B, UCHL1 and LRRK2 were also reported to be potential therapeutic targets for neurodegenerative diseases and inhibitors had been designed against them [36–38]. The remaining 5 common genes showing no indication yet were all related with key processes in neurodegeneration. CASP3, FAS, SQSTM1 and YWHAZ participate in cell apoptosis [39, 40], which are activated in neurodegenerative diseases [41]. TFAM, playing a role in organizing and compacting mitochondrial DNA, is related with the mitochondrial dysfunction in neurodegenerative disorders [42]. The 10 common genes acquired here might be a good starting point to find overlapped pathogenic mechanisms underlying the three diseases, facilitating efforts to discover potential drug targets for neurodegenerative diseases.
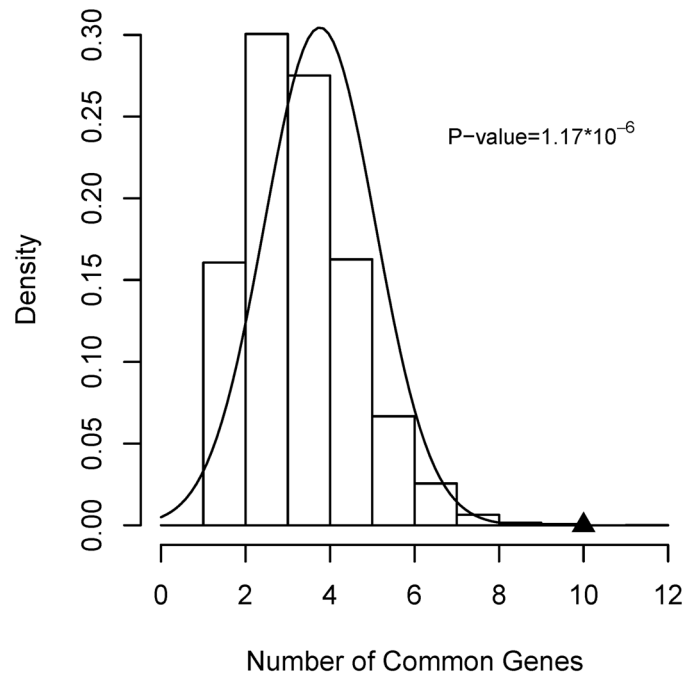
**Fig 3. Probability density for obtaining common genes.** The observed value is marked with a filled triangle.

## Eighty-one KEGG pathways were enriched with common susceptibility genes and their nearest neighbors in hPPIN

The CFNN covers 1294 human genes with 21679 interactions. 81 KEGG pathways were enriched with adjusted p-value < 0.05. 574 genes were found to be associated with CFNN and enriched KEGG pathways, called the **associated genes** (see Fig 2). The list of enriched KEGG pathways and their associated genes can be found in S2 File.

**Table 2. Indications of common susceptibility genes of AD, PD and HD.**

| Gene symbol | Protein | Indication[a] |
|---|---|---|
| ESR2 | Estrogen receptor beta | **Successful target:** Vasomotor symptoms |
| PARP1 | Poly [ADP-ribose] polymerase-1 | **Successful target:** Inflammatory skin conditions; **Clinical trial target:** Malignant melanoma, Triple negative breast cancer, Non small cell lung cancer, Brain cancer, Stroke, Myocardial infarction |
| GSK3B | Glycogen synthase kinase-3 beta | **Clinical trial target:** Non Hodgkin lymphoma, Glioblastoma multiforme, Acute promyelocytic leukemia, Brain and central nervous system tumors; **Research target**: AD, Type II diabetes |
| UCHL1 | Ubiquitin carboxyl-terminal hydrolase isozyme L1 | **Research target:** Cancer, AD and PD |
| LRRK2 | Leucine-rich repeat serine/ threonine-protein kinase 2 | **Research target:** PD |

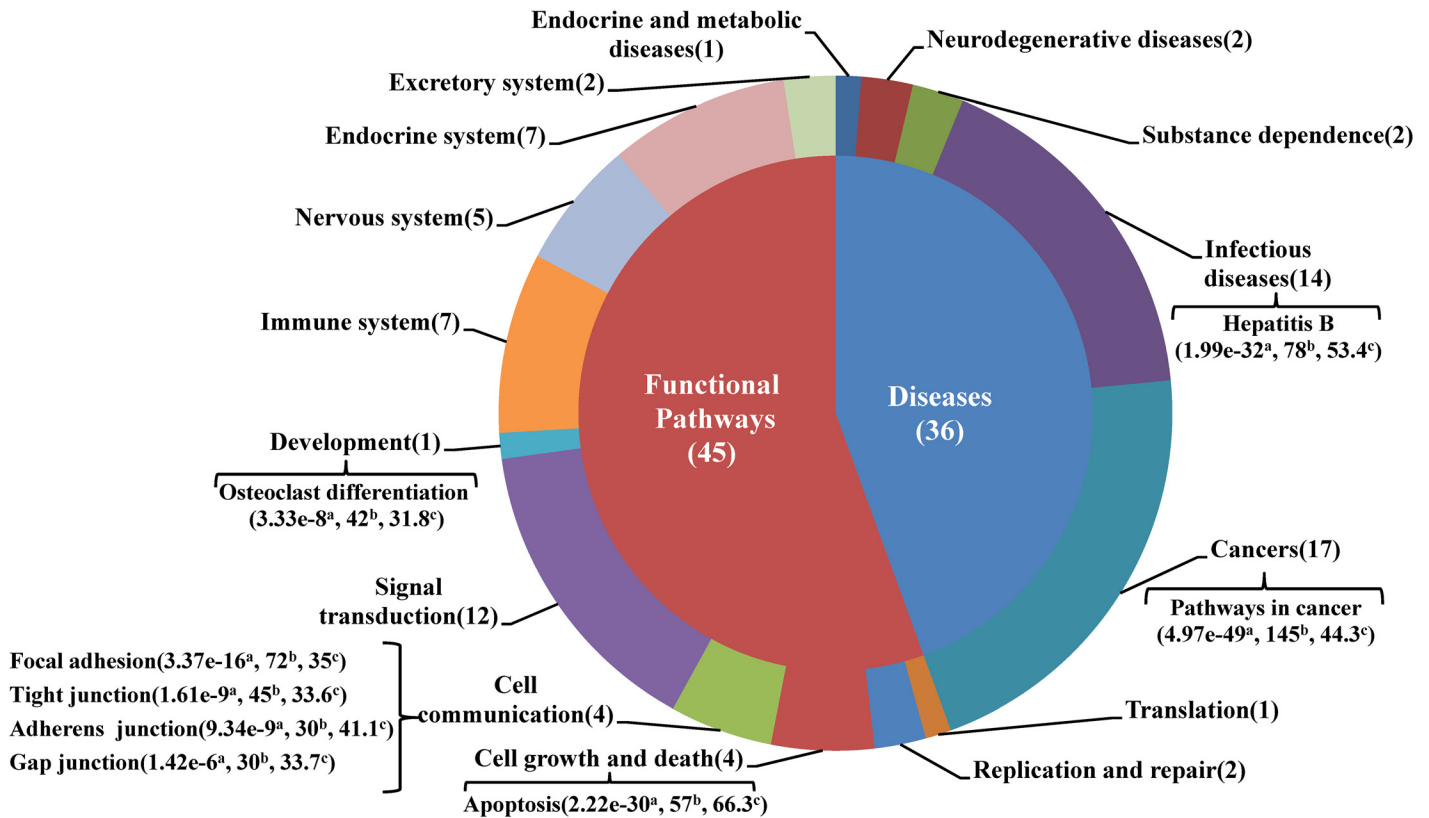[a]: Source from Therapeutic Target Database [43].

**Fig 4. Subclasses of enriched KEGG pathways.** The number of pathways belong to each subclass is shown in parentheses. Pathways mentioned in section 3.2 are labeled. [a]: Benjamini and Hochberg adjusted p-value; [b] and [c]: Number and percentage of associated genes in each pathway. Percentage represents the proportion of associated genes in all known genes involved in a pathway.

The enriched pathways belonged to two categories: functional pathways and diseases (Fig 4). Thirty six were human disease pathways, which belonged to 5 types of diseases: cancers, infectious diseases, neurodegenerative diseases, endocrine and metabolic diseases, and substance dependence. Among those, cancers and infectious diseases were the two largest subclasses, which had 17 and 14 disease pathways respectively (Fig 4). The two most significantly enriched human disease pathways were pathways in cancer and hepatitis B, with adjusted p-values of $4.97 \times 10^{-49}$ and $1.99 \times 10^{-32}$ respectively (Fig 4). Pathways in cancer is a KEGG overview pathway which integrates all specific KEGG cancer pathways' signaling networks. It is actually not surprising to see many cancers and infectious diseases related to neurodegeneration. Although neurodegenerative disease and cancer are two distinct pathological disorders, past epidemiological studies suggest that sufferers of neurodegenerative disorders have reduced incidence for most cancers [44–46]. Moreover, a growing body of evidence shows that these two types of diseases share common mechanisms of genetic and molecular abnormalities, which involve regulation of cell cycle, DNA repair, protein turnover, oxidative stress, and autophagy [47]. Many studies have also shown that viral and bacterial infections can induce significant neuronal dysfunction and degeneration of specific neuronal populations [48]. It was reported that viruses could induce brain dysfunction by either direct cytolytic effects or bystander inflammatory reactions, especially by neurotropic viruses (for example, measles, herpesviridae and influenza) [49]. Recently, Deleidiet al. raised a hypothesis that viral infections

and inflammation prime neurons and immune cells in the brain, rendering neuronal populations vulnerable to degeneration in the face of subsequent insults [50].

The remaining 45 were miscellaneous functional pathways, which could be divided into 10 subclasses: signal transduction, immune system, endocrine system, nervous system, cell communication, cell growth and death, excretory system, replication and repair, translation, and development (Fig 4). Pathway apoptosis was found with an very high p-value of $2.22 \times 10^{-30}$ (Fig 4). It is known that neuronal death underlies the symptoms of many neurodegenerative disorders including Alzheimer's, Parkinson's and Huntington's diseases. Early research had shown that apoptosis, involving oxidative stress, perturbed calcium homeostasis, mitochondrial dysfunction and activation of cysteine proteases called caspases, is a shared pathway of AD, PD and HD [51]. The newly discovered immune channel of brain [52] suggests possible critical role of immune system in etiology of neurodegenerative disorders. In fact, immune system was found to be a main subclass of functional pathways enriched with genes of neurodegeneration diseases (Fig 4). Immune system's role in the initiation of neuronal degeneration has been documented for HD, and activation of microglia (brain macrophages) is associated with cognitive dysfunction [53, 54]. Immune activation has also been indicated in the early phases of AD [55]. Moreover, several studies in rodent models of PD demonstrated that neuroinflammation can precipitate PD-like pathology [56–61].

Interestingly, correlation was also found between osteoclast differentiation and neurodegenerative disorders. Osteoclast differentiation was the only pathway in development subclass that was enriched. There were 42 associated genes (nearly one-third of osteoclast differentiation genes) and the adjusted p-value was $3.33 \times 10^{-8}$ (Fig 4). The osteoclasts, multinuclear cells originating from the hematopoietic monocyte-macrophage lineage, are responsible for bone resorption. Epidemiological studies showed that patients with AD had an increased risk of developing osteoporotic hip fractures [62]. Quite recently, it was found that amyloid beta peptide in patients with AD was elevated in osteoporotic bone tissues and enhances osteoclast function [63]. Our findings, combined with previously published results, suggest that osteoclast differentiation pathway may be a common factor for both osteoporosis and neurodegeneration.

Focal adhesion and gap junction, members of the cell communication group, had been reported to be related to neurodegenerative diseases [64–66]. In the case of the remaining two pathways in the cell communication group (Fig 4), i.e., adherens junction and tight junction, little research was found on their relationship with neurodegeneration. Our results, however, showed adherens junction and tight junction also had significant correlation with neurodegenerative disorders. The number of associated genes of adherens junction and tight junction were 30 and 45, with p-values of $9.34 \times 10^{-9}$ and $1.61 \times 10^{-9}$, respectively (Fig 4).

## Gene expression analysis confirmed that adherens and tight junctions were indeed correlated with neurodegeneration

After gene expression analysis, 927 significantly differentially expressed genes for PD and 2063 for AD were obtained. The list of differentially expressed genes can be found in S3 File.

Each of adherens and tight junctions had 50 pathway components (see S4 File for details and section 2.5 for the definition of "component"). For the PD differentially expressed gene set, adherens and tight junction had 10 and 9 differentially expressed components, respectively. For AD, the numbers of differentially expressed components were 12 and 14. For adherens junction, p-values for obtaining the number of components in PD and AD were $1.96 \times 10^{-6}$ and $1.32 \times 10^{-4}$ (Fig 5(A) and 5(B)). For tight junction, the p-values were $3.52 \times 10^{-3}$ and $5.82 \times 10^{-3}$ (Fig 5(C) and 5(D)). The small p-values imply that the number of differentially expressed
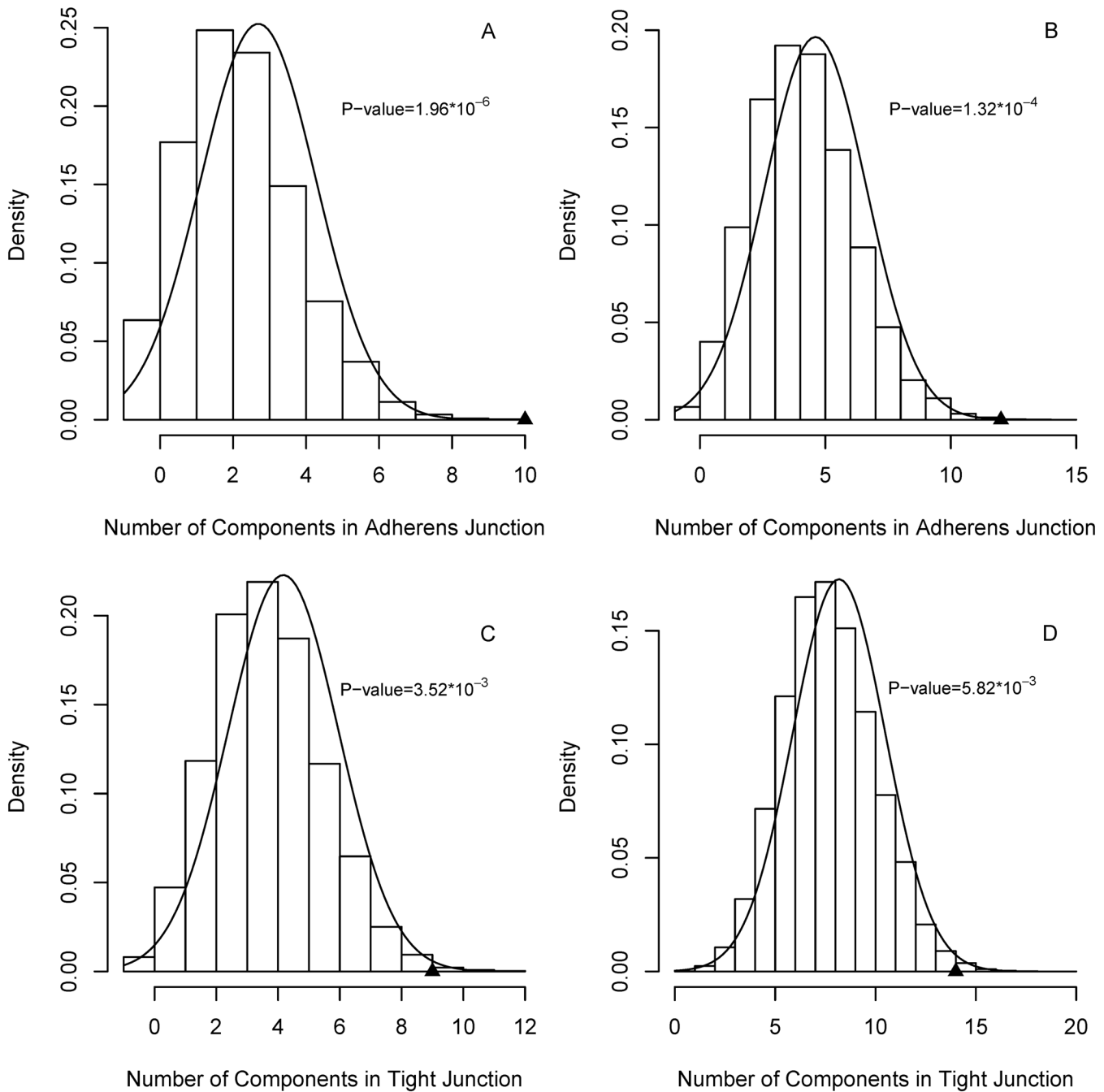
**Fig 5. Probability density for obtaining the number of components in adherens and tight junction pathways.** (A) and (C) for PD, (B) and (D) for AD; the observed value is marked with a filled triangle.

components is statistically significant for the two junction pathways. Pathway enrichment (section 3.2) and gene expression analysis together indicated that adherens and tight junction are very likely related to neurodegenerative diseases. Actually, adherens and tight junction were found to be involved in maintaining blood-brain barrier (BBB) integrity [67]. It had been shown that changes in BBB existed in AD and PD patients [68]. The two junction pathways may deserve more attention for better understanding of neurodegenerative processes.

## Common modules behind the enriched pathways were identified through clustering analysis

Section 3.2 had shown that the enriched pathways were interconnected, such as infectious diseases and immune system. Seeking out the molecular connections among those pathways might help to illustrate their relationship with neurodegenerative diseases, lead to deeper sight into the pathogenic process of neurodegeneration, which could then facilitate designing of effective synergistic treatment strategies. Clustering analysis was utilized to explore internal connections of the enriched pathways. Fig 6 shows result of the hierarchical clustering based on the cosine similarity of associated gene vectors. 14 clusters were finally acquired, which showed significant differences from the KEGG categories. Some clusters were composed of functional pathways and diseases, e.g., cluster 1 and cluster 2. For others, pathways belonged to different subclasses were clustered together, e.g., cluster 3, cluster 4 and cluster 10. The common associated genes within each cluster and their interaction network, called as common module, were extracted. The extracted common module was also a part of CFNN, because the associated genes were obtained by taking intersection of CFNN and the enriched pathways (Fig 2). Those modules were connected denser than CFNN. The mean clustering coefficient of them was 0.65 (Fig 6), while clustering coefficient of CFNN was only 0.38. The found modules could thus be the local cores within CFNN and might play a bridging role between pathways in a cluster and neurodegeneration. Elucidating working mechanisms of the modules, how they control those related pathways, may provide a fruitful strategy for understanding neurodegenerative disorders.

As an example, Fig 7 showed the acquired common module from cluster 2. The module's relationship with common susceptibility genes of AD, PD and HD was also shown. The common module, which happens to be a fully connected network, was composed of RELA,
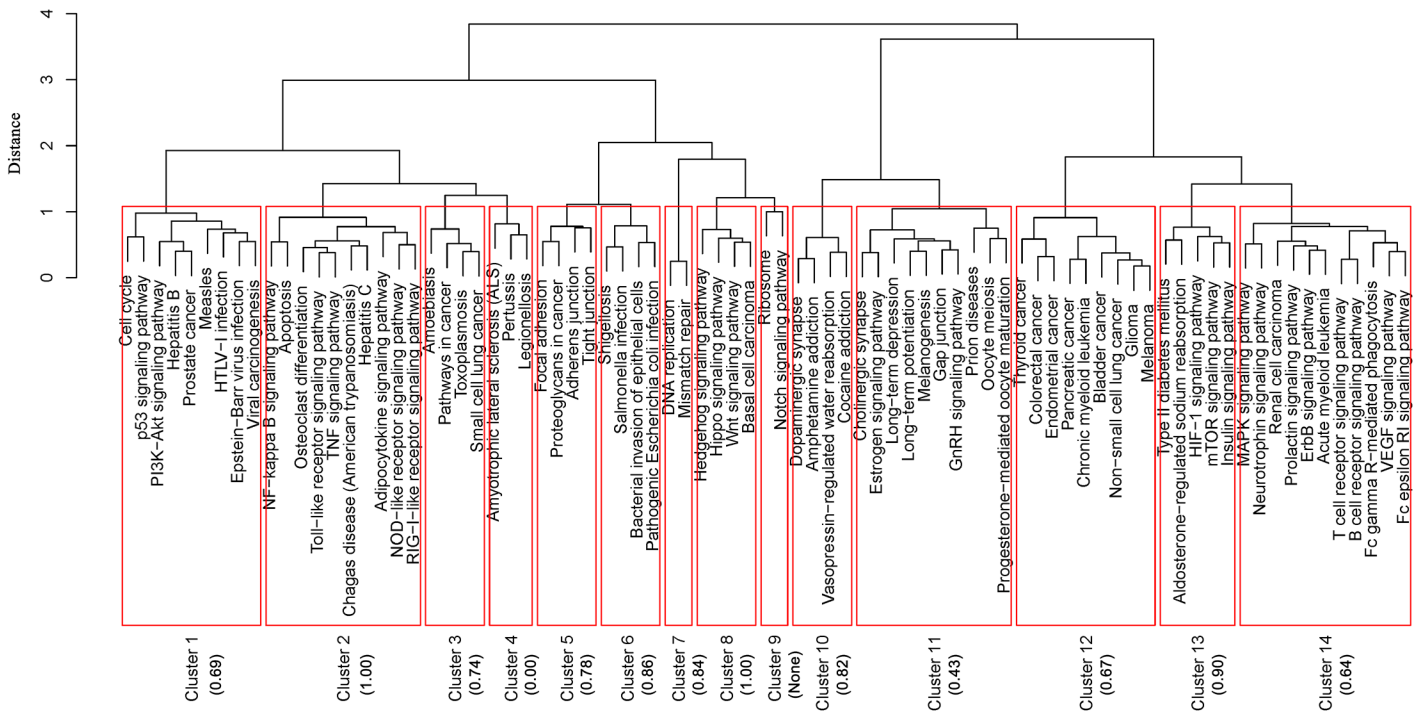


**Fig 6. Hierarchical clustering of the enriched pathways.** Each cluster is encircled by a rectangle and numbered. Clustering coefficient of common module in each cluster is shown in parentheses. None: represent the fact that there are no common associated genes within the corresponding cluster.
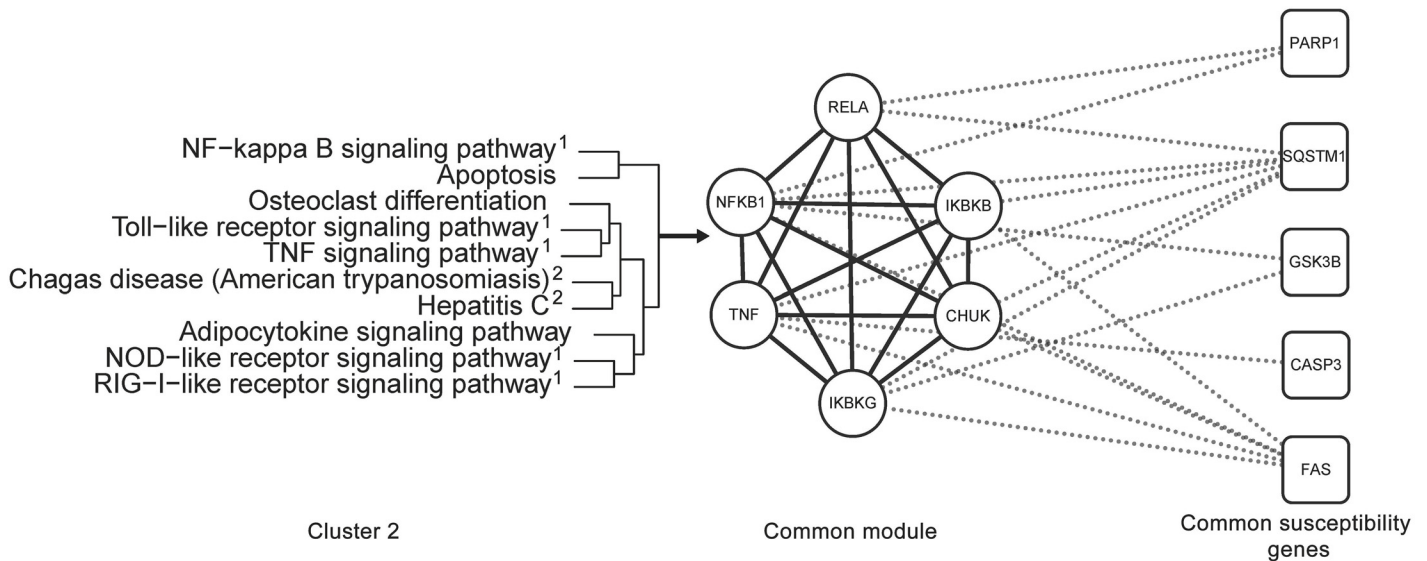
doi:10.1371/journal.pone.0143045.g006

**Fig 7. The common module from cluster 2.** Their relationship with common susceptibility genes of AD, PD and HD is depicted. Module inner interactions are labeled by solid line, its relationship with susceptibility genes by dashed line. Gene names are marked on the nodes. [1]: pathways related to inflammation, [2]: pathways involved in infectious diseases.

NFKB1, IKBKB, TNF, CHUK and IKBKG. Half of the pathways in cluster 2 ([Fig 7](#)) had been found to be directly related to inflammation. Chagas disease and Hepatitis C were involved in infectious diseases. Inflammation and infectious diseases had been shown to be correlated with neurodegeneration. Our study also showed that Osteoclast differentiation might be a common pathway for both osteoporosis and neurodegeneration (section 3.2). The extracted common module's dysfunction, caused by dysregulation of common susceptibility genes, may be a key contributing factor for neurodegenerative disorders, inflammation, infectious diseases and osteoporosis. The found module role in neurodegeneration could thus deserve more in-depth research. Detailed information about other common modules can be found in [S1 Fig](#).

## Conclusion

The traditional drug discovery paradigm of attempting to design precise drugs hitting single targets has seen itself challenged for treatment of complex diseases. The less than perfect efficacy of the single target, single drug approach is mainly due to drug promiscuity, off-target effects, and biological pathway redundancy/robustness. Apparent similarities among groups of closely related disorders hint at possible overlaps in their underlying mechanisms. Figuring out common factors and network modules shared within a group of distinct but related diseases may allow us to pinpoint the fundamental factors responsible for the group of disorders. Computed relationship among pathways of related diseases can assist understanding of their etiology; correlations between the shared pathways with other biological processes/disorders can facilitate drug discovery efforts by suggesting possible treatment candidates for drugs already approved (drug repositioning).

Neurodegenerative disorders including AD, PD and HD were taken as a case study. Their susceptibility genes were collected to compute biological pathways related with neurodegeneration. 81 KEGG pathways were found to be enriched with neurodegenerative genes. Those pathways were involved in cancers, infectious diseases, apoptosis, osteoclast differentiation, and immune system. Sufficient evidences exist for the found correlation between neurodegeneration and the

aforementioned pathways. Our work also showed that adherens and tight junctions, part of the cell communication process, were also correlated with neurodegeneration. Gene expression analysis confirmed that the two junction pathways were indeed correlated with neurodegeneration. The approach applied in this paper can thus be expected to find non-obvious pathways related with a group of closely related disorders. All of these show that a combination of common susceptibility genes and hPPIN is an effective method to study shared pathways involved in a group of related diseases. Not only the functional pathways related with them, but their relationships with other diseases. Moreover, the computed shared pathways can provide mechanistic hypotheses which can guide confirmatory testing to deepen our understanding of the diseases. Common modules bridging distinct pathways were identified by clustering analysis. Those bridging modules may be key points in linking together neurodegeneration and other pathways. Detailed study of the modules may provide potential targets to treat groups of related disorders simultaneously.

## Supporting Information

**S1 File. List of collected gene expression data sets of AD, PD and HD.**
(XLS)

**S2 File. Detailed information of enriched pathways, including the list of associated genes.**
(XLS)

**S3 File. List of significantly expressed genes in AD and PD.**
(XLS)

**S4 File. The components of adherens and tight junction.**
(XLS)

**S1 Fig. The common modules in each cluster.**
(TIFF)

## Author Contributions

Conceived and designed the experiments: PL JY. Performed the experiments: PL. Analyzed the data: PL JY. Contributed reagents/materials/analysis tools: PL YN JY. Wrote the paper: PL JY.

## References

1. Harper S. Economic and social implications of aging societies. Science. 2014; 346(6209):587–91. doi: 10.1126/science.1254405 PMID: 25359967.

2. Tenreiro S, Eckermann K, Outeiro TF. Protein phosphorylation in neurodegeneration: friend or foe? Frontiers in molecular neuroscience. 2014; 7:42. doi: 10.3389/fnmol.2014.00042 PMID: 24860424; PubMed Central PMCID: PMC4026737.

3. Keith CT, Borisy AA, Stockwell BR. Multicomponent therapeutics for networked systems. Nature reviews Drug discovery. 2005; 4(1):71–8. doi: 10.1038/nrd1609 PMID: 15688074.

4. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nature reviews Genetics. 2011; 12(1):56–68. doi: 10.1038/nrg2918 PMID: 21164525; PubMed Central PMCID: PMC3140052.

5. Sun X, Vilar S, Tatonetti NP. High-throughput methods for combinatorial drug discovery. Science translational medicine. 2013; 5(205):205rv1. doi: 10.1126/scitranslmed.3006667 PMID: 24089409.

6. Stavitsky K, Brickman AM, Scarmeas N, Torgan RL, Tang MX, Albert M, et al. The progression of cognition, psychiatric symptoms, and functional abilities in dementia with Lewy bodies and Alzheimer disease. Archives of neurology. 2006; 63(10):1450–6. doi: 10.1001/archneur.63.10.1450 PMID: 17030662.

7. Dewey RB Jr., Taneja A, McClintock SM, Cullum CM, Dewey RB 3rd, Bernstein I, et al. Motor symptoms at onset of Parkinson disease and risk for cognitive impairment and depression. Cognitive and

behavioral neurology: official journal of the Society for Behavioral and Cognitive Neurology. 2012; 25 (3):115–20. doi: 10.1097/WNN.0b013e31826dfd62 PMID: 22960435; PubMed Central PMCID: PMC3477612.

8. Kirkwood SC, Su JL, Conneally P, Foroud T. Progression of symptoms in the early and middle stages of Huntington disease. Archives of neurology. 2001; 58(2):273–8. PMID: 11176966.

9. Navlakha S, Kingsford C. The power of protein interaction networks for associating genes with diseases. Bioinformatics. 2010; 26(8):1057–63. doi: 10.1093/bioinformatics/btq076 PMID: 20185403; PubMed Central PMCID: PMC2853684.

10. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research. 2002; 30(1):207–10. PMID: 11752295; PubMed Central PMCID: PMC99122.

11. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, et al. The BioGRID interaction database: 2013 update. Nucleic acids research. 2013; 41(Database issue):D816–23. doi: 10. 1093/nar/gks1158 PMID: 23203989; PubMed Central PMCID: PMC3531226.

12. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human Protein Reference Database—2009 update. Nucleic acids research. 2009; 37(Database issue):D767–72. doi: 10.1093/nar/gkn892 PMID: 18988627; PubMed Central PMCID: PMC2686490.

13. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic acids research. 2014; 42(Database issue):D358–63. doi: 10.1093/nar/gkt1115 PMID: 24234451; PubMed Central PMCID: PMC3965093.

14. Persico M, Ceol A, Gavrila C, Hoffmann R, Florio A, Cesareni G. HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. BMC bioinformatics. 2005; 6 Suppl 4:S21. doi: 10.1186/1471-2105-6-S4-S21 PMID: 16351748; PubMed Central PMCID: PMC1866386.

15. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nature genetics. 2004; 36(5):431–2. doi: 10.1038/ng0504-431 PMID: 15118671.

16. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, et al. The Comparative Toxicogenomics Database: update 2013. Nucleic acids research. 2013; 41(Database issue): D1104–14. doi: 10.1093/nar/gks994 PMID: 23093600; PubMed Central PMCID: PMC3531134.

17. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic acids research. 2005; 33 (Database issue):D514–7. doi: 10.1093/nar/gki033 PMID: 15608251; PubMed Central PMCID: PMC539987.

18. Qiu C, Kivipelto M, von Strauss E. Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. Dialogues in clinical neuroscience. 2009; 11(2):111–28. PMID: 19585947; PubMed Central PMCID: PMC3181909.

19. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. Lancet neurology. 2006; 5(6):525–35. doi: 10.1016/S1474-4422(06)70471-9 PMID: 16713924.

20. Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. Movement disorders: official journal of the Movement Disorder Society. 2012; 27(9):1083–91. doi: 10.1002/mds.25075 PMID: 22692795.

21. Kohler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. American journal of human genetics. 2008; 82(4):949–58. doi: 10.1016/j.ajhg.2008.02. 013 PMID: 18371930; PubMed Central PMCID: PMC2427257.

22. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009; 25(8):1091–3. doi: 10.1093/bioinformatics/btp101 PMID: 19237447; PubMed Central PMCID: PMC2666812.

23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000; 28(1):27–30. PMID: 10592173; PubMed Central PMCID: PMC102409.

24. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011; 27(3):431–2. doi: 10.1093/bioinformatics/btq675 PMID: 21149340; PubMed Central PMCID: PMC3031041.

25. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995; 57(1):289–300. WOS:A1995QE45300017.

26. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, et al. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. PLoS

genetics. 2007; 3(6):e98. doi: 10.1371/journal.pgen.0030098 PMID: 17571925; PubMed Central PMCID: PMC1904362.

27. Moran LB, Duke DC, Deprez M, Dexter DT, Pearce RK, Graeber MB. Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. Neurogenetics. 2006; 7(1):1–11. doi: 10.1007/s10048-005-0020-2 PMID: 16344956.

28. Zhang Y, James M, Middleton FA, Davis RL. Transcriptional analysis of multiple brain regions in Parkinson's disease supports the involvement of specific protein processing, energy metabolism, and signaling pathways, and suggests novel disease mechanisms. American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics. 2005; 137B(1):5–16. doi: 10.1002/ajmg.b.30195 PMID: 15965975.

29. Zheng B, Liao Z, Locascio JJ, Lesniak KA, Roderick SS, Watt ML, et al. PGC-1alpha, a potential therapeutic target for early intervention in Parkinson's disease. Science translational medicine. 2010; 2 (52):52ra73. doi: 10.1126/scitranslmed.3001059 PMID: 20926834; PubMed Central PMCID: PMC3129986.

30. Wang S, Qaisar U, Yin X, Grammas P. Gene expression profiling in Alzheimer's disease brain microvessels. Journal of Alzheimer's disease: JAD. 2012; 31(1):193–205. doi: 10.3233/JAD-2012-120454 PMID: 22531426.

31. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004; 5(10):R80. doi: 10.1186/gb-2004-5-10-r80 PMID: 15461798; PubMed Central PMCID: PMC545600.

32. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, et al. A comparison of background correction methods for two-colour microarrays. Bioinformatics. 2007; 23(20):2700–7. doi: 10.1093/bioinformatics/btm412 PMID: 17720982.

33. Smyth GK, Speed T. Normalization of cDNA microarray data. Methods. 2003; 31(4):265–73. PMID: 14597310.

34. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology. 2004; 3:Article3. doi: 10.2202/1544-6115.1027 PMID: 16646809.

35. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007; 23(14):1846–7. doi: 10.1093/bioinformatics/btm254 PMID: 17496320.

36. Mitsui T, Hirayama K, Aoki S, Nishikawa K, Uchida K, Matsumoto T, et al. Identification of a novel chemical potentiator and inhibitors of UCH-L1 by in silico drug screening. Neurochemistry international. 2010; 56(5):679–86. doi: 10.1016/j.neuint.2010.01.016 PMID: 20144674.

37. Li T, Yang D, Zhong S, Thomas JM, Xue F, Liu J, et al. Novel LRRK2 GTP-binding inhibitors reduced degeneration in Parkinson's disease cell and mouse models. Human molecular genetics. 2014. doi: 10.1093/hmg/ddu341 PMID: 24993787.

38. Hu S, Begum AN, Jones MR, Oh MS, Beech WK, Beech BH, et al. GSK3 inhibitors show benefits in an Alzheimer's disease (AD) model of neurodegeneration but adverse effects in control animals. Neurobiology of disease. 2009; 33(2):193–206. doi: 10.1016/j.nbd.2008.10.007 PMID: 19038340; PubMed Central PMCID: PMC4313761.

39. DiMauro S, Schon EA. Mitochondrial disorders in the nervous system. Annual review of neuroscience. 2008; 31:91–123. doi: 10.1146/annurev.neuro.30.051606.094302 PMID: 18333761.

40. Aitken A. 14-3-3 proteins: a historic overview. Seminars in cancer biology. 2006; 16(3):162–72. doi: 10.1016/j.semcancer.2006.03.005 PMID: 16678438.

41. Vila M, Przedborski S. Targeting programmed cell death in neurodegenerative diseases. Nat Rev Neurosci. 2003; 4(5):365–75. doi: 10.1038/Nrn1100 WOS:000182665100013. PMID: 12728264

42. Chaturvedi RK, Flint Beal M. Mitochondrial diseases of the brain. Free radical biology & medicine. 2013; 63:1–29. doi: 10.1016/j.freeradbiomed.2013.03.018 PMID: 23567191.

43. Qin C, Zhang C, Zhu F, Xu F, Chen SY, Zhang P, et al. Therapeutic target database update 2014: a resource for targeted therapeutics. Nucleic acids research. 2014; 42(Database issue):D1118–23. doi: 10.1093/nar/gkt1129 PMID: 24265219; PubMed Central PMCID: PMC3964951.

44. Inzelberg R, Jankovic J. Are Parkinson disease patients protected from some but not all cancers? Neurology. 2007; 69(15):1542–50. doi: 10.1212/01.wnl.0000277638.63767.b8 WOS:000250007000011. PMID: 17699801

45. Bennett DA, Leurgans S. Is there a link between cancer and Alzheimer disease? Neurology. 2010; 74 (2):100–1. doi: 10.1212/Wnl.0b013e3181cbb89a WOS:000273563600001. PMID: 20032287

46. Sorensen SA, Fenger K, Olsen JH. Significantly lower incidence of cancer among patients with Huntington disease—An apoptotic effect of an expanded polyglutamine tract? Cancer. 1999; 86(7):1342–6.

doi: 10.1002/(Sici)1097-0142(19991001)86:7<1342::Aid-Cncr33>3.0.Co;2-3 WOS:000082749200033. PMID: 10506723

47. Morris LGT, Veeriah S, Chan TA. Genetic determinants at the interface of cancer and neurodegenerative disease. Oncogene. 2010; 29(24):3453–64. doi: 10.1038/Onc.2010.127 WOS:000278835400001. PMID: 20418918

48. Amor S, Peferoen LA, Vogel DY, Breur M, van der Valk P, Baker D, et al. Inflammation in neurodegenerative diseases—an update. Immunology. 2014; 142(2):151–66. doi: 10.1111/imm.12233 PMID: 24329535; PubMed Central PMCID: PMC4008224.

49. van den Pol AN. Viral infections in the developing and mature brain. Trends Neurosci. 2006; 29(7):398–406. doi: 10.1016/j.tins.2006.06.002 WOS:000239909600008. PMID: 16806513

50. Deleidi M, Isacson O. Viral and Inflammatory Triggers of Neurodegenerative Diseases. Science translational medicine. 2012; 4(121). ARTN 121ps3 doi: 10.1126/scitranslmed.3003492 WOS:000300732600001. PMID: 22344685

51. Mattson MP. Apoptosis in neurodegenerative disorders. Nat Rev Mol Cell Bio. 2000; 1(2):120–9. doi: 10.1038/35040009 WOS:000165765300016.

52. Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD, et al. Structural and functional features of central nervous system lymphatic vessels. Nature. 2015. doi: 10.1038/nature14432 PMID: 26030524.

53. Bjorkqvist M, Wild EJ, Thiele J, Silvestroni A, Andre R, Lahiri N, et al. A novel pathogenic pathway of immune activation detectable before clinical onset in Huntington's disease. J Exp Med. 2008; 205 (8):1869–77. doi: 10.1084/Jem.20080178 WOS:000258528500016. PMID: 18625748

54. Politis M, Pavese N, Tai YF, Kiferle L, Mason SL, Brooks DJ, et al. Microglial activation in regions related to cognitive function predicts disease onset in Huntington's Disease: A multimodal imaging study. Hum Brain Mapp. 2011; 32(2):258–70. doi: 10.1002/Hbm.21008 WOS:000286281300010. PMID: 21229614

55. Song F, Poljak A, Smythe GA, Sachdev P. Plasma biomarkers for mild cognitive impairment and Alzheimer's disease. Brain Res Rev. 2009; 61(2):69–80. doi: 10.1016/j.brainresrev.2009.05.003 WOS:000271169000003. PMID: 19464319

56. Koprich JB, Reske-Nielsen C, Mithal P, Isacson O. Neuroinflammation mediated by IL-1beta increases susceptibility of dopamine neurons to degeneration in an animal model of Parkinson's disease. Journal of neuroinflammation. 2008; 5:8. doi: 10.1186/1742-2094-5-8 PMID: 18304357; PubMed Central PMCID: PMC2292163.

57. Gao HM, Kotzbauer PT, Uryu K, Leight S, Trojanowski JQ, Lee VM. Neuroinflammation and oxidation/ nitration of alpha-synuclein linked to dopaminergic neurodegeneration. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2008; 28(30):7687–98. doi: 10.1523/JNEUROSCI. 0143-07.2008 PMID: 18650345; PubMed Central PMCID: PMC2702093.

58. Frank-Cannon TC, Tran T, Ruhn KA, Martinez TN, Hong J, Marvin M, et al. Parkin deficiency increases vulnerability to inflammation-related nigral degeneration. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2008; 28(43):10825–34. doi: 10.1523/JNEUROSCI.3001-08.2008 PMID: 18945890; PubMed Central PMCID: PMC2603252.

59. Deleidi M, Hallett PJ, Koprich JB, Chung CY, Isacson O. The Toll-like receptor-3 agonist polyinosinic: polycytidylic acid triggers nigrostriatal dopaminergic degeneration. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2010; 30(48):16091–101. doi: 10.1523/JNEUROSCI. 2400-10.2010 PMID: 21123556; PubMed Central PMCID: PMC3075577.

60. Pott Godoy MC, Tarelli R, Ferrari CC, Sarchi MI, Pitossi FJ. Central and systemic IL-1 exacerbates neurodegeneration and motor symptoms in a model of Parkinson's disease. Brain: a journal of neurology. 2008; 131(Pt 7):1880–94. doi: 10.1093/brain/awn101 PMID: 18504291; PubMed Central PMCID: PMC2442423.

61. Chung CY, Koprich JB, Siddiqi H, Isacson O. Dynamic changes in presynaptic and axonal transport proteins combined with striatal neuroinflammation precede dopaminergic neuronal loss in a rat model of AAV alpha-synucleinopathy. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2009; 29(11):3365–73. doi: 10.1523/JNEUROSCI.5427-08.2009 PMID: 19295143; PubMed Central PMCID: PMC2693917.

62. Weller I, Schatzker J. Hip fractures and Alzheimer's disease in elderly institutionalized Canadians. Annals of epidemiology. 2004; 14(5):319–24. doi: 10.1016/j.annepidem.2003.08.005 PMID: 15177270.

63. Li SF, Liu B, Zhang LM, Rong LM. Amyloid beta peptide is elevated in osteoporotic bone tissues and enhances osteoclast function. Bone. 2014; 61:164–75. doi: 10.1016/j.bone.2014.01.010 WOS:000332815700022. PMID: 24473375

64. Caltagarone J, Jing Z, Bowser R. Focal adhesions regulate A beta signaling and cell death in Alzheimer's disease. Bba-Mol Basis Dis. 2007; 1772(4):438–45. doi: 10.1016/j.bbadis.2006.11.007 WOS:000245694800008.

65. Blanc EM, Bruce-Keller AJ, Mattson MP. Astrocytic gap junctional communication decreases neuronal vulnerability to oxidative stress-induced disruption of Ca2+ homeostasis and cell death. Journal of neurochemistry. 1998; 70(3):958–70. WOS:000072132700009. PMID: 9489715

66. Ozog MA, Siushansian R, Naus CCG. Blocked gap junctional coupling increases glutamate-induced neurotoxicity in neuron-astrocyte co-cultures. Journal of neuropathology and experimental neurology. 2002; 61(2):132–41. WOS:000173768100004. PMID: 11855382

67. Abbott NJ, Patabendige AA, Dolman DE, Yusof SR, Begley DJ. Structure and function of the blood-brain barrier. Neurobiology of disease. 2010; 37(1):13–25. doi: 10.1016/j.nbd.2009.07.030 PMID: 19664713.

68. Desai BS, Monahan AJ, Carvey PM, Hendey B. Blood-brain barrier pathology in Alzheimer's and Parkinson's disease: implications for drug therapy. Cell transplantation. 2007; 16(3):285–99. PMID: 17503739.