Methods 131 (2017) 93–103

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Detecting pathway relationship in the context of human protein-protein interaction network and its application to Parkinson's disease

Ying Hu [a], Yichen Yang [a], Zhonghai Fang [a], Yan-Shi Hu [a], Lei Zhang [b,*], Ju Wang [a,*]

[a] School of Biomedical Engineering, Tianjin Medical University, Tianjin 300070, China
[b] School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 15 April 2017
Received in revised form 31 July 2017
Accepted 3 August 2017
Available online 5 August 2017

Keywords:
PPI network
Pathway relationship
Pathway interaction network
Parkinson's disease

ABSTRACT

In human physiological conditions like complex diseases, a large number of genes/proteins, as well as their interactions, are involved. Thus, detecting the biochemical pathways enriched in these genes/proteins and identifying the pathway relationships is critical to understand the molecular mechanisms underlying a disease and can also be valuable in selecting the potential molecular targets for further exploration. In this study, we proposed a method to measure the relationship between pathways based on their distribution in the human PPI network. By representing each pathway as a gene module in the PPI network, a distance was calculated to measure the closeness of two pathways. For the pathways in the KEGG database, a total of 2143 pathway pairs with close connections were identified. Additional evaluations indicated the pathway relationship built via such approach was consistent with available evidence. Further, based on the genes and pathways potentially associated with the pathogenesis of Parkinson's disease (PD), we analyzed the pathway relationship and identified the major pathways related to this disorder via the new method. Also, by analyzing the pathway interaction network constructed by the identified major pathways, we explored the potential pathway targets that may be important in the etiology and development of PD. In summary, we proposed an approach to measure the relationship between pathways, which could provide a more systematic profile on pathways involved in a phenotype, and may also help to improve the result of pathway enrichment analysis.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

In biological system, a pathway is a series of actions or interactions among genes or genes products that leads to the generation of a certain product or a change in a cell. As most biological functions are performed by multiple genes and their products cooperatively, pathways are fundamental for the proper function of a biological system and their dysregulation is often related with dysfunction or diseases. Thus, identifying the molecules involved in a biological process and uncovering the pathways associated with them has become essential in studies aiming at deciphering the molecular mechanisms underlying a phenotype or disease. Especially, with the advancement and wide application of high-throughput techniques such as DNA microarray, proteomic profiling and RNA sequencing, we are able to analyze the expression of many genes or their products simultaneously in a single experiment, through which we often obtain a list of genes/proteins whose expression significantly changed between two conditions. For a given set of candidate genes associated with a phenotype or disease, a quick and effective way to uncover the underlying biological themes is to identify the pathways enriched in the genes, i.e., pathways whose expression patterns or functions are potentially regulated in the conditions or phenotypes under study. Currently, numerous methods have been developed to identify the pathways enriched in a given gene list [1–6], which usually are based on over-representation analysis, functional class scoring or topology analysis.

Similar to the molecular components such as genes and proteins, in a biological process, pathways are also not independent to each other; instead, they usually work together in a highly orchestrated fashion and the function of two pathways can be cooperative, compensatory or alternative. For this reason, the regulation or dysfunction of one pathway may affect other pathways directly or indirectly, and the dysregulation of the pathway interactions may also lead to detrimental consequences in cells; at the same time, modification in a physiological condition is often associated with changes or adaptations in pathway dependencies, with the hub pathways particularly important to these phenotypic changes [7,8]. For instance, synthesized inhibitors targeting core

* Corresponding authors.
E-mail addresses: lzhang@tju.edu.cn (L. Zhang), wangju@tmu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.ymeth.2017.08.001
1046-2023/© 2017 Elsevier Inc. All rights reserved.

proteins in some defected pathways in tumors have drawn much attention for cancer therapy. However, in many cases, the clinical outcomes are not effective as expectation. It is suggested that the complex interactions among the underlying pathways are responsible for such observations, because when a pathway is blockaded by an inhibitor, a compensatory pathway may be modified and the function of the cascade can be recovered or maintained [7]. Thus, after the enriched pathways having been identified from a given gene list, detecting the relationships between pathways is essential for us to obtain a more comprehensive and systematic view about the biological conditions under study. It is not only important for us to properly interpret the biological function of the detected pathways, but also can provide useful insight on the collective behavior and effects of individual genes included in the pathways.

In a biological process, the relationship between pathways may be direct crosstalk, in which two or more pathways share some common components or are involved in a cellular event in a time order. For example, the crosstalk between cAMP and MAPK signaling plays important roles in a series of biological processes [9–11]. While cAMP signaling pathway is involved in the cellular response to various extracellular signals, MAPK signaling pathway can communicate extracellular signals to gene transcription in the nucleus and produces changes in the cell. The two pathways interact with each other via ERK, a protein in the MAPK signaling that can be phosphorylated by PKA, the major intracellular receptor of cAMP. The interactions between some pathways are indirect connection because of the spatial or temporal separation. For example, some components of the signaling pathways inside the cell can affect the structure of the cytoskeleton and thereby the cell's interaction with the extracellular matrix (ECM); on the other hand, the interactions between the ECM and cell can trigger responses within the cell by coordinating the signaling pathways that control the cellular behavior. Through such a transmembrane extracellular matrix-cytoskeleton crosstalk, pathways in a cell can not only affect the properties of the ECM, but also may trigger responses within the neighboring cells [12,13].

To detect the direct crosstalks among pathways, a simple and straightforward way is to compare the genes/proteins and their interactions common to the pathways. For a pair of pathways, a contingency table based on the number of common genes/proteins and common interactions is constructed, and then *p*-value is computed by *Fisher*'s exact test to indicate the statistical significance of such overlap compared to random effect [14–16]. The gene expression profiles can also be used to detect the crosstalk between pathways [8]. In the case of two pathways sharing few or none components, if some genes/proteins in one pathway have direct interaction with the members in the other pathway, protein-protein interaction (PPI) information can be used to identify the pathway relationship. Such an approach is based on assumption that two pathways are likely to interact with or influence each other if significantly more protein interactions are detected between these two pathways than expected by chance [17–22].

However, as mentioned above, the relationship between pathways in real biological systems can be more complex than the direct crosstalk and interaction [7,12,13], which makes these methods less useful for detecting and measuring the more general and complex indirect connection between pathways.

Nowadays, protein-protein interaction network has been recognized as a powerful tool in understanding of how genes perform their biological function. For instance, it is suggested that genes related to a disease often interact with each other and have a tendency to aggregate into a cluster or module in the PPI network [23,24]. Based on such observation, the relationships among the human diseases can be measured by the separation of disease-related module in the human PPI network [25]. As the function of a pathway depends on its molecular components and their interactions, a pathway naturally forms a network of related genes/proteins at the molecular network level, which means the pathway localizations and relationships can be measured and quantified on the basis of PPI network.

In this study, we proposed a method to measure the relationship between pathways based on their distribution in the human PPI network. We first extracted the pathways and the genes included in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and represented each pathway as a gene module in the PPI network and distances between the modules were calculated to measure the closeness of pathways. Then, we applied the proposed methods to analyze the relationships between pathways related to the Parkinson's disease (PD). As we know, PD is a severe neurodegenerative disorder with multiple genes and their mutations involved in the etiology of this disorder [26–30]. The pathogenetic mechanism and potential therapeutic targets underlying PD is still unclear.

2. Methods

2.1. Data source

To explore the correlation and interaction among the pathways and their genes, we compiled a comprehensive protein-protein interaction (PPI) network, based on which the network topological properties of the pathways were calculated and analyzed. Briefly, the human PPI data were obtained from the Protein Interaction Network Analysis (PINA) database [31] by pooling and curating the unique physical interaction information from six main public protein interaction databases, i.e., BioGRID, IntAct, DIP, MINT, MIPS/MPact, and HPRD. After excluding the protein pairs collected by Negatome database (it consists of protein and domain pairs that are unlikely to engage in direct physical interactions) [32], we constructed a human PPI network with 15,202 nodes (genes in our case) and 161,994 edges (connections between genes).

The pathway annotations and genes included in each pathway were extracted from the KEGG database [33]. In the KEGG database, each pathway is assigned to one of seven major categories, i.e., metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and drug development. These major categories are further divided into 56 sub-categories, including carbohydrate metabolism, lipid metabolism, signal transduction, signaling molecules and interaction, etc. In the category of drug development, pathways mainly contain the molecular networks on the chemical structures or chemical components of drug active ingredients, as well as their molecular interaction information. In these pathways, genes and proteins usually are not the major components, so they were not included in this study. Altogether, 299 human-related pathways were collected. It was suggested that a gene set with fewer than 25 genes could hardly form an observable module in PPI network [25]; thus, the sixty pathways with fewer than 25 genes mapped to human PPI network were excluded from further analysis. Finally, we mapped a total of 6614 unique human genes in the 239 pathways to the PPI network. The pathways were from the six categories, i.e., metabolism (38 pathways), genetic information processing (18), environmental information processing (29), cellular processes (15), organismal systems (67), and human diseases (72).

Genes related to Parkinson's disease were collected from two sources. By reviewing the publications deposited in PUBMED, Hu et al. [34] identified 242 genes genetically associated with PD, which provided a comprehensive coverage of genetic factors related to the pathogenetic mechanism underlying PD. In another widely used PD-related genes set, PDGene database [35,36], genes

associated with PD were identified based on a meta-analysis on 15 independent GWAS datasets. Merging the two datasets resulted in a list of 860 unique genes. Of these genes, 703 could be mapped to the human PPI network, and the corresponding gene set was referred to as the PD-related genes list (PDglist) in this study.

2.2. Modularization of pathways in the human PPI network

Since the genes related to a phenotype or biological process often have a tendency to aggregate in the PPI network, it is expected that the genes in a pathway may cluster into one or more modules when they are mapped onto the PPI network. A parameter, module size (S), was employed to measure the modularity of a pathway in the human PPI network, where S was defined as the number of genes in the module with the largest number of connected members of the pathway. For a given pathway with N genes mapped to the PPI network, the statistical significance of the S was estimated by comparing its measured value with those of a set of random control samples. In brief, a simulated pathway with N genes was randomly extracted from the PPI network and its module size S was calculated. This randomization procedure was repeated 10,000 times, and the z-score statistic was used to evaluate the statistical significance:

$$z_S - score = \frac{S - \overline{S^{rand}}}{\sigma(S^{rand})} \quad (1)$$

where $\overline{S^{rand}}$ and $\sigma(S^{rand})$ were the mean value and standard deviation of module size of the simulated pathways, respectively. A significance threshold of $z - score \geq 1.65$ was selected, which corresponded to a significance of $P\text{-}value < 0.05$.

2.3. Detection of pathway-pathway relationship

In the PPI network, the distance between two genes was defined as the length of the shortest path between them, i.e., the smallest number of edges needed to connect the two genes. For a pathway A, we used d_{AA} to denote the mean of the distance of each gene to its closest neighboring gene in the same pathway. For pathway A and B, we used d_{AB} to denote the mean of the distance of each gene in pathway A or B to its closest neighboring gene in the other pathway. Then, a separation value D was used to measure the relationship of two pathways on the PPI network:

$$D_{AB} = d_{AB} - \frac{d_{AA} + d_{BB}}{2} \quad (2)$$

the value of D_{AB} could be positive or negative, with a negative value indicating the existence of overlap between the two pathways, while a positive value implicating the two pathways were connected via some intermediate genes.

To assess the statistical significance of the separation of a pathway pair, D_{AB} was compared with the network separation values obtained from a set of simulated pathway pairs. In brief, two simulated pathways with the same number of genes as pathway A or B were randomly extracted from the PPI. Then, the pathway separation value corresponding to the simulated pathway pair was calculated. This randomization procedure was repeated 1000 times, and the z-score statistic was used to evaluate the statistical significance:

$$z_D - score = \frac{D_{AB} - \overline{D_{AB}^{rand}}}{\sigma(D_{AB}^{rand})} \quad (3)$$

where $\overline{D_{AB}^{rand}}$ and $\sigma(D_{AB}^{rand})$ were the mean value and standard deviation of pathway separation value for the random samples. A threshold of $z_D - score \leq -1.65$ was analytically calculated as significance.

2.4. Pathway enrichment analysis of PDglist

To identify the significantly dysregulated pathways involved in PD, ToppGene [37] was applied to perform the pathway enrichment analysis of the PDglist. Briefly, the symbols of genes in PDglist were uploaded into ToppGene server and compared with the genes included in each pathway in KEGG database; the significantly enriched pathways in PDglist were identified via the Fisher's exact test and p-value was assigned to each pathway. Thereafter, the multiple testing correction p-value (FDR) was calculated with the method of Benjamin and Hockberg [38] and a threshold of FDR < 0.01 was adopted to select the significantly enriched pathways.

2.5. Pathway crosstalk analysis

For the enriched pathways enriched in PDglist, we also performed pathway crosstalk analysis to explore the interactions among pathways. To describe the overlap between a given pair of pathways, we adopted two measurements [39,40], i.e., the Jaccard Coefficient(JC) $= |\frac{A \cap B}{A \cup B}|$ and the Overlap Coefficient(OC) $= \frac{|A \cap B|}{\min(|A|,|B|)}$ where A and B are the lists of genes included in the two tested pathways. To construct the pathway crosstalk, we implemented the following procedure:

1) Select the significantly enriched pathways in PDglist (i.e., FDR < 0.01) for crosstalk analysis. Meanwhile, the pathways containing less than 5 candidate genes were removed because pathways with too few genes may have insufficient biological information.
2) Count the number of shared candidate genes between any pair of pathways. Pathway pair with less than 3 overlapped genes was removed.
3) Calculate the overlap of all pathway pairs and rank them. All the pathway pairs were ranked according to the average of their JC and OC values. Only the pathway pairs with crosstalk scores in the top 10% were chosen.

The parameters were chosen for a balance between an appropriate number of pathways and crosstalk events [39].

3. Results and discussion

3.1. Pathways modularization in human PPI network

For the human-related pathways in the KEGG pathway database, there were 239 pathways with more than 25 mapped genes on the human PPI network. To check whether a pathway could form an observable module in the PPI network, we calculated the module size S (i.e., the size of the largest connected components) of the sub network formed by the genes in each pathway. According to S value, there were 205 out of 239 pathways with $z_S - score \geq 1.65$, suggesting that more than 85% (205/239) pathways have a significant observable module in human PPI network. The module size S was in the range 2–349 and were larger than 100 for 20% (47/239) of the pathways. As a comparison, for 98% (234/239) of the pathways, the mean module sizes of the simulated pathways were smaller than 20 (Fig. 1a). Besides, with the increase of the size of the pathways (the number of genes in a pathway could be mapped to the PPI network), module sizes for the real pathways increased more rapidly than the simulated pathways (Fig. 1b).

There were 34 pathways without significant observable module size, which meant the genes in these pathways could not form connected modules in the PPI network. Most of these pathways were

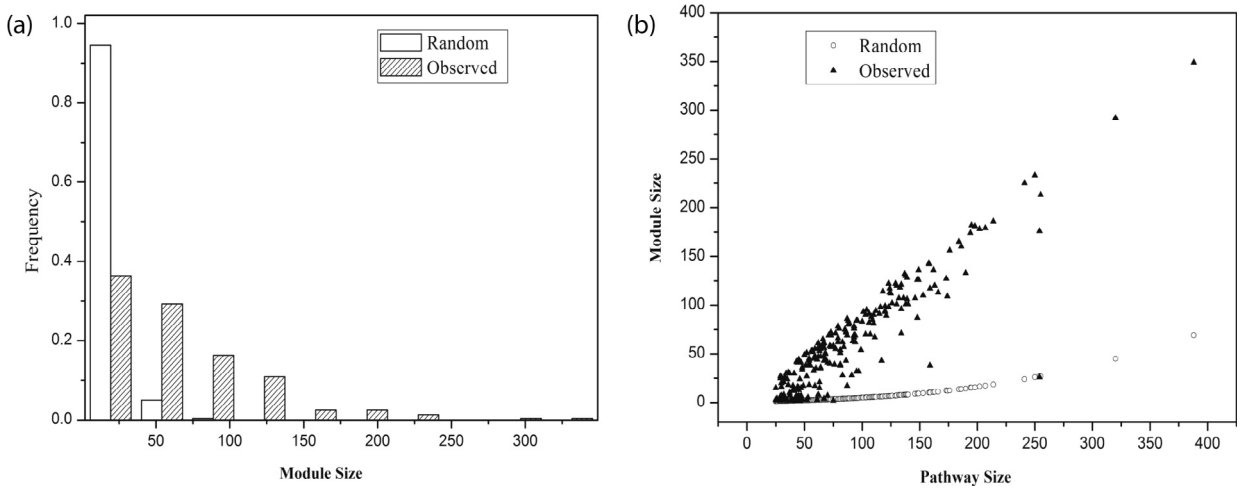

Fig. 1. Modularization of pathways in the human PPI network. The genes contained in a pathway tend to aggregate in the PPI network and cluster into one or more modules. Modularity of a pathway on the PPI network can be measured by its module size. For each pathway, 10,000 simulated pathways are generated, and each simulated pathway is generated by randomly selecting genes from the human PPI network with the same number of genes as the corresponding real pathway mapped to the PPI network. The module size of the simulated pathways corresponding to a real pathway is the average of the module sizes of all the simulated pathways. (a) Distribution of module size (largest connected component) *S* of KEGG pathways (denoted as observed in the figure) and simulated pathways (denoted as random in the figure) in the human PPI network. Compared with the simulated pathways, the module sizes of the real pathways have a much wider range and are also more evenly distributed, suggesting the genes in a real pathway are more likely to be connected with each other. (b) The correlation between module size and the pathway size on the human PPI network. For real pathways, the module size increases almost linearly with the increase of pathway size (number of genes of a pathway mapped to the PPI network). For the simulated pathways, however, there is little or no increase in module size when the pathway size increases.

metabolism related. Actually, for the 38 pathways categorized as metabolism in KEGG database, 23 pathways were among these 34 pathways. A close check on these pathways showed that the metabolism pathways in KEGG database had some features different than other pathways. While pathways in the other categories usually were composed of genes and their products, most metabolism pathways showed the relation of genes (such as enzymes) and metabolism products. Since many genes in metabolism pathways only specifically functioned on certain chemical substances and did not have explicit connections with other genes, they tended to be more 'loosely' connected on the PPI network. Actually, the modularity of a pathway could also be measured by the mean of the distance of each gene to its closest neighboring gene in the same pathway (d_{AA} for pathway A). Compared with the simulated pathways, all the 239 pathways had p-value < 0.05 when the values of d_{AA} were compared, indicating the genes in the real pathways were more closely connected.

3.2. Pathways relationships

For the 205 pathways with significant module size on the human PPI network, their relationships were further evaluated. For each pathway pair, the network-based separation value D was calculated and its significance level was evaluated based on the z-score. Altogether, there were 20,910 pathway pairs ($205 \times 204/2$). Of these pathway pairs, 2150 (10.28%) had D smaller than 0, which meant they had overlap neighborhood. Of them, 2143 pairs showed statistical significance with p-value < 0.05 ($z_D - score \leq -1.65$). For the remaining 18,760 pathway pairs (89.72%), the D values were positive and 2298 of them were statistically significant.

As specified earlier, in the KEGG pathway database, the pathways were grouped into six major categories based on the biological processes they were involved, i.e., metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems and human diseases, with pathways in each category more closely related. To test whether the network-based separation value could reflect the relationship between pathways, the mean of D was calculated for pathways within each category and category pair (Table 1). Compared with those in other categories, most of pathways in each of the six categories had smaller average D values, consistent with the assumption that pathways within the same category were more similar to each other than those in different categories. We also compared the mean of the shortest distance (d) for pathways within each category and between two categories (Table 1). Similar to the average D values, the d values for pathways within each of category were relatively small compared to those in other categories.

In KEGG pathway database, the pathways in each category are further clustered into more specific sub-categories. For each of the six categories, we also compared the mean of D for pathways within each sub-category and between every two sub-categories.

Table 1
The average network-based separation values and shortest distance values within and between different pathway categories.

	Metabolism	Genetic information processing	Environmental information processing	Cellular processes	Organismal systems	Human diseases
Metabolism	0.30 (1.52)*	0.59 (1.91)	0.62 (1.98)	0.57 (1.92)	0.57 (1.97)	0.56 (1.92)
Genetic information processing		0.60 (1.12)	0.67 (1.84)	0.63 (1.78)	0.68 (1.87)	0.63 (1.79)
Environmental information processing			0.30 (1.20)	0.36 (1.55)	0.33 (1.56)	0.32 (1.51)
Cellular processes				0.37 (1.19)	0.40 (1.65)	0.37 (1.55)
Organismal systems					0.27 (1.27)	0.34 (1.57)
Human diseases						0.27 (1.19)

* The values inside the parentheses are the mean of shortest distance (d_{AA} or d_{AB}) for pathways within each category or between two categories.

In most cases, the average D values were smaller for pathways in the same sub-category than those in different sub-categories. A comparison of the D values for human diseases related pathways was shown in Table 2. For pathways in nine of the ten sub-categories, average D values for pathways within the sub-categories were smaller than those between sub-categories. For the sub-category 'endocrine and metabolic diseases', the average D value for pathways in this sub-category was slightly larger ($D = 0.17$) than that for pathways between 'endocrine and metabolic diseases' and 'cancers: overview' (0.15), 'endocrine and metabolic diseases' and 'cancers: specific types' (0.16), as well as 'endocrine and metabolic diseases' and 'infectious diseases: viral' (0.12). This result implied that a dysfunction of the endocrine and metabolic system was likely to be accompanied by other disorders, which was consistent with previous studies [41–43]. The values inside the parentheses were the mean of shortest distance (d) for pathways within each sub-category and between every two sub-categories. These results demonstrated that pathways within the same category clustered more closely in the PPI network.

For each pathway pair, the value of D reflected their relationship in function. The pathway pairs with smaller D values tended to be more closely related in function. For example, for the 20 pathway pairs with the smallest D values (Table 3), most (15/20) pairs consisted of pathways from the same category (e.g., cardiovascular diseases and dilated cardiomyopathy, cocaine addiction and amphetamine addiction, Parkinson's disease and Alzheimer's disease). The pathways in these pairs belonged to the same KEGG pathway categories and had similar biological functions. Some pathway pairs with interesting relationship could also be found in the list. For example, the two pathways, type I diabetes mellitus and allograft rejection, were in the two seemingly unrelated sub-category 'endocrine and metabolic diseases' and 'immune diseases'; however, the small separation value between the two pathways suggested they were closely related. Actually, previous studies showed that post-transplant diabetes was a major complication after kidney or liver transplantation [44,45]. Of these pathway pairs, there were also some (5/20) from different categories, i.e., hedgehog signaling pathway and basal cell carcinoma, retrograde endocannabinoid signaling and morphine addiction, GABAergic synapse and morphine addiction, oxidative phosphorylation and Parkinson's disease, VEGF signaling pathway and Fc epsilon RI signaling pathway. Although these pathway pairs are clustered into different categories in the KEGG pathway database, available studies suggest that the pathways in each pair may be involved in the same diseases or biological processes. Take morphine addiction as an example, evidences demonstrate that morphine exposure can modulate the effect of long-term potentiation in GABAergic synapse [46] and have a major influence on the GABAergic transmission [47]. The altered GABAergic synapse, in turn, can induce or accelerate the process of the morphine addiction [48]. Besides, those pathway pairs from categories of organismal system and human diseases may suggest the organismal systems suffering the functional dysregulation in a specific human disease. Additionally, the two pathways with close connections may have compensatory relationships and influence the effective of targeted therapies. For example, PD325901, a MAP/ERK kinase (MEK) inhibitor, can be used to treat prostate cancer potentiation by regulating the expression several components of PI3K-Akt signaling pathway and extracellular signal-regulated kinase (ERK) signaling pathway [49]. In line with this observation, MAPK signaling pathway and PI3K-Akt signaling pathway has a significantly closely pathways ($Dab = -0.04$) in our list. In addiction, MAPK signaling pathway was found to be closely related to Rap1 signaling pathway ($Dab = -0.167$) and Ras signaling pathway ($Dab = -0.311$). These alternative pathways parallel to the 'BRAF-

Table 2
The average network-based separation values and shortest distance values within and between different sub-categories of pathways related to human diseases*.

	Cancers: overview	Cancers: specific types	Immune diseases	Neurodegenerative diseases	Substance dependence	Cardiovascular diseases	Endocrine and metabolic diseases	Infectious diseases: bacterial	Infectious diseases: viral	Infectious diseases: parasitic
Cancers: overview	−0.05 (1.20)**	−0.02 (1.11)	0.51 (1.76)	0.30 (1.53)	0.32 (1.59)	0.36 (1.57)	0.15 (1.36)	0.25 (1.44)	0.02 (1.17)	0.21 (1.42)
Cancers: specific types		−0.23 (1.06)	0.56 (1.75)	0.35 (1.51)	0.40 (1.60)	0.43 (1.58)	0.16 (1.29)	0.27 (1.39)	0.11 (1.19)	0.25 (1.39)
Immune diseases			0.04 (1.31)	0.60 (1.88)	0.68 (2.01)	0.50 (1.76)	0.34 (1.60)	0.44 (1.68)	0.33 (1.54)	0.25 (1.57)
Neurodegenerative diseases				0.04 (1.23)	0.33 (1.63)	0.47 (1.71)	0.27 (1.50)	0.38 (1.59)	0.29 (1.47)	0.36 (1.61)
Substance dependence					−0.07 (1.35)	0.46 (1.74)	0.44 (1.72)	0.45 (1.71)	0.39 (1.61)	0.45 (1.73)
Cardiovascular diseases						−0.24 (1.24)	0.40 (1.62)	0.44 (1.64)	0.43 (1.59)	0.37 (1.59)
Endocrine and metabolic diseases							0.17 (1.21)	0.31 (1.50)	0.12 (1.28)	0.17 (1.39)
Infectious diseases: Bacterial								0.20 (1.17)	0.20 (1.33)	0.17 (1.37)
Infectious diseases: viral									−0.2 (1.10)	0.08 (1.25)
Infectious diseases: parasitic										−0.07 (1.23)

* In KEGG pathway database, the human diseases related pathways are clustered into 12 sub-categories. For the 205 pathways with significant observable module in human PPI network, 71 belong to human diseases. These pathways are from the ten sub-categories listed in this table.

** The values inside the parentheses are the mean of shortest distance (d_{AA} or d_{AB}) for pathways within each sub-category or between two sub-categories.

Table 3
Pathway pairs with closely relationships (top 20 pathway pairs).

Pathway A			Pathway B			D_{AB}
KEGG ID	Name	Category*	KEGG ID	Name	Category*	
hsa05410	Hypertrophic cardiomyopathy	Human diseases (Cardiovascular diseases)	hsa05414	Dilated cardiomyopathy	Human Diseases (Cardiovascular diseases)	−0.993
hsa05320	Autoimmune thyroid disease	Human diseases (Immune diseases)	hsa05330	Allograft rejection	Human Diseases (Immune diseases)	−0.869
hsa05330	Allograft rejection	Human diseases (Immune diseases)	hsa05332	Graft-versus-host disease	Human Diseases (Immune diseases)	−0.834
hsa04940	Type I diabetes mellitus	Human diseases (Endocrine and metabolic diseases)	hsa05332	Graft-versus-host disease	Human Diseases (Immune diseases)	−0.833
hsa04940	Type I diabetes mellitus	Human diseases (Endocrine and metabolic diseases)	hsa05330	Allograft rejection	Human Diseases (Immune diseases)	−0.824
hsa04340	Hedgehog signaling pathway	Environmental information processing (Signal transduction)	hsa05217	Basal cell carcinoma	Human disease (Cancers: specific types)	−0.813
hsa04727	GABAergic synapse	Organismal systems (Nervous system)	hsa05032	Morphine addiction	Human diseases (Substance dependence)	−0.782
hsa04723	Retrograde endocannabinoid signaling	Organismal systems (Nervous system)	hsa04727	GABAergic synapse	Organismal systems (Nervous system)	−0.771
hsa00190	Oxidative phosphorylation	Metabolism (Energy metabolism)	hsa05012	Parkinson's disease	Human diseases (Neurodegenerative diseases)	−0.757
hsa04723	Retrograde endocannabinoid signaling	Organismal systems (Nervous system)	hsa05032	Morphine addiction	Human diseases (Substance dependence)	−0.729
hsa05030	Cocaine addiction	Human diseases (Substance dependence)	hsa05031	Amphetamine addiction	Human diseases (Substance dependence)	−0.712
hsa05410	Hypertrophic cardiomyopathy	Human diseases (Cardiovascular diseases)	hsa05412	Arrhythmogenic right ventricular cardiomyopathy	Human diseases (Cardiovascular diseases)	−0.709
hsa05214	Glioma	Human disease (Cancers: specific types)	hsa05223	Non-small cell lung cancer	Human disease (Cancers: specific types)	−0.697
hsa05412	Arrhythmogenic right ventricular cardiomyopathy	Human diseases (Cardiovascular diseases)	hsa05414	Dilated cardiomyopathy	Human diseases (Cardiovascular diseases)	−0.679
hsa05010	Alzheimer's disease	Human diseases (Neurodegenerative diseases)	hsa05012	Parkinson's disease	Human diseases (Neurodegenerative diseases)	−0.652
hsa04370	VEGF signaling pathway	Environmental information processing (Signal transduction)	hsa04664	Fc epsilon RI signaling pathway	Organismal systems (Immune system)	−0.639
hsa05320	Autoimmune thyroid disease	Human diseases (Immune diseases)	hsa05332	Graft-versus-host disease	Human diseases (Immune diseases)	−0.635
hsa00230	Purine metabolism	Metabolism (Nucleotide metabolism)	hsa00240	Pyrimidine metabolism	Metabolism (Nucleotide metabolism)	−0.623
hsa04713	Circadian entrainment	Organismal systems (Environmental adaptation)	hsa04723	Retrograde endocannabinoid signaling	Organismal systems (Nervous system)	−0.618
hsa04911	Insulin secretion	Organismal systems (Endocrine system)	hsa04925	Aldosterone synthesis and secretion	Organismal systems (Endocrine system)	−0.608

* The sub-category of the pathway is included in the parentheses.

MEK-ERK' signaling pathway has been demonstrated to contribute the acquired resistance of PLX4032, a selective BRAF inhibitor in melanoma therapeutic [50].

Thus, the pathway separation value *D*, could be a useful measurement of the relationship between pathways.

3.3. The relationships among pathways involved in Parkinson's disease

3.3.1. Pathway enrichment analysis in PDglist

Identifying biological pathways enriched in the candidate genes may provide meaningful information to give insight on the molecular mechanism underlying Parkinson's disease. We identified the significantly enriched pathways in the PDglist by ToppGene and found 79 significant enrichment pathways with FDR < 0.01 (Supplementary Table 1). The top 20 significantly pathways were shown in Table 4. Among these pathways, several pathways associated with human diseases were included, including neurodegenerative diseases, infectious diseases, immune diseases, as well as endocrine and metabolic diseases. These results were consistent with the previous studies; for instance, a case-control study suggests that infection is a risk factor for Parkinson's disease [51], and a recent analysis shows that Parkinson's disease has a closely correlation with immune diseases such as asthma [52] or inflammatory bowel disease [53]. In addition, signal transduction pathways, e.g., calcium signaling pathway, PI3K-Akt signaling pathway, MAPK signaling pathway and HIF-1 signaling pathway were found to be enriched in PDglist. We also identified pathways-related to neurotransmitters, such as dopaminergic synapse, serotonergic synapse, and cholinergic synapse, pathway playing important roles in various biological processes in the nervous system. Besides, immune system- and endocrine system-related pathways, such as the T cell receptor signaling pathway, chemokine signaling pathway, estrogen signaling pathway, were enriched in PDglist, which were consistent with available results [34]. Further, pathways related to transport and catabolism such as apoptosis and focal adhesion were found to be enriched in the PDglist, suggesting that programmed cellular death were critical in the etiology and pathological process of Parkinson's disease, in line with prior knowledge [54]. In summary, as a complex disease, the occurrence and development of Parkinson's disease are accompanied by several dysfunctional pathways, mainly including pathways associated with immune, endocrine and metabolic, neurodevelopment, infectious and cellular processes. Of note, the pathway enrichment analysis results also provided some explanations to the comorbidity of PD and other diseases.

3.3.2. The relationships among PD-related pathways

Then, the relationships among the enriched pathways were further evaluated. By calculating the separation value of PD-related enriched pathways in the human PPI network, the pathway pairs with significantly close relationships were obtained (Table 5). Among the pathway pairs with the smallest *D* values (20 pairs),

Table 4
Pathways enriched in the Parkinson's diseases-related genes (PDglist) (Top 20 pathways).

Pathway ID	Pathway Name	p-value	FDR	Genes in PDglist included in the pathway
hsa05012	Parkinson's disease	4.48×10^{-25}	1.19×10^{-22}	APAF1, CASP3, CASP9, GPR37, HTRA2, LRRK2, MT-ATP6, MT-ATP8, MT-CO1, MT-CO2, MT-CO3, MT-CYB, MT-ND1, MT-ND2, MT-ND3, MT-ND4, MT-ND5, MT-ND6, NDUFA1, NDUFA10, NDUFA6, NDUFA7, NDUFA8, NDUFB4, NDUFB7, NDUFB8, NDUFB9, NDUFS1, NDUFS2, NDUFS4, NDUFS7, NDUFS8, NDUFV2, PARK2, PARK7, PINK1, SLC18A2, SLC6A3, SNCA, SNCAIP, TH, UBE2L3, UCHL1, UQCRH
hsa05010	Alzheimer's disease	3.27×10^{-20}	4.35×10^{-18}	APAF1, APOE, BAD, CAPN2, CASP3, CASP8, CASP9, CDK5R1, GRIN1, GRIN2A, GRIN2B, GSK3B, IDE, IL1B, LRP1, MAPT, MT-ATP6, MT-ATP8, MT-CO1, MT-CO2, MT-CO3, MT-CYB, NDUFA1, NDUFA10, NDUFA6, NDUFA7, NDUFA8, NDUFB4, NDUFB7, NDUFB8, NDUFB9, NDUFS1, NDUFS2, NDUFS4, NDUFS7, NDUFS8, NDUFV2, NOS1, PLCB4, PPP3CA, SNCA, TNF, TNFRSF1A, UQCRH
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	1.15×10^{-18}	1.02×10^{-16}	AKT1, BAX, CASP3, CASP8, CDC42, CXCL8, CYP2E1, GSK3B, IL1A, IL1B, IL6, INS, JUN, MT-CO1, MT-CO2, MT-CO3, MT-CYB, NDUFA1, NDUFA10, NDUFA6, NDUFA7, NDUFA8, NDUFB4, NDUFB7, NDUFB8, NDUFB9, NDUFS1, NDUFS2, NDUFS4, NDUFS7, NDUFS8, NDUFV2, NFKB1, PIK3R3, PRKAA2, SREBF1, TGFB1, TNF, TNFRSF1A, UQCRH
hsa05152	Tuberculosis	7.47×10^{-15}	4.97×10^{-13}	AKT1, APAF1, ATP6V0B, BAD, BAX, BCL2, CASP3, CASP8, CASP9, CD14, CR1, CREB1, CTSD, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, HSPA9, IFNG, IFNGR2, IL10, IL18, IL1A, IL1B, IL6, ITGB2, LAMP1, NFKB1, NFYC, NOD2, NOS2, PPP3CA, TGFB1, TGFB2, TNF, TNFRSF1A, VDR
hsa05016	Huntington's disease	4.88×10^{-14}	2.60×10^{-12}	APAF1, BAX, BDNF, CASP3, CASP8, CASP9, CREB1, DCTN1, GRIN1, GRIN2B, MT-ATP6, MT-ATP8, MT-CO1, MT-CO2, MT-CO3, MT-CYB, NDUFA1, NDUFA10, NDUFA6, NDUFA7, NDUFA8, NDUFB4, NDUFB7, NDUFB8, NDUFB9, NDUFS1, NDUFS2, NDUFS4, NDUFS7, NDUFS8, NDUFV2, PLCB4, PPARGC1A, RCOR1, SOD2, TBP, TFAM, UQCRH
hsa05323	Rheumatoid arthritis	3.66×10^{-13}	1.62×10^{-11}	ATP6V0B, CCL2, CCL5, CTSL, CXCL8, FLT1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, ICAM1, IFNG, IL18, IL1A, IL1B, IL6, ITGB2, JUN, MMP1, MMP3, TGFB1, TGFB2, TNF, VEGFA
hsa04612	Antigen processing and presentation	9.92×10^{-13}	3.77×10^{-11}	B2M, CALR, CANX, CREB1, CTSB, CTSL, HLA-A, HLA-B, HLA-C, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, HSP90AA1, HSPA1A, HSPA1L, HSPA5, HSPA8, IFNG, NFYC, PDIA3, TNF
hsa05321	Inflammatory bowel disease (IBD)	1.28×10^{-12}	4.26×10^{-11}	HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, IFNG, IFNGR2, IL10, IL18, IL1A, IL1B, IL2, IL4, IL6, JUN, NFKB1, NOD2, RORA, TGFB1, TGFB2, TNF
hsa05204	Chemical carcinogenesis	4.49×10^{-12}	1.33×10^{-10}	ADH1B, ADH4, ADH7, ARNT, CHRNA7, CYP1A1, CYP1A2, CYP1B1, CYP2C19, CYP2C9, CYP2E1, EPHX1, GSTA4, GSTM1, GSTM3, GSTO1, GSTO2, GSTP1, GSTT1, MGST2, NAT1, NAT2, PTGS2
hsa05145	Toxoplasmosis	8.92×10^{-12}	2.37×10^{-10}	AKT1, BAD, BCL2, CASP3, CASP8, CASP9, CCR5, GNAI3, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, HSPA1A, HSPA1L, HSPA8, IFNG, IFNGR2, IL10, MAP2K6, NFKB1, NOS2, PIK3R3, TGFB1, TGFB2, TNF, TNFRSF1A
hsa00190	Oxidative phosphorylation	1.46×10^{-11}	3.54×10^{-10}	ATP6V0B, MT-ATP6, MT-ATP8, MT-CO1, MT-CO2, MT-CO3, MT-CYB, MT-ND1, MT-ND2, MT-ND3, MT-ND4, MT-ND5, MT-ND6, NDUFA1, NDUFA10, NDUFA6, NDUFA7, NDUFA8, NDUFB4, NDUFB7, NDUFB8, NDUFB9, NDUFS1, NDUFS2, NDUFS4, NDUFS7, NDUFS8, NDUFV2, UQCRH
hsa05140	Leishmaniasis	8.88×10^{-11}	1.97×10^{-9}	CR1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, IFNG, IFNGR2, IL10, IL1A, IL1B, IL4, ITGB2, JUN, NFKB1, NOS2, PTGS2, TGFB1, TGFB2, TNF
hsa05030	Cocaine addiction	1.50×10^{-10}	3.08×10^{-9}	BDNF, CDK5R1, CREB1, DDC, DRD1, DRD2, GNAI3, GRIN1, GRIN2A, GRIN2B, JUN, MAOA, MAOB, NFKB1, SLC18A2, SLC6A3, TH
hsa05142	Chagas disease (American trypanosomiasis)	2.47×10^{-10}	4.70×10^{-9}	ACE, AKT1, CALR, CASP8, CCL2, CCL5, CXCL8, GNAI3, IFNG, IFNGR2, IL10, IL1B, IL2, IL6, JUN, NFKB1, NOS2, PIK3R3, PLCB4, PPP2R2B, TGFB1, TGFB2, TNF, TNFRSF1A
hsa05134	Legionellosis	8.22×10^{-10}	1.46×10^{-8}	APAF1, CASP3, CASP8, CASP9, CD14, CR1, CXCL8, HSPA1A, HSPA1L, HSPA8, IL18, IL1B, IL6, ITGB2, NFKB1, TNF, VCP
hsa05014	Amyotrophic lateral sclerosis (ALS)	2.03×10^{-9}	3.33×10^{-8}	APAF1, BAD, BAX, BCL2, CASP3, CASP9, CAT, GRIN1, GRIN2A, GRIN2B, MAP2K6, NEFL, NOS1, PPP3CA, TNF, TNFRSF1A
hsa05169	Epstein-Barr virus infection	2.13×10^{-9}	3.33×10^{-8}	AKT1, BCL2, BST1, CR2, CSNK2A1, CSNK2A2, CSNK2B, FCER2, GSK3B, HLA-A, HLA-B, HLA-C, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, HSPA1A, HSPA1L, HSPA8, ICAM1, IFNG, IL10, JUN, MAP2K6, MAP3K14, NFKB1, PIK3R3, PSMC4, SND1, TBP, YWHAH
hsa05332	Graft-versus-host disease	2.52×10^{-9}	3.73×10^{-8}	HLA-A, HLA-B, HLA-C, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, IFNG, IL1A, IL1B, IL2, IL6, TNF
hsa05166	HTLV-I infection	3.95×10^{-9}	5.53×10^{-8}	AKT1, ATR, BAX, CALR, CANX, CDC20, CDC27, CDKN2C, CREB1, CREM, GSK3B, HLA-A, HLA-B, HLA-C, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5, ICAM1, IL2, IL2RA, IL6, ITGB2, JUN, MAP3K14, MRAS, NFKB1, PIK3R3, PPP3CA, RRAS2, TBP, TGFB1, TGFB2, TNF, TNFRSF1A, WNT3
hsa00982	Drug metabolism-cytochrome P450	4.25×10^{-9}	5.66×10^{-8}	ADH1B, ADH4, ADH7, CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP2E1, GSTA4, GSTM1, GSTM3, GSTO1, GSTO2, GSTP1, GSTT1, MAOA, MAOB, MGST2

8 pairs were involved in neurodegenerative diseases, which were in line with the knowledge on the pathology of PD. Based on their *D* values, the pathway pairs were linked to construct a PD-related pathway network, which included 61 enriched nodes (pathways) and 311 edges (pathway pairs) (Fig. 2). According to its topological structure, the PD-related pathways network could be roughly divided into four major modules, each of which included pathways had closer relationships compared with other pathways and might likely be participated in the same or similar biological process. The first module mainly consisted of pathways associated with neurodegenerative diseases, including pathways like Parkinson's disease, Alzheimer's disease and Huntington's disease. Of note, as a pathway related to energy metabolism, oxidative phosphorylation has closely relationships with neurodegenerative diseases' pathways. This result was consistent with previous reports that oxidative stress had an ubiquitous role in neurodegenerative diseases [55]. The second module was primarily composed of neurodevelopment-related signaling pathways, such as dopaminergic synapse, serotonergic synapse, cholinergic synapse and substance dependence related pathways. The third module was dominated by pathways associated with immune diseases including asthma, autoimmune thyroid disease and rheumatoid arthritis.

Table 5
Parkinson's disease-related pathway pairs with close relationships (top 20 pathway pairs).

Pathway A		Pathway B		D_{AB}
KEGG ID	Name	KEGG ID	Name	
hsa05320	Autoimmune thyroid disease	hsa05330	Allograft rejection	−0.869
hsa05330	Allograft rejection	hsa05332	Graft-versus-host disease	−0.834
hsa04940	Type I diabetes mellitus	hsa05332	Graft-versus-host disease	−0.833
hsa04940	Type I diabetes mellitus	hsa05330	Allograft rejection	−0.824
hsa00190	Oxidative phosphorylation	hsa05012	Parkinson's disease	−0.757
hsa05030	Cocaine addiction	hsa05031	Amphetamine addiction	−0.712
hsa05010	Alzheimer's disease	hsa05012	Parkinson's disease	−0.652
hsa05320	Autoimmune thyroid disease	hsa05332	Graft-versus-host disease	−0.635
hsa05010	Alzheimer's disease	hsa05016	Huntington's disease	−0.590
hsa05012	Parkinson's disease	hsa05016	Huntington's disease	−0.579
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	hsa05010	Alzheimer's disease	−0.576
hsa00190	Oxidative phosphorylation	hsa05010	Alzheimer's disease	−0.569
hsa04940	Type I diabetes mellitus	hsa05320	Autoimmune thyroid disease	−0.555
hsa05210	Colorectal cancer	hsa05212	Pancreatic cancer	−0.516
hsa04014	Ras signaling pathway	hsa04015	Rap1 signaling pathway	−0.512
hsa05310	Asthma	hsa05330	Allograft rejection	−0.506
hsa00190	Oxidative phosphorylation	hsa05016	Huntington's disease	−0.505
hsa04932	Non-alcoholic fatty liver disease (NAFLD)	hsa05012	Parkinson's disease	−0.480
hsa04728	Dopaminergic synapse	hsa05031	Amphetamine addiction	−0.473
hsa04725	Cholinergic synapse	hsa04915	Estrogen signaling pathway	−0.472

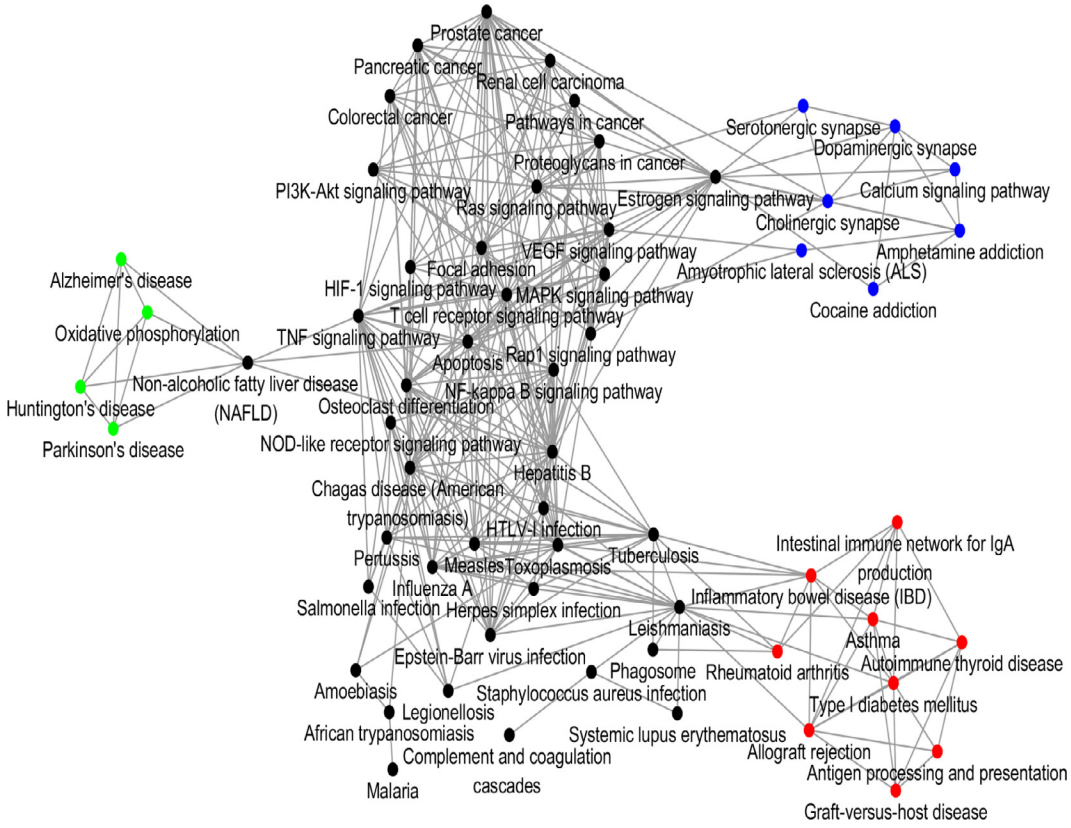

Fig. 2. Pathway interaction network of the PD-related pathways. The PD-related pathway pairs with significantly close relationships were linked to construct a PD-related pathway network, which included 61 enriched nodes (pathways) and 311 edges (pathway pairs). Nodes represent enriched pathways and edges represent closely relationships between pathways. Node size corresponds to the negative logarithm of FDR for each pathway.

The forth module was the central and the most complex with comprising of pathways associated with multiple systems such as immune, endocrine and metabolic, transport and catabolism. Among these central pathways, we found that pathways associated with signal transduction and cellular processes including T cell receptor signaling pathway, HIF-1 signaling pathway, MAPK signaling pathway, NF-kappa B signaling pathway and apoptosis, focal adhesion were the core pathways which connected to other pathways with the most frequency, suggesting that these pathways might play important roles in the development of Parkinson's disease. Meanwhile, pathways related to infectious diseases and cancers were involved in central module. As mentioned above, several previous reports pointed out that there was a close relationship between infection and Parkinson's disease and infection could be regarded as a risk factor for PD [51]. These infectious diseases-related pathways were mainly linked to NF-kappa B signaling

pathway, T cell receptor signaling pathway, NOD-like receptor signaling pathway and apoptosis, implying that infection events might trigger the dysfunction of immune system and play a role in the development of PD. Currently, the relationship between PD and cancer is controversial [56]. Indeed, the pathways associated with cancer were connected to various systems' pathways in PD-related pathway relationship network. Of note, all the modules were not isolated; instead, the central module was connected with the other three periphery modules via a couple of pathways. Such results could shed new light on the molecular mechanism underlying this disease and provide meaningful information on predicting the novel molecular targets that may be important in the etiology and development of the disease.

For a comparison, we also performed pathway crosstalk analysis on the same set of pathways significantly enriched in PDglist. Crosstalk scores among 78 enriched pathways containing 5 or more genes (79 significantly enriched pathways in total) in PDglist were calculated. There were a total of 1206 connections meeting the crosstalk criteria between any two of these pathways. We then ranked the pathway pairs according to their crosstalk scores and 20 pathway pairs with high crosstalk scores were shown in Table 6. Compared with the pathways pairs with close relationship defined by our method (Table 5), there were five pathway pairs detected by both two methods, i.e., oxidative phosphorylation and Parkinson's disease, cocaine addiction and amphetamine addiction, type I diabetes mellitus and autoimmune thyroid disease, asthma and allograft rejection, and type I diabetes mellitus and allograft rejection. However, for eleven of the twenty pathway pairs detected by the new method, one or both pathways were neuronal function or neuronal disease related, while only two pathway pairs by crosstalk analysis were neuronal function or neuronal disease related (Tables 5 and 6). Considering that PD is mainly a neuronal disorder, such result suggests the pathway relationships identified by the new method may be more reasonable.

Since in pathway crosstalk analysis, no stringent criterion was defined for selecting the significantly connected pathway pairs, we selected the top 10% pairs arbitrarily [39], which resulted a list of 120 pathway pairs. With the new method, 311 pathway pairs with significantly close relationships were identified based on their *D* values, of which 63 pathway pairs were shared with the pathway crosstalk analysis. A close inspection indicated that, compared with pathway crosstalk analysis, more pathway pairs closely related to the neurodevelopment or neurological signaling transduction (e.g., serotonergic synapse and dopamine synapse, cholinergic synapse and serotonergic synapse, cholinergic synapse and dopamine synapse, calcium signaling pathway and dopamine synapse) were ranked in the top of the list given by the new method. Additionally, we tested the robustness of the two methods by randomly removed 10% and 20% genes from the 79 enriched pathways and re-calculated the crosstalk scores and separation values between the pathways. While the pathway pairs with close relationship were largely consistent for the new method, there was a more significant disturbance in the result of pathway crosstalk analysis. Taken together, compared to pathway crosstalk analysis, the new method based on network-based separation value, could give more reasonable and robust pathway relationships.

In addition to PD, we also employed our approach on a couple of other diseases, including nicotine addiction, Alzheimer's disease, and ovarian cancer co-occurrence with depression. For ovarian cancer co-occurrence with depression, 219 differentially expressed genes in ovarian cancers from patients with high vs. low bio-behavioral risk profiles were obtained [57]. From these genes, 28 dysregulated pathways were identified by ToppGene. With our method, 72 pathway pairs with significantly close relationships were identified, with which a pathway network associated with ovarian cancer co-occurrence with depression was constructed. In this network, pathways such as antigen processing and presentation, intestinal immune network for IgA production, as well as phagosome were at the hub positions, which were consistent with the available knowledge on this disease.

Undeniably, there are some limitations in this study. Currently, there are quite several pathway databases available. For instance, ConsensusPathDB (CPDB; http://cpdb.molgen.mpg.de/) integrates pathways from 12 pathway databases (e.g., BioCarta, Reactome, KEGG, and WikiPathways). In the current work, we only calculated relationships among pathways from KEGG database and pathways from other databases have not been included. A preliminary evaluation indicated that the performance of our method was consistent across different pathway databases. However, a comprehensive and detailed evaluation is still necessary. Although KEGG pathways are acceptable for their high accuracy, the less abundant of pathways is an inherent drawback of KEGG database. On the other hand, due to the definition of pathways as well as their relevant genes in different pathway database may be differ-

Table 6
Crosstalk in Parkinson's disease-related pathways (top 20 pathway pairs).

Pathway A		Pathway B		Crosstalk score
KEGG ID	Name	KEGG ID	Name	
hsa05330	Allograft rejection	hsa05320	Autoimmune thyroid disease	0.929
hsa05332	Graft-versus-host disease	hsa04940	Type I diabetes mellitus	0.904
hsa05204	Chemical carcinogenesis	hsa00980	Metabolism of xenobiotics by cytochrome P450	0.826
hsa05330	Allograft rejection	hsa05310	Asthma	0.821
hsa05012	Parkinson's disease	hsa00190	Oxidative phosphorylation	0.794
hsa05332	Graft-versus-host disease	hsa05330	Allograft rejection	0.782
hsa04940	Type I diabetes mellitus	hsa05330	Allograft rejection	0.782
hsa00982	Drug metabolism-cytochrome P450	hsa00980	Metabolism of xenobiotics by cytochrome P450	0.774
hsa05320	Autoimmune thyroid disease	hsa05310	Asthma	0.752
hsa05030	Cocaine addiction	hsa05031	Amphetamine addiction	0.744
hsa05321	Inflammatory bowel disease (IBD)	hsa05140	Leishmaniasis	0.732
hsa04672	Intestinal immune network for IgA production	hsa05310	Asthma	0.730
hsa05140	Leishmaniasis	hsa05310	Asthma	0.714
hsa05332	Graft-versus-host disease	hsa05320	Autoimmune thyroid disease	0.711
hsa04940	Type I diabetes mellitus	hsa05320	Autoimmune thyroid disease	0.711
hsa05204	Chemical carcinogenesis	hsa00982	Drug metabolism-cytochrome P450	0.705
hsa05321	Inflammatory bowel disease (IBD)	hsa05310	Asthma	0.705
hsa05200	Pathways in cancer	hsa05210	Colorectal cancer	0.684
hsa05310	Asthma	hsa05322	Systemic lupus erythematosus	0.680
hsa05332	Graft-versus-host disease	hsa05168	Herpes simplex infection	0.674

ent, it is still difficult to merge the information in different pathway database, and the calculation results may be biased [58]. Further, although the quantity and quality of PPI data has been greatly improved, the human PPI network is still far from complete. Also, due to the limitation of current technology, there may be some false positives in the PPI data [56]. Such potential biases associated with human PPI network may affect the performance of the method and our interpretation of the results.

4. Conclusion

In summary, we proposed to use the distances between genes within and between pathways to measure the localization characters of pathways in the human protein-protein interaction network and to analyze the pathways relationships. It was based on the observation that genes associated with complex phenotypes can form observed modules in the human PPI network, and the relationships between phenotypes can be identified based on their modular distribution. In our approach, the localization characters of pathways in the human PPI network were analyzed based on the largest connected component of each pathway-related sub network. We found that the majority of the pathways had a significantly observable module in human PPI network, which made it possible to detect the relationships among pathways by evaluating the separation of pathway modules in PPI network. Unlike the methods depending on the overlap of pathways or the integrity of the PPI network information, the new method estimates relationships of pathways through their network modules separation on the PPI network. Via this method, we can study the correlation between a disease and the dysregulated pathways by analyzing the relationships of pathways. Based on the relationship of pathways, a more comprehensive and systematic view of dysfunctional pathways underlying disease can be obtained.

As an application, we analyzed the relationship between pathways related to Parkinson's disease and constructed a PD-related pathways interaction network. Via such a network, pathways that may be important in the occurrence and development of PD were predicted.

Acknowledgement

This study was supported in part by grants from National Key Research and Development Program of China (No. 2016YFC0906300), National Natural Science Foundation of China (Nos. 31271411 & 61202379). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2017.08.001.

References

[1] A.L. Tarca, S. Draghici, P. Khatri, S.S. Hassan, P. Mittal, et al., A novel signaling pathway impact analysis, Bioinformatics 25 (2009) 75–82.
[2] M.G. Hong, Y. Pawitan, P.K. Magnusson, J.A. Prince, Strategies and issues in the detection of pathway enrichment in genome-wide association studies, Hum. Genet. 126 (2009) 289–301.
[3] I. Rivals, L. Personnaz, L. Taing, M. Potier, Enrichment or depletion of a GO category within a class of genes: which test?, Bioinformatics 23 (2007) 401–407
[4] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[5] D.A. Hosack, G. Dennis Jr., B.T. Sherman, H.C. Lane, R.A. Lempicki, Identifying biological themes within lists of genes with EASE, Genome Biol. 4 (2003) R70.
[6] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15545–15550.
[7] J.S. Logue, D.K. Morrison, Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy, Genes Dev. 26 (2012) 641–650.
[8] M.F. Sharpnack, K. Huang, Detecting cancer pathway crosstalk with distance correlation, AMIA Jt. Summits Transl. Sci. Proc. 2015 (2015) 41–45.
[9] P.J. Stork, J.M. Schmitt, Crosstalk between cAMP and MAP kinase signaling in the regulation of cell proliferation, Trends Cell Biol. 12 (2002) 258–266.
[10] N. Sengupta, P.K. Vinod, K.V. Venkatesh, Crosstalk between cAMP-PKA and MAP kinase pathways is a key regulatory design necessary to regulate FLO11 expression, Biophys. Chem. 125 (2007) 59–71.
[11] Q. Shu, W. Li, J. Li, W. Wang, C. Liu, et al., Cross-talk between cAMP and MAPK pathways in HSD11B2 induction by hCG in placental trophoblasts, PLoS One 9 (2014) e107938.
[12] S. Huveneers, E.H. Danen, Adhesion signaling – crosstalk between integrins, Src and Rho, J. Cell Sci. 122 (2009) 1059–1069.
[13] B. Geiger, A. Bershadsky, R. Pankov, K.M. Yamada, Transmembrane crosstalk between the extracellular matrix–cytoskeleton crosstalk, Nat. Rev. Mol. Cell Biol. 2 (2001) 793–805.
[14] J. Sun, Z. Zhao, Functional features, biological pathways, and protein interaction networks of addiction-related genes, Chem. Biodivers. 7 (2010) 1153–1162.
[15] P. Jia, J.M. Ewers, Z. Zhao, Prioritization of epilepsy associated candidate genes by convergent analysis, PLoS One 6 (2011) e17162.
[16] P. Jia, C.F. Kao, P.H. Kuo, Z. Zhao, A comprehensive network and pathway analysis of candidate genes in major depressive disorder, BMC Syst. Biol. 5 (Suppl 3) (2011) S12.
[17] Y. Li, P. Agarwal, D. Rajagopalan, A global pathway crosstalk network, Bioinformatics 24 (2008) 1442–1447.
[18] C. Ogris, D. Guala, T. Helleday, E.L. Sonnhammer, A novel method for crosstalk analysis of biological networks: improving accuracy of pathway annotation, Nucleic Acids Res. 45 (2017) e8.
[19] C. Ogris, T. Helleday, E.L. Sonnhammer, PathwAX: a web server for network crosstalk based pathway annotation, Nucleic Acids Res. 44 (2016) W105–109.
[20] A.N. Tegge, N. Sharp, T.M. Murali, Xtalk: a path-based approach for identifying crosstalk between signaling pathways, Bioinformatics 32 (2016) 242–251.
[21] J. Han, C. Li, H. Yang, Y. Xu, C. Zhang, et al., A novel dysregulated pathway-identification analysis based on global influence of within-pathway effects and crosstalk between pathways, J. R. Soc. Interface 12 (2015) 20140937.
[22] T. McCormack, O. Frings, A. Alexeyenko, E.L. Sonnhammer, Statistical assessment of crosstalk enrichment between gene groups in biological networks, PLoS One 8 (2013) e54945.
[23] A. Barabasi, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat. Rev. Genet. 12 (2011) 56–68.
[24] S.D. Ghiassian, J. Menche, A. Barabasi, A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome, PLOS Comput. Biol. 11 (2015).
[25] J. Menche, A. Sharma, M. Kitsak, S.D. Ghiassian, M. Vidal, et al., Uncovering disease-disease relationships through the incomplete interactome, Science 347 (2015) 1257601.
[26] M. Volta, A.J. Milnerwood, M.J. Farrer, Insights from late-onset familial parkinsonism on the pathogenesis of idiopathic Parkinson's disease, Lancet Neurol. 14 (2015) 1054–1064.
[27] A.-G. Sun, J. Wang, Y.-Z. Shan, W.-J. Yu, X. Li, et al., Identifying distinct candidate genes for early Parkinson's disease by analysis of gene expression in whole blood, Neuro. Endocrinol. Lett. 35 (2014) 398–404.
[28] A. Dumitriu, J.C. Latourelle, T.C. Hadzi, N. Pankratz, D. Garza, et al., Gene expression profiles in Parkinson disease prefrontal cortex implicate FOXO1 and genes under its transcriptional regulation, PLoS Genet. 8 (2012) e1002794.
[29] T. Botta-Orfila, E. Tolosa, E. Gelpi, A. Sànchez-Pla, M.-J. Martí, et al., Microarray expression analysis in idiopathic and LRRK2-associated Parkinson's disease, Neurobiol. Dis. 45 (2012) 462–468.
[30] F. Simunovic, M. Yi, Y. Wang, L. Macey, L.T. Brown, et al., Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease pathology, Brain 132 (2009) 1795–1809.
[31] M.J. Cowley, M. Pinese, K.S. Kassahn, N. Waddell, J.V. Pearson, et al., PINA v2.0: mining interactome modules, Nucleic Acids Res. 40 (2012) D862–865.
[32] P. Smialowski, P. Pagel, P. Wong, B. Brauner, I. Dunger, et al., The Negatome database: a reference set of non-interacting protein pairs, Nucleic Acids Res. 38 (2010) D540–D544.
[33] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.
[34] Y. Hu, Z. Pan, Y. Hu, L. Zhang, J. Wang, Network and pathway-based analyses of genes associated with Parkinson's disease, Mol. Neurobiol. (2016) 1–14.
[35] C.M. Lill, J.T. Roehr, M.B. McQueen, F.K. Kavvoura, S. Bagade, et al., Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database, PLoS Genet. 8 (2012) e1002548.

[36] M.A. Nalls, N. Pankratz, C.M. Lill, C.B. Do, D.G. Hernandez, et al., Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease, Nat. Genet. 46 (2014) 989–993.
[37] J. Chen, E.E. Bardes, B.J. Aronow, A.G. Jegga, ToppGene Suite for gene list enrichment analysis and candidate gene prioritization, Nucleic Acids Res. 37 (2009).
[38] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Royal Stat. Soc. Ser. B (Methodological) (1995) 289–300.
[39] P. Jia, C.-F. Kao, P.-H. Kuo, Z. Zhao, A comprehensive network and pathway analysis of candidate genes in major depressive disorder, BMC Syst. Biol. 5 (2011) S12.
[40] M. Liu, R. Fan, X. Liu, F. Cheng, J. Wang, Pathways and networks-based analysis of candidate genes associated with nicotine addiction, PLoS One 10 (2015) e0127438.
[41] V. Quagliariello, S. Rossetti, C. Cavaliere, R. Di Palo, E. Lamantia, et al., Metabolic syndrome, endocrine disruptors and prostate cancer associations: biochemical and pathophysiological evidences, Oncotarget 8 (2017) 30606–30616.
[42] C. Alves, L. Dourado, Endocrine and metabolic disorders in HTLV-1 infected patients, Braz. J. Infect. Dis. 14 (2010) 613–620.
[43] N.L. Levy, A.L. Notkins, Viral Infections and diseases of the endocrine system, J. Infect. Dis. 124 (1971) 94–103.
[44] E. Hathout, E.M. Alonso, R. Anand, K. Martz, E. Imseis, et al., Post-transplant diabetes mellitus in pediatric liver transplantation, Pediatr. Transplant. 13 (2009) 599–605.
[45] B.L. Kasiske, J.J. Snyder, D. Gilbertson, A.J. Matas, Diabetes mellitus after kidney transplantation in the United States, Am. J. Transplant. 3 (2003) 178–185.
[46] H. Wang, H. Wei, B. Chen, Y. Zhou, Chronic morphine exposure impairs short-term synaptic depression of geniculo-cortical visual pathway in vivo, Neurosci. Lett. 410 (2006) 228.
[47] M. Bajo, M. Roberto, S.G. Madamba, G.R. Siggins, Neuroadaptation of GABAergic transmission in the central amygdala during chronic morphine treatment, Addict. Biol. 16 (2011) 551–564.
[48] J.A. Kauer, R.C. Malenka, Synaptic plasticity and addiction, Nat. Rev. Neurosci. 8 (2007) 844–858.
[49] D. Gioeli, W. Wunderlich, J. Sebolt-Leopold, S. Bekiranov, J.D. Wulfkuhle, et al., Compensatory pathways induced by MEK inhibition are effective drug targets for combination therapy against castration-resistant prostate cancer, Mol. Cancer Ther. 10 (2011) 1581–1590.
[50] R. Nazarian, H. Shi, Q. Wang, X. Kong, R.C. Koya, et al., Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation, Nature 468 (2010) 973–977.
[51] H. Vlajinac, E. Dzoljic, J. Maksimovic, J. Marinkovic, S. Sipetic, et al., Infections as a risk factor for Parkinson's disease: a case-control study, Int. J. Neurosci. 123 (2013) 329–332.
[52] C. Cheng, Y. Wu, S.J. Tsai, Y.M. Bai, J.W. Hsu, et al., Risk of developing Parkinson's disease among patients with asthma: a nationwide longitudinal study, Allergy 70 (2015) 1605–1612.
[53] J.C. Lin, C.S. Lin, C. Hsu, C.L. Lin, C.H. Kao, Association between Parkinson's disease and inflammatory bowel disease, Inflamm. Bowel Dis. 22 (2016) 1049–1055.
[54] K. Venderova, D.S. Park, Programmed cell death in Parkinson's disease, Cold Spring Harb. Perspect. Med. 2 (2012).
[55] V. Shukla, S.K. Mishra, H.C. Pant, Oxidative stress in neurodegeneration, Adv. Pharmacol. Sci. 2011 (2011). 572634–572634.
[56] D.D. Feng, W. Cai, X. Chen, The associations between Parkinson's disease and cancer: the plot thickens, Transl. Neurodegener. 4 (2015). 20–20.
[57] S.K. Lutgendorf, K. DeGeest, C.Y. Sung, J.M. Arevalo, F. Penedo, et al., Depression, social support, and beta-adrenergic transcription control in human ovarian cancer, Brain Behav. Immun. 23 (2009) 176–183.
[58] L. Wadi, M. Meyer, J. Weiser, L. Stein, J. Reimand, Impact of outdated gene annotations on pathway enrichment analysis, Nat. Methods 13 (2016) 705–706.