

# Omnigenic Model: The Evidence from Neurodegenerative Diseases

Libing Shen<sup>1\*</sup>, Qili Shi<sup>1</sup>

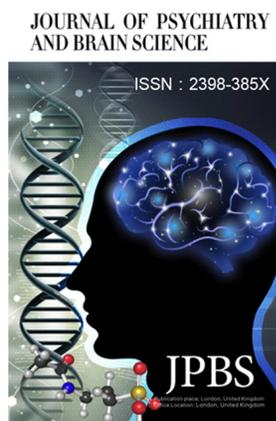
<sup>1</sup> Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, 200032, P. R. China.

\*Correspondence: Libing Shen, Ph.D.; Email: shenlibing@sioc.ac.cn.

## ABSTRACT

Recently, a seminal model, called the omnigenic model, is proposed for understanding complex traits such as schizophrenia. In this study, we examined this model in Alzheimer's disease and Parkinson's disease from the perspectives of expression spectrum, shared disease-associated genes, common biological pathways, organ specificities, and network properties. Our results support the arguments brought forward by the omnigenic model. Although we only provided the limited evidence for the omnigenic model in neurodegenerative diseases, we hope that our effort can improve the understanding of these diseases and thus spur new ideas on how to prevent and treat them. The most important information we try to convey in our study is that there are more genes than we expected playing a role in the pathogenesis of neurodegenerative diseases and it is insufficient to study these diseases only focusing on some core genes or genetic pathways.

**Keywords:** Omnigenic model; Neurodegenerative diseases; Alzheimer's disease; Parkinson's disease



<http://jpbs.qingres.com>

 OPEN ACCESS

DOI: 10.20900/jpbs.20170019

Received: October 24, 2017

Accepted: October 24, 2017

Published: December 25, 2017

**Copyright:** ©2017 Cain *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Complex traits such as height and intelligence are always a fascinating research topic for geneticists. Two central questions on complex traits are how many genes are exactly behind a complex trait and whether they equally contribute to such a trait. In the era of genome project and next-generation sequencing, we seem to be approaching to the answers for these two questions. Nevertheless, the big data of this era only bring us confounding results instead of definitive answers. The study of complex traits with common SNPs shows that all autosomal SNPs contribute to the 45 % variance in height and the variance

explained by each chromosome is proportional to its length<sup>[1]</sup>. In another word, each gene in an individual's genome could probably make its tiny contribution to his or her height. For complex diseases, such answer is obviously far beyond satisfaction, because we are trying to seek possible treatment, not probable statement.

In a recently published seminal paper, Boyle *et al.* proposed a new omnigenic model for understanding complex diseases<sup>[2]</sup>. The omnigenic model proposes that all genes expressed in disease-relevant cells have the influence on the functions of core disease-related genes and thus the genetic propensity for complex diseases can lay outside core gene pathways. The hypothesis behind the omnigenic model is that all genes expressed in a cell are tightly interconnected by an omnibus gene regulatory network and thus each of them can

affect the development of disease. We found this model both troubling and enlightening and tested it with Alzheimer's disease and Parkinson's disease microarray data. Both of them are neurodegenerative diseases, whose causes are still poorly understood<sup>[3, 4]</sup>.

First, we downloaded the expression data of Alzheimer's disease and Parkinson's disease from GEO database and then identified the differentially expressed genes (DEGs) between cases and controls in each dataset. The DEGs are classified into two categories --- down-regulated genes and up-regulated genes (Table 1). Our result shows the number of down-regulated genes and up-regulated genes varies from dataset to dataset, which is ranging from a few of thousands to a couple of hundreds. Accordingly, the gene expression spectra are widely different among different disease datasets.

**Table 1. The number of differentially expressed genes (DEGs) in each microarray dataset.**

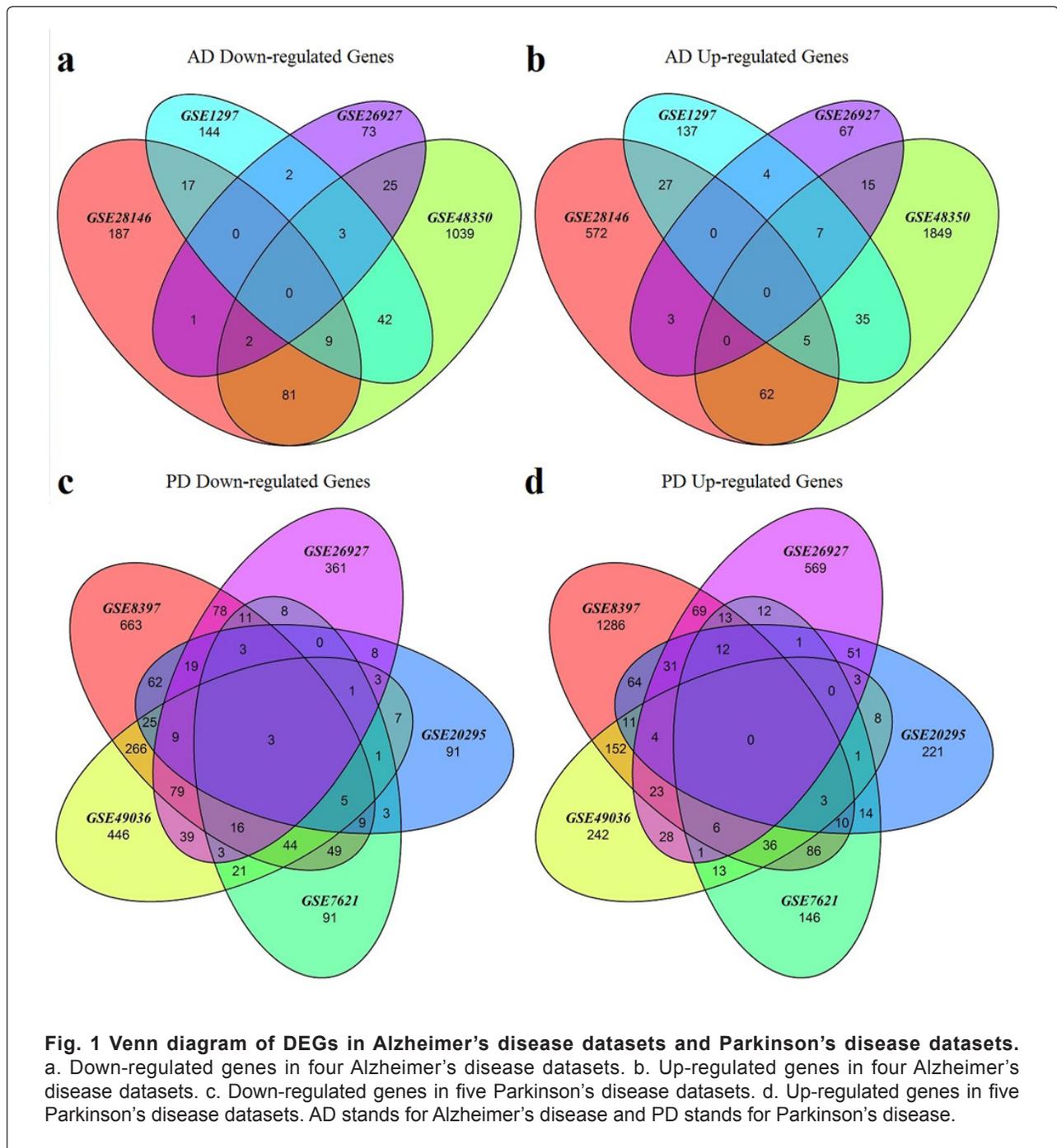
Disease (GEO accession number)	Number of differentially expressed genes (DEGs)	
	Down-regulated	Up-regulated
Alzheimer's Disease (GSE28146)	298	669
Alzheimer's Disease (GSE48350)	1201	1973
Alzheimer's Disease (GSE1297)	217	215
Alzheimer's Disease (GSE26927)	106	97
Parkinson's disease (GSE7621)	268	354
Parkinson's disease (GSE8397)	1342	1806
Parkinson's disease (GSE49036)	968	531
Parkinson's disease (GSE20295)	249	435
Parkinson's disease (GSE26927)	641	823

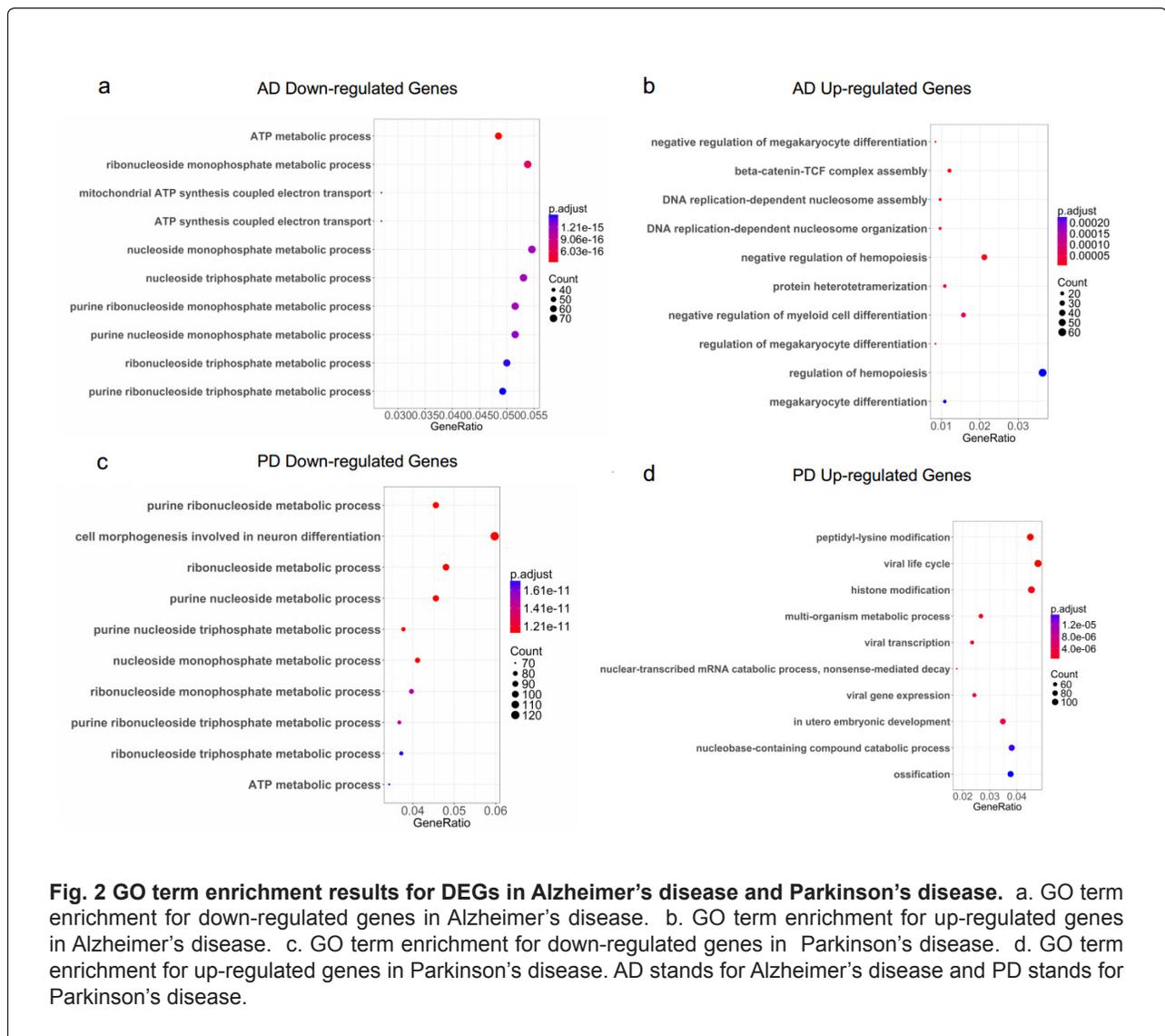
Second, we tried to find the common DEGs in down-regulated gene category and up-regulated gene category for Alzheimer's disease and Parkinson's disease, respectively. Among four Alzheimer's disease datasets, just dozens of common DEGs were found in both down-regulated

genes and up-regulated genes (Fig. 1a and 1b). Among five Parkinson's disease datasets, only three common DEGs were found in down-regulated genes whereas no common DEG was found in up-regulated genes. Certainly, there exists a possibility that these disease cases share some prominent

common core gene pathways, although they share very few common genes. To for or against such possibility, we performed the gene-GO term enrichment analyses for 1300 down-regulated genes and 1651 up-regulated genes in Alzheimer's disease and 2042 down-regulated genes and 2469 up-regulated genes in Parkinson's disease. The gene-GO term enrichment results are shown in Fig. 2. Although a couple of thousands of genes were used for each gene-GO term enrichment analysis, the most significant biological processes only contain a dozens of genes in each GO analysis. For the

down-regulated genes in Alzheimer's disease and Parkinson's disease, their significant biological processes mainly involve in metabolic processes but not conspicuous cell degenerative pathways such as apoptotic process as we expected. Furthermore, the down-regulation of cell metabolism is a result of cell death but not a cause for it<sup>[5]</sup>. Apparently, there are very few shared genes among different Alzheimer's disease datasets or Parkinson's disease datasets and our GO analyses failed to detect any common initiation for the neuron degeneration process in Alzheimer's disease and Parkinson's disease.

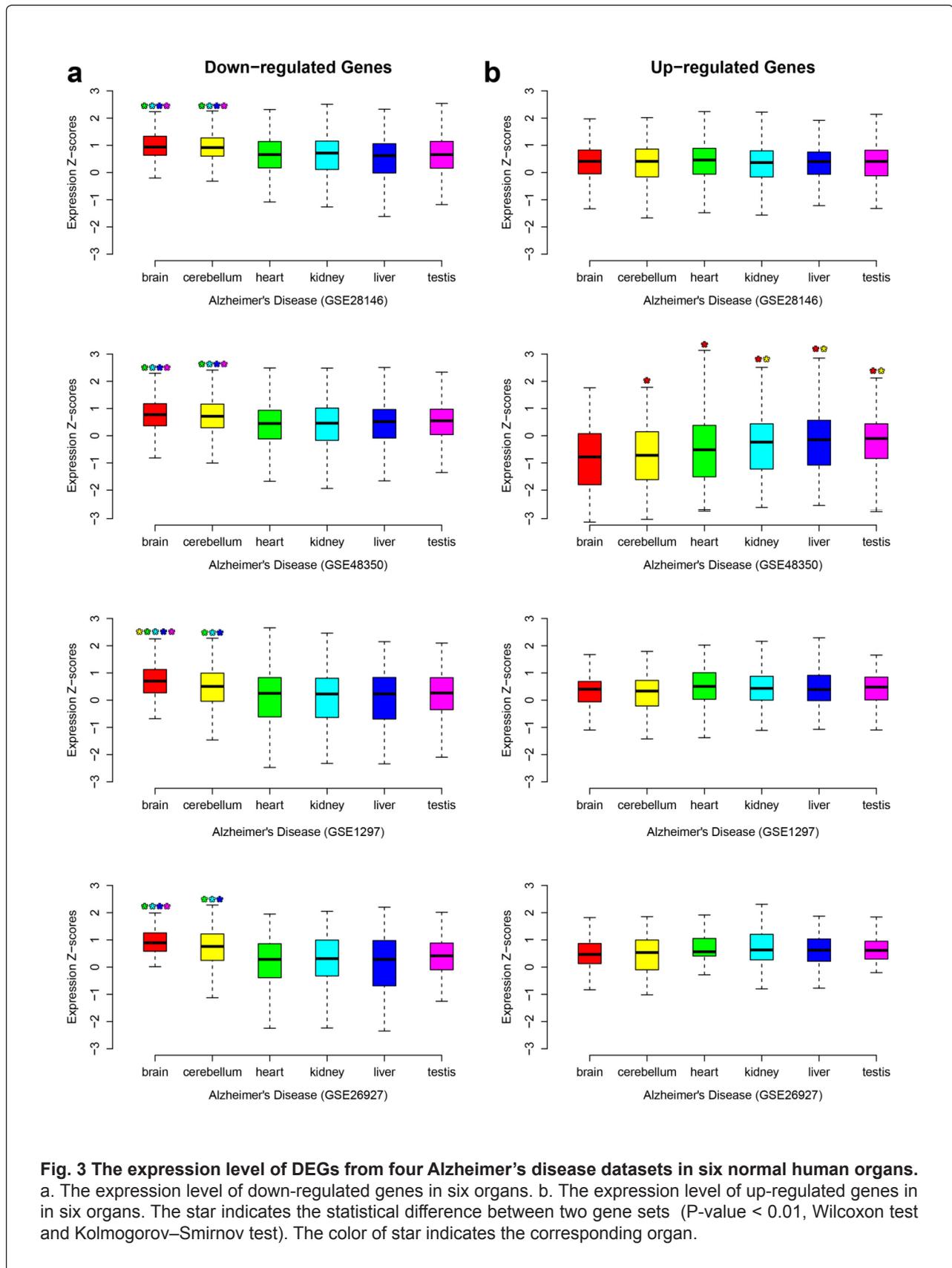




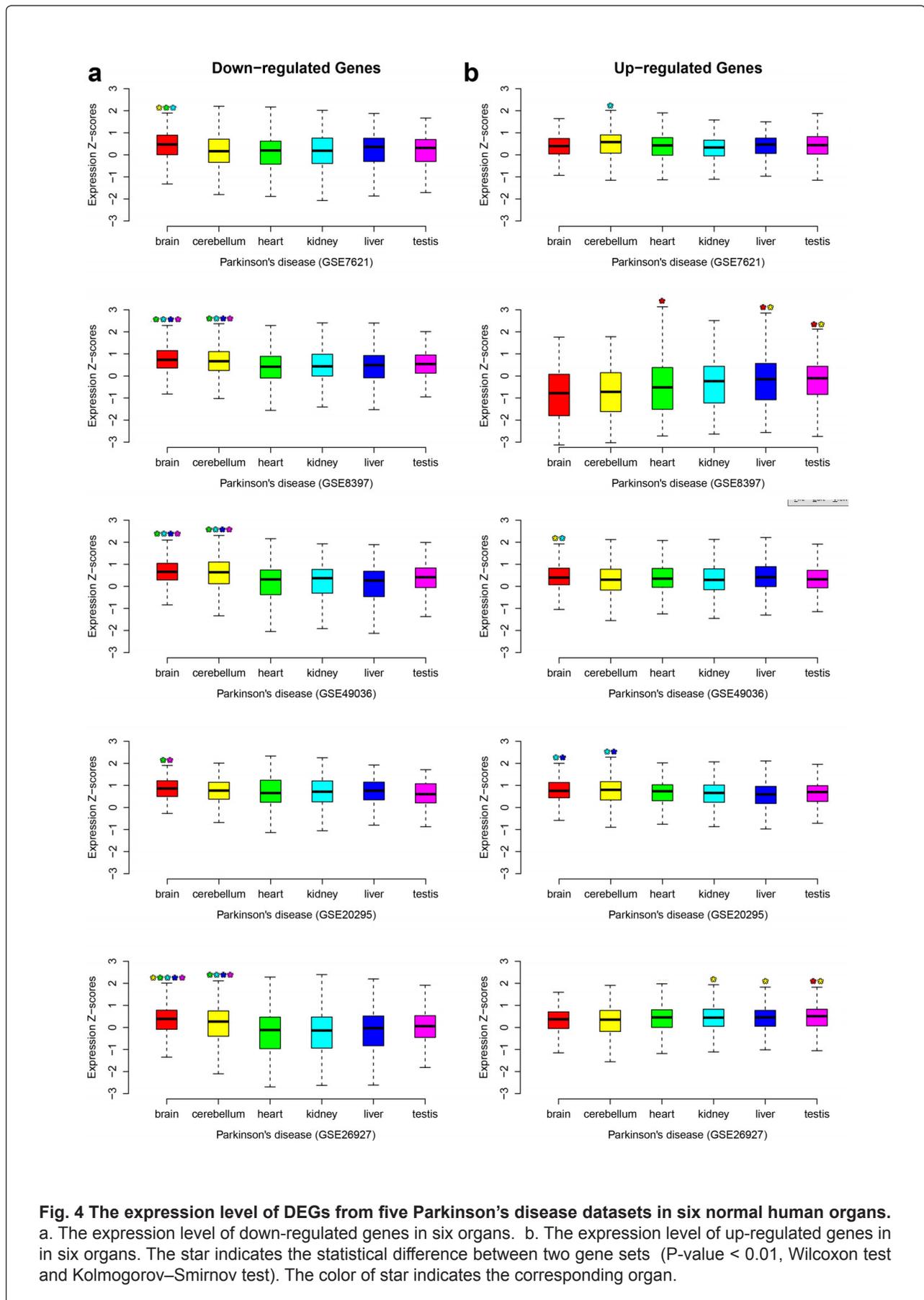
**Fig. 2 GO term enrichment results for DEGs in Alzheimer's disease and Parkinson's disease.** a. GO term enrichment for down-regulated genes in Alzheimer's disease. b. GO term enrichment for up-regulated genes in Alzheimer's disease. c. GO term enrichment for down-regulated genes in Parkinson's disease. d. GO term enrichment for up-regulated genes in Parkinson's disease. AD stands for Alzheimer's disease and PD stands for Parkinson's disease.

Third, we checked the expression levels of DEGs in normal organs, since the omnigenic model states that all genes expressed in disease-relevant cells can have the influence on the disease development process. And vice versa, the disease must have the impact on their expression level in disease-relevant organs. Fig. 3 shows the expression levels of down-regulated genes and up-regulated genes in four Alzheimer's disease dataset among six normal human organs. The down-regulated genes in four Alzheimer's disease datasets exhibit a statistically higher expression level in brain and cerebellum than heart, liver, kidney, liver and testis while the up-regulated genes don't (Fig. 3a and 3b). In five Parkinson's disease datasets, the down-regulated genes show a similarly higher expression pattern in

brain and cerebellum, although such pattern is less obvious in dataset GSE20295 (Fig. 4a). Moreover, the up-regulated genes in five Parkinson's disease datasets show a more diversified expression pattern (Fig. 4b), which suggests that the pathogenic process might be more complex in Parkinson's disease than in Alzheimer's disease. Fig. 3 and 4 actually show that these down-regulated genes are specifically expressed in normal central nervous system (CNS), because they have a higher expression level in brain and/or cerebellum compared to the other four organs. That the CNS-specifically-expressed genes are down-regulated in neurodegenerative diseases is consistent with the assumption of the omnigenic model.

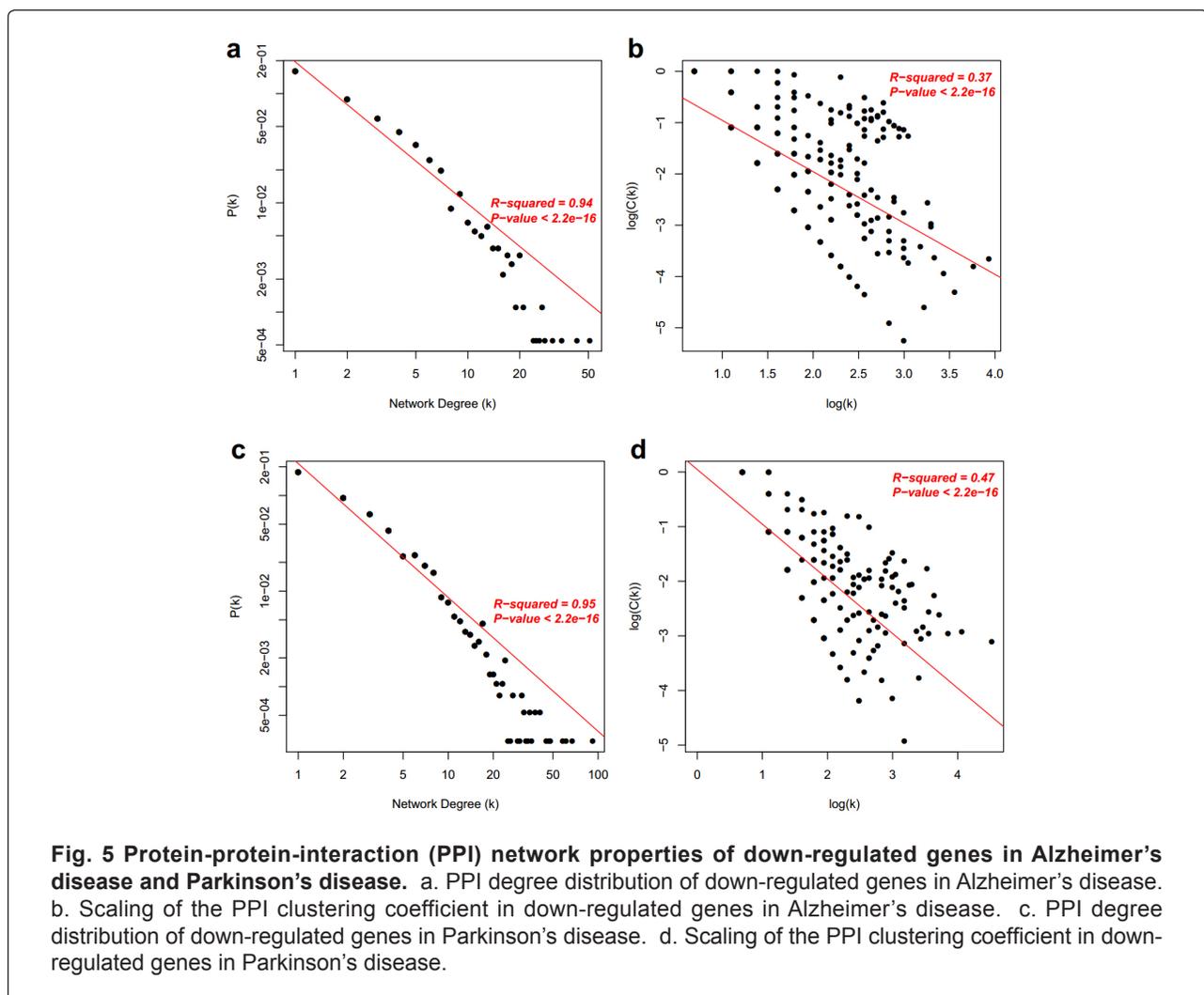


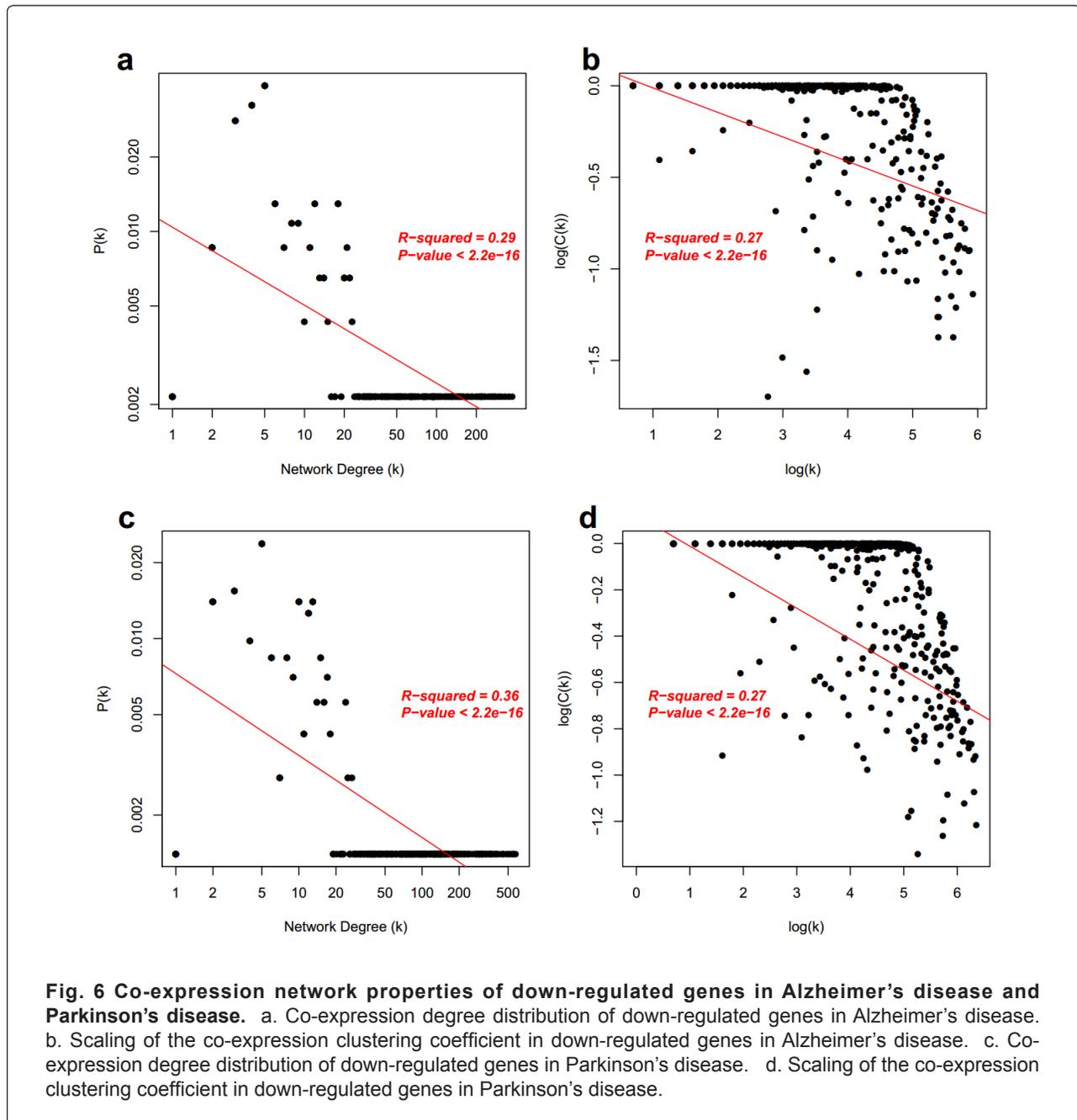
**Fig. 3** The expression level of DEGs from four Alzheimer's disease datasets in six normal human organs. a. The expression level of down-regulated genes in six organs. b. The expression level of up-regulated genes in six organs. The star indicates the statistical difference between two gene sets ( $P$ -value  $< 0.01$ , Wilcoxon test and Kolmogorov–Smirnov test). The color of star indicates the corresponding organ.



Finally, we examined the network properties of the down-regulated genes in Alzheimer's disease and Parkinson's disease using both protein-protein-interaction information and gene co-expression information. The key point in omnigenic model is that all genes expressed in disease-relevant cells are sufficiently interconnected by gene regulatory networks. Because our result shows that the down-regulated genes in neurodegenerative diseases are also CNS-specifically-expressed genes, we infer that the down-regulated genes in neurodegenerative diseases must constitute a biological network which can be characterized by the power-law degree distribution and hierarchical structure [6]. Fig. 5a shows that the degree distribution of the down-regulated genes in Alzheimer's disease to the probability  $P(k)$  fits the power law in protein-protein-interaction networks (the power-law fit is shown as a red line,  $R^2 = 0.94$ ,  $P\text{-value} < 2 \times 10^{-16}$ ). Fig. 5b shows that the scaling of the clustering coefficient of the down-regulated genes in Alzheimer's disease follows  $C(K) \sim K^{-1}$  in protein-protein-interaction networks (which is shown in a red straight line with negative slope,  $R^2 = 0.37$ ,  $P\text{-values} < 2 \times 10^{-16}$ ). Our

network property analyses also show that the down-regulated genes in Parkinson's disease have the same network properties as the ones in Alzheimer's disease in protein-protein-interaction network (Fig. 5c and 5d). The network analysis based on gene co-expression information shows a similar result for the down-regulated genes in Alzheimer's disease and Parkinson's disease (Fig. 6a, 6b, 6c and 6d). That a gene set's degree distribution fits the power law indicates that they constitute a scale-free network (there exist a small number of highly connected nodes, known as hubs, within the network). And the scaling of the clustering coefficients of two sets of down-regulated genes show that they have a hierarchical network structure as a protein-protein interaction network (sparsely connected nodes are part of highly clustered areas and different clustered areas are communicated by a few hubs). The hierarchical network structure exhibits the small-world property of disease-related genes, which is predicted by Boyle *et al.*'s paper [2]. Thus, the results of our network analyses further support the omnigenic model.



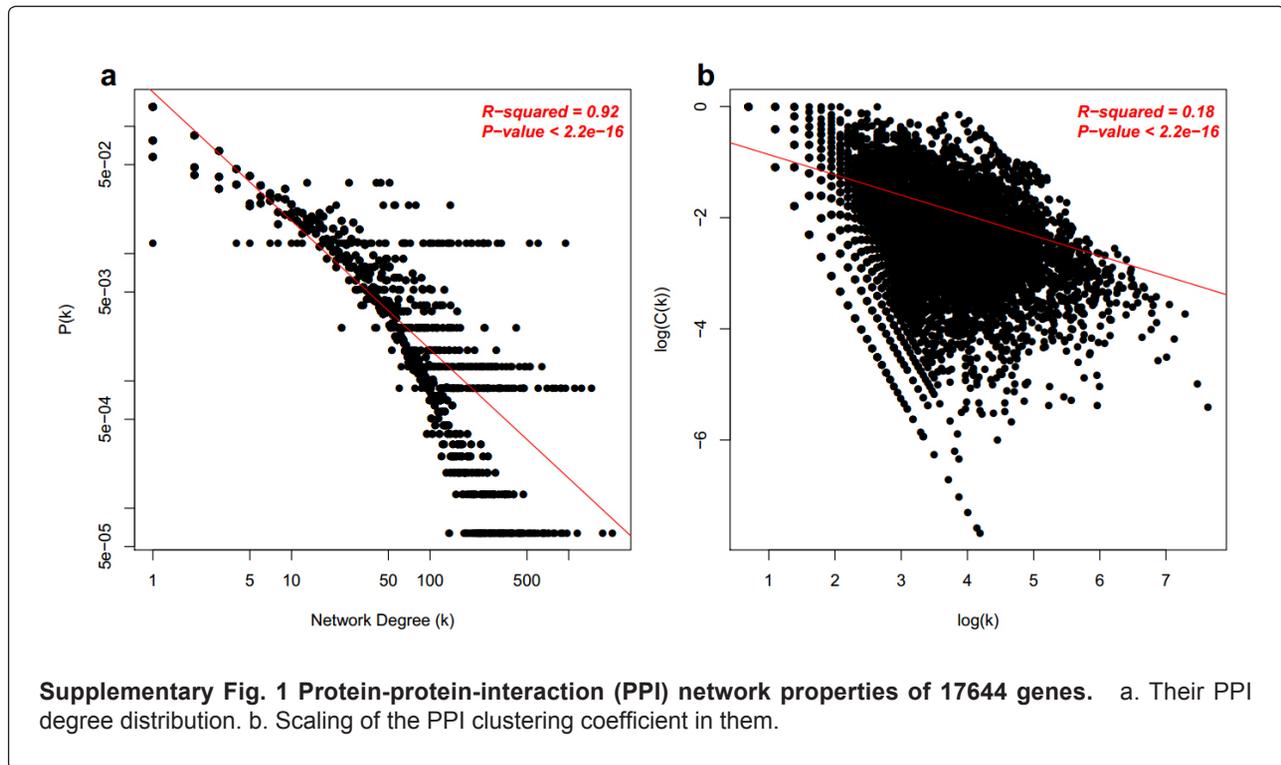


There is a major difference between two network analysis results. The coefficient of determinations ( $R^2$ ) are smaller in gene co-expression network analyses than in protein-protein-interaction ones. We find that gene's average degree and clustering coefficient is much higher in gene co-expression networks than in protein-protein-interaction networks and the genes with high degree and clustering coefficient in gene co-expression networks also outnumber their counterparts in protein-protein-interaction networks. For example, for the down-regulated genes in Alzheimer's disease, the protein-

protein-interaction network analysis shows that they have the highest degree of 50 while the gene co-expression one shows that they have the highest degree of 250 (Fig. 5a and Fig. 6a), which suggests that the protein-protein-interaction information might be limited. This limitation could lead to bias in our network analyses, although the bias is favorable towards our conclusion. In order to rule out such bias in our study, we performed the network property analysis for all genes with protein-protein interaction information and used them as the background (total 17644 genes). The result shows that all genes with

protein-protein interaction information also constitute a genuine biological network (scale-free and highly modular, Supplementary Fig. 1). Consequently, our network analysis in Figure 5 is less likely to be biased since it is based on a small protein-protein-interaction network from a much bigger one. In

addition, the large average degree and clustering coefficient from gene co-expression network analysis propose that the down-regulated genes in disease samples are often expressed in a modular fashion in normal brain regions.



In this comment, we used the expression data from Alzheimer's disease and Parkinson's disease to verify the omnigenic model in neurodegenerative diseases. We originally felt some qualm about this model, because it provides more challenge for those like us who work in the field of complex diseases. To our surprise, our results fully support it, although we have to say that our evidence is circumstantial. However, acquiring direct evidence for this model would not be an easy task. First, we should know how many genes are exactly expressed in a disease-relevant cell. Then, we have to thoroughly understand the cellular network in order to separate peripheral genes from core genes, since they play different roles in disease pathogenesis. In our view, the most important information that the omnigenic model conveys is that new methods, both experimental and bioinformatic ones, are needed for studying the complex diseases such as Alzheimer's disease and Parkinson's disease, because there are

more genes than we expected participating in the process of neurodegenerative disease development.

## 1 MATERIALS AND METHODS

### 1.1 Microarray data and detection of differentially expressed genes

All microarray datasets used in this study were downloaded from GEO database. Four datasets are Alzheimer's disease, which are GSE28146 (15 cases and 8 controls), GSE48350 (19 cases and 43 controls), GSE1297 (15 cases and 9 controls), and GSE26927 (11 cases and 7 controls). Five datasets are Parkinson's disease, which are GSE7621 (16 cases and 9 controls), GSE8397 (29 cases and 16 controls), GSE49036 (16 cases and 8 controls), GSE20295 (12 cases and 18 controls), and GSE26927 (10 cases and 10 controls). The detailed

disease information for each dataset has been shown in previous studies<sup>[7-13]</sup>.

In each dataset, the expression counts for each gene were normalized with the quantile algorithm in the limma package under R environment<sup>[14]</sup>. Differentially expressed genes (DEGs) were detected by limma package using linear models with a cutoff P-value < 0.01.

## 1.2 GO term enrichment analysis

To test whether there exist certain common biological pathways in both diseases, we used the clusterProfiler package to perform gene-GO term enrichment analyses for the combined DEGs in Alzheimer's disease or Parkinson's disease<sup>[15]</sup>. To visualize the GO results, we used the ggplot2 package for GO result display.

## 1.3 RNA-Seq data for six human organs

The RNA-Seq data of six human organs (brain, cerebellum, heart, kidney, liver, and testis) were downloaded from the supplementary information of Brawand *et al.*<sup>[16]</sup>. We calculated the RPKM (Reads Per Kilobase per Million mapped reads) value for each gene based on the downloaded data (unique read coverage per exon). Due to the uneven number of samples in some organs from some species, we used the mean RPKM value if multiple RPKM values were available for each human organ. We transformed the RPKM values into the log<sub>2</sub>(RPKM) values and then calculated the Z-score for every log<sub>2</sub>(RPKM) value within each organ, in order to render the gene expression values comparable among different organs.

## 1.4 Statistical analysis and data display

The R package (version 3.2.4) was used to perform statistical analyses in this study. Both Wilcoxon test and Kolmogorov–Smirnov test were employed to compare the expression level of the DEGs in Alzheimer's disease or Parkinson's disease among six human organs and P-value smaller than 0.01 was viewed as statistically significant. The R package VennDiagram was used for Venn diagram plotting<sup>[17]</sup>.

## 1.5 Network property analysis

We downloaded the protein-protein interaction (PPI) information from mentha database and the gene co-expression information from www.brainExp.org, a website hosting the gene co-expression data from different brain regions, age stages, and genders<sup>[18]</sup>. These information were used to perform the network property analysis in this study. We filtered non-human proteins for PPI network analysis and only used a node gene's top 5 % of positively correlated co-expression gene partners (weighted correlation > 0.3) for co-expression network analysis.

We used the down-regulated genes in Alzheimer's disease or Parkinson's disease for network property analysis, because they are specifically expressed in central nervous system (brain and cerebellum) while the down-regulated ones don't exhibit a uniform tissue-specific expression pattern. According to the omnigenic model, a cellular regulatory network is made of the genes specifically expressed in a cell type.

For each down-regulated gene, we calculated its degree  $k$  and clustering coefficient  $C(k)$  using both protein-protein interaction information and gene co-expression information. Degree is a measure of a node's connectivity in a network. In protein-protein interaction network, it measures how many interaction neighbors (neighbor genes) a specific gene has. In gene co-expression network, it measures how many co-expression gene partners a specific gene has. Clustering coefficient measures a node's modularity in a network, i.e. the degree to which nodes in a network tend to cluster together. In cellular network, modularity implies certain biological function.

The clustering coefficient is mathematically defined as follows<sup>[6]</sup>:

where  $n$  is the number of direct links among a specific gene's neighbors or its co-expression partners and  $k(k-1)/2$  is the total possible number of direct links among its neighbors or partners.

## AUTHOR CONTRIBUTIONS

The first author and second author contribute equally to this article.

## REFERENCES

1. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA, Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME, Visscher PM. Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics*. 2011; 43(6): 519-525.
2. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017; 169(7): 1177-1186.
3. Burns A, Iliffe S. Alzheimer's disease. *BMJ*. 2009; 338: b158.
4. Kalia LV, Lang AE. Parkinson's disease. *Lancet*. 2015; 386(9996): 896-912.
5. Mason EF, Rathmell JC. Cell metabolism: an essential link between cell growth and apoptosis. *Biochimica et biophysica acta*. 2011; 1813(4): 645-654.
6. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature reviews Genetics*. 2004; 5(2): 101-113.
7. Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, Landfield PW. Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(7): 2173-2178.
8. Berchtold NC, Cribbs DH, Coleman PD, Rogers J, Head E, Kim R, Beach T, Miller C, Troncoso J, Trojanowski JQ, Zielke HR, Cotman CW. Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc Natl Acad Sci U S A*. 2008 Oct 7; 105(40): 15605-15610.
9. 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.
10. 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.
11. Dijkstra AA, Ingrassia A, de Menezes RX, van Kesteren RE, Rozemuller AJ, Heutink P, van de Berg WD. Evidence for Immune Response, Axonal Dysfunction and Reduced Endocytosis in the Substantia Nigra in Early Stage Parkinson's Disease. *PloS one*. 2015; 10(6): e0128651.
12. Lesnick TG, Papapetropoulos S, Mash DC, Ffrench-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE, Maraganore DM. A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS genetics*. 2007; 3(6): e98.
13. Durrenberger PF, Fernando FS, Kashefi SN, Bonnert TP, Seilhean D, Nait-Oumesmar B, Schmitt A, Gebicke-Haerter PJ, Falkai P, Grünblatt E, Palkovits M, Arzberger T, Kretschmar H, Dexter DT, Reynolds R. Common mechanisms in neurodegeneration and neuroinflammation: a BrainNet Europe gene expression microarray study. *J Neural Transm (Vienna)*. 2015; 122(7): 1055-1068.
14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015; 43(7): e47.
15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012 May; 16(5): 284-287.
16. Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, Kaessmann H. The evolution of gene expression levels in mammalian organs. *Nature*. 2011; 478(7369): 343-348.
17. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics*. 2011; 12: 35.
18. Calderone A, Castagnoli L, Cesareni G. mentha: a resource for browsing integrated protein-interaction networks. *Nature methods*. 2013; 10(8): 690-691.