African Journal of Biotechnology Vol. 10(69), pp. 15656-15661, 7 November, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.2121
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

A graph-clustering approach to search important molecular markers and pathways of Parkinson's disease

Bo Diao[1], Ying Liu[2], Yi Zhang[1] and Guo-zheng Xu[3]*

[1]Department of Medical Experiment, Wuhan General Hospital of Guangzhou Military Command of Chinese PLA, Wuhan 430070, Hubei Province, China.
[2]Department of Laboratory Technology, Wuhan General Hospital of Guangzhou Military Command of Chinese PLA, Wuhan 430070, Hubei Province, China.
[3]Department of Neurosurgery, Wuhan General Hospital of Guangzhou Military Command of Chinese PLA, Wuhan 430070, Hubei Province, China.

Accepted 7 September, 2011

Parkinson disease is the second most common neurodegenerative disorder. Therefore, it is worthwhile to search for important molecular markers and pathways that hold great promise for further treatment of patients with Parkinson's disease. DNA-microarray-based technologies allow simultaneous analysis of expression of thousands of genes. Here, we performed a comprehensive gene level assessment of Parkinson's disease using 16 colorectal cancer samples and nine normal samples. The results show that SLC6A3, SLC18A2, and EN1, etc., are related to Parkinson's disease. Besides, we further mined the underlying molecular mechanism within these different genes. The results indicate that tyrosine metabolism pathway and Parkinson's disease pathway were two significant pathways, with hope to provide insights into the development of novel therapeutic targets and pathways.

Key words: Microarrays, graph-cluster, Parkinson's disease, gene ontology (GO), pathway.

INTRODUCTION

Parkinson's disease (PD) is the second most common neurodegenerative disorder in the elderly after Alzheimer's disease (de Lau and Breteler, 2006). PD affects more than 5% of the population, with a higher prevalence in men. It is characterized clinically by severe motor symptoms including resting tremor, rigidity, bradykinesia, postural instability, and gait impairment (Shulman et al., 2010). Pathologically, disruption of motor abilities is due to decreased striatal dopamine levels, arising from selective loss of dopaminergic cells within the substantia nigra pars compacta, and locus coeruleus of the midbrain. In addition, this pathological feature is also accompanied by the presence of intracellular proteinaceous inclusions known as Lewy bodies, in surviving dopaminergic cells (Tan and Skipper, 2007).

Despite a well-described clinical and pathological phenotype, the molecular mechanisms remain elusive. Recently, genetic studies of PD indicated that many signal genes mutation are involved in neurodegeneration, such as alpha-synuclein (SNCA), leucine-rich repeat kinase 2 (LRRK2), parkin (PRKN), PTEN-induced putative kinase 1 (PINK1), ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) and DJ-1 (Vila and Przedborski, 2004). The most extensively studied PD-related genes are SNCA and LRRK2. Three missense mutations (A53T, A30P and E46K) were identified in several families of PD (Farrer, 2006). Recent results suggest that SNCA dynamically modulates neurotransmitter vesicle function. SNCA expression in *Drosophila melanogaster* flies could lead to selective loss of dopaminergic neurons, Lewy body-like inclusions and a movement disorder (Chen and Feany, 2005). LRRK2 mutations, especially in G2019S, as a cause of PD were reported in several kindred, including

*Corresponding author. E-mail: huayanl77@gmail.com. Tel: +86-13601914530.

cases with Lewy body pathology, tangle pathology, or like the original linkage report, with cell loss in the absence of obvious protein deposition (Hardy et al., 2006).

The technique of cDNA expression array is being extensively used to study global changes in gene expression in disease, model systems and in response to drug treatment (Grünblatt et al., 2001; Spies et al., 2002; Beal et al., 2005). A microarray experiment has also been designed to analyze genetic expression patterns and identify potential genes to target for PD (Anantharam et al., 2007). In this study, we used a graph-clustering approach to identify gene expression profiles that distinguish PD patients from normal samples. Furthermore, the relevant gene ontology (GO) terms in the network were analyzed to explain potential mechanisms in response to PD.

MATERIALS AND METHODS

Affymetrix microarray data and differentially expressed genes (DEGs) analysis

Microarray analysis was performed in 16 PD patients' samples and nine control samples to identify differential genes. The GSE7621 expression profile was obtained from a public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/), which are based on the Affymetrix Human Genome U133 Plus 2.0 Array.

Statistical analysis

For the GSE7621 dataset, the limma method (Smyth, 2004) was used to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the RMA method with the default settings implemented in bioconductor, and then the linear model was constructed. The DEGs only with the fold change >2 and p-value <0.05 were selected. For demonstrating the potential connection, the Spearman rank correlation (r) was used for comparative target genes correlations. The significance level was set at r >0.8 and local false discovery rate (fdr) <0.05 (Strimmer, 2008). All statistical tests were performed with the R program (http://www.r-project.org/).

Network analysis and graph clustering

To identify co-expressed groups, we used DPClus (Altaf-Ul-Amin et al., 2006), a graph clustering algorithm that can extract densely connected nodes as a cluster. It is based on density-and periphery tracking of clusters. DPClus is freely available from http://kanaya.naist.jp/DPClus/. In this study, we used the overlapping-mode with the DPClus settings. We set the parameter settings of cluster property cp; density values were set to 0.5 (Fukushima et al., 2011) and minimum cluster size was set to 5.

Gene ontology (GO) and pathway enrichment analysis

The gene ontology (Ashburner et al., 2000) project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases.

The pathway (Kanehisa, 2002) database records networks of molecular interactions in the cells, and variants of them specific to particular organisms (http://www.genome.jp/kegg/). The database for annotation, visualization and integrated discovery (DAVID) (Huang da et al., 2009) was used to identify over-represented GO terms in biological process and pathways. The P-value <0.05 is as the threshold for the analysis using the hypergeometric distribution.

RESULTS

DEGs selection and correlation network construction

We obtained publicly available microarray dataset GSE7621 from GEO. After microarray analysis, 1383 differentially expressed genes (DEGs) with the fold change >2 and p-value < 0.05 were selected using the limma method. To get the relationships among DEGs, r > 0.8 and fdr <0.05 were chosen as the threshold. The expression profiles of the 70 DEGs are shown in Figure 1.

Graph clustering identified significant modules

At r ≥0.8, DPClus (Altaf-Ul-Amin et al., 2006) identified nine clusters in the correlation network for PD, which ranged in size from 5 to 25 genes. Part of the graph clustering results is presented in Figure 2. Using the DPClus algorithm (threshold r ≥ 0.8), we extracted nine clusters in PD (only six clusters were displayed). The internal nodes of the clusters are connected by green edges; neighboring clusters are connected by red edges. The nodes refer to modules.

GO and pathway enrichment analysis

To assess the significance of the clusters, we used the over-represented GO terms and KEGG pathways in the clusters. Enrichment analysis by using the hyper-geometrical distribution is to find the significant GO terms and pathways. Some of GO terms were enriched among these genes in the correlation network, including MAPKKK cascade, catecholamine biosynthetic process and camera-type eye development, etc. (Table 1). Moreover, significant pathways, such as Parkinson's disease pathway and tyrosine metabolism were detected as shown in the Table 2.

DISCUSSION

According to our results, we found that many target genes and pathways closely related with PD had been linked by our method. Among them were SLC6A3, SLC18A2, and EN1 in the correlation network suggesting that these genes might play important roles in PD. The relationship between PD and identified genes was therefore discussed based on previous reports.

SLC6A3, known as dopamine transporter (DAT) gene,

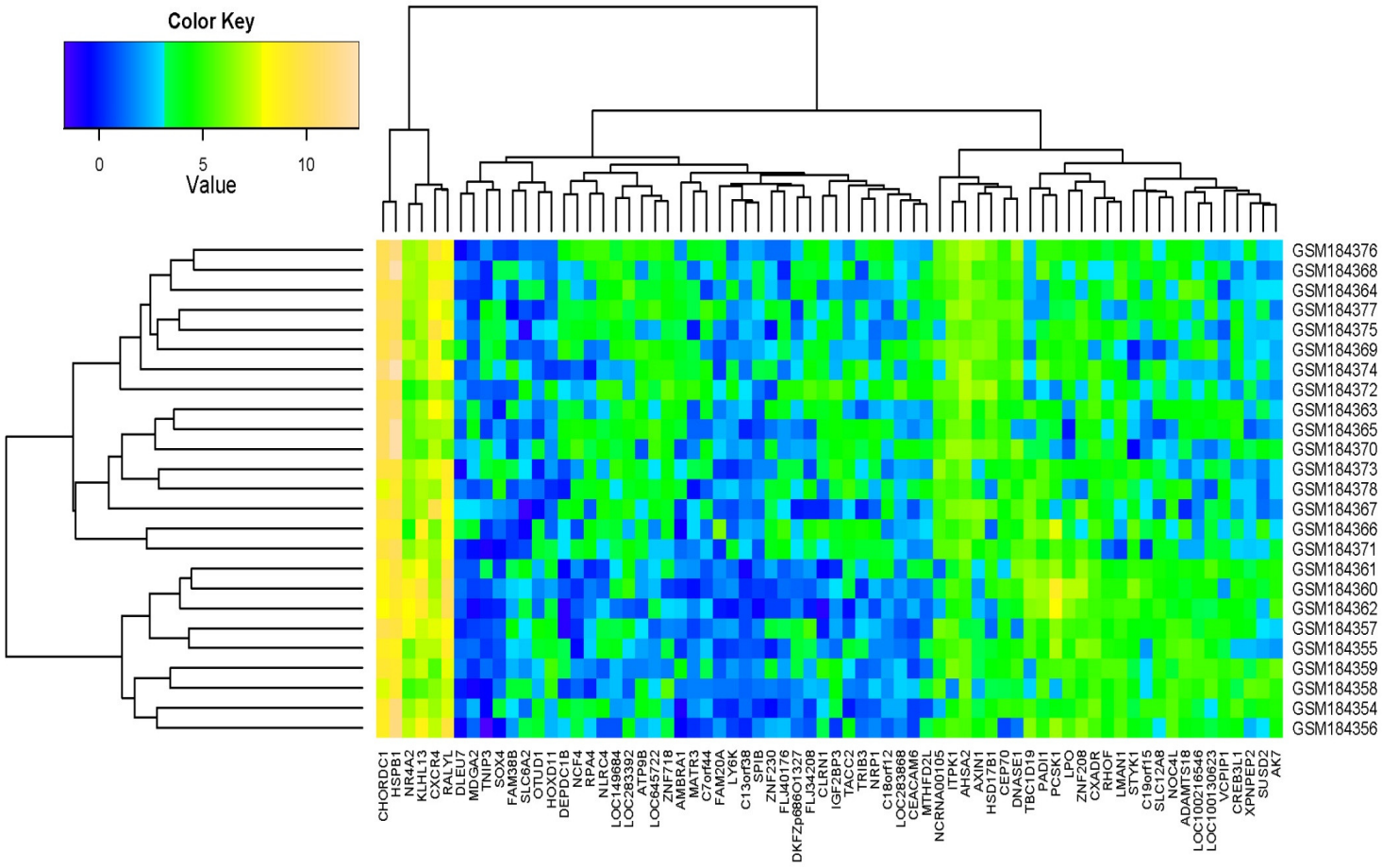

Figure 1. Expression profile clustering. The x-axis list the DEGs and the y-axis list the microarrays.

is a strong biological candidate gene for PD, which plays a critical role in dopaminergic neurotransmission and serves to terminate the synaptic signal by transporting dopamine back into the neuron from the synaptic cleft after its release (Uhl, 2003). The coding region of SLC6A3 is well conserved, but non-coding regions are more variable, most notably a variable number of tandem repeats (VNTR) polymorphism in the 3' untranslated region (Kelada et al., 2006). Some study indicated that decreased DAT availability may be necessary for, but not invariably associated with the development of anxiety and depression symptoms in PD (Weintraub et al., 2005; Dissanayaka et al., 2009).

SLC18A2, known as vesicular monoamine transporter 2 (VMAT2), sequesters cytoplasmic dopamine into synaptic vesicles for storage and release. Thus, disruption of dopamine storage has been hypothesized to damage the dopamine neurons that are lost in PD. VMAT2 mRNA has been identified with reduction in platelets from PD patients suggesting the existence of a systemic impairment of this transporter (Sala et al., 2010). Further study indicated that VMAT2-deficient mice showed obvious Parkinson's symptoms, and display progressive neurodegeneration in the substantia nigra, locus coeruleus and dorsal raphe (Taylor et al., 2011).

Engrailed 1 gene (EN1) is homeodomain-containing transcription factor necessary for the development and maintenance of mesencephalic dopaminergic neurons. Deletion in the EN1 has been shown to affect the survival of mesencephalic dopaminergic neurons both during development and in the adult (Sgado et al., 2008). EN1+/– transgenic mice were suggested to act as a potential model for PD because of the pathological and clinical resemblance exhibiting progressive mDA neuron loss as well as PD-like motor and non-motor behaviors (Haubenberger et al., 2011).

GO terms were further analyzed among these genes in the correlation network, including MAPKKK cascade, catecholamine biosynthetic process and camera-type eye development etc. MAPK cascades are well-defined signaling pathways that obey an ordered structure, consisting of a MAPKKK upstream of a MAPKK, which is in turn upstream of a MAPK. Three major MAPK pathways—ERK1/2, p38 MAPK, and JNK— have been linked to PD. Post-mortem studies showed ERK1/2, p38 MAPK and JNK to be activated in PD brains (Ferrer et al.,

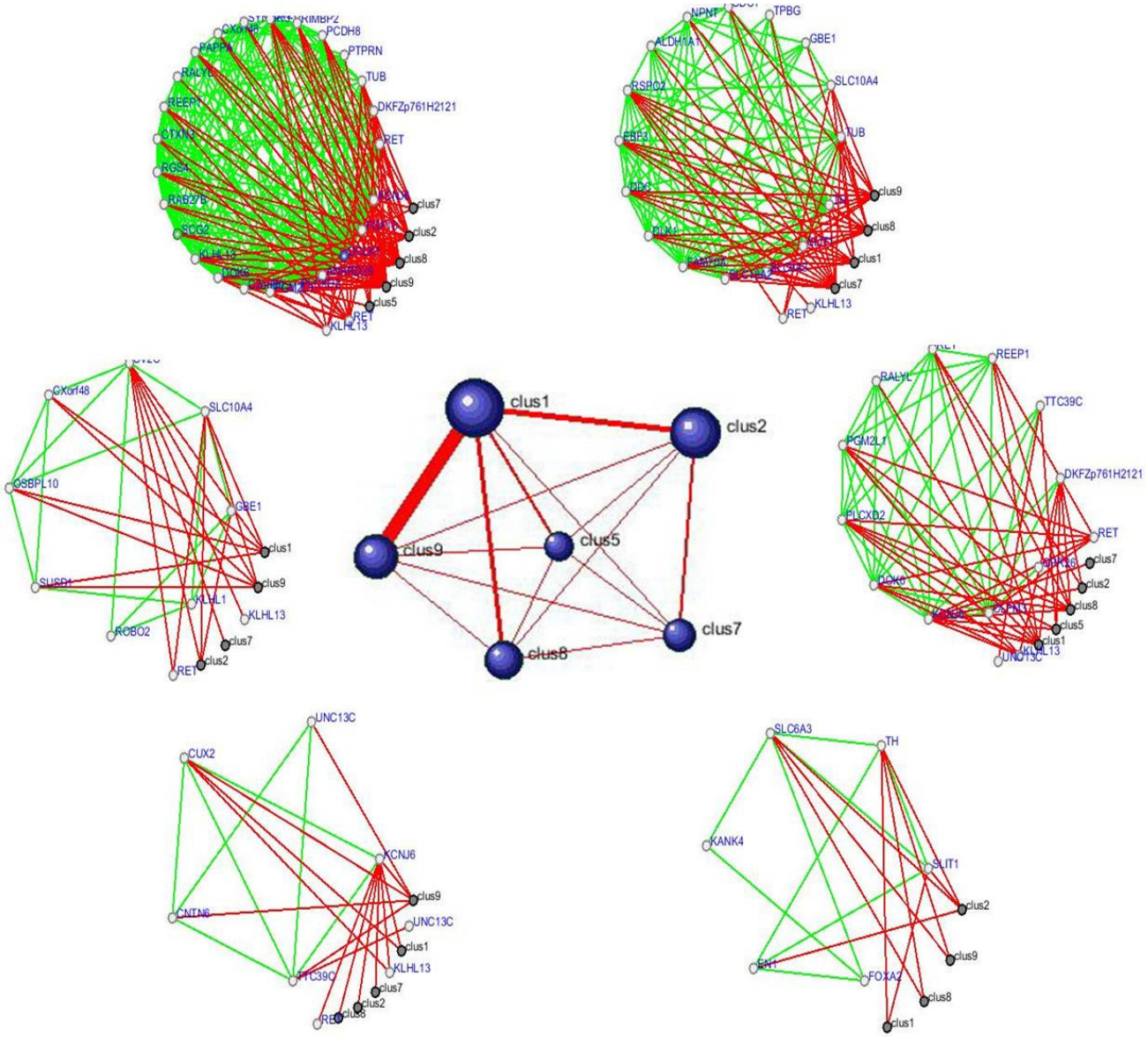

Figure 2. Graph clustering of correlated modules in PD. Using the DPClus algorithm (threshold r ≥ 0.8), we extracted nine clusters in PD (only six clusters were displayed). The internal nodes of the clusters are connected by green edges; neighboring clusters are connected by red edges. The nodes refer to modules.

Table 1. List of enriched GO term in cluster 1 to 9 detected by DPClus.

Cluster	Term	Count	P Value	Benjamini
1	GO:0000165~MAPKKK cascade	4	9.99E - 04	0.182743
2	GO:0042423~catecholamine biosynthetic process	3	4.67E - 05	0.010305
3	GO:0043010~camera-type eye development	2	0.023543	0.981731
4	GO:0009409~response to cold	2	0.007739	0.432847
5	NA	-	-	-
6	GO:0008380~RNA splicing	2	0.061672	0.98839
7	GO:0030182~neuron differentiation	4	1.32E - 04	0.028285
8	GO:0031175~neuron projection development	2	0.091118	0.999604
9	NA	-	-	-

NA, No term was detected.

Table 2. List of enriched pathways in cluster1 to 9 detected by DPClus.

Cluster	Term	Count	P Value	Benjamini
1	NA	-	-	-
2	hsa05012:Parkinson's disease	3	0.012153	0.115097
2	hsa00350:Tyrosine metabolism	2	0.059054	0.262397
3,4,5 and 6	NA	-	-	-
7	hsa05012:Parkinson's disease	2	0.073645	0.263608
8,9	NA	-	-	-

NA, No term was detected.

2001). ERK1/2 activity is central to cell death elicited by dopamine treatment (Liao et al., 2009). Over-expressed ERK1, p38a or JNK1 are sufficient to elicit apoptosis in dopaminergic neurons. Intriguingly, in all cases, this can be rescued by co-transfection of Parkin, encoded by PARK2, which is disrupted in recessive forms of early onset PD (Ren et al., 2009). MAPK pathways also could mediate the effects of pathogenic LRRK2 mutation (Berwick et al., 2011).

Catecholamine (CA) implies DA and its metabolic products, noradrenaline (NA) and adrenaline (A). These CAs are synthesized from the amino acid L-tyrosine (L - Tyr) in a common biosynthetic pathway that uses six enzymes. Among them, tyrosine hydroxylase (TH) is the rate-limiting enzyme for CA biosynthesis (Flatmark, 2000). PD is characterized by decreasing neuromelanin in the SN. Neuromelanin synthesis results from excess cytosolic CAs. Dopaminergic neuron in the midbrain could express TH. TH-negative pigmented neurons are present in these nuclei of patients with PD. Thus, it is likely that TH negativity is a manifestation of a diseased cell with decreased functional ability to synthesize dopamine (Mori et al., 2006).

To identify the relevant pathways changed in each cluster, we used the hypergeometric distribution approach on pathway level. Significance analysis with the p-value <0.05, shows that only Parkinson's disease pathway and tyrosine metabolism pathway were as significant. Tyrosine metabolism is mainly regulated by tyrosinase. Tyrosinase in the brain catalyzes the hydroxylation of tyrosine to L-DOPA and the consequent oxidation of L-DOPA to form melanin in the melanin biosynthetic pathway of peripheral melanocytes, thereby preventing slow progressive cell damage by auto-oxidation of DA, thus maintaining DA levels. Besides, tyrosinase also possesses catecholamine-synthesizing activity to synthesize DA in the absence of TH as above mentioned. These findings imply that tyrosinase may be involved in DA synthesis and oxidation, and the dysfunction of this activity may contribute to the neurodegeneration associated with Parkinson's disease (Lücking and Brice, 2000; Asanuma et al., 2003).

Conclusion

So far, we have used network analysis as a conceptual framework to explore the pathobiology of PD based on the assumption that PD is a contextual attribute of distinct patterns of interactions between multiple genes. The salient results of our study include SLC6A3 gene, SLC18A2 gene, EN1 gene, tyrosine metabolism pathway, MAPK cascades GO terms and catecholamine synthesis GO terms, which were all related with PD in direct or indirect manner. However, further experiments are still indispensable to confirm our conclusion.

REFERENCES

Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K,Kanaya S (2006). Development and implementation of an algorithm for detection of protein complexes in large interaction networks. BMC Bioinformatics, 7: p. 207.

Anantharam V, Lehrmann E, Kanthasamy A, Yang Y, Banerjee P, Becker KG, Freed WJ, Kanthasamy AG (2007). Microarray analysis of oxidative stress regulated genes in mesencephalic dopaminergic neuronal cells: relevance to oxidative damage in Parkinson's disease. Neurochem. Int. 50(6): 834-847.

Asanuma M, Miyazaki I,Ogawa N (2003). Dopamine-or L-DOPA-induced neurotoxicity: the role of dopamine quinone formation and tyrosinase in a model of Parkinson¡¯s disease. Neurotoxicity Res. 5(3): 165-176.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25(1): 25-29.

Beal MJ, Falciani F, Ghahramani Z, Rangel C, Wild DL (2005). A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. Bioinformatics, 21(3): p. 349.

Berwick DC, Harvey K (2011). LRRK2 signaling pathways: the key to unlocking neurodegeneration? Trends Cell Biol. 21(5): 257-265.

Chen L, Feany MB (2005). Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a Drosophila model of Parkinson disease. Nat. Neurosci. 8(5): 657-663.

de Lau LML, Breteler M (2006). Epidemiology of Parkinson's disease. Lancet Neurol. 5(6): 525-535.

Dissanayaka NN, Silburn PA, O'Sullivan JD, Mellick GD (2009). Serotonin and dopamine transporter genes do not influence depression in Parkinson's disease. Movement disorders : Official J. Movement Disorder Soc., 24(1): 111-115.

Farrer MJ (2006). Genetics of Parkinson disease: paradigm shifts and

future prospects. Nat. Rev. Genet., 7(4): 306-318.
Ferrer I, Blanco R, Carmona M, Puig B, Barrachina M, Gomez C, Ambrosio S (2001). Active, phosphorylation-dependent mitogen-activated protein kinase (MAPK/ERK), stress-activated protein kinase/c-Jun N-terminal kinase (SAPK/JNK), and p38 kinase expression in Parkinson's disease and Dementia with Lewy bodies. J. Neural Trans. 108(12): 1383-1396.
Flatmark T (2000). Catecholamine biosynthesis and physiological regulation in neuroendocrine cells. Acta Physiologica Scandinavica, 168(1): 1-17.
Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2011). Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. BMC Syst. Biol. 5: p. 1.
Grünblatt E, Mandel S, Maor G, Youdim MBH (2001). Gene expression analysis in N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine mice model of Parkinson's disease using cDNA microarray: effect of R-apomorphine. J. Neurochem. 78(1): 1-12.
Hardy J, Cai H, Cookson MR, Gwinn-Hardy K,Singleton A (2006). Genetics of Parkinson's disease and parkinsonism. Ann. Neurol. 60(4): 389-398.
Haubenberger D, Reinthaler E, Mueller JC, Pirker W, Katzenschlager R, Froehlich R, Bruecke T, Daniel G, Auff E, Zimprich A (2011). Association of transcription factor polymorphisms PITX3 and EN1 with Parkinson's disease. Neurobiol. Aging. 32(2): 302-307.
Huang da W, Sherman BT, Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4(1): 44-57.
Kanehisa M (2002). The KEGG database. Novartis Found Symp. 247: 91-101; discussion pp. 101-103, 119-128, 244-152.
Kelada SN, Checkoway H, Kardia SL, Carlson CS, Costa-Mallen P, Eaton DL, Firestone J, Powers KM, Swanson PD, Franklin GM, Longstreth Jr. WT, Weller TS, Afsharinejad Z, Costa LG (2006). 5' and 3' region variability in the dopamine transporter gene (SLC6A3), pesticide exposure and Parkinson's disease risk: a hypothesis-generating study. Hum. Mol. Genet. 15(20): 3055-3062.
Lücking C, Brice A (2000). Alpha-synuclein and Parkinson's disease. Cellular Mol. Life Sci. 57(13): 1894-1908.
Liao Y, Wei Y, Zhou X, Yang JY, Dai C, Chen YJ, Agarwal NK, Sarbassov D, Shi D, Yu D, Hung MC (2009). Peptidyl-prolyl cis/trans isomerase Pin1 is critical for the regulation of PKB/Akt stability and activation phosphorylation. Oncogene, 28(26): 2436-2445.
Mori F, Nishie M, Kakita A, Yoshimoto M, Takahashi H, Wakabayashi K (2006). Relationship among alpha-synuclein accumulation, dopamine synthesis, and neurodegeneration in Parkinson disease substantia nigra. J. Neuropathol. Exp. Neurol. 65(8): 808-815.
Ren Y, Jiang H, Yang F, Nakaso K, Feng J (2009). Parkin protects dopaminergic neurons against microtubule-depolymerizing toxins by attenuating microtubule-associated protein kinase activation. J. Biol. Chem. 284(6): 4009-4017.
Sala G, Brighina L, Saracchi E, Fermi S, Riva C, Carrozza V, Pirovano M, Ferrarese C (2010). Vesicular monoamine transporter 2 mRNA levels are reduced in platelets from patients with Parkinson's disease. J. Neural. Trans. 117(9): 1093-1098.
Sgado P, Viaggi C, Fantacci C, Corsini GU (2008). Characterization of the Engrailed mutant mice as experimental models for Parkinson's disease. Parkinsonism related disorders, 14(2): 103-106.
Shulman J, De Jager P, Feany M (2010). Parkinson's Disease: Genetics and Pathogenesis. Annu. Rev. Pathol. 2011 Feb 28;6:193-222. Review.
Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications Genet. Mol. Biol. 3: p. 3.
Spies M, Dasu MR, Svrakic N, Nesic O, Barrow RE, Perez-Polo JR, Herndon DN (2002). Gene expression analysis in burn wounds of rats. Am. J. Physiol. Regul. Integr. Comp. Physiol. 283(4): 918-930.
Strimmer K (2008). fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics, 24(12): 1461-1462.
Tan EK, Skipper LM (2007). Pathogenic mutations in Parkinson disease. Human mutation, 28(7): 641-653.
Taylor TN, Caudle WM, Miller GW (2011). VMAT2-Deficient Mice Display Nigral and Extranigral Pathology and Motor and Nonmotor Symptoms of Parkinson's Disease. Parkinson's Disease. 2011.
Uhl GR (2003). Dopamine transporter: basic science and human variation of a key molecule for dopaminergic function, locomotion, and parkinsonism. Movement disorders: Official Journal of the Movement Disorder Society. 18(7): 71-80.
Vila M, Przedborski S (2004). Genetic clues to the pathogenesis of Parkinson's disease. Nat. Med., 10: p. 58.
Weintraub D, Newberg AB, Cary MS, Siderowf AD, Moberg PJ, Kleiner-Fisman G, Duda JE, Stern MB, Mozley D, Katz IR (2005). Striatal dopamine transporter imaging correlates with anxiety and depression symptoms in Parkinson's disease. J. Nuclear medicine : Official Publication, Society Nuclear Medicine, 46(2): 227-232.