

Identification of crucial genes associated with Parkinson's disease using microarray data

YONGQI SUN, LINLIN YE, YONGHUI ZHENG and ZICHAO YANG

Department of Neurology, The Fourth Affiliated Hospital of Harbin Medical University,
Harbin, Heilongjiang 150001, P.R. China

Received March 18, 2016; Accepted February 23, 2017

DOI: 10.3892/mmr.2017.8305

Abstract. The present study aimed to examine potential crucial genes associated with Parkinson's disease (PD) in addition to the interactions and regulators of these genes. The chip data (GSE7621) were obtained from the Gene Expression Omnibus and standardized using the robust multi-array average in the Affy package of R software. The differentially expressed genes (DEGs) were then screened using the Samr package with a false discovery rate (FDR) <0.05 and \log_2 fold change (FC) >1. Crucial PD-associated genes were predicted using the Genetic Association Database in the Database for Annotation, Visualization and Integrated Discovery and sequence alignment. Furthermore, transcription factors (TFs) of the crucial PD-associated genes were predicted, and protein-protein interactions (PPIs) between the crucial PD-associated genes were analyzed using the Search Tool for the Retrieval of Interacting Genes/Proteins. Additionally, another dataset of PD was used to validate the expression of crucial PD-associated genes. A total of 670 DEGs (398 upregulated and 272 downregulated genes) were identified in the PD samples. Of these, 10 DEGs enriched in pathways associated with the nervous system were predicted to be crucial in PD, including C-X-C chemokine receptor type 4 (*CXCR4*), deleted in colorectal cancer (*DCC*) and NCL adaptor protein 2 (*NCK2*). All 10 genes were associated with neuron development and differentiation. They were simultaneously modulated by multiple TFs, including *GATA*, *E2F* and E4 promoter-binding protein 4. The PPI networks showed that *DCC* and *CXCR4* were hub proteins. The *DCC*-netrin 1-roundabout guidance receptor 2-slit guidance ligand 1 interaction pathway, and several genes, including TOX high mobility group box family member 4, kinase insert domain receptor and zymogen granule protein 16B, which interacted with *CXCR4*, were novel findings. Additionally,

CXCR4 and *NCK2* were upregulated in another dataset (GSE8397) of PD. These genes, interactions of proteins and TFs may be important in the progression of PD.

Introduction

Parkinson's disease (PD) is a common chronic neurodegenerative disorder, which affects >0.1% of individuals >40 years of age (1,2). There is a loss of dopaminergic nigrostriatal neurons in the substantia nigra (SN) pars compacta in PD (3,4), and clinical characteristics for the majority of patients with PD include motor disorder, slowness of movement, tremor at rest, rigidity and disturbances in balance. However, it is difficult to diagnose PD in its early stages due to the high rate of misdiagnosis with other movement disorders, which can share certain features with PD, including essential tremor, multiple system atrophy (MSA) and progressive supranuclear palsy (5). Therefore, it is necessary and urgent to screen diagnostic or therapeutic biomarkers, which can be utilized in the diagnosis or treatment of PD.

In previous years, numerous significant factors have been correlated with the development of PD, which has improved current understanding of PD. For example, the level of S100 calcium binding protein B (S100B) is higher in the post-mortem SN of patients with PD, compared with control tissue, and the ablation of S100B results in neuroprotection, reduces microgliosis and decreases the expression of tumor necrosis factor- α and the receptor for advanced glycation end products (6). The reduced expression level of lysosomes-associated membrane protein 2A in dopaminergic cell lines leads to reduced chaperone-mediated autophagy and decreases the metabolism of wild-type α -synuclein (*SNCA*), which is a cause of PD (7). In addition, a set of molecules, which offer potential in identifying individuals at risk for PD, and assist in the diagnosis and management of PD have been identified, including *SNCA* (8) and DJ-1 (9). Furthermore, based on microarray analysis, Papapetropoulos *et al* identified 11 candidate genes with deregulated expression in brain regions from patients with PD, and the significant role of mitochondria ribosomal protein S6 (*MRPS6*) in the oxidative phosphorylation system was reported (10). In addition, a previous study identified a set of differentially expressed genes (DEGs) in PD, and the candidate compound, alvespimycin, was selected as a lead compound for PD, based on the gene expression data from

Correspondence to: Dr Zichao Yang, Department of Neurology, The Fourth Affiliated Hospital of Harbin Medical University, 37 Yiyuan Street, Harbin, Heilongjiang 150001, P.R. China
E-mail: yujunpanpan122@163.com

Key words: Parkinson's disease, differentially expressed gene, transcription factor, protein-protein interaction network

the study by Papapetropoulos *et al* (10,11). In organisms, each protein exerts its functions via interactions with other proteins and regulation of a set of regulators. However, interactions and regulators of DEGs in PD, which assist in obtaining a more comprehensive understanding of the pathogenesis of PD, were not fully investigated in the studies of Papapetropoulos *et al* and Gao *et al* (10,11); there may be additional potential key genes associated with PD, which remain undetected.

The present study aimed to examine additional potential key genes associated with PD, in addition to examining their interactions and regulation. The microarray data of Papapetropoulos *et al* (10) were used to identify DEGs between patients with PD and those without PD. The potential crucial genes associated with PD were then examined. Subsequently, transcription factors (TFs) and protein-protein interactions (PPIs) of the DEGs were analyzed. A second microarray dataset (12) of PD, which included additional post-mortem brain samples, was selected to validate the expression of key genes in PD. The results may provide novel information for investigation of the molecular mechanisms underlying the pathogenesis of PD.

Materials and methods

Gene expression data of PD. The GSE7621 gene expression data (10) were obtained from the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) based on [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array platform (GPL570; Affymetrix, Inc., Santa Clara, CA, USA). A total of 25 post-mortem SN samples, including 16 samples from patients with idiopathic PD and nine samples from aged controls without PD, were included in the GSE7621 expression data. For the patients with PD, the ratio of men to women was 13:3, and the mean age was 74 years, ranging between 60 and 88. For the controls, the ratio of men to women was 4:5, and the mean age was 78.2 years, ranging between 46 and 90. The age, post-mortem interval (PMI), and brain pH of the two subject groups were matched as closely as possible. Analysis of the quality control parameters showed no significant differences in age, gender, brain pH, PMI or RNA QC values between the aged control and PD groups (10).

Another gene expression dataset, GSE8397 (12), was used for validation, which was also downloaded from the GEO, which was based on the platforms of Affymetrix Human Genome U133A Array (GPL96) and Affymetrix Human Genome U133B Array (GPL97; Affymetrix, Inc.). A total of 47 individual tissue samples analyzed using one A and one B GeneChip per sample [15 medial SN (MSN) samples and nine lateral SN (LSN) samples obtained from patients with idiopathic PD, eight MSN samples and seven LSN samples obtained from controls without PD, five superior frontal gyrus samples from patients with PD, and three superior frontal gyrus samples from controls] were included in this dataset. The LSN and MSN samples were obtained from the same cohort of patients with idiopathic PD or controls. For the patients with PD, the ratio of men to women was 9:6, and the mean age was 80 years, ranging between 68 and 89; for the controls, the ratio of men to women was 6:2, and the mean age was 70.6 years, ranging between 46 and 81. The subjects in the PD and control groups were well matched in age, PMI and brain

pH (12). Among the 47 brain tissue samples, only the data of 16 LSN samples and 23 MSN samples from the patients with PD and controls produced by the GPL96 platform were used in the present study.

Data preprocessing and identification of DEGs. The CEL files and probe annotation files were downloaded from the database. The raw data were preprocessed via background correction, quantile normalization, and calculating expression using the robust multi-array average (RMA) algorithm (13). The probes were then mapped into corresponding gene names. The probe, which corresponded to a plurality of genes was removed. If one gene symbol was matched by multiple probe IDs, the mean expression value was selected as the expression level of this gene. The Samr package in R software (version 3.1; <http://cran.r-project.org/web/packages/samr/index.html>) (14) was used to select the DEGs between the PD and control samples. The P-value for each gene was calculated using an unpaired t-test, and adjusted into a false discovery rate (FDR) using the Benjamini-Hochberg (BH) method (15). Only the genes with FDR<0.05 and \log_2 fold change (FC)>1, as commonly used criteria, were selected as DEGs.

Prediction of crucial PD-associated genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (<http://www.kegg.jp/kegg/pathway.html>) is a collection of manually drawn pathway maps representing current knowledge on the molecular interaction and reaction networks for metabolism, cellular processes and human diseases (16). Firstly, the online Database for Annotation, Visualization and Integrated Discovery (DAVID; <https://david.ncifcrf.gov/>), which integrates a set of functional annotation tools to understand the biological meaning of numerous genes (17), was used to perform KEGG pathway enrichment analysis for the identified DEGs. The P-value calculated by Fisher's exact test of P<0.05, and a gene count >2 were set as the cut-off criteria.

Secondly, in order to obtain the PD-associated genes, the DEGs were submitted to the Genetic Association Database (<http://geneticassociationdb.nih.gov/>), which is a database of genetic association data from complex diseases and disorders. PD-associated genes with P<0.05 were considered significant.

Finally, to identify the crucial PD-associated genes, sequence alignment of the protein sequences of the genes, which were involved in the nervous system-associated KEGG pathways and PD-associated genes was performed using the BLASTP program (version 2.2.31; [https://blast.ncbi.nlm.nih.gov/Blast.cgi?; E<10⁻¹⁰](https://blast.ncbi.nlm.nih.gov/Blast.cgi?; E<10<sup>-10</sup)). The smaller the E value was, the more reliable the alignment result. If the protein sequence of the gene involved in the nervous system-associated KEGG pathways was matched to multiple protein sequences of PD-associated genes, and had the lowest E-value, this gene was considered to be a crucial PD-associated gene.

Gene ontology (GO) enrichment analysis of crucial PD-associated genes. The GO database collects biological data in the format of gene-to-annotation, and is a suitable and commonly used method for high-throughput bioinformatics scanning for enrichment analysis (18). GO enrichment analysis of the predicted crucial PD-associated genes was performed

using DAVID. Only the GO terms with the common criterion of $FDR < 0.01$ were considered significant.

Prediction of TFs for the crucial PD-associated genes. TF binding sites (TFBSs) are short DNA motifs in genes, which bind with TFs. TFBSs often occur in close proximity to each other forming *cis*-regulatory modules, which suggest the existence of a combinatorial code for transcriptional regulation (19). Based on the known TFBS information in the University of California Santa Cruz (UCSC) database (<http://genome.ucsc.edu/>) in DAVID (20), the TFs in the UCSC database targeting the crucial PD-associated genes were identified. $FDR < 0.01$ was set as the cut-off criterion.

PPI network analysis. PPI analysis of the screened DEGs was performed using the online Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; version 10.0; <http://string-db.org/>) software, which integrates known and predicted protein interaction data (21) (combined score > 0.4). The PPI networks containing the predicted crucial PD-associated genes were then extracted and visualized using Cytoscape software (version 3.2.0; <http://www.cytoscape.org/>), which is an open source software project for integrating biomolecular interaction networks (22). In a network, a node represents a protein (gene), and lines represent interactions of the proteins. The 'degree' of each node is equal to the number of nodes interacting with this node. The higher the degree is, the closer the connections with other nodes are, indicating a higher importance of the node in the network.

Validation of the expression level of crucial PD-associated genes. The DEGs were identified in the PD and control samples from another dataset, GSE8397 (Moran *et al.* 2006) using the same method as that used for GSE7621. The identified DEGs were then compared with those in the GSE7621 dataset.

Results

Analyses of DEGs and crucial PD-associated genes. Following normalization of the raw data in the GSE7621 dataset, a total of 670 DEGs were identified in the SN of PD, including 398 upregulated genes and 272 downregulated genes (data not shown). Based on the criteria of $P < 0.05$ and gene count > 2 , these DEGs were associated with the pathways of axon guidance ($P = 0.003347$; gene count, 12; http://www.genome.jp/dbget-bin/show_pathway?hsa04360+1630+4893+8440+3265+219699+80031+7852+56920+6092+9423+4775+6585) and endocytosis ($P = 0.017984$; gene count=13; http://www.genome.jp/dbget-bin/show_pathway?hsa04144+5979+3265+79643+80230+3303+3791+3305+7852+2870+57154+10938+8218+22905), which were associated with the nervous system. The DEGs, including deleted in colorectal cancer (*DCC*), NCL adaptor protein 2 (*NCK2*) and C-X-C chemokine receptor type 4 (*CXCR4*), were significantly involved in the axon guidance pathway, and a set of DEGs, including *CXCR4* and clathrin heavy chain like 1 (*CLTCL1*), were associated with the endocytosis pathway (Table I).

A total of 10 DEGs, which may be correlated with PD were identified from the Genetic Association Database, including heat shock 70 kDa protein 1-like (*HSPA1L*), Cytochrome P450

Table I. Kyoto Encyclopedia of Genes and Genomes pathways associated with the nervous system for differentially expressed genes.

Term	P-value	Genes
hsa04360 Axon guidance	0.003347	<i>DCC</i> , <i>NRAS</i> , <i>NCK2</i> , <i>HRAS</i> , <i>UNC5B</i> , <i>SEMA6D</i> , <i>CXCR4</i> , <i>SEMA3G</i> , <i>ROBO2</i> , <i>NTN1</i> , <i>NFATC3</i> , <i>SLIT1</i>
hsa04144 Endocytosis	0.017984	<i>RET</i> , <i>HRAS</i> , <i>CHMP6</i> , <i>RUFY1</i> , <i>HSPA1A</i> , <i>KDR</i> , <i>HSPA1L</i> , <i>CXCR4</i> , <i>GRK6</i> , <i>SMURF1</i> , <i>EHD1</i> , <i>CLTCL1</i> , <i>EPN2</i>

family 27 subfamily A1 (*CYP27A1*), dopamine receptor D2 (*DRD2*), solute carrier family 6 member 3 (*SLC6A3*), PARK2 coregulated (*PACRG*), tyrosine hydroxylase (*TH*), nuclear receptor subfamily 4 A2 (*NR4A2*), solute carrier family 18 member A2 (*SLC18A2*), luteal phase S-transferase ζ 1 (*GSTZ1*) and heat shock protein family A member 1A (*HSPA1A*). Based on the DEGs enriched in the pathways of axon guidance and endocytosis, in addition to DEGs correlated with PD, which were identified using the Genetic Association Database, a total of 10 genes were predicted to be closely associated with PD via sequence alignment, including *CXCR4*, *DCC*, EH domain-containing 1 (*EHD1*), epsin-2 (*EPN2*), G protein-coupled receptor kinase 6 (*GRK6*), *NCK2*, roundabout guidance receptor 2 (*ROBO2*), RUN and FYVE domain containing 1 (*RUFY1*), semaphorin 6D (*SEMA6D*) and *SLIT1*. These were defined as crucial PD-associated genes.

GO enrichment analysis. To further reveal the biological functions of the predicted crucial PD-associated genes, GO enrichment analysis was performed. The majority of GO terms of the predicted crucial PD-associated genes, including *NCK2*, *CXCR4*, *SLIT1*, *ROBO2* and *DCC*, were involved in neuron development and differentiation, and cell morphogenesis. *EHD1*, *RUFY1*, and *EPN2* were associated with GO terms associated with endocytosis (Table II). No significant GO terms were enriched by the other two crucial PD-associated genes (*SEMA6D* and *GRK6*).

Analyses of TFs and PPI interaction networks. To further investigate the regulatory factors and interactions of the crucial PD-associated genes, TFs targeting the genes were predicted. A total of 55 TFs were found to regulate the 10 crucial PD-associated genes. *DCC*, *CXCR4*, *NCK2*, *SLIT1*, *ROBO2* and *GRK6* were all regulated by a set of TFs, including *GATA*, *E2F* and E4 promoter-binding protein 4 (*E4BP4*) (Table III).

Furthermore, the PPIs of the crucial PD-associated genes were searched, and a PPI network was constructed. In the PPI networks, the downregulated *DCC* and upregulated *CXCR4* were the hub nodes, interacting with multiple genes (Fig. 1). *DCC*, *CXCR4* and *NCK2* interacted with breast cancer anti-estrogen resistance 1 (*BCAR1*), and netrin 1 (*NTN1*) interacted with *ROBO2*, which interacted with *SLIT1*.

Table II. Gene ontology enrichment analysis of the crucial Parkinson's disease-associated genes.

Gene	Gene Ontology terms
<i>SLIT1</i>	Cell morphogenesis, neuron development, neuron projection development, cell part morphogenesis, cell projection morphogenesis, cell morphogenesis involved in differentiation, neuron projection morphogenesis, cell morphogenesis involved in neuron differentiation, axonogenesis, cell motion, cell projection organization, axon guidance, neuron differentiation, cellular component morphogenesis
<i>RUFY1</i>	Regulation of endocytosis, membrane organization, membrane invagination, endocytosis
<i>ROBO2</i>	Cell morphogenesis, neuron development, neuron projection development, cell part morphogenesis, cell projection morphogenesis, cell morphogenesis involved in differentiation, neuron projection morphogenesis, cell morphogenesis involved in neuron differentiation, axonogenesis, cell motion, cell projection organization, axon guidance, neuron projection, neuron differentiation, cellular component morphogenesis
<i>NCK2</i>	Cell motion, cell projection organization, cell motility, localization of cell, cell migration
<i>EPN2</i>	Endocytic vesicle, regulation of endocytosis, membrane organization, membrane invagination, endocytosis
<i>EHD1</i>	Endocytic vesicle, membrane organization, membrane invagination, endocytosis
<i>DCC</i>	Cellular component morphogenesis, cell morphogenesis, neuron development, neuron projection development, cell part morphogenesis, cell projection morphogenesis, cell morphogenesis involved in differentiation, neuron projection morphogenesis, cell morphogenesis involved in neuron differentiation, axonogenesis, cell motion, cell projection organization, axon guidance, neuron migration, neuron projection, cell motility, localization of cell, cell migration, neuron differentiation
<i>CXCR4</i>	Cellular component morphogenesis, cell morphogenesis, neuron development, neuron projection development, cell part morphogenesis, cell projection morphogenesis, cell morphogenesis involved in differentiation, neuron projection morphogenesis, cell morphogenesis involved in neuron differentiation, axonogenesis, cell motion, cell projection organization, axon guidance, neuron migration, neuron projection, cell motility, localization of cell, cell migration, neuron differentiation

False discovery rate of all terms is <0.01.

Validation of the expression level of crucial PD-associated genes. To confirm the expression level of the potential crucial PD-associated genes, DEGs in another microarray dataset, GSE8397 (12), were identified, and the mean probe expression levels of each common DEG in the two datasets were compared manually. In the other microarray dataset of PD (12), two crucial PD-associated genes (*CXCR4* and *NCK2*) were also upregulated in the PD samples, compared with the controls, which was consistent with that in the first dataset (10). *CXCR4* was significantly upregulated in the LSN and MSN samples (Fig. 2A and B), and *NCK2* was upregulated in the MSN samples (Fig. 2B).

Discussion

In the present study, based on the analysis of the GSE7621 dataset, 398 upregulated genes and 272 downregulated genes were identified in the SN from the patients with PD, compared with the controls. A total of 10 DEGs (*CXCR4*, *DCC*, *EHD1*, *EPN2*, *GRK6*, *NCK2*, *ROBO2*, *RUFY1*, *SEMA6D* and *SLIT1*) were predicted to be closely associated with PD. Among these 10 genes, downregulated *DCC* and upregulated *CXCR4* were the hub proteins, which interacted with multiple genes in the PPI network. These two proteins were distinctly enriched in the axon guidance pathway, and GO terms associated with neuron development and differentiation.

In the PPI network, *DCC* interacted with *NTN1*. *DCC* encodes an NTN1 receptor, a member of the immunoglobulin superfamily, which mediates the axon guidance of neuronal

growth cones towards sources of NTN1 ligand (23). *NTN1* binds to the *DCC* receptor and signals through downstream proteins, which regulate cytoskeletal reorganization, causing growth cone extension and neurite growth (24). Another study reported that *NTN1* and *DCC* are able to attract axons to the nervous system midline, and mutations in *DCC* lead to midline guidance defects in humans (25). There is also evidence that *DCC*-mediated NTN1 signaling affects the formation/maintenance of mesocorticolimbic DA topography (26). Polymorphisms in *NTN1* and *DCC* are associated with the loss of multiple functions and increased the risk of PD (27,28). These results suggested that the decreased expression of *DCC* and *NTN1* may lead to a loss of dopaminergic nigrostriatal neurons in the SN, resulting in the onset of PD. In the PPI network, *NTN1* interacted with *ROBO2*, which interacted with *SLIT1*. The *ROBO2* protein belongs to the Robo family, which is important for axon guidance across the midline during central nervous system development (29). *ROBO2* has also previously been found to be deregulated in PD (30). Loss of Slit1 expression results in an abnormal course of the nigrostriatal pathway through the diencephalon (31). Although the genes of the *DCC-NTN1-ROBO2-SLIT1* interaction pathway have been previously reported to be associated with the nervous system, their interactions in the present study are a novel finding.

CXCR4 encodes a CXC chemokine receptor specific for stromal cell-derived factor-1 (32). In the SN of patients with PD, higher expression of *CXCR4* and its natural ligand, *CXCL12*, have been detected, compared with control subjects (28,33),

Table III. List of the 55 transcription factors targeting the crucial Parkinson's disease-associated genes.

Gene	Transcription factors
<i>DCC</i>	<i>IK2, SP1, HNF3B, MYOGNF1, TATA, TCF11, GFII, NKX25, FAC1, PAX2, MEIS1B, HOXA9, OLF1, ARP1, HEN1, CEBPB, GATA1, HOXA3, HAND1E47, S8, CDP, RSRFC4, SOX5, NFY, CDP, CR3HD, SREBP1, MZF1, NRSF, MYC, MAX, AHRARNT, PAX5, MSX1, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, PAX3, AP2REP, AREB6, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</i>
<i>EPN2</i>	<i>IK2, SP1, HNF3B, TCF11, GFII, NKX25, PAX2, ARP1, HEN1, GATA1, HAND1E47, S8, RSRFC4, CDP, CR3HD, MZF1, NRSF, MYC, MAX, PAX5, MYB, AP4, AP2REP, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, AML1, LMO2COM, E4BP4</i>
<i>CXCR4</i>	<i>IK2, MYOGNF1, TATA, TCF11, GFII, NKX25, FAC1, PAX2, MEIS1B, HOXA9, HAND1E47, S8, CDP, SOX5, NRSF, MYC, MAX, AHRARNT, MSX1, MYB, GATA2, IRF1, GATA, PAX3, AP2REP, AREB6, SRY, MEF2, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, E4BP4</i>
<i>NCK2</i>	<i>IK2, HNF3B, MYOGNF1, TATA, TCF11, NKX25, PAX2, MEIS1B, HOXA9, CEBPB, GATA1, HOXA3, S8, CDP, RSRFC4, SOX5, NFY, CDP, CR3HD, MZF1, NRSF, MYC, MAX, PAX5, MSX1, IRF1, AP4, GATA, MRF2, AP2REP, AREB6, MEF2, GRE, USF, BRACH, COMP1, AML1, E2F, E4BP4</i>
<i>SLIT1</i>	<i>IK2, HNF3B, MYOGNF1, TATA, TCF11, GFII, NKX25, FAC1, PAX2, MEIS1B, HOXA9, OLF1, ARP1, HEN1, CEBPB, GATA1, HAND1E47, S8, CDP, RSRFC4, SOX5, CDP, CR3HD, SREBP1, NRSF, MYC, MAX, AHRARNT, PAX5, MSX1, TAX, CREB, MYB, COUP, GATA2, AP4, GATA, MRF2, PAX3, AP2REP, AREB6, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</i>
<i>SEMA6D</i>	<i>IK2, HNF3B, TATA, TCF11, GFII, NKX25, FAC1, PAX2, MEIS1B, HOXA9, ARP1, HEN1, CEBPB, GATA1, HOXA3, HAND1E47, S8, CDP, NFY, CDP, CR3HD, SREBP1, MZF1, MYC, MAX, PAX5, MSX1, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, AREB6, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</i>
<i>RUFY1</i>	<i>SP1, HNF3B, TATA, NKX25, PAX2, MEIS1B, HOXA9, OLF1, ARP1, CEBPB, GATA1, HOXA3, HAND1E47, S8, CDP, RSRFC4, NFY, CDP, CR3HD, SREBP1, MZF1, NRSF, AHRARNT, PAX5, MSX1, TAX, CREB, IRF1, AP4, GATA, MRF2, AP2REP, AREB6, SRY, MEF2, GRE, USF, BRACH, PAX6, AML1, LMO2COM, YY1, E4BP4</i>
<i>EHD1</i>	<i>SP1, MYOGNF1, TATA, TCF11, GFII, PAX2, OLF1, ARP1, HEN1, GATA1, HOXA3, SOX5, NFY, SREBP1, MZF1, NRSF, MYC, MAX, AHRARNT, PAX5, TAX, CREB, MYB, COUP, IRF1, AP4, PAX3, AP2REP, AREB6, MEF2, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</i>
<i>ROBO2</i>	<i>HNF3B, MYOGNF1, TATA, TCF11, GFII, NKX25, FAC1, PAX2, MEIS1B, HOXA9, OLF1, HEN1, CEBPB, GATA1, HOXA3, HAND1E47, S8, CDP, RSRFC4, SOX5, NFY, CDP, CR3HD, SREBP1, MZF1, NRSF, MYC, MAX, AHRARNT, MSX1, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, PAX3, AP2REP, AREB6, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</i>
<i>GRK6</i>	<i>MYOGNF1, TATA, TCF11, GFII, FAC1, PAX2, MEIS1B, HOXA9, OLF1, ARP1, HEN1, CEBPB, GATA1, HOXA3, HAND1E47, S8, NFY, SREBP1, MZF1, NRSF, MYC, MAX, AHRARNT, PAX5, MSX1, TAX, CREB, COUP, AP4, PAX3, AREB6, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</i>

which is consistent with the results of the present study. The activation of *CXCR4* has been shown to reduce dopamine levels in the striatum, and induce cell death and tissue loss in the SN and striatum (34,35). Furthermore, *CXCL12* and *CXCR4* are neurotoxic to dopamine neuronal populations through the release of neurotoxins from microglia (36). Therefore, *CXCR4* may be pivotal in the etiology of PD by reducing dopamine levels in the striatum and/or releasing neurotoxins from microglia to induce the death of dopamine neurons. In the PPI network, *CXCR4* interacted with other gene proteins, including TOX high mobility group box family member 4 (*TOX4*), *YY1*, *GRK6*, kinase insert domain receptor (*KDR*) and zymogen granule protein 16B (*ZG16B*). Among these genes, *YY1* has been reported to regulate the expression of *SNCA*, which is an established susceptibility gene for PD (37). In the present study, *YY1* was predicted as a TF targeting *DCC*, *ROBO2*, *SLIT1* and *GRK6*. Therefore, *YY1* may also be involved in PD via regulating the expression of these genes. In an animal model of PD, the expression of

GRK6 was found to be altered in the caudate/putamen and ventral striatum (38), and the altered expression of *GRK6* can result in the dysregulation of DA receptors, thus contributing to the core motor deficits observed in PD (39). Other genes, including *TOX4*, *KDR* and *ZG16B*, and their interactions with *CXCR4*, have not been correlated with PD in previous studies; thus, these are novel potential genes, which may be associated with PD.

In the present study, *NCK2* was upregulated in the SN of PD based on the analysis of the two datasets. *NCK2* encodes a member of the NCK family of adaptor proteins, which are involved in the regulation of receptor protein tyrosine kinases (40). It has been reported that *NCK2* is expressed in neurons and glia (41). There is evidence that *NCK2* interacts with particularly interesting cysteine histidine-rich protein and Dock180 protein, which are involved in neuropathological processes (42,43). In the PPI network, *NCK2*, *CXCR4* and *DCC* all interacted with *BCAR1*. *BCAR1* is involved in various cellular events, including migration, survival and transformation (44).

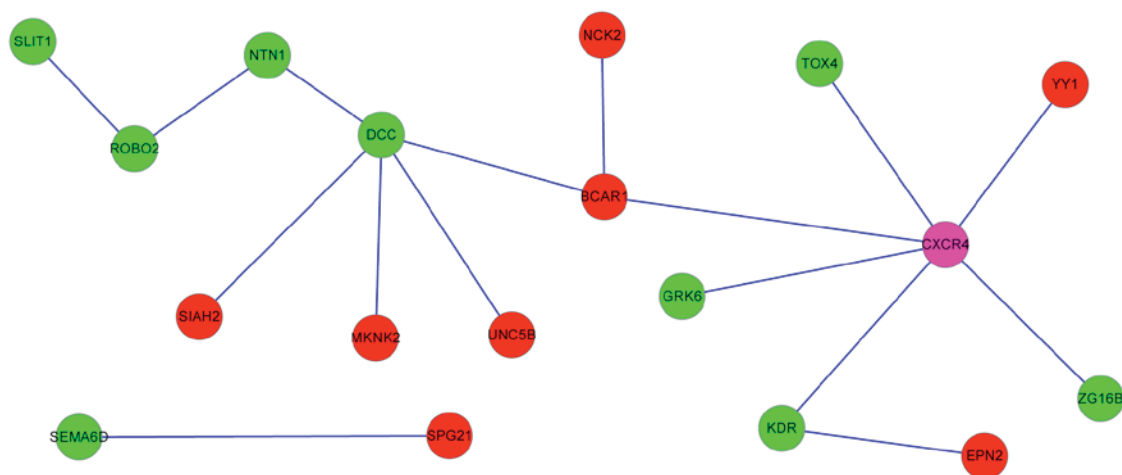


Figure 1. Protein-protein interaction networks for the crucial Parkinson's disease-associated genes. Red nodes represent upregulated genes; green nodes represent downregulated genes; purple node represents the hub node, interacting with the majority of genes.

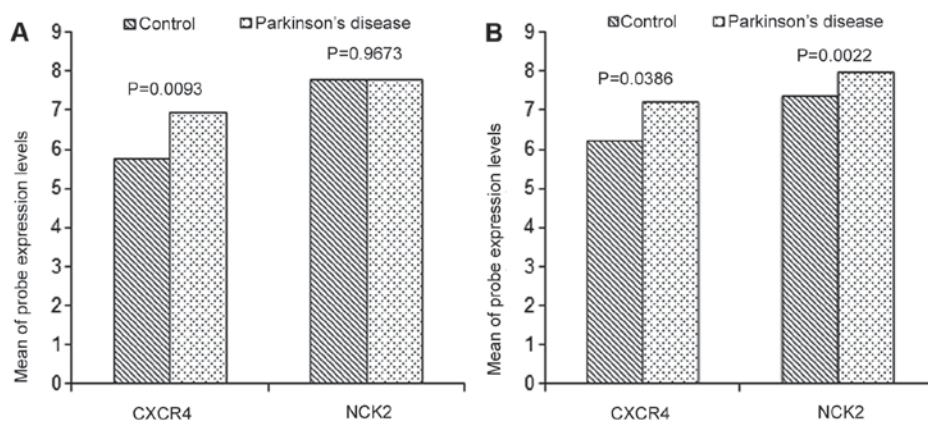


Figure 2. Mean probe expression levels of *CXCR4* and *NCK2*. Probe expression was calculated using the robust multi-array average algorithm. (A) Mean probe expression levels of *CXCR4* and *NCK2* in the lateral SN from patients with PD and controls. (B) Mean probe expression levels of *CXCR4* and *NCK2* in the medial SN samples from patients with PD and controls. *CXCR4*, chemokine (C-X-C motif) receptor 4; *NCK2*, NCK adaptor protein 2; PD, Parkinson's disease; SN, substantia nigra.

A previous study has reported the differential expression of *BCAR1* in PD (45). The cerebellum may have certain roles in the pathophysiology of PD (46). p130Cas/*BCAR1* was found to be involved in cellular mechanisms regulating different stages of cerebellar development (47), suggesting that the abnormal expression of *BCAR1* may be associated with PD. However, there is no evidence to confirm the association of *NCK2* with the progression of PD. Collectively, these findings indicated that the upregulation of *NCK2* may be responsible for the progression of PD, via interaction with *BCAR1* or regulation by TFs.

In addition, according to the TF analysis for crucial PD-associated genes in the present study, a series of TFs were predicted to target the 10 crucial PD-associated genes. *DCC*, *CXCR4*, *NCK2*, *SLIT1*, *ROBO2* and *GRK6* were all regulated by a set of TFs, including *GATA*, *E2F* and *E4BP4*. A previous study reported that *GATA-1* and *GATA-2* regulate the expression of *SNCA* dopaminergic cells (48). The pRb/E2F cell-cycle pathway mediates cell death in PD, and *E2F-1*-deficient mice are significantly more resistant to dopaminergic cell death, compared with their wild-type

littermates (49). *E4BP4*, also known as nuclear factor, interleukin 3 regulated is a transcriptional regulator of the human interleukin-3 promoter (50). A previous study reported that *E4BP4* is reportedly associated with neuron disease (51), and is upregulated in the SN tissue from patients with PD (52). Collectively, the *DCC*, *CXCR4* and *NCK2* genes are involved in the pathogenesis of PD, not only via the axon guidance pathway, but also likely via the regulation of multiple TFs, including *GATA*, *E2F* and *E4BP4*.

Of note, although the present study reanalyzed the public data deposited by others, the analytical procedure of the present study was different from that of previous studies. The criteria of the $FDR < 0.05$ and $|\log_2 FC| > 1$ were more strict, compared with those ($P < 0.05$; $|FC| \geq 1.3$) used in the previous study by Papapetropoulos *et al* (10), indicating that the DEGs identified in the present study show a higher level of differential expression. In the present study, the P-value of each gene was adjusted using the BH method to reduce false positives, however, the P-value was not adjusted in the previous study. Furthermore, in the present study, the protein interactions of DEGs and the TFs regulating DEGs were analyzed, which

assisted in providing a more comprehensive understanding of the pathogenesis of PD, whereas these were not investigated in the previous study. Despite the significance, there were certain limitations in the present study. For example, only two datasets were analyzed, whereas performing a meta-analysis using multiple datasets, including multiple types of data, may produce additional significant results. This is an important aim of further investigations.

In conclusion, the present study identified a set of potential crucial PD-associated genes, including *CXCR4*, *DCC* and *NCK2*, which were associated with neuron development and differentiation. The *DCC-NTN1-ROBO2-SLIT1* interaction pathway and several genes, including *TOX4*, *KDR* and *ZG16B*, which interacted with *CXCR4*, were novel findings. Genes, including *DCC*, *CXCR4*, *NCK2*, *SLIT1* and *ROBO2* were also regulated by multiple TFs, including *GATA*, *E2F* and *E4BP4*. Additionally, *CXCR4* and *NCK2* were detected to be upregulated in the SN of patients with PD following the analysis of two datasets. These genes, interactions of proteins and TFs may be key in the progression of PD. These findings contribute to an improved understanding of the pathogenesis of PD. However, further experiments are required to confirm the results of the present study and further elucidate the pathogenesis of PD.

References

- Dawson TM and Dawson VL: Molecular pathways of neurodegeneration in Parkinson's disease. *Science* 302: 819-822, 2003.
- Armentero MT, Pinna A, Ferré S, Lanciego JL, Müller CE and Franco R: Past, present and future of A(2A) adenosine receptor antagonists in the therapy of Parkinson's disease. *Pharmacol Ther* 132: 280-299, 2011.
- Michell AW, Lewis SJ, Foltynie T and Barker RA: Biomarkers and Parkinson's disease. *Brain* 127: 1693-1705, 2004.
- Berendse HW, Booij J, Francot CM, Bergmans PL, Hijman R, Stoof JC and Wolters EC: Subclinical dopaminergic dysfunction in asymptomatic Parkinson's disease patients' relatives with a decreased sense of smell. *Ann Neurol* 50: 34-41, 2001.
- Galvin JE, Lee VM and Trojanowski JQ: Synucleinopathies: Clinical and pathological implications. *Arch Neurol* 58: 186-190, 2001.
- Sathe K, Maetzler W, Lang JD, Mounsey RB, Fleckenstein C, Martin HL, Schulte C, Mustafa S, Synofzik M, Vukovic Z, et al: S100B is increased in Parkinson's disease and ablation protects against MPTP-induced toxicity through the RAGE and TNF- α pathway. *Brain* 135: 3336-3347, 2012.
- Alvarez-Erviti L, Rodriguez-Oroz MC, Cooper JM, Caballero C, Ferrer I, Obeso JA and Schapira AH: Chaperone-mediated autophagy markers in Parkinson disease brains. *Arch Neurol* 67: 1464-1472, 2010.
- Malek N, Swallow D, Grosset KA, Anichtchik O, Spillantini M and Grosset DG: Alpha-synuclein in peripheral tissues and body fluids as a biomarker for Parkinson's disease-a systematic review. *Acta Neurol Scand* 130: 59-72, 2014.
- Devic I, Hwang H, Edgar JS, Izutsu K, Presland R, Pan C, Goodlett DR, Wang Y, Armaly J, Tumas V, et al: Salivary α -synuclein and DJ-1: Potential biomarkers for Parkinson's disease. *Brain* 134: e178, 2011.
- Papapetropoulos S, French-Mullen J, McCorquodale D, Qin Y, Pablo J and Mash DC: Multiregional gene expression profiling identifies MRPS6 as a possible candidate gene for Parkinson's disease. *Gene Expr* 13: 205-215, 2006.
- Gao L, Zhao G, Fang JS, Yuan TY, Liu AL and Du GH: Discovery of the neuroprotective effects of alvespimycin by computational prioritization of potential anti-Parkinson agents. *FEBS J* 281: 1110-1122, 2014.
- Moran LB, Duke DC, Deprez M, Dexter DT, Pearce RK and Graeber MB: Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. *Neurogenetics* 7: 1-11, 2006.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264, 2003.
- Tibshirani R, Chu G, Narasimhan B and Li J: SAM: Significance Analysis of Microarrays. Version 2.0. The Comprehensive R Archive Network, 2011.
- Glueck DH, Mandel J, Karimpour-Fard A, Hunter L and Muller KE: Exact calculations of average power for the Benjamini-Hochberg procedure. *Int J Biostat* 4: Article 11, 2008.
- Kanehisa M, Goto S, Sato Y, Furumichi M and Tanabe M: KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109-D114, 2012.
- Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC and Lempicki RA: The DAVID gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 8: R183, 2007.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25-29, 2000.
- Levine M and Tjian R: Transcription regulation and animal diversity. *Nature* 424: 147-151, 2003.
- Martin TM, Plautz SA and Pannier AK: Network analysis of endogenous gene expression profiles after polyethyleneimine-mediated DNA delivery. *J Gene Med* 15: 142-154, 2013.
- Serrato-Combe A: Lindebmayr Systems-experimenting with software string rewriting as an assist to the study and generation of architectural form. In: Proceedings of the 9th Iberoamerican Congress of Digital Graphics. SIGRADI, Lima, pp161-166, 2005.
- Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27: 431-432, 2011.
- Keino-Masu K, Masu M, Hinck L, Leonardo ED, Chan SS, Culotti JG and Tessier-Lavigne M: Deleted in colorectal cancer (DCC) encodes a netrin receptor. *Cell* 87: 175-185, 1996.
- Xu B, Goldman JS, Rymar VV, Forget C, Lo PS, Bull SJ, Vereker E, Barker PA, Trudeau LE, Sadikot AF and Kennedy TE: Critical roles for the netrin receptor deleted in colorectal cancer in dopaminergic neuronal precursor migration, axon guidance, and axon arborization. *Neuroscience* 169: 932-949, 2010.
- Engle EC: Human genetic disorders of axon guidance. *Cold Spring Harb Perspect Biol* 2: a001784, 2010.
- Manitt C, Mimee A, Eng C, Pokinko M, Stroth T, Cooper HM, Kolb B and Flores C: The netrin receptor DCC is required in the pubertal organization of mesocortical dopamine circuitry. *J Neurosci* 31: 8381-8394, 2011.
- Lin L, Lesnick TG, Maraganore DM and Isacson O: Axon guidance and synaptic maintenance: Preclinical markers for neurodegenerative disease and therapeutics. *Trends Neurosci* 32: 142-149, 2009.
- Lesnick TG, Papapetropoulos S, Mash DC, French-Mullen J, Shehadeh L, de Andrade M, Henley JR, Rocca WA, Ahlskog JE and Maraganore DM: A genomic pathway approach to a complex disease: Axon guidance and Parkinson disease. *PLoS Genet* 3: e98, 2007.
- Sundaresan V, Mambetisaeva E, Andrews W, Annan A, Knöll B, Tear G and Bannister L: Dynamic expression patterns of Robo (Robo1 and Robo2) in the developing murine central nervous system. *J Comp Neurol* 468: 467-481, 2004.
- Bossers K, Meerhoff G, Balesar R, van Dongen JW, Kruse CG, Swaab DF and Verhaagen J: Analysis of gene expression in Parkinson's disease: Possible involvement of neurotrophic support and axon guidance in dopaminergic cell death. *Brain Pathol* 19: 91-107, 2009.
- Bagri A, Mari'n O, Plump AS, Mak J, Pleasure SJ, Rubenstein JL and Tessier-Lavigne M: Slit proteins prevent midline crossing and determine the dorsoventral position of major axonal pathways in the mammalian forebrain. *Neuron* 33: 233-248, 2002.
- Wegner SA, Ehrenberg PK, Chang G, Dayhoff DE, Sleeker AL and Michael NL: Genomic organization and functional characterization of the chemokine receptor CXCR4, a major entry co-receptor for human immunodeficiency virus type 1. *J Biol Chem* 273: 4754-4760, 1998.
- Shimoji M, Pagan F, Heaton EB and Mocchetti I: CXCR4 and CXCL12 expression is increased in the nigro-striatal system of Parkinson's disease. *Neurotox Res* 16: 318-328, 2009.

34. Bachis A, Aden SA, Nosheny RL, Andrews PM and Mocchetti I: Axonal transport of human immunodeficiency virus type 1 envelope protein glycoprotein 120 is found in association with neuronal apoptosis. *J Neurosci* 26: 6771-6780, 2006.
35. Nosheny RL, Bachis A, Aden SA, De Bernardi MA and Mocchetti I: Intrastratial administration of human immunodeficiency virus-1 glycoprotein 120 reduces glial cell-line derived neurotrophic factor levels and causes apoptosis in the substantia nigra. *J Neurobiol* 66: 1311-1321, 2006.
36. Bezzi P, Domercq M, Brambilla L, Galli R, Schols D, De Clercq E, Vescovi A, Bagetta G, Kollias G, Meldolesi J and Volterra A: CXCR4-activated astrocyte glutamate release via TNF α : Amplification by microglia triggers neurotoxicity. *Nat Neurosci* 4: 702-710, 2001.
37. Mizuta I, Takafuji K, Ando Y, Satake W, Kanagawa M, Kobayashi K, Nagamori S, Shinohara T, Ito C, Yamamoto M, *et al*: YY1 binds to α -synuclein 3'-flanking region SNP and stimulates antisense noncoding RNA expression. *J Hum Genet* 58: 711-719, 2013.
38. Bezard E, Gross CE, Qin L, Gurevich VV, Benovic JL and Gurevich EV: L-DOPA reverses the MPTP-induced elevation of the arrestin2 and GRK6 expression and enhanced ERK activation in monkey brain. *Neurobiol Dis* 18: 323-335, 2005.
39. Managò F, Espinoza S, Salahpour A, Sotnikova TD, Caron MG, Premont RT and Gainetdinov RR: The role of GRK6 in animal models of Parkinson's disease and L-DOPA treatment. *Sci Rep* 2: 301, 2012.
40. Tu Y, Li F and Wu C: Nck-2, a novel Src homology2/3-containing adaptor protein that interacts with the LIM-only protein PINCH and components of growth factor receptor kinase-signaling pathways. *Mol Biol Cell* 9: 3367-3382, 1998.
41. Ritchie MD: Using prior knowledge and genome-wide association to identify pathways involved in multiple sclerosis. *Genome Med* 1: 65, 2009.
42. Rearden A, Hurford R, Luu N, Kieu E, Sandoval M, Perez-Liz G, Del Valle L, Powell H and Langford TD: Novel expression of PINCH in the central nervous system and its potential as a biomarker for human immunodeficiency virus-associated neurodegeneration. *J Neurosci Res* 86: 2535-2542, 2008.
43. Shi L: Dock protein family in brain development and neurological disease. *Commun Integr Biol* 6: e26839, 2013.
44. Brinkman A, van der Flier S, Kok EM and Dorsers LC: BCAR1, a human homologue of the adapter protein p130Cas, and anties-trogen resistance in breast cancer cells. *J Natl Cancer Inst* 92: 112-120, 2000.
45. Hourani M, Mendes A, Berretta R and Moscato P: Genetic biomarkers for brain hemisphere differentiation in Parkinson's disease. *AIP Conference Proceedings* 952: 207-216, 2007.
46. Wu T and Hallett M: The cerebellum in Parkinson's disease. *Brain* 136: 696-709, 2013.
47. Furuichi T, Shiraishi-Yamaguchi Y, Sato A, Sadakata T, Huang J, Shinoda Y, Hayashi K, Mishima Y, Tomomura M, Nishibe H and Yoshikawa F: Systematizing and cloning of genes involved in the cerebellar cortex circuit development. *Neurochem Res* 36: 1241-1252, 2011.
48. Scherzer CR, Grass JA, Liao Z, Pepivani I, Zheng B, Eklund AC, Ney PA, Ng J, McGoldrick M, Mollenhauer B, *et al*: GATA transcription factors directly regulate the Parkinson's disease-linked gene alpha-synuclein. *Proc Natl Acad Sci USA* 105: 10907-10912, 2008.
49. Höglinger GU, Breunig JJ, Depboylu C, Rouaux C, Michel PP, Alvarez-Fischer D, Boutillier AL, Degregori J, Oertel WH, Rakic P, *et al*: The pRb/E2F cell-cycle pathway mediates cell death in Parkinson's disease. *Proc Natl Acad Sci USA* 104: 3585-3590, 2007.
50. Zhang W, Zhang J, Kornuc M, Kwan K, Frank R and Nimer SD: Molecular cloning and characterization of NF-IL3A, a transcriptional activator of the human interleukin-3 promoter. *Mol Cell Biol* 15: 6055-6063, 1995.
51. Hulme DJ, Blair IP, Dawkins JL and Nicholson GA: Exclusion of NFIL3 as the gene causing hereditary sensory neuropathy type I by mutation analysis. *Hum Genet* 106: 594-596, 2000.
52. Hu WC: Parkinson disease is a TH17 dominant autoimmune disorder against accumulated alpha-synuclein. *Nature Preced*: 6176, 2011.