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 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 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), 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 therapeutical 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...
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 llog_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 numerious 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). 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+219699+80031+7852+56920+6092+9423+4775+6585) and heat shock protein family A member 1A (HSPA1A), KDR, HSPA1L, CXCR4, GRK6, SMURF1, EHD1, CLTCL1, EPN2 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), 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, RET, HRAS, CHEMP6, RUFY1, HSPA1A, KDR, HSPA1L, CXCR4, GRK6, SMURF1, EHD1, CLTCL1, EPN2.

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

<table>
<thead>
<tr>
<th>Term</th>
<th>P-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04360</td>
<td>0.003347</td>
<td>DCC, NRAS, NCK2, HRAS, UNC5B, SEMA6D, CXCR4, SEMA3G, ROBO2, NTN1, NFATC3, SLIT1</td>
</tr>
<tr>
<td>Axon guidance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa04144</td>
<td>0.017984</td>
<td>RET, HRAS, CHEMP6, RUFY1, HSPA1A, KDR, HSPA1L, CXCR4, GRK6, SMURF1, EHD1, CLTCL1, EPN2</td>
</tr>
</tbody>
</table>
Table II. Gene ontology enrichment analysis of the crucial Parkinson's disease-associated genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Ontology terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLIT1</td>
<td>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</td>
</tr>
<tr>
<td>RUFY1</td>
<td>Regulation of endocytosis, membrane organization, membrane invagination, endocytosis</td>
</tr>
<tr>
<td>ROBO2</td>
<td>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</td>
</tr>
<tr>
<td>NCK2</td>
<td>Cell motion, cell projection organization, cell motility, localization of cell, cell migration</td>
</tr>
<tr>
<td>EPN2</td>
<td>Endocytic vesicle, regulation of endocytosis, membrane organization, membrane invagination, endocytosis</td>
</tr>
<tr>
<td>EHD1</td>
<td>Endocytic vesicle, membrane organization, membrane invagination, endocytosis</td>
</tr>
<tr>
<td>DCC</td>
<td>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</td>
</tr>
<tr>
<td>CXCR4</td>
<td>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</td>
</tr>
</tbody>
</table>

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, RU3FY1, 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),
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 genes proteins, including TOX high mobility group 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).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Transcription factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCC</td>
<td>IK2, SP1, HNF3B, MYOGF1, TATA, TCF11, GF11, NXK25, FAC1, PA2X, MEIS1B, HOX9A, OLFI1, ARPI1, HEN1, CEBPB, GATA1, HOX3, HAND1E47, S8, CDP, RSRCF4, SOX5, NPY, CDP, CR3H, SRGPB1, MZF1, NRSF, MYC, MAX, AHRNRT, PAX5, MXS1, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, PA2REP, ARE6B, GRE, MEF2, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>EPN2</td>
<td>IK2, SP1, HNF3B, TCF11, GF11, NXK25, FAC1, PA2X, MEIS1B, HOX9A, HAND1E47, S8, CDP, SOX5, NRSF, MYC, MAX, AHRNRT, MXS1, MYB, GATA2, IRF1, GATA, PA2X, PA2REP, ARE6B, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>CXCR4</td>
<td>IK2, MYOGF1, TATA, TCF11, GF11, NXK25, FAC1, PA2X, MEIS1B, HOX9A, CEBPB, GATA1, HOX3, S8, CDP, RSRCF4, SOX5, NPY, CDP, CR3H, MZF1, NRSF, MYC, MAX, PAX5, MXS1, IRF1, AP4, GATA, MRF2, PA2REP, ARE6B, MEF2, GRE, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>NCK2</td>
<td>IK2, HNF3B, MYOGF1, TATA, TCF11, GF11, NXK25, FAC1, PA2X, MEIS1B, HOX9A, OLFI1, APRI1, HEN1, CEBPB, GATA1, HAND1E47, S8, CDP, RSRCF4, SOX5, CDP, CR3H, SRGPB1, NRSF, MYC, MAX, AHRNRT, PAX5, MXS1, TAX, CREB, MYB, COUP, GATA2, AP4, GATA, MRF2, PA2X, PA2REP, ARE6B, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>SLIT1</td>
<td>IK2, HNF3B, MYOGF1, TATA, TCF11, GF11, NXK25, FAC1, PA2X, MEIS1B, HOX9A, APRI1, HEN1, CEBPB, GATA1, HOX3, HAND1E47, S8, CDP, NPY, CDP, CR3H, MZF1, NRSF, MYC, MAX, PAX5, MXS1, IRF1, AP4, GATA, MRF2, PA2REP, ARE6B, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>SEMA6D</td>
<td>IK2, HNF3B, MYOGF1, TATA, TCF11, GF11, NXK25, FAC1, PA2X, MEIS1B, HOX9A, APRI1, HEN1, CEBPB, GATA1, HOX3, HAND1E47, S8, CDP, NPY, CDP, CR3H, MZF1, NRSF, MYC, MAX, AHRNRT, PAX5, MXS1, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, ARE6B, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>RUFIY</td>
<td>SP1, HNF3B, TATA, NXK25, PA2X, MEIS1B, HOX9A, OLFI1, APRI1, CEBPB, GATA1, HOX3, HAND1E47, S8, CDP, RSRCF4, SOX5, NPY, CDP, CR3H, MZF1, NRSF, MYC, MAX, AHRNRT, PAX5, MXS1, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, PA2REP, ARE6B, SRY, MEF2, GRE, USF, BRACH, PAX6, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>EHD1</td>
<td>SP1, MYOGF1, TATA, TCF11, GF11, PX2, OLFI1, APRI1, HEN1, GATA1, HOX3, SOX5, NPY, SRGPB1, MZF1, NRSF, MYC, MAX, AHRNRT, PAX5, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, PA2X, PA2REP, ARE6B, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>ROBO2</td>
<td>HNF3B, MYOGF1, TATA, TCF11, GF11, NXK25, FAC1, PA2X, MEIS1B, HOX9A, OLFI1, HEN1, CEBPB, GATA1, HOX3, HAND1E47, S8, CDP, RSRCF4, SOX5, NPY, CDP, CR3H, SRGPB1, MZF1, NRSF, MYC, MAX, AHRNRT, PAX5, TAX, CREB, MYB, COUP, GATA2, IRF1, AP4, GATA, MRF2, PA2X, PA2REP, ARE6B, SRY, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
<tr>
<td>GRK6</td>
<td>MYOGF1, TATA, TCF11, GF11, FAC1, PA2X, MEIS1B, HOX9A, OLFI1, APRI1, HEN1, CEBPB, GATA1, HOX3, HAND1E47, S8, NPY, SRGPB1, MZF1, NRSF, MYC, MAX, AHRNRT, PAX5, MXS1, TAX, CREB, MYB, COUP, AP4, AP4, PA2X, ARE6B, MEF2, GRE, MIF1, USF, BRACH, PAX6, COMP1, AML1, LMO2COM, E2F, YY1, E4BP4</td>
</tr>
</tbody>
</table>
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 |log2 FC|>1 were more strict, compared with those (P<0.05; |FC|≥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


