Multimodal Analysis Reveals New Genetic Effectors in Parkinson'S Disease-Related Gene LRRK2 Gly2019Ser Mutation

Longping Yao (✉ Loupe_yao@163.com)

Maastricht Universitair Medisch Centrum+ <https://orcid.org/0000-0002-7239-1392>

Shizhong Zhang

Zhujiang Hospital

Research article

Keywords: LRRK2, PD, mutation, Gly2019Ser, hub genes

Posted Date: November 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1094033/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Mutations in the LRRK2 gene, which encodes leucine-rich repeat kinase 2 (LRRK2), generate one of the most prevalent monogenic forms of Parkinson's disease (PD). Patients with autosomal dominant PD and apparent sporadic PD, who are clinically indistinguishable from those with idiopathic PD, are found to have LRRK2 mutations, particularly the most prevalent variant Gly2019Ser. Nonetheless, potential effectors of Gly2019Ser remain unknown.

Methods

We used the GEO database to undertake and evaluate a multitiered bioinformatic investigation to look into the gene expression implicated in the development of Parkinson's disease. Individual differences in gene expression were then confirmed in whole blood samples collected in the clinic. These genetic factors were also subjected to an interaction analysis and prediction.

Results

In total, 607 genes in the LRRK2 Gly2019Ser mutation group expressed differently from those in the wild group. The following 10 top hub genes were discovered in protein-protein interaction (PPI) networks: CD44, CTGF, THBS1, VEGFA, SPP1, EGF, VCAM1, MMP3, CXCR4, and LOX. The gene expression of CD44, CTGF, THBS1, SPP1, EGF, and LOX was considerably higher in the LRRK2 Gly2019Ser mutant group than in the LRRK2 wild group. Meanwhile, CXCR4 gene expression in the LRRK2 Gly2019Ser mutant group was significantly lower than in the LRRK2 wild group. We then confirmed the expression of the hub genes in LRRK2 Gly2019Ser mutated iPSC-induced DA cells. As a result, the levels of CD44, CTGF, THBS1, VEGFA, SPP1 were positively correlated to the mutation of LRRK2, displayed promising effectors for discriminating the pathogenesis of PD.

Conclusions

We identified CD44, CTGF, THBS1, VEGFA, and SPP1 as the potential genetic effectors responding to the mutation of LRRK2. They could be a promising mechanism for discriminating the PD and potential factors contributing to the disease's development.

Background

The aberrant deposits of α -synuclein aggregates in the brain and the loss of dopaminergic neurons in the substantia nigra pars compacta (SNpc) are pathological hallmarks of Parkinson's disease (PD) [1, 2]. The most common cause of autosomal-dominant hereditary PD is missense mutations in the LRRK2 gene, which encodes leucine-rich repeat kinase 2 (LRRK2) [3]. The gene has emerged as one of the most important to familial and sporadic PD [4]. Because the lifetime risk of LRRK2 mutations is predicted to be 22–32% in clinical populations, the penetrance of LRRK2 mutations in PD is incomplete, implying

substantial modifiers of LRRK2 illness [4]. Patients with autosomal dominant PD and apparent sporadic PD, who are clinically indistinguishable from those with idiopathic PD, are found to have LRRK2 mutations, particularly the most prevalent variant Gly2019Ser [3]. The Gly2019Ser LRRK2 mutation is located in the majority of patients with LRRK2-PD, and it is linked to neuropathological changes that are similar to those seen in idiopathic PD [5, 6]. These changes include typical α -synuclein aggregates in the form of Lewy bodies and neurites, as well as cell loss in vulnerable areas [7].

LRRK2 is a 286 kDa protein having repetitive motifs on the N-terminal half and kinase and GTPase domains surrounded by protein-protein interaction domains on the C-terminal half [8]. The majority of pathogenic mutations are found in the two catalytic domains, but they all cause the kinase to become hyperactive in cells [9]. Reduced LRRK2 activity or expression has been shown to be neuroprotective in preclinical research; small-molecule LRRK2 inhibitors and antisense oligonucleotides have been produced and are now regarded safe for clinical use [5]. However, it is yet unknown how LRRK2 kinase activity is regulated at the molecular and cellular levels and how aberrant LRRK2 kinase activity causes PD.

In this study, we used the datasets from Gene Expression Omnibus (GEO) and validated the top gene targets affected by the LRRK2 mutation. We utilized weighted gene co-expression network analysis (WGCNA), Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analyses (GOFEA, KEGGFEA), Gene Ontology (GO), and STRING protein-interacting network to examine the variable genes and confirmed the expression of hub interacting proteins/genes, indicating that the focal adhesion and mitogen-activated protein kinase (MAPK) signalling pathway were closely linked in pathogenic and pathogenesis.

Materials And Methods

Data acquisition and processing

Datasets were employed from GENE EXPRESSION OMNIBUS (GEO) public database (<https://www.ncbi.nlm.nih.gov/gds>) for the keywords "LRRK2". The R software has been employed for background correction of sample expression matrix, data downgrade, normalization, $\log_2$ transformation, and probe reannotation. RMA, devtools, AnnoProbe, limma, Biobase, affyPLM, and GEOquery were among the R packages used in this research. The gene's expression level was calculated by averaging the values of these probes if a single gene in the chip corresponded to many probes. Using Fisher's combined probability test, P values were pooled, then adjusted for multiple comparison correction using the false discovery rate (FDR). $FDR \leq 0.05$ was chosen as the significance level (**Figure 1**).

WGCNA and hub genes selection

WGCNA is a common data mining method for analyzing biological networks that employs pairwise correlations between variables. While it can be used to analyze various high-dimensional data sets, it is most commonly utilized in genomics. It can be used to construct modules (clusters), intramodular hubs, and network nodes based on module membership, as well as investigate co-expression module

connections interactions and compare network topologies (differential network analysis). Correlation networks make it easier to uncover prospective biomarkers or treatment targets using network-based gene screening methodologies. The WGCNA, an R package, was used to perform the weighted correlation network analysis. Pearson's correlation matrices and the average linkage technique were employed first for every pairwise Gene, and then a weighted adjacency matrix was constructed using a power function $A_{mn}=|C_{mn}|^\beta$ (C_{mn} = Pearson's correlation between Gene_m and Gene_n; A_{mn} =adjacency between Gene_m and Gene_n). β was a soft-thresholding parameter that might amplify strong gene-gene interactions while penalizing weak ones. The adjacency was transformed into a topological overlap matrix (TOM), which could be used to quantify Gene's network connection, which was defined as the sum of its adjacency with all other Genes for network generation, and the corresponding dissimilarity (1-TOM) was calculated after choosing the power of 7. Using the TOM-based dissimilarity metric and a minimum size (Gene group) for the Genes dendrogram, average linkage hierarchical clustering was utilized to classify Genes with similar expression profiles into Gene modules. To investigate the modules further, we calculated their dissimilarity, selected a cut line for the module dendrogram, and merged some modules. We additionally merged modules with a distance of less than 0.35, resulting in a total of 5 co-expression modules.

Cluster analysis

Cluster analysis was carried out using ConsensusClusterPlus[10] was used to perform cluster analysis, which included agglomerative pam clustering with 1-Pearson correlation distances and resampling 80 percent of the samples for 10 repetitions. The empirical cumulative distribution function plot was used to find the appropriate number of clusters.

Functional enrichment analysis

To determine the most highly enriched pathways, we utilized DAVID (<https://david.ncifcrf.gov/home.jsp>) to run GO and KEGG functional enrichment investigations, indicating the module's potential biological importance. To obtain statistically significant gene sets with meaningful functional annotations and signalling pathways, the *P* values were determined using Fisher's exact probability technique. *P* < 0.05 was used as the significant level.

Protein-protein interaction network and hub genes selection in the Key Module

The STRING 11.5 database (https://string-db.org/cgi/input?sessionId=boTaerIW4Ew9&input_page_show_search=on) was used to map the protein-protein interaction (PPI) network and predict the interactions of all of the proteins discovered in the study. The PPI activity is a primary focus of cellular biology research and serves as a prerequisite for system biology [11]. Proteins interact with other proteins within the cell to carry out their functions, and the information supplied by a PPI network improves the perception of the protein's function [12]. We used the STRING database's data analysis mode and an integrated confidence value of 0.4 to generate a network map of PPI. The confidence score, which is a medium confidence score, was calculated using the STRING platform. For a

better visual depiction of the network and to identify hub genes, the obtained PPI was assessed using the software Cytoscape (<https://cytoscape.org/>).

Reprogramming of human fibroblasts to induced pluripotent stem cells and their differentiation to dopaminergic neurons

Three samples were taken from three wild healthy volunteers (WT) and three from PD patients who had the LRRK2 Gly2019Ser S mutation verified. WT fibroblast and PD patient fibroblasts with LRRK2-Gly2019Ser mutation were reprogrammed to induced pluripotent stem cells (iPSCs) as described in the previous report [13].

The iPSCs were cultured in DMEM/F12, 2% of B27 without vitamin A (12587-010 Gibco), supplemented with: 10 μ M of Y-27632 (ROCK inhibitor; Tocris Biochem), 1% of N-2 Supplement (17502-048 Gibco), 20 μ M of SB431542 (SMAD inhibitor; S4317-5MG Sigma), 500 ng/ml of noggin (120-10C Peprotech), and bFGF 2 ng/ml. As previously described, the iPSCs were differentiated into neuroectodermal spheres (NES) [14]. The NESs were then placed in dopaminergic (DA) neural patterning media after being cultivated for four days following passaging: B27 supplement (1:50), DMEM/F12, N2 supplement (1:100), 100 U/ml penicillin-streptomycin (all from Invitrogen), 100 ng/ml FGF8, 100 ng/ml SHH, 3 μ M CHIR99021 (Tocris), 1 μ M purmorphamine (Tocris), and 0.2 mM 2-Phospho-L-ascorbic acid (Sigma). After four days culture, the medium were changed to terminal differentiation medium: DMEM/F12, B27 supplement (1:50), N2 supplement (1:100), 20 ng/ml BDNF, 100 U/ml penicillin-streptomycin, 0.5 mM dibutyryl-cAMP (Enzo, Farmingdale, NY, USA), 20 ng/ml GDNF, and 0.2 mM 2-Phospho-L-ascorbic acid (Sigma). The cells were cultured at 37°C in a humidified incubator with 95% air:5% CO₂.

Immunofluorescence staining and confocal laser scanning microscopy

The protocols for double immunofluorescence staining were carried out as described previously in Kavanagh et al [15]. In brief, free-floating 30- μ m sections of the midbrain from each group were incubated with rabbit-derive tyrosine hydroxylase antibody (1:1000; Novus Biologicals) at 4°C overnight followed by a goat anti-mouse IgG conjugated with Alexa Fluor 488 (GB25301, 1:400; Servicebio) for 1 h at room temperature (22 $\pm$ 2°C). Immunoreactivity fluoresced green under an LSM 880 laser scanning confocal microscope (Carl Zeiss, Oberkochen, Germany). ZEN light software was used to capture and analyze confocal images (Carl Zeiss).

Reverse transcription-quantitative real-time polymerase chain reaction

The cells were collected, and total RNA was extracted according to the manufacturer's instructions using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). For the quantification of messenger RNA (mRNA) of protein-encoding genes, RNA was reverse transcribed to complementary DNA (cDNA) with a random primer (Sangon Biotech, Shanghai, China) using a Reverse Transcription Kit (RR047A; Takara, Dalian, China) and the mRNA levels were determined using reverse transcription-quantitative real-time polymerase chain reaction (RT-qPCR). A human glyceraldehyde-3-phosphate dehydrogenase (GAPDH)

primer pair served as an internal control. The Ct technique $\Delta\Delta C_t$ method was used to calculate and normalize relative gene expressions. All the sequences of the primers used are listed in Table 1.

Statistical analysis

The results were provided as the mean SD of three separate studies. A two-tailed Student's t-test was used in the statistical analysis. A statistically significant difference was defined as one with $P < 0.05$.

Results

Consensus clustering presents quantitative evidence for estimating the number of possible clusters and their membership in the datasets.

The GEO database has 2 LRRk2-related microarray transcriptome datasets and was conducted in three cell lines. With the threshold $FDR \leq 0.05$ criterion, a total of 607 differentially expressed genes (DEGs) were confirmed. The differential expression profiling and distribution of these DEGs were shown (Figure 2A). Cluster analysis was performed using the ConsensusClusterPlus program, which includes agglomerative pam clustering with 1-Pearson correlation distances and resampling 80% of the samples for 10 times. The optimal cluster number was found using the cumulative distribution function (CDF). When $k = 2$, the cluster was discovered to be the most stable clustering based on the CDF Delta area curve (Figure 2B, C, and D). We obtained four Immune Subtype (IS) results using $k = 2$. (Fig. 2E and F).

WGCNA evaluated the genes that had varied expression patterns.

Different databases contribute to the overall number of genes differently, with varying degrees of database overlap (**Figure 3A and B**). The 607 DEGs were hierarchically classified into 18 categories using WGCNA analysis, respectively: brown2, firebrick4, darkgreen, coral2, greenyellow, lightsteelblue, darkolivegreen4, darkred, darkolivegreen, brown4, yellow4, darkseagreen4, plum2, lightyellow, lightcoral, darkorange, royalblue, darkgrey, and black (**Figure 3C**). The horizontal axis was the network's average connectivity (**Figure 3D**), and the vertical axis was the scale-free topology fitting index R^2 (values in the SFT.R Sq column in the statistical data) (**Figure 3E**). As a consequence of respringing the cluster tree, we detected 18 modules in the heat map of association between the tree diagram of gene expression and module attributes (**Figure 3F**). The relationship of modules and a phenotype heat map was exhibited in the grouping of LRRK2 wild and mutation. The blue module showed the most robust inverse link with the LRRK2 mutation phenotype. (**Figure 3G**).

The most highly enriched pathways were found using GO and KEGG.

To learn more about the biological processes and pathways involved, the DAVID was utilized to conduct GO and KEGG analyses (**Table 3**). The 607 genes were primarily enriched in the biological progress (BP) category for system development, anatomical structure morphogenesis, regulation of the multicellular organismal process, nervous system development, cellular developmental process and cell

differentiation; the cellular component (CC) category for the extracellular region, extracellular region part, vesicle, endomembrane system, extracellular space, plasma membrane part, extracellular vesicle, extracellular organelle and extracellular exosome (**Figure 4A, B and C**). Many transcription factors were enriched, which have been associated with the LRRK2 mutation (**Figure 4D**) (**Table 4**). The 607 genes were primarily enriched in focal adhesion, proteoglycans in cancer, ECM-receptor interaction, MAPK signalling pathway, PI3K-Akt signalling pathway, amoebiasis, axon guidance, vascular smooth muscle contraction and Ras signalling pathway according to KEGG pathway analysis (**Figure 4E**).

The PPI network reveals interacting proteins and hub genes

The frequent DEGs were entered using STRING 11.5, and the file created from the analysis was reintroduced into Cytoscape for further exploration, including hub gene discovery. The PPI network looked at gene-gene and protein-protein interactions, finding 600 links with a medium confidence interaction score (0.4) (**Figure 5A**). We chose 50 genes for the heat map that indicated apparent differences in PD compared to controls (**Figure 5B**). The PPI network was utilized with a medium confidence interaction score to investigate gene-gene/protein-protein interactions with inflammation-related genes (0.4). Then, hub genes and essential modules were detected based on the PPIs network in **Figure 5A**. We selected the top 10 top hub genes of CD44, CTGF, THBS1, VEGFA, SPP1, EGF, VCAM1, MMP3, CXCR4, LOX and analyzed their interactions (**Figure 5C**), and the relative expressions of the interacting proteins were shown in **Figure 5D**. Moreover, the gene expression of CD44, CTGF, THBS1, SPP1, EGF, and LOX was considerably higher in the LRRK2 Gly2019Ser mutant group than in the LRRK2 wild group (**Figure 5E**), and operating characteristic (ROC) analysis was used in **Figure 5F**. The expressions between CD44 and CTGF, CD44 and THBS1, CD44 and SPP1, had a significant positive correlation (**Figure 5G**).

The validation of hub genes in iPSC-induced DA cells.

We enrolled three wild healthy volunteers who had not been diagnosed with a central nervous system disease and three from PD patients who had the LRRK2 Gly2019Ser mutation verified. The WT fibroblast and PD patient fibroblasts with LRRK2- Gly2019Ser mutation were reprogrammed to iPSCs. Further, the iPSCs were differentiated into DA neurons. We validated the DA neurons differentiation with tyrosine hydroxylase (TH) staining (**Figure 6A**). RT-qPCR analyzed the peripheral blood samples of PD and healthy control to measure the gene expression. We measured the 10 hub genes expression. As a result, the levels of CD44, CTGF, THBS1, VEGFA and SPP1 were positively correlated to the mutation of LRRK2, displayed promising effectors for discriminating the pathogenesis of PD (**Figure 6B, C, D, E and F**).

The PPI network reveals interaction for the effectors responding to LRRK2 mutation.

A PPI network for the genes of CD44, CTGF, THBS1, VEGFA and SPP1 was built using the STRING online tool (**Figure 7A, B, C, D, and E**). The PPIs network contains a number of for CD44: number of nodes: 41, number of edges: 352, average node degree: 17.2; for CTGF: number of nodes: 41, number of edges: 327, average node degree: 16; for THBS1: number of nodes: 41, number of edges: 262, average node degree:

12.8; for VEGFA: number of nodes: 41, number of edges: 341, average node degree: 16.6, and for SPP1: number of nodes: 41; number of edges: 317; average node degree: 15.5.

Discussion

Increasing evidence demonstrated a role for the mutation of LRRK2 in the pathogenesis of PD. However, it is still unclear how LRRK2 kinase activity is regulated at the molecular and cellular levels, as well as how abnormal LRRK2 kinase activity leads to PD. Specifically, recent technological advancements have made it possible to identify molecular changes in human diseases on a worldwide scale. However, these methods are constrained by the high complexity of data, which leads to many testing challenges, the often small number of samples examined, and the low technical and statistical reproducibility of the produced results [16]. In this study, we performed and tested a multitiered bioinformatic analysis using the GEO database to investigate global alterations that responded to the mutation of LRRK2. We introduced the recently developed WGCNA methodology, a commonly used data mining method based on pairwise correlations between variables and is especially useful for investigating biological networks [17]. Interaction analysis and prediction were also performed on these genetic variables.

A total of 607 DEGs were validated using the $FDR \leq 0.05$ criterion. To learn more about the biological processes and pathways involved, the 607 genes were primarily enriched in the BP category for system development, anatomical structure morphogenesis, regulation of the multicellular organismal process, nervous system development, cellular developmental process and cell differentiation; the CC category for the extracellular region, extracellular region part, vesicle, endomembrane system, extracellular space, plasma membrane part, extracellular vesicle, extracellular organelle and extracellular exosome. Moreover, according to KEGG pathway analysis, the 607 genes mainly were abundant in focal adhesion, cancer proteoglycans, ECM-receptor interaction, MAPK signalling pathway, PI3K-Akt signalling pathway, amoebiasis, axon guidance, vascular smooth muscle contraction, and Ras signalling pathway. Then, top hub genes and essential modules were detected based on the PPIs network. Finally, the top 10 top hub genes were selected: CD44, CTGF, THBS1, VEGFA, SPP1, EGF, VCAM1, MMP3, CXCR4 and LOX. What's more, we further validated distinct differences in gene's alteration in iPSCs-induced DA cells from LRRK2 wild and mutated human samples. As a result, the levels of CD44, CTGF, THBS1, VEGFA and SPP1 were positively correlated to the mutation of LRRK2, displayed promising effectors for discriminating the pathogenesis of PD.

Meanwhile, we discovered that several transcription factors are involved in the pathogenesis of LRRK2 mutation. Among them, NF1 and HNF3 are in the highest priority position of the representative enriched terms. Nuclear factor I (NF1) has been reported to regulate neuronal apoptosis, inflammatory response, and oxidative stress in PD [18]. Therefore, we speculate that NF1 may be related to regulating downstream gene expression by increased kinase activity after LRRK2 mutation. To date, most preclinical research has used transgenic rodent models to assess the efficacy of LRRK2 inhibitors in various PD models [19]. As a result, investigating how LRRK2 influences the process of NF1 regulating PD could be therapeutic. Meanwhile, hepatocyte nuclear factor 3 (HNF3) was the first identified member of this

extensive family of transcription factors, which binds to cis-regulatory elements in hundreds of genes encoding gluconeogenic and glycolytic enzymes, serum proteins and hormones [20]. To date, there is currently no information on how the LRRK2 mutation affects HNF3. However, we believe that this could lead to new study areas.

However, more long-term research is needed to see if gene expression levels can be used to track clinical development, including the transcription factors and the effectors responding to LRRK2 mutation that we validated in this study.

Conclusion

We performed multimodal approaches to identify 607 genes expressed differently in LRRK2- Gly2019Ser mutated iPSCs-induced DA cells than in the LRRK2 wild control group. They could be a promising mechanism for discriminating the PD and potential factors contributing to the disease's development.

Abbreviations

LRRK2: leucine-rich repeat kinase 2

PD: Parkinson's disease

PPI: protein-protein interaction

SNpc: substantia nigra pars compacta

WGCNA: weighted gene co-expression network analysis

KEGG: Kyoto Encyclopedia of Genes and Genomes

GO: Gene Ontology

MAPK: mitogen-activated protein kinase

GEO: GENE EXPRESSION OMNIBUS

FDR: false discovery rate

TOM: topological overlap matrix

WT: healthy volunteers

iPSCs: induced pluripotent stem cells

DA: dopaminergic

mRNA: messenger RNA

cDNA: complementary DNA

GAPDH: glyceraldehyde-3-phosphate dehydrogenase

DEGs: differentially expressed genes

CDF: Cumulative Distribution Function

BP: biological process

CC: cellular component

ROC: operating characteristic

TH: tyrosine hydroxylase

NF1: Nuclear factor I

HNF3: hepatocyte nuclear factor 3

Declarations

Availability of data and material

All data generated or analysed during this study are included in the present article.

Author Contributions

YLP and ZSZ conceived and designed the experiments. YLP performed the literature search, analyzed the data, and predominantly contributed to the writing of the article. YLP and ZSZ organized the validation experiment. Guangdong Provincial Key Laboratory on Brain Function Repair and Regeneration provided the experimental technical support. ZSZ gave the supervision and revised the paper. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (81671240).

Ethics approval

Ethical approval was obtained from the ethics committee of Zhujiang Hospital of Southern Medical University. To take part in this study, all participants signed a written informed consent form.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Acknowledgements

Not applicable.

References

1. Yao L, Ye Y, Mao H, Lu F, He X, Lu G, Zhang S: **MicroRNA-124 regulates the expression of MEKK3 in the inflammatory pathogenesis of Parkinson's disease.** *Journal of neuroinflammation* 2018, **15**:13.
2. Yao L, Zhu Z, Wu J, Zhang Y, Zhang H, Sun X, Qian C, Wang B, Xie L, Zhang S, Lu G: **MicroRNA-124 regulates the expression of p62/p38 and promotes autophagy in the inflammatory pathogenesis of Parkinson's disease.** *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2019, **33**:8648-8665.
3. Tolosa E, Vila M, Klein C, Rascol O: **LRRK2 in Parkinson disease: challenges of clinical trials.** *Nat Rev Neurol* 2020, **16**:97-107.
4. Yao L, Wu J, Koc S, Lu G: **Genetic Imaging of Neuroinflammation in Parkinson's Disease: Recent Advancements.** *Front Cell Dev Biol* 2021, **9**:655819.
5. Di Maio R, Hoffman EK, Rocha EM, Keeney MT, Sanders LH, De Miranda BR, Zharikov A, Van Laar A, Stepan AF, Lanz TA, et al: **LRRK2 activation in idiopathic Parkinson's disease.** *Science translational medicine* 2018, **10**.
6. Ho PW, Leung CT, Liu H, Pang SY, Lam CS, Xian J, Li L, Kung MH, Ramsden DB, Ho SL: **Age-dependent accumulation of oligomeric SNCA/ α -synuclein from impaired degradation in mutant LRRK2 knockin mouse model of Parkinson disease: role for therapeutic activation of chaperone-mediated autophagy (CMA).** *Autophagy* 2020, **16**:347-370.
7. Lees AJ, Hardy J, Revesz T: **Parkinson's disease.** *Lancet (London, England)* 2009, **373**:2055-2066.
8. Kalia LV, Lang AE: **Parkinson's disease.** *Lancet (London, England)* 2015, **386**:896-912.
9. Watanabe R, Buschauer R, Böhning J, Audagnotto M, Lasker K, Lu TW, Boassa D, Taylor S, Villa E: **The In Situ Structure of Parkinson's Disease-Linked LRRK2.** *Cell* 2020, **182**:1508-1518.e1516.
10. Wilkerson MD, Hayes DN: **ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.** *Bioinformatics* 2010, **26**:1572-1573.
11. Taz TA, Ahmed K, Paul BK, Kawsar M, Aktar N, Mahmud SMH, Moni MA: **Network-based identification genetic effect of SARS-CoV-2 infections to Idiopathic pulmonary fibrosis (IPF) patients.** *Brief Bioinform* 2021, **22**:1254-1266.
12. Ben-Hur A, Noble WS: **Kernel methods for predicting protein–protein interactions.** *Bioinformatics* 2005, **21**:i38-i46.

13. Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, Okamoto S, Hong H, Nakagawa M, Tanabe K, Tezuka K, et al: **A more efficient method to generate integration-free human iPS cells.** *Nature methods* 2011, **8**:409-412.
14. Borgs L, Peyre E, Alix P, Hanon K, Grobarczyk B, Godin JD, Purnelle A, Krusy N, Maquet P, Lefebvre P, et al: **Dopaminergic neurons differentiating from LRRK2 G2019S induced pluripotent stem cells show early neuritic branching defects.** *Scientific reports* 2016, **6**:33377.
15. Kavanagh E, Burguillos MA, Carrillo-Jimenez A, Oliva-Martin MJ, Santiago M, Rodhe J, Joseph B, Venero JL: **Deletion of caspase-8 in mouse myeloid cells blocks microglia pro-inflammatory activation and confers protection in MPTP neurodegeneration model.** *Aging* 2015, **7**:673.
16. Maver A, Peterlin B: **Positional integratomic approach in identification of genomic candidate regions for Parkinson's disease.** *Bioinformatics* 2011, **27**:1971-1978.
17. Su R, Liu X, Xiao G, Wei L: **Meta-GDBP: a high-level stacked regression model to improve anticancer drug response prediction.** *Brief Bioinform* 2020, **21**:996-1005.
18. Yan L, Li L, Lei J: **Long noncoding RNA small nucleolar RNA host gene 12/microRNA-138-5p/nuclear factor I/B regulates neuronal apoptosis, inflammatory response, and oxidative stress in Parkinson's disease.** *Bioengineered* 2021.
19. Zhao Y, Dzamko N: **Recent Developments in LRRK2-Targeted Therapy for Parkinson's Disease.** *Drugs* 2019, **79**:1037-1051.
20. Kaestner KH: **The hepatocyte nuclear factor 3 (HNF3 or FOXA) family in metabolism.** *Trends Endocrinol Metab* 2000, **11**:281-285.

Tables

Table1. Primer sequences used

genes	Primer sequences, 5'–3'	
	Forward	Reverse
CD44	CCAGAAGGAACAGTGGTTTGGC	ACTGTCCTCTGGGCTTGGTGTT
CTGF	CTTGCGAAGCTGACCTGGAAGA	CCGTCGGTACATACTCCACAGA
THBS1	GCTGGAAATGTGGTGCTTGTCC	CTCCATTGTGGTTGAAGCAGGC
SPP1	CGAGGTGATAGTGTGGTTTATGG	GCACCATTCAACTCCTCGCTTTC
VEGFA	TTGCCTTGCTGCTCTACCTCCA	GATGGCAGTAGCTGCGCTGATA
GAPDH	AGAAGGCTGGGGCTCATTTG	AGGGGCCATCCACAGTCTTC

Table 2 Datasets information

Database	No.	Published date	Samples
GEO	GSE33298	Submitted in 2011, updated in 2018	Fibroblast iPSC-induced neural cells
GEO	GSE36321	2014	H9 hESC derived NSC

Table 3. Top 20 clusters with their representative enriched terms (one per cluster). The number of genes in the user-provided lists that belong to the supplied ontology term is called "count." "%" is the percentage of all of the user-provided genes that are found in the given ontology term (only input genes with at least one ontology term annotation are included in the calculation). " $\text{Log}_{10}(P)$ " is the P -value in log base 10. " $\text{Log}_{10}(q)$ " is the multi-test adjusted p-value in log base 10.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0001568	GO Biological Processes	blood vessel development	76	11.55	-27.86	-23.5
GO:0048729	GO Biological Processes	tissue morphogenesis	65	9.88	-25.35	-21.29
GO:0007507	GO Biological Processes	heart development	63	9.57	-24.78	-20.89
R-HSA-9006934	Reactome Gene Sets	Signalling by Receptor Tyrosine Kinases	59	8.97	-24.53	-20.77
GO:0032989	GO Biological Processes	cellular component morphogenesis	70	10.64	-24.24	-20.58
GO:0070848	GO Biological Processes	response to growth factor	68	10.33	-23.57	-19.99
R-HSA-1474244	Reactome Gene Sets	Extracellular matrix organization	44	6.69	-22.82	-19.31
WP5087	WikiPathways	Malignant pleural mesothelioma	50	7.6	-21.38	-17.97
GO:0007169	GO Biological Processes	transmembrane receptor protein tyrosine kinase signalling pathway	60	9.12	-21.05	-17.82
GO:0009611	GO Biological Processes	response to wounding	56	8.51	-20.97	-17.81
WP5094	WikiPathways	Orexin receptor pathway	31	4.71	-20.35	-17.26
GO:0061061	GO Biological Processes	muscle structure development	59	8.97	-20.07	-17.01
GO:0060322	GO Biological Processes	head development	66	10.03	-19.88	-16.84
GO:0008285	GO Biological Processes	negative regulation of cell population proliferation	65	9.88	-19.72	-16.71
GO:0048732	GO Biological Processes	gland development	48	7.29	-19.45	-16.47
GO:0030155	GO Biological Processes	regulation of cell adhesion	63	9.57	-19.24	-16.29
GO:0007423	GO Biological Processes	sensory organ development	53	8.05	-18.71	-15.81
GO:0010942	GO Biological Processes	positive regulation of cell death	56	8.51	-18.66	-15.77

GO:0034330	GO Biological Processes	cell junction organization	59	8.97	-18.25	-15.41
GO:0060485	GO Biological Processes	mesenchyme development	37	5.62	-17.15	-14.45

Table 4 Summary of enrichment analysis in Transcription Factor Targets

GO	Description	Count	%	Log10(P)	Log10(q)
M572	TGCCAAR NF1 Q6	63	9.6	-20	-17
M4764	TGTTTGY HNF3 Q6	60	9.1	-17	-15
M17712	WGTTNNNNNAAA UNKNOWN	50	7.6	-17	-14
M6517	AAAYWAACM HFH4 01	34	5.2	-16	-13
M17420	WGGAATGY TEF1 Q6	37	5.6	-13	-11
M14357	TATA C	31	4.7	-13	-10
M14351	AFP1 Q6	30	4.6	-13	-10
M8720	STAT5A 03	30	4.6	-12	-10
M16291	YNGTTNNNATT UNKNOWN	35	5.3	-12	-9.9
M1328	WTGAAAT UNKNOWN	45	6.8	-12	-9.3
M16417	FOXO4 02	28	4.3	-11	-9
M822	YTATTTTNR MEF2 02	48	7.3	-11	-8.9
M946	GGATTA PITX2 Q2	43	6.5	-11	-8.9
M9612	OCT1 01	28	4.3	-11	-8.7
M11838	FOXD3 01	24	3.6	-11	-8.5
M9329	SRY 02	27	4.1	-11	-8.5
M4815	WTTGKCTG UNKNOWN	39	5.9	-11	-8.4
M399	SMTTTTGT UNKNOWN	34	5.2	-11	-8.4
M1031	FREAC2 01	27	4.1	-10	-8.3
M19327	CDC5 01	27	4.1	-10	-8.2

Figures

Figure 1

Study flow diagram

Figure 2

Consensus clustering presents quantitative evidence for estimating the number of possible clusters and their membership in the datasets. A. volcano plots. B. CDF curve. C-D. The relative change in area under the cumulative distribution function (CDF) curve for each category number k compared to $k-1$ is shown in the consensus clustering delta area curve (C) and column chart (D). The horizontal axis displays the category number k , while the vertical axis depicts the relative change in area under the CDF curve. E-F. D

Soft Threshold (Power), while the vertical axis shows the network's average connectivity. E. Soft Threshold (Power) represents the weight, and the scale-free topology fitting index R^2 is indicated on the vertical axis. F. The heat maps of module correlations display genes and samples in each cell, along with correlation coefficients and P values. G. Genes and samples are classified into PD and control in the heat maps of correlations between modules, with correlation coefficients and P values in each cell.

Figure 4

GO and KEGG to identify the most highly enriched pathways. A. Network of enriched terms: (a) coloured by cluster-ID, where nodes that share the same cluster-ID are typically close to each other; (b) coloured by p-value, where terms containing more genes tend to have a more significant P-value. B. The biological process (BP) category identifies gene list enrichments, and C. in the cellular component (CC) category. All genes in the genome were used as an enrichment background. The enrichment factor is the ratio between the observed counts and the counts expected by chance. Terms having a p-value of less than 0.01, a minimum count of three, and an enrichment factor of more than 1.5 are aggregated and categorized into clusters based on membership commonality. D. Enrichment analysis in transcription factor targets. E. KEGG biological pathways analysis of 607, which involved in the mutation of LRRK2.

Figure 5

The PPI network reveals interacting proteins. A. The STRING online tool constructed a PPI network for the genes in DEGs. There were a total of 600 protein-protein interactions. B. A heat map represented the 50 genes that demonstrated evident changes in the LRRK2 mutated group. C. The STRING web program created a PPI network for the hub genes to show the interaction between the proteins with the top hub genes. D. The relative hub genes levels of CD44, CTGF, THBS1, VEGFA, SPP1, EGF, VCAM1, MMP3,

CXCR4, and LOX in LRRK2 wild and mutation group. E. The relative gene expression of CD44, CTGF, THBS1, SPP1, EGF, and LOX was considerably higher in the LRRK2 Gly2019Ser mutant group than in the LRRK2 wild group. *P < 0.05, **P < 0.01, ***P < 0.001. F. We used the R software package pROC (version 1.17.0.1) to perform ROC analysis to obtain the area under the ROC curve (AUC). H. Pearson correlation coefficients among individual the 10 hub genes. *P < 0.05, **P < 0.01, ***P < 0.001.

Figure 6

The validation of hub genes in iPSC-induced DA cells. A. A confocal image supported by immunofluorescence confirmed the expression of TH. Green: anti-TH (antibody labelling DA). Blue: DAPI. Scale bar, 100 μ m. B-F. The mRNA levels of CD44 (B), CTGF (C), THBS1 (D), VEGFA (E), SPP1 (F) were determined using real-time qRT-PCR in the LRRK2 wild and mutated group.

Figure 7

The PPI network reveals interaction for the effectors responding to LRRK2 mutation. The STRING web program created a PPI network for the for the effectors responding to LRRK2 mutation: A. CD44. B. CTGF. C. THBS1, D. VEGFA. E. SPP1.