Science Advances
AAAS

Supplementary Materials for

Characterization of altered molecular mechanisms in Parkinson's disease through cell type–resolved multiomics analyses

Andrew J. Lee *et al.*

Corresponding author: Inkyung Jung, ijung@kaist.ac.kr; Eliezer Masliah, eliezer.masliah@nih.gov

Sci. Adv. **9**, eabo2467 (2023)
DOI: 10.1126/sciadv.abo2467

The PDF file includes:

Figs. S1 to 14
Legends for tables S1 to S10

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S10

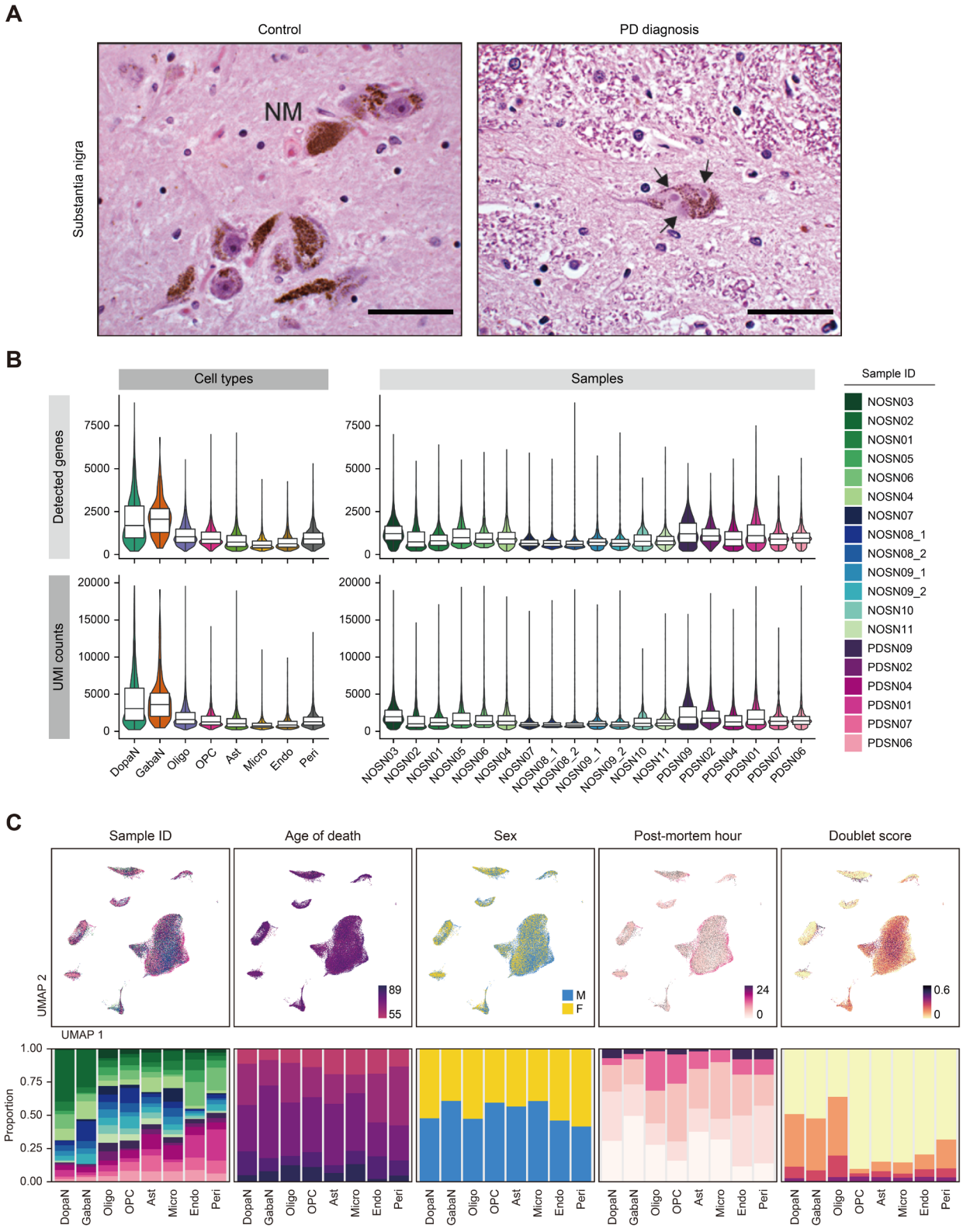

Fig. S1. Pathological examination of human SN tissues and quality assessment of snRNA-seq clusters.

(A) Histopathological analysis of the brains used in the experiment. Staining of SN sections by Hematoxylin and Eosin (H&E) shows loss of neurons, diminished neuromelanin (NM), and Lewy bodies in dopaminergic neurons in PD (PDSN02) compared to control (NOSN01). Arrows indicate Lewy body. Scale bars, 40 μm. **(B)** Distributions of genes detected (top) and UMI counts (bottom) with respect to snRNA-seq cell types (left) and samples (right). Box boundaries and line correspond to the interquartile range (IQR) and median respectively. Whiskers extend to the lowest or highest data points that are no further than 1.5 times the IQR from the box boundaries. **(C)** Top: UMAP embeddings of snRNA-seq nuclei mapped to potential confounding variables (samples, age of death, sex, post-mortem hour, and doublet score). Bottom: bar plots illustrating the normalized proportion of each variable with respect to the cell types.

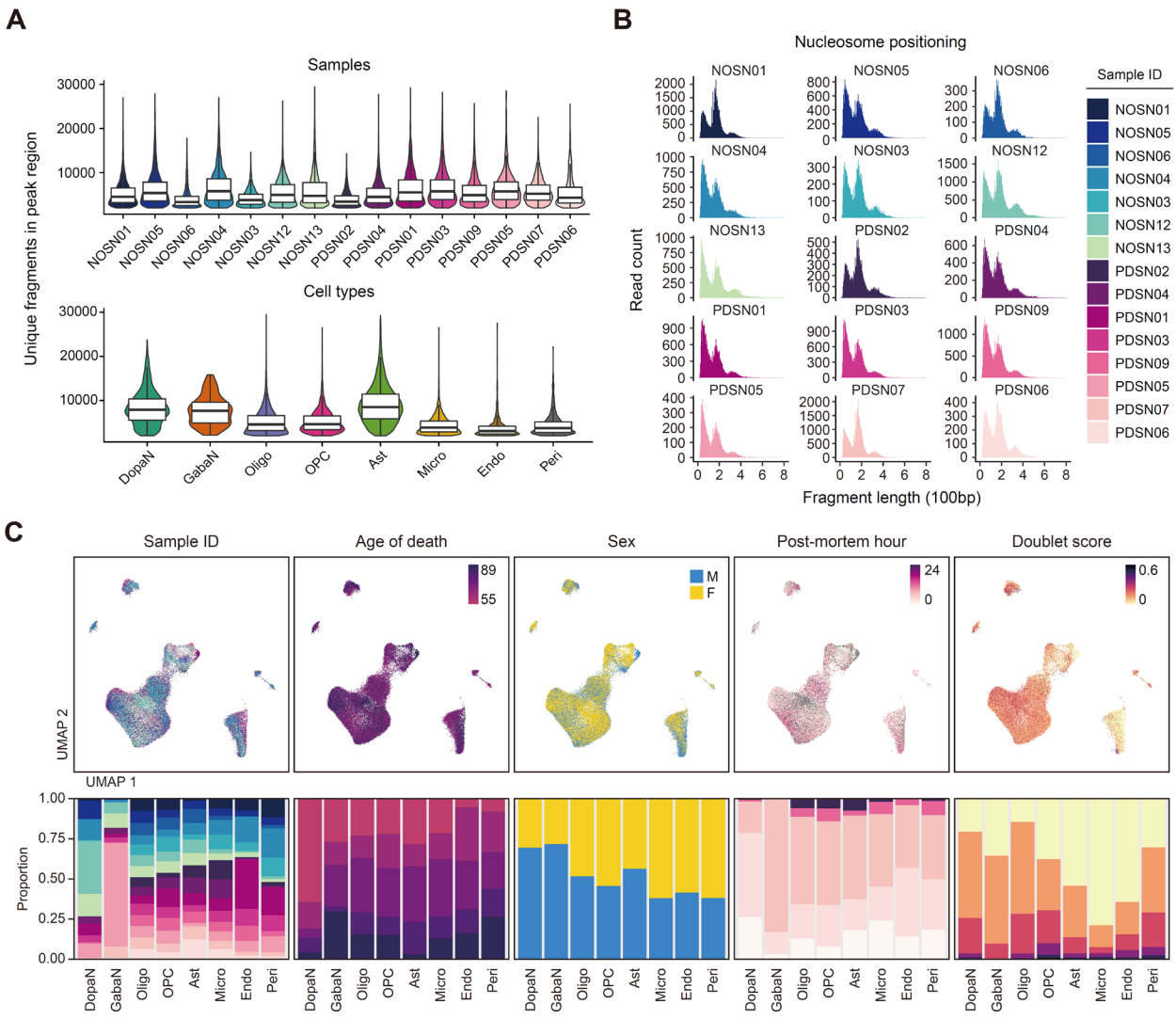

Fig. S2. Data quality assessment of snATAC-seq clusters.

(A) Distributions of unique fragments with respect to snATAC-seq samples (top) and cell types (bottom). Box boundaries and line correspond to the interquartile range (IQR) and median respectively. Whiskers extend to the lowest or highest data points that are no further than 1.5 times the IQR from the box boundaries. **(B)** Histograms of nucleosome positioning in each biological sample. **(C)** Top: UMAP embeddings of snATAC-seq nuclei mapped to potential confounding variables (samples, age of death, sex, post-mortem hour, and doublet score).

Bottom: bar plots illustrating the normalized proportion of each variable with respect to the cell types.

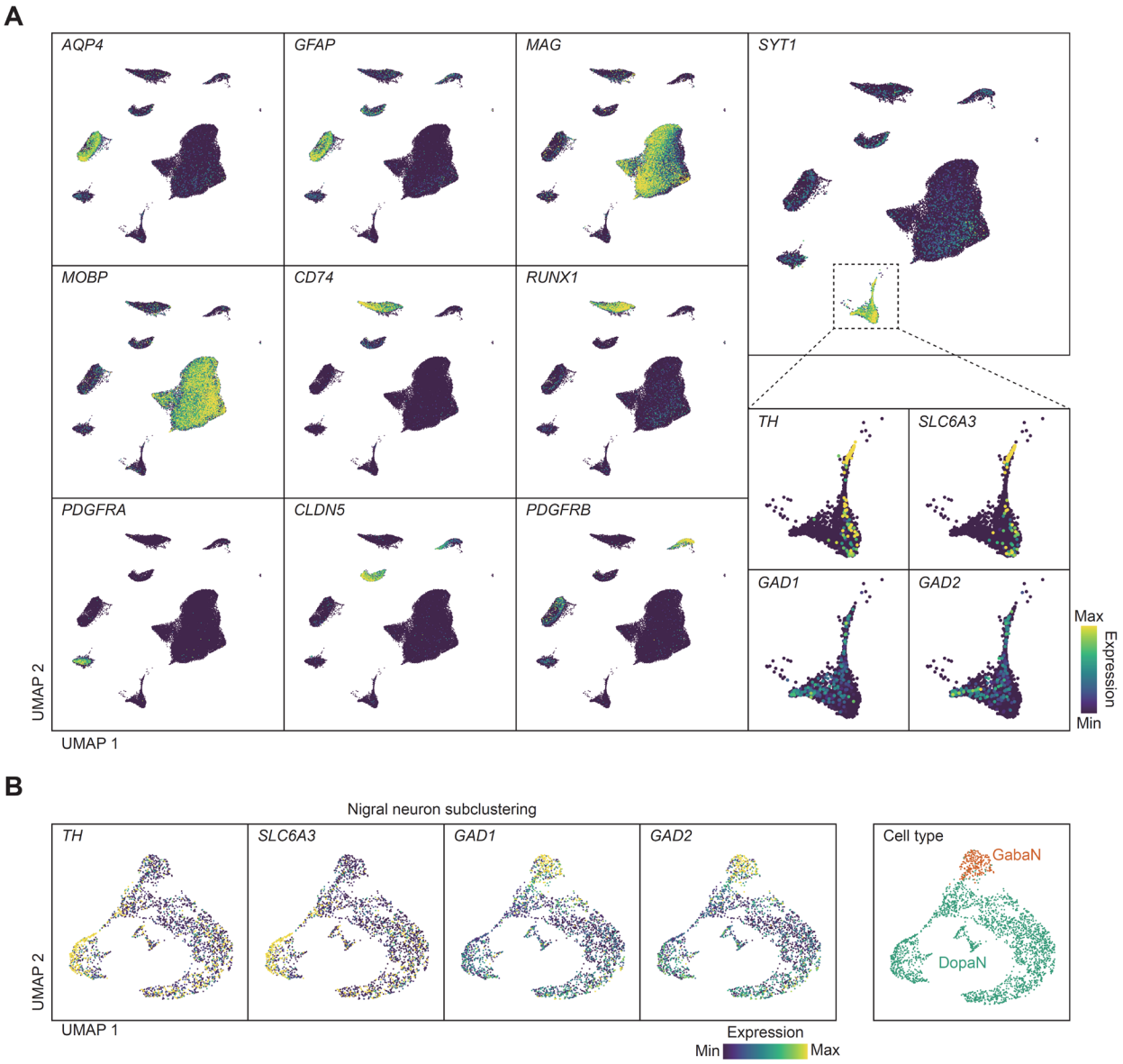

Fig. S3. Cell type annotation for snRNA-seq clusters in the human SN.

(A) UMAP embeddings of snRNA-seq nuclei illustrating the expression level of cell type markers to identify the major cell types present in the human SN; neurons (*SYT1*), oligodendrocytes (*MAG* and *MOBP*), oligodendrocyte precursor cells (*PDGFRA*), astrocytes (*AQP4* and *GFAP*), microglia (*CD74* and *RUNX1*), endothelial cells (*CLDN5*), and pericytes (*PDGFRB*). **(B)** UMAP embeddings of snRNA-seq nuclei illustrating the sub-annotation of

neuronal populations into dopaminergic neurons (*TH* and *SLC6A3*) and GABAergic neurons (*GAD1* and *GAD2*).

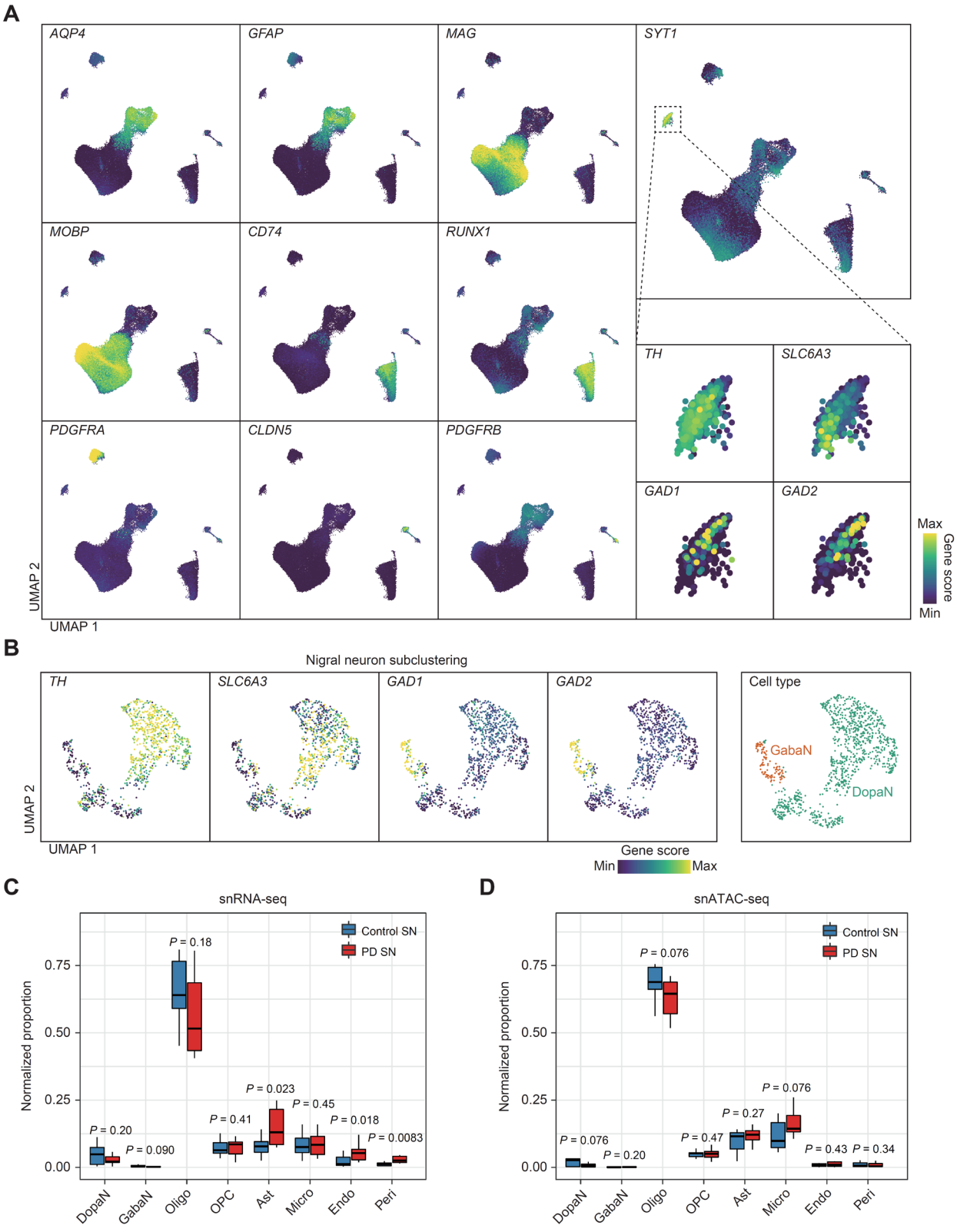
A
AQP4
GFAP
MAG
SYT1
MOBP
CD74
RUNX1
PDGFRA
CLDN5
PDGFRB
TH
SLC6A3
GAD1
GAD2
Max
Gene score
Min
UMAP 2
UMAP 1
B
Nigral neuron subclustering
TH
SLC6A3
GAD1
GAD2
Cell type
GabaN
DopaN
UMAP 2
UMAP 1
Gene score
Min
Max
C
snRNA-seq
Control SN
PD SN
Normalized proportion
0.75
0.50
0.25
0.00
P = 0.20
P = 0.090
P = 0.18
P = 0.41
P = 0.023
P = 0.45
P = 0.018
P = 0.0083
DopaN
GabaN
Oligo
OPC
Ast
Micro
Endo
Peri
D
snATAC-seq
Control SN
PD SN
Normalized proportion
0.75
0.50
0.25
0.00
P = 0.076
P = 0.076
P = 0.20
P = 0.47
P = 0.27
P = 0.076
P = 0.43
P = 0.34
DopaN
GabaN
Oligo
OPC
Ast
Micro
Endo
Peri

Fig. S4. Cell type assignment for snATAC-seq clusters and cellular compositions in the human SN.

(A) UMAP embeddings of snATAC-seq nuclei illustrating the gene activity score of cell type markers to identify the major cell types present in the human SN; neurons (*SYT1*), oligodendrocytes (*MAG* and *MOBP*), oligodendrocyte precursor cells (*PDGFRA*), astrocytes (*AQP4* and *GFAP*), microglia (*CD74* and *RUNX1*), endothelial cells (*CLDN5*), and pericytes (*PDGFRB*). **(B)** UMAP embeddings of snATAC-seq nuclei illustrating the sub-annotation of neuronal populations into dopaminergic neurons (*TH* and *SLC6A3*) and GABAergic neurons (*GAD1* and *GAD2*). **(C-D)** Box plots showing the proportion of nuclei mapped to each cell type and sample, compared between PD (red) and control (blue) SN for snRNA-seq (C) and snATAC-seq (D). Two-sided Wilcoxon tests are performed to measure the statistical significance of the difference in cellular compositions between PD and control cases. The box boundaries and line correspond to the interquartile range (IQR) and median, respectively. Whiskers extend to the lowest or highest data points that are no further than 1.5 times the IQR from the box boundaries.

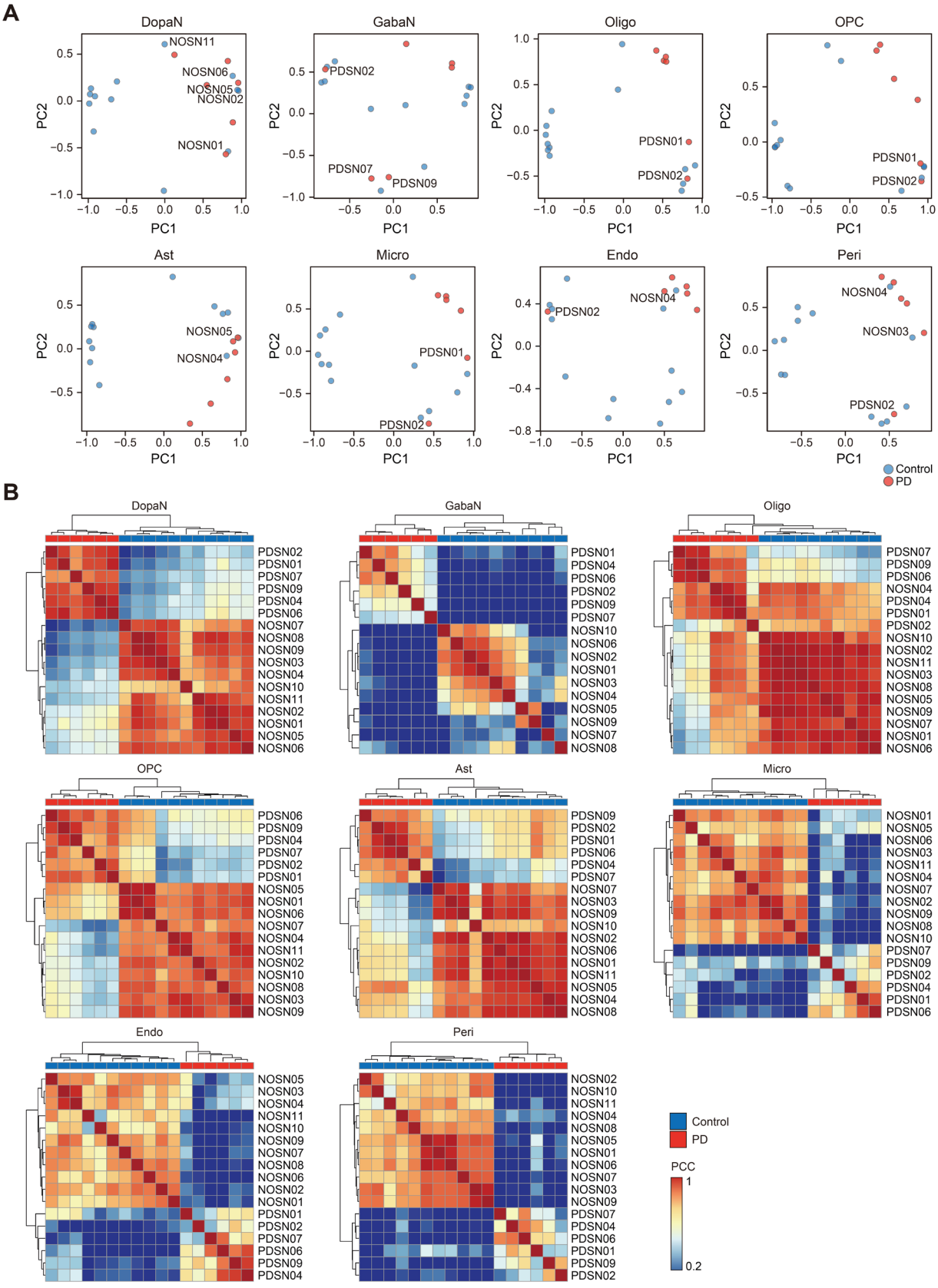
A
DopaN
GabaN
Oligo
OPC
Ast
Micro
Endo
Peri
PC1
PC2
NOSN11
NOSN06
NOSN05
NOSN02
NOSN01
PDSN02
PDSN07
PDSN09
PDSN01
NOSN04
NOSN03
Control
PD
B
PDSN02
PDSN01
PDSN07
PDSN09
PDSN04
PDSN06
NOSN07
NOSN08
NOSN09
NOSN03
NOSN04
NOSN10
NOSN11
NOSN02
NOSN01
NOSN05
NOSN06
PDSN04
PDSN06
PDSN01
PDSN02
PDSN07
PDSN09
NOSN04
NOSN10
NOSN02
NOSN11
NOSN03
NOSN08
NOSN05
NOSN09
NOSN07
NOSN01
PDSN09
PDSN02
PDSN01
PDSN06
PDSN04
PDSN07
NOSN07
NOSN03
NOSN09
NOSN10
NOSN02
NOSN06
NOSN01
NOSN11
NOSN05
NOSN04
NOSN08
NOSN01
NOSN05
NOSN06
NOSN03
NOSN11
NOSN04
NOSN07
NOSN02
NOSN09
NOSN08
NOSN10
PDSN07
PDSN09
PDSN02
PDSN04
PDSN01
PDSN06
NOSN05
NOSN03
NOSN04
NOSN11
NOSN10
NOSN09
NOSN07
NOSN08
NOSN06
NOSN02
NOSN01
PDSN01
PDSN02
PDSN07
PDSN06
PDSN09
PDSN04
NOSN02
NOSN10
NOSN11
NOSN04
NOSN08
NOSN05
NOSN01
NOSN06
NOSN07
NOSN03
NOSN09
PDSN07
PDSN04
PDSN06
PDSN01
PDSN09
PDSN02
Control
PD
PCC
1
0.2

Fig. S5. Sample reproducibility among PD and control donors based on snRNA-seq data. **(A)** PCA plots illustrating the overall similarity among PD and control snRNA-seq samples based on the top 1,000 variable genes obtained from each cell type. **(B)** Pearson correlation coefficients heatmap illustrating the reproducibility of PD and control cases based on PD-associated DEGs in individual cell types.

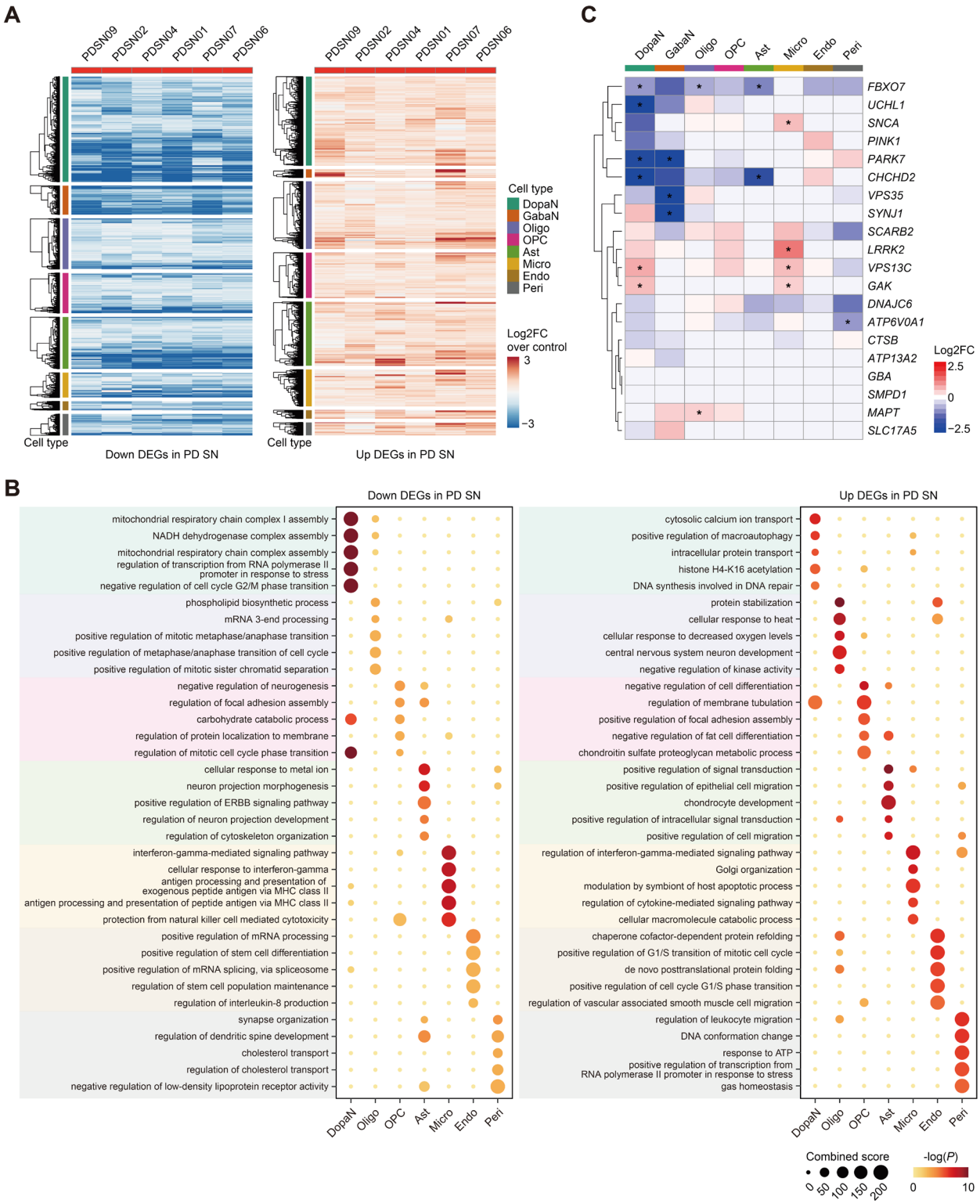

Fig. S6. Assessment of PD-associated DEGs at the individual and cell type level.

(A) A heatmap showing the log2 fold change of cell type-resolved RNA-seq reads for each PD case compared to the mean of control cases for down-regulated DEGs (left) and up-regulated DEGs (right). **(B)** Enriched GO pathway analysis for down- (left) and up- (right) regulated DEGs showing the top 5 enriched biological processes for each cell type based on the significance level. **(C)** A heatmap showing the log2 fold change of cell type-resolved RNA-seq reads for PD SN compared to control SN across 20 known PD risk genes in a cell type-specific manner. The asterisk represents snRNA-seq DEGs identified based on the following criteria: EdgeR-LRT BH-adjusted $P < 0.05$, log2 fold change > 0.5, and minimum expressed cells > 0.1.

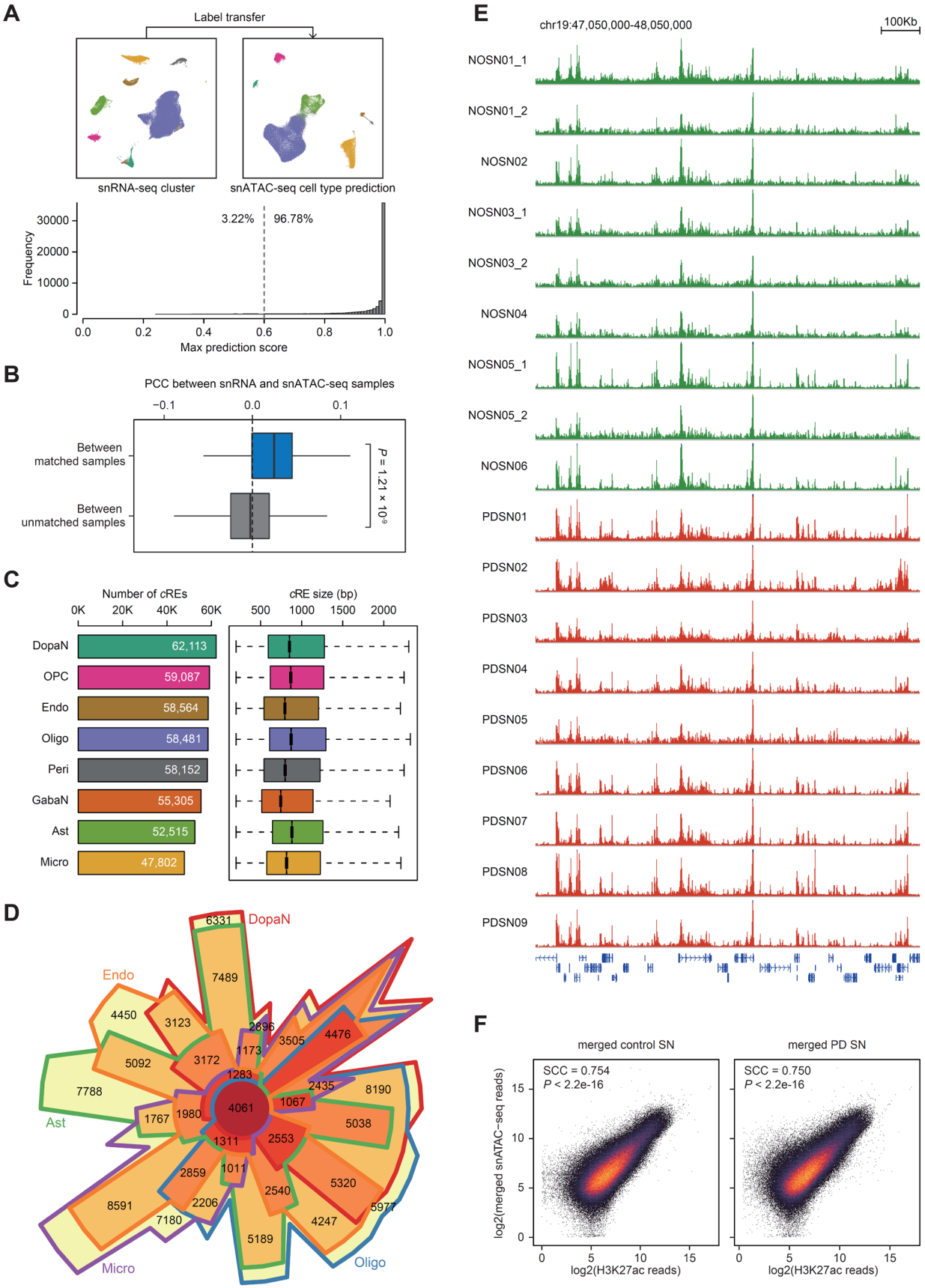
A
Label transfer
snRNA-seq cluster
snATAC-seq cell type prediction
3.22%
96.78%
Frequency
Max prediction score
B
PCC between snRNA and snATAC-seq samples
Between matched samples
Between unmatched samples
P = 1.21 × 10⁻⁹
C
Number of cREs
cRE size (bp)
DopaN 62,113
OPC 59,087
Endo 58,564
Oligo 58,481
Peri 58,152
GabaN 55,305
Ast 52,515
Micro 47,802
D
DopaN
Endo
Ast
Micro
Oligo
E
chr19:47,050,000-48,050,000
100Kb
NOSN01_1
NOSN01_2
NOSN02
NOSN03_1
NOSN03_2
NOSN04
NOSN05_1
NOSN05_2
NOSN06
PDSN01
PDSN02
PDSN03
PDSN04
PDSN05
PDSN06
PDSN07
PDSN08
PDSN09
F
merged control SN
merged PD SN
SCC = 0.754
P < 2.2e-16
SCC = 0.750
P < 2.2e-16
log2(merged snATAC-seq reads)
log2(H3K27ac reads)

Fig. S7. Characterization of cell type-resolved *cis*-regulatory landscape in the human SN. **(A)** Cell type label transfer and snATAC-seq cell type prediction using the label transfer algorithms. UMAP embeddings of cell type annotated snRNA-seq clusters and the predicted cell type in snATAC-seq clusters (up), and a histogram showing the distribution of max prediction score across snATAC-seq nuclei (down). **(B)** Boxplots showing the Pearson correlation coefficients between snRNA-seq and snATAC-seq data at the individual and cell type level comparing between matched and unmatched samples. The significance in difference between matched correlations and unmatched correlations was calculated based on two-sided Wilcoxon rank sum test. **(C)** Bar plots showing the number of cell type-resolved cREs (left), and boxplots illustrating the overall size of cell type-resolve *c*REs (right) for each cell type. Box represents the upper and lower interquartile range (IQR), and whiskers correspond to the highest and lowest points of 1.5× the IQR. **(D)** A Chow-Ruskey plot illustrating the association of cell type-resolved *c*REs among dopaminergic neurons (DopaN), oligodendrocytes (Oligo), astrocytes (Ast), microglia (Micro), and endothelial cells (Endo). **(E)** Epigenome browser tracks for H3K27ac ChIP-seq signals in a genomic locus (chr19:47,050,000-48,050,000) showing the peak quality for individual samples. The ChIP-seq signals were normalized by the total reads mapped in *c*REs. **(F)** Scatter plots showing the correlation of log2 merged ChIP-seq reads for PD (right) and control (left) SN against log2 merged snATAC-seq reads across the *c*REs determined based on Spearman's correlation coefficient (SCC).

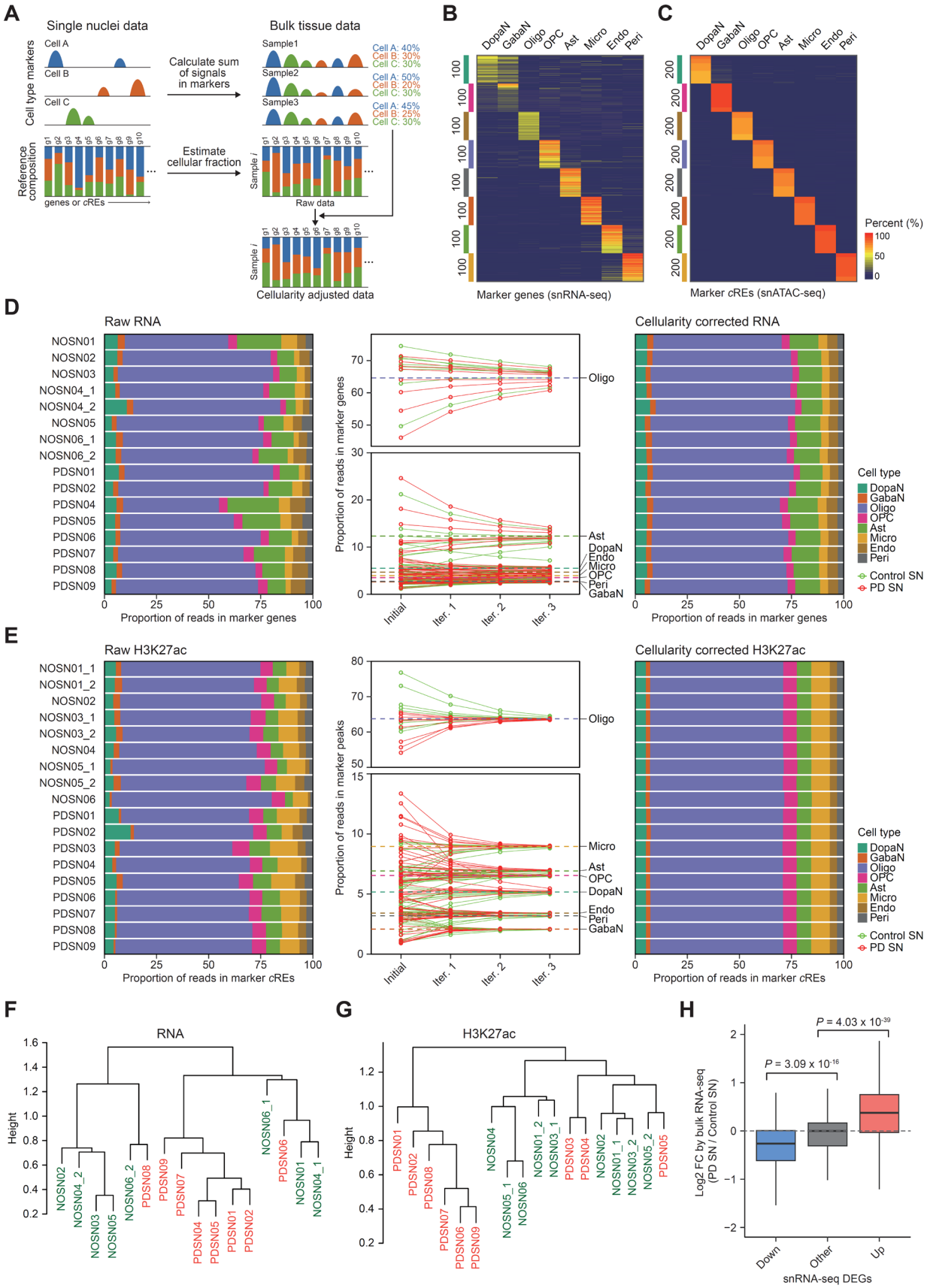
A
Single nuclei data
Bulk tissue data
Cell type markers
Cell A
Cell B
Cell C
Calculate sum of signals in markers
Sample1
Sample2
Sample3
Cell A: 40%
Cell B: 30%
Cell C: 30%
Cell A: 50%
Cell B: 20%
Cell C: 30%
Cell A: 45%
Cell B: 25%
Cell C: 30%
Reference composition
genes or cREs
Estimate cellular fraction
Sample i
Raw data
Cellularity adjusted data
B
DopaN GabaN Oligo OPC Ast Micro Endo Peri
Marker genes (snRNA-seq)
C
Marker cREs (snATAC-seq)
Percent (%)
D
Raw RNA
Cellularity corrected RNA
Proportion of reads in marker genes
Oligo
Ast
DopaN
Endo
Micro
OPC
Peri
GabaN
Initial
Iter. 1
Iter. 2
Iter. 3
Cell type
Control SN
PD SN
E
Raw H3K27ac
Cellularity corrected H3K27ac
Proportion of reads in marker peaks
Proportion of reads in marker cREs
F
RNA
Height
G
H3K27ac
H
P = 4.03 x 10-39
P = 3.09 x 10-16
Log2 FC by bulk RNA-seq (PD SN / Control SN)
Down
Other
Up
snRNA-seq DEGs

Fig. S8. Cellularity evaluation and adjustment for bulk RNA-seq and ChIP-seq data.

(A) Overview of an iterative cellularity correction approach. First, the sum of reads in unique marker genes or *c*REs is accounted for individual bulk assay samples in each cell type. Then, cellular heterogeneity for each sample is adjusted toward the average cellular compositions based on the cell type-resolved transcriptome and chromatin accessibility reference across the genes or *c*REs. **(B-C)** Heatmaps showing the percentage of reads across cell types over marker genes (B) and marker *c*REs (C). **(D)** Relative ratio of reads in marker genes before (left) and after (right) iterative cellularity correction for bulk RNA-seq samples. **(E)** Relative ratio of reads in marker *c*REs before (left) and after (right) iterative cellularity correction for bulk H3K27ac ChIP-seq samples. **(F-G)** Unbiased hierarchical clustering of bulk transcriptome (F) and H3K27ac activity (G) using the cellularity-corrected data across the PD and control samples. PD and control cases are highlighted by red and green, respectively. **(H)** Box plots illustrating the log2 fold change of cellularity-corrected bulk RNA-seq expression (mean PD SN over mean control SN) based on PD-associated DEGs identified from snRNA-seq data.

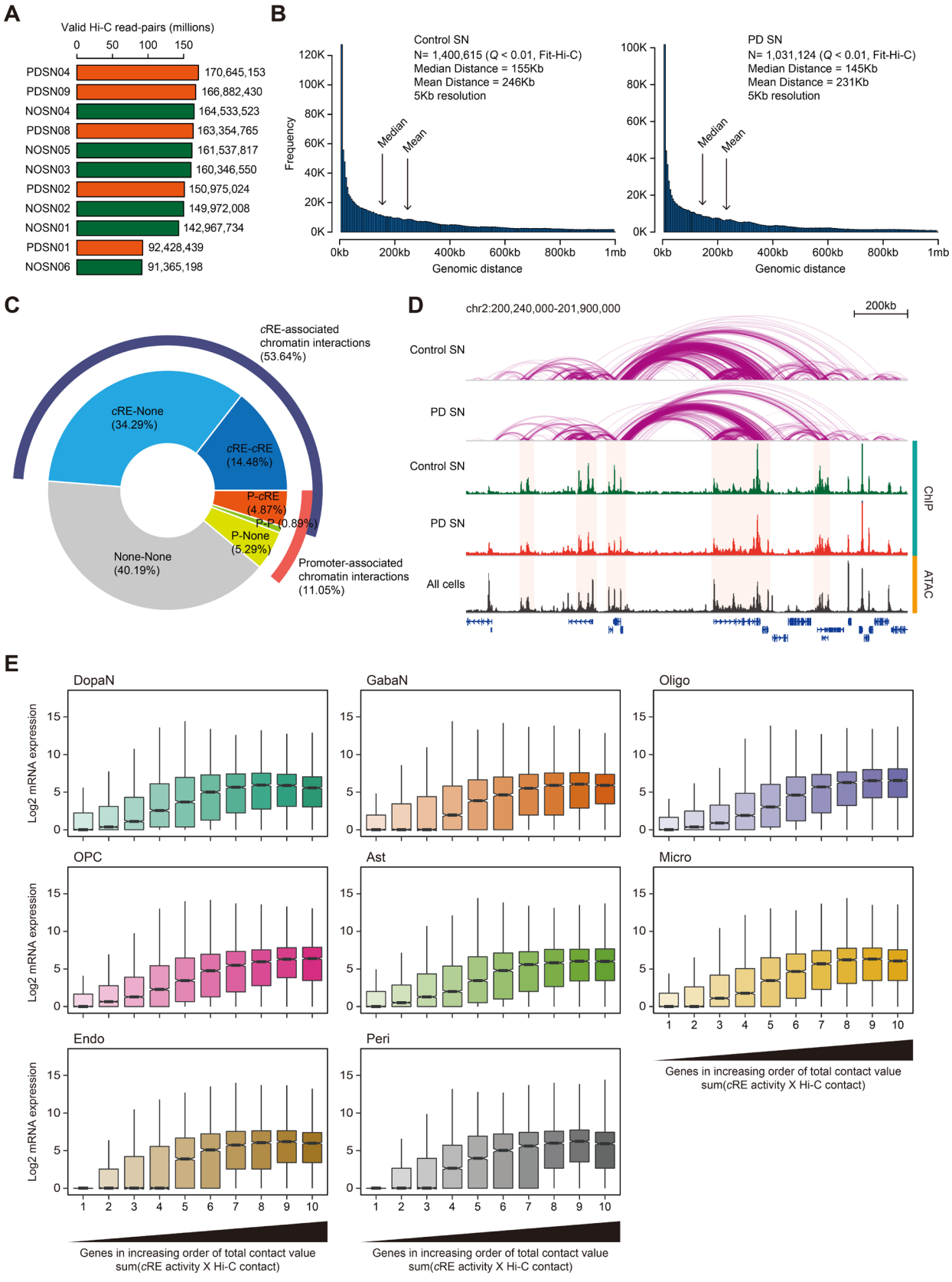
A
Valid Hi-C read-pairs (millions)
0
50
100
150
PDSN04 170,645,153
PDSN09 166,882,430
NOSN04 164,533,523
PDSN08 163,354,765
NOSN05 161,537,817
NOSN03 160,346,550
PDSN02 150,975,024
NOSN02 149,972,008
NOSN01 142,967,734
PDSN01 92,428,439
NOSN06 91,365,198
B
Control SN
N= 1,400,615 (Q < 0.01, Fit-Hi-C)
Median Distance = 155Kb
Mean Distance = 246Kb
5Kb resolution
PD SN
N= 1,031,124 (Q < 0.01, Fit-Hi-C)
Median Distance = 145Kb
Mean Distance = 231Kb
5Kb resolution
Frequency
Median
Mean
Genomic distance
0kb
200kb
400kb
600kb
800kb
1mb
C
cRE-associated chromatin interactions (53.64%)
cRE-None (34.29%)
cRE-cRE (14.48%)
P-cRE (4.87%)
P-P (0.89%)
P-None (5.29%)
None-None (40.19%)
Promoter-associated chromatin interactions (11.05%)
D
chr2:200,240,000-201,900,000
200kb
Control SN
PD SN
Control SN
PD SN
All cells
ChIP
ATAC
E
DopaN
GabaN
Oligo
OPC
Ast
Micro
Endo
Peri
Log2 mRNA expression
Genes in increasing order of total contact value sum(cRE activity X Hi-C contact)

Fig. S9. Characterization of high-resolution 3D chromatin interactions in human SN.

(A) Distribution of valid and unique Hi-C read-pairs generated by each biological sample. **(B)** Histograms showing the distribution of genomic distances of significant chromatin interactions identified by control (left) and PD (right) SN. **(C)** A donut plot illustrating the ratio of long-range chromatin interactions among promoters, *c*REs, and non-regulatory regions. **(D)** Representative chromatin interaction profiles at a genome coordinate (chr2:200,240,000-201,900,000) with genome browser tracks of H3K27ac ChIP-seq reads for control (green) and PD (red) SN, along with merged pseudobulk snATAC-seq reads. The signals for ChIP-seq and pseudobulk snATAC-seq were normalized by the total reads mapped in *c*REs. Regions with enriched H3K27ac signals (translucent orange) show extensive chromatin interactions. **(E)** Notched box plots illustrating the correlation between gene expression and overall chromatin contacts connected to gene promoter in all cell types. In notched box plots, the center line is the median, notch displays the 95% CI of median, box represents the upper and lower IQR, and whiskers correspond to the highest and lowest points of 1.5× the IQR.

Fig. S10. Evaluation of genetic sequence perturbation mediated by CRISPR-Cas9 genome editing in SH-SY5Y neuroblastoma.

Illustration of the guide-RNA target site for CRISPR-Cas9 experiment in the parental SH-SY5Y genome and three mutant strains. Two PD GWAS-SNPs (rs2072368, $P = 3.59 \times 10^{-11}$; rs2072369, $P = 4.29 \times 10^{-11}$) are located within a target *c*RE (chr7:23144778-23146888). Three

mutant clones from three individual guide-RNAs having a different target PAM sequence contained mutations with a varying size ranging from 20-50 base-pairs, confirmed by paired-end Sanger sequencing.

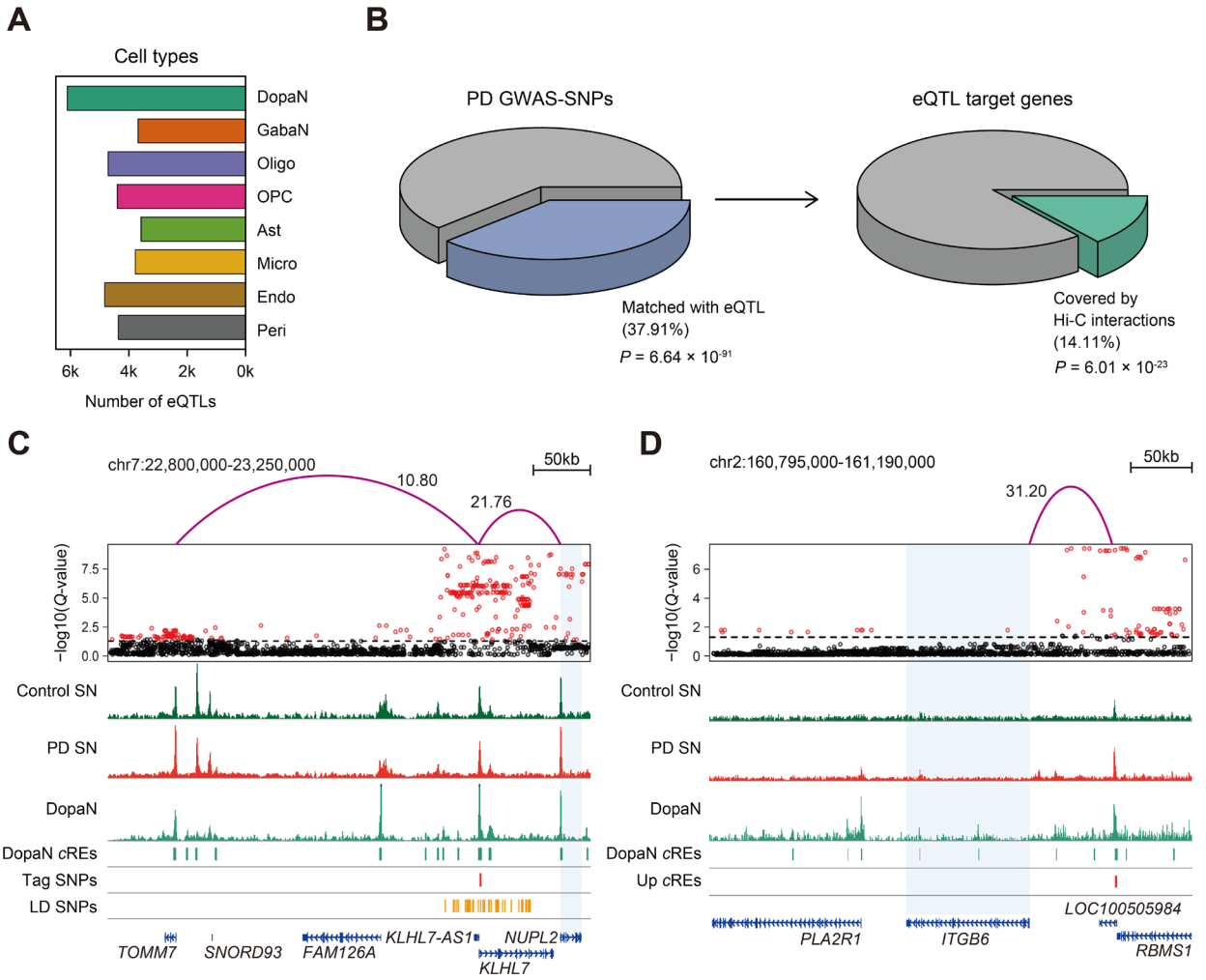

Fig. S11. Integrative analysis of eQTL data set with cell type-resolved *cis*-regulatory landscape of the human SN.

(A) Bar plots showing the overlap of significant eQTLs with cell type-resolved *c*RE in the human SN. **(B)** 3D pie charts illustrating the portion of LD-expanded PD GWAS-SNPs that matched with eQTLs (left), and the portion of eQTL-target gene association overlapped with Hi-C interactions by the matched GWAS-SNPs (right). The *P*-value and odd ratio were calculated using two-sided Fisher exact test for eQTL enrichment, and the significance of target gene overlap between eQTL associations and Hi-C interactions were calculated based on hypergeometric test (***: $P < 0.001$). **(C-D)** Example of a PD GWAS-SNP (C) and an up-

regulated *c*RE (D) in dopaminergic neurons whose significant chromatin interactions to a target gene promoter are supported by eQTL associations, with H3K27ac ChIP-seq tracks for PD and control SN and pseudobulk chromatin accessibility for dopaminergic neurons (DopaN). The signals for ChIP-seq and cell type-resolved snATAC-seq are normalized by the total reads mapped in *c*REs. Significant long-range chromatin interactions shown in purple arcs with the corresponding ABC score.

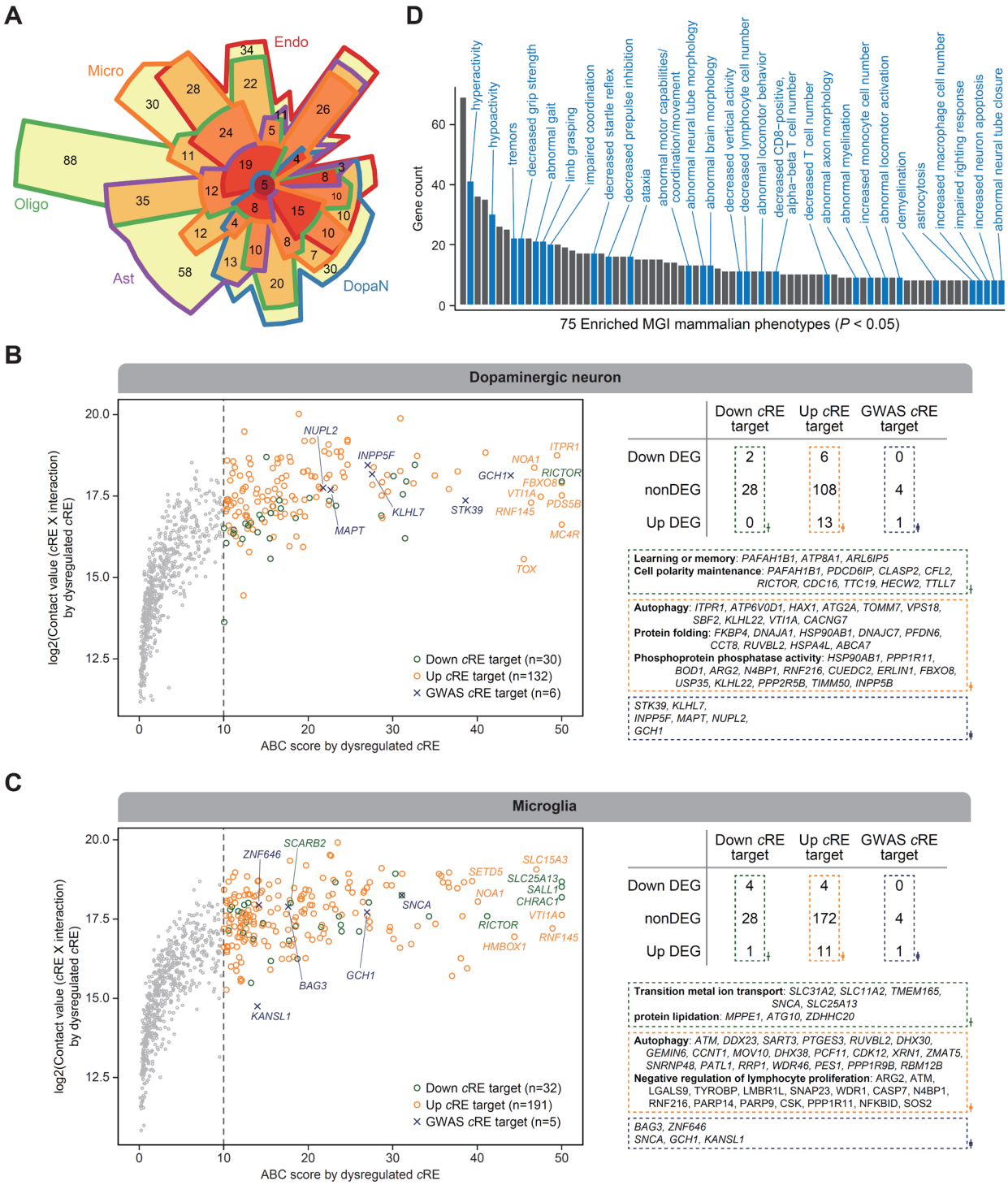

Fig. S12. Characterization of putative target genes of dysregulated *c*REs and PD GWAS-SNPs.

(A) A Chow-Ruskey plot for putative target genes of dysregulated *c*REs and PD GWAS-SNPs for dopaminergic neurons (DopaN), oligodendrocytes (Oligo), astrocytes (Ast), microglia (Micro), and endothelial cells (Endo). **(B-C)** Left: scatterplots illustrating the putative target genes for dopaminergic neurons (B), and microglia (C) with an ABC score (x-axis) threshold of 10, with the annotation of top 10 genes with the highest ABC score, along with target genes of GWAS-SNPs. Putative target genes were grouped into three categories, whether they are anchored by down- (green circles), up-regulated (orange circles) *c*REs, or GWAS-SNP harboring *c*REs (navy saltire). Right: putative target genes were categorized based on DEG status and the representative GO biological pathways. **(D)** Bar plots showing 75 enriched mammalian phenotypes ($P < 0.05$ and the number of associated genes ≥ 8) linked to putative targets of dysregulated *c*REs and PD GWAS-SNPs. 28 phenotype ontologies related to neurological, movement, and immune symptoms potentially associated with PD are highlighted in blue. The full list of enriched phenotype ontologies and their associated genes are provided in table S6.

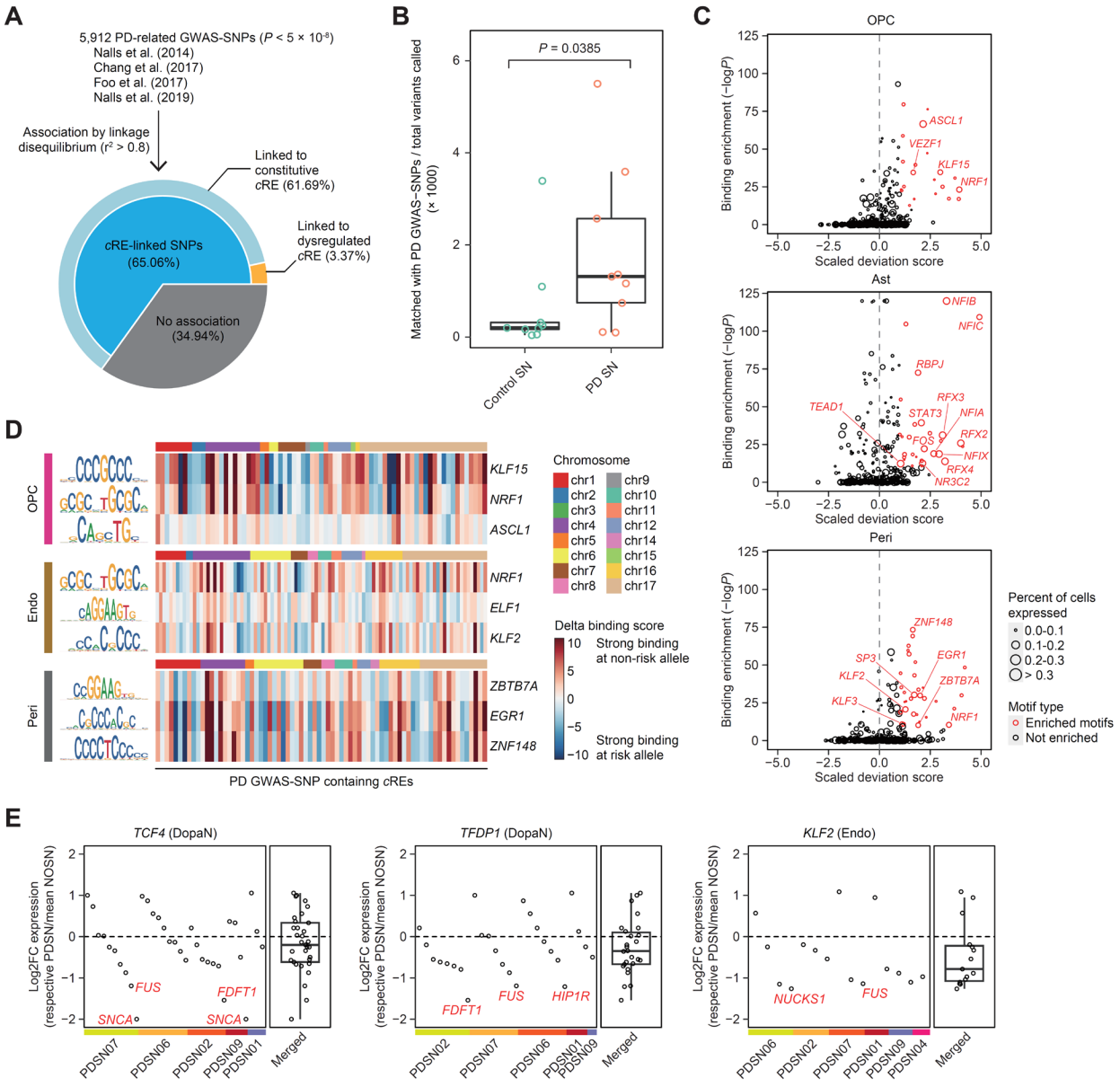

Fig. S13. Characterization of PD GWAS-SNPs based on *cis*-regulatory landscape of the human SN.

(A) A stacked donut chart showing the proportion of 5,912 PD GWAS-SNPs associated with *c*REs identified in the human SN based on LD structure ($r^2 > 0.8$). **(B)** Box plot illustrating the level of individual SNPs matched in PD GWAS-SNPs between PD and control groups. The difference in the number of GWAS-matched SNPs between PD and control groups was calculated through one-sided Wilcoxon rank sum test. **(C)** Scatter plots illustrating the -log10

hypergeometric enrichment of TFs and the scaled chromVAR deviation score, depicting TF motif activity, for OPC, astrocyte, and pericyte. The size of each data point represents the percent of nuclei expressed, and TFs with minimum expression detected greater than 10% were selected as enriched TFs. **(D)** Heat maps showing the delta binding scores of individual PD GWAS-SNP-containing *c*REs across the genome, for enriched TFs in OPC, endothelial cell, and pericyte. **(E)** Scatter plots describing the log2 fold change expression (a respective PD donor over the mean of control samples) for target genes of *c*REs with disruption in binding motifs for respective TFs by GWAS-matched donor variants based on cumulative ABC score greater than 1. The name of target genes with log2 fold change less than -1 is labeled in red.

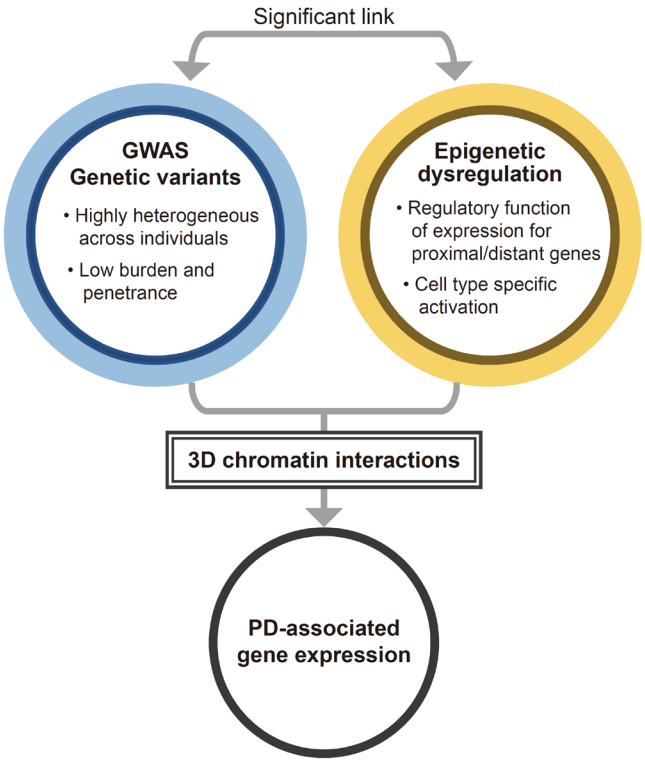

Fig. S14. Contribution of genetic and epigenetic perturbations in sporadic PD pathogenesis. A schematic overview illustrating the effects of GWAS-variants and *cis*-regulatory dysregulation in PD-specific gene expression.

Supplementary Tables:

Table S1. Sample information.

Table S2. List of nuclei obtained from single-nucleus data and sample demultiplexing.

Table S3. List of PD-associated differentially expressed genes identified by snRNA-seq.

Table S4. List of dysregulated *c*REs in PD SN.

Table S5. List of putative target genes for individual cell types.

Table S6. List of enriched MGI mouse phenotypes in putative target genes.

Table S7. Source information for GWAS summary statistics.

Table S8. List of enriched transcription factors and motifs.

Table S9. List of altered TF binding scores for GWAS-SNP-containing *c*REs.

Table S10. Enriched pathways of putative target gene clusters.