Molecular and spatial transcriptomic classification of midbrain dopamine neurons and their alterations in a LRRK2^{G2019S} model of Parkinson's disease

Zachary Gaertner[1,5#], Cameron Oram[2#], Amanda Schneeweis[1,5], Elan Schonfeld[1], Cyril Bolduc[2], Chuyu Chen[3,5], Daniel Dombeck[4,5], Loukia Parisiadou[3,5], Jean-Francois Poulin[2]*, Rajeshwar Awatramani[1,5]*

1 Northwestern University Feinberg School of Medicine, Dept of Neurology, Chicago, IL 60611

2 McGill University (Montreal Neurological Institute), Faculty of Medicine and Health Sciences, Dept of Neurology and Neurosurgery, Montreal (QC), Canada

3 Northwestern University Feinberg School of Medicine, Dept of Pharmacology, Chicago, IL 60611

4 Northwestern University, Dept of Neurobiology, Evanston, IL 60201

5 Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, USA

# equal contribution

* Corresponding authors

ZG: 0000-0002-1760-6549

CO: 0009-0003-2778-6031

AS: 0000-0003-4141-6064

ES: 0000-0001-7368-1562

CB: 0009-0003-6104-8462

CC: 0000-0001-5666-5173

DD: 0000-0003-2576-5918

LP: 0000-0002-2569-4200

JFP: 0000-0002-1039-4985

RA: 0000-0002-0713-2140

Short title: Taxonomy of midbrain DA neurons: Implications for PD

Teaser: Using snRNASeq and MERFISH we identified midbrain DA subtypes, mapped their spatial location, and identified alterations in a LRRK2 model

Abstract

Several studies have revealed that midbrain dopamine (DA) neurons, even within a single neuroanatomical area, display heterogeneous properties. In parallel, studies using single cell profiling techniques have begun to cluster DA neurons into subtypes based on their molecular signatures. Recent work has shown that molecularly defined DA subtypes within the substantia nigra (SNc) display distinctive anatomic and functional properties, and differential vulnerability in Parkinson's disease (PD). Based on these provocative results, a granular understanding of these putative subtypes and their alterations in PD models, is imperative. We developed an optimized pipeline for single-nuclear RNA sequencing (snRNA-seq) and generated a high-resolution hierarchically organized map revealing 20 molecularly distinct DA neuron subtypes belonging to three main families. We integrated this data with spatial MERFISH technology to map, with high definition, the location of these subtypes in the mouse midbrain, revealing heterogeneity even within neuroanatomical sub-structures. Finally, we demonstrate that in the preclinical LRRK2^{G2019S} knock-in mouse model of PD, subtype organization and proportions are preserved. Transcriptional alterations occur in many subtypes including those localized to the ventral tier SNc, where differential expression is observed in synaptic pathways, which might account for previously described DA release deficits in this model. Our work provides an advancement of current taxonomic schemes of the mouse midbrain DA neuron subtypes, a high-resolution view of their spatial locations, and their alterations in a prodromal mouse model of PD.

Introduction

Midbrain DA neurons are traditionally organized in three main anatomical areas – the ventral tegmental area (VTA), substantia nigra pars compacta (SNc) and retrorubral area (RR)[1], with further sub-divisions therein[2]. To explain the disparate roles of DA in normal behavior and disease, additional heterogeneity of midbrain neurons within these areas has been postulated[3]. Recent studies have indeed revealed diversity within clusters in electrophysiological properties as well as responses during various behavioral paradigms[4-15]. Complementing these studies, recent evidence using single cell classification has opened the possibility that DA neurons can be clustered based on their molecular signatures[16-25]. Early studies have begun to suggest that molecularly distinct DA populations may have distinctive anatomical projection patterns, as well as functionally distinct activity patterns[4, 25-27]. For example, in the SNc, activity in a ventral Anxa1+ population is correlated to acceleration in mice on a treadmill, whereas activity in dorsal Calb1+ or lateral Vglut2+ populations is correlated to deceleration[25]. These results suggest that molecularly defined DA neurons need to be considered as a key variable for electrophysiological and behavioral studies.

Based on these exciting results, and the fact that the functional interrogation of DA neurons is happening at rapid speed often agnostic to DA subpopulations, there is a clear need to define DA neuron subtypes with more granularity. Previous studies using single cell sequencing in mice have been limited by the number of cells analyzed, the quality of cDNA library, as well as bias during isolation, thus providing an incomplete picture of DA neuron subtypes[17, 19-23]. Furthermore, the anatomical distribution of these subtypes in the midbrain remains only partially elucidated. Additionally, given the closely related nature of these subtypes, the stability of these subtypes across pathological conditions remains unclear. In other words, are some of these “subtypes” a representation of cell state rather than cell type?

The unbiased study of the transcriptomic landscape of DA neuron subtypes in the midbrain can inform about the molecular drivers of selective dysfunction of DA neurons in PD. Pathogenic mutations which increase leucine-rich repeat kinase 2 (LRRK2) activity are one of the most common causes of autosomal dominant PD and clinically similar to sporadic cases[28, 29,30]. Additionally, even in patients with idiopathic PD, LRRK2 kinase activity is increased in DA neurons[31], providing justification for studying LRRK2-driven mechanisms in DA neurons. Structural and functional synaptic changes are a recurring theme with LRRK2 mutations[32-34]. Accordingly, impaired DA transmission is observed with several LRRK2 mouse models including knock-in (KI) mouse lines expressing the most common LRRK2 mutation (G2019S) at physiological levels[35-37], suggestive of functional synaptic deficits in dopaminergic circuits in the absence

of overt DA neuron loss. Thus, LRRK2 KI mouse models are valuable for investigating early PD mechanisms in vulnerable DA neurons. However, despite the clinical relevance, the cell-autonomous functions of LRRK2 in nigral DA neurons are largely unknown.

Single-nuclear RNA sequencing (snRNA-seq) offers an avenue to begin to understand downstream transcriptomic alterations within DA neuron subtypes. We have used snRNA-seq to surveil DA neurons, which allows greater acquisition of DA numbers than whole-cell approaches and minimizes isolation bias[25]. Here, we extend previous studies by optimizing our isolation protocol, enabling sequencing of large numbers of DA nuclei. We provide a high-resolution view of DA subtypes and an online portal to interrogate this dataset. We systematically map DA subtype distribution using MERFISH, providing unprecedented spatial resolution. Finally, in a LRRK2^{G2019S} preclinical mouse model of PD, we reveal molecular pathways that are altered in locomotion-relevant DA neuron subtypes.

Results

A large snRNA-seq dataset of midbrain dopaminergic neurons reveals novel molecular subtypes

We performed snRNA-seq from dopaminergic nuclei (defined by Dat-Cre expression) in LRRK2^{G2019S} heterozygous knock-in mutants and control littermates using a similar technique to our previously published dataset[25] but modified to utilize a chip-based microfluidic sorting method which minimizes nuclear stress. This improved protocol enabled increased yield and sample quality (Fig 1A, Fig S1A). Doing so enabled us to generate a dataset with 28,532 profiled cells (after quality control filtering) with a median UMI count of 7750.5 and median of 3056 genes. Downstream clustering resulted in 22 unique clusters (Fig 1B), of which two clusters (16 and 21, 911 total cells) were determined to represent possible dopaminergic/glial doublets based on co-expression of glial markers in these populations and were excluded from downstream analyses (shown in grey in Fig 1B, Fig S1F). The 20 remaining clusters all showed expression of key pan-dopaminergic markers including *Slc18a2* (*Vmat2)*, *Ddc*, *Th*, and *Slc6a3 (Dat)* (Fig 1C). These clusters were observed in both males and females, and across genotypes, in similar proportions (Fig S1B-E). These subtypes showed distinct expression signatures of many previously described markers of DA neuron subtypes (Fig 1D)[17]. However, our increased yield and sample quality also allowed explicit detection of known populations that have eluded most prior single-cell studies, such as DA neurons expressing *Vip*[17, 38], thereby demonstrating the strength of this pipeline in detecting rare DA neuron subtypes. Importantly, all the described subtypes showed high expression of midbrain floorplate markers (Fig 1E), with exception of clusters 0 and 17, which possibly represent *Slc6a3+* neurons in the

supramammillary and/or premamillary regions[39, 40]. Thus, our increased number of subtypes over prior studies represents increased granularity among classic midbrain DA neuron subtypes, rather than detection of non-floorplate derived populations that express DA markers.

Analysis of differentially expressed genes reveals families of subtypes and relationships to prior studies

Given the large number of clusters discovered in our dataset, we first sought to organize these subtypes into groups in an unbiased manner. By creating a cluster dendrogram (Fig 2A), we were able to map an approximation of the relative relatedness of different subtypes, which segregated into three main families based on major branch points. By running differential expression on each node of the cluster dendrogram (i.e., exploring differentially expressed genes (DEGs) between the two arms emerging from a given node), we were then able to create a stepwise expression code that can uniquely define any individual cluster (Fig 2A, G). The earliest division point among our clusters was defined by genes *Prkcd* and *Lef1*, which are highly expressed only in cluster 17 (Log2 fold changes = 6.279 and 7.246 respectively, BH-corrected p values ≈ 0; note, such p values are calculated using default Seurat differential expression functions which treat each cell as an independent sample thus potentially inflating statistical significance). Next, a family of clusters defined by *Gad2*/*Kctd16* (deemed Gad2 family, Log2 fold changes = 5.318 and 4.014 respectively, BH-corrected p values ≈ 0) separated from all other clusters, which expressed significantly higher levels of *Ddc* and *Slc6a3* (Log2 fold changes = 1.419 and 1.643 respectively, BH-corrected p values ≈ 0). The first major branch point between *Ddc*-high/*Slc6a3*-high clusters was then defined by cells expressing *Sox6/Col25a1* (Log2 fold changes = 2.720 and 2.305 respectively, BH-corrected p values ≈ 0) versus those expressing *Calb1/Ndst3* (Log2 fold changes = 2.025 and 3.368 respectively, BH-corrected p values ≈ 0), thus creating the Sox6 family and Calb1 families respectively. This division between *Sox6* and *Calb1* is consistent with our previous work[38], as well as a prior large single cell profiling study that found this division to be the central branch point among DA neurons across several species including humans[41]. Cluster divisions can be visualized by plotting the co-expression of DEGs taken from the branch points labeled in red (Fig 2C-F) on the UMAP plot. Of note, while these branch point markers describe the overall trends of gene expression, these genes are not perfectly distinct or universal among the members of these cluster families. For example, clusters 1, 11 and 13 show expression of both *Sox6* and *Calb1* (Fig 1D). The co-expression of these markers has previously been described in a fraction of mouse DA neurons[38, 42, 43]. Furthermore, because the calculations of differential expression at each node take into account only the descendants of that branch point, markers defining a branch point may also be strongly expressed in clusters outside this comparison. Thus, while the final marker(s) listed for each subtype in Fig 2A are not

unique to that population, following the sequential branches leading to this marker in a stepwise fashion will specifically enrich for this population (Fig 2G). By doing so, we provide a strategy for potential genetic access to any individual DA subtype using intersectional logic gates.

To enable comparisons of DA populations across existing and future literature, we next sought to better define genetic signatures for each of the individual putative subtypes and to provide a common language for discussing these populations. To do so, we utilized two complementary approaches. First, we explored the top differentially expressed genes with positive relative expression in each cluster (Fig 2B). While many of these markers are expressed in more than one population, some (e.g. *Vip*) are almost entirely unique to a single cluster. As a complementary approach, we performed differential expression among clusters within each of our three subtype families, thereby creating a shorthand identity based on the cluster family and top DEGs (Fig 2G). Within the Sox6 family, these clusters are Sox6^{Tmem132d} (listed as cluster 1 in Fig 1 and 2), Sox6^{Arhgap28} (cluster 2), Sox6^{March3} (cluster 3), Sox6^{Tafa1} (cluster 4), Sox6^{Kcnmb2} (cluster 8), and Sox6^{Vcan} (cluster 10); in the Calb1 family, these are Calb1^{Pde11a} (cluster 5), Calb1^{Kctd8} (cluster 6), Calb1^{Ptprt} (cluster 7), Calb1^{Sulf1} (cluster 9), Calb1^{Stac} (cluster 11), Calb1^{Chrm2} (cluster 12), Calb1^{Sox6} (cluster 13), Calb1^{Lpar1} (cluster 14), Calb1^{Ccdc192} (cluster 18), Calb1^{Gipr} (cluster 19); in the Gad2 family, these are Gad2^{Syndig1} (cluster 0), Gad2^{Egfr} (cluster 15), and Gad2^{Ebf2} (cluster 20). Cluster 17 (defined by Lef1) did not have an associated cluster family (Fig 2A). These shorthand identities provide a simple nomenclature for our clusters, which we have utilized throughout the remainder of this paper.

After establishing genetic signatures for each putative subtype, we were next able to correlate our cluster identities to those described in another recent study[25], as well as the consensus subtypes proposed in Poulin et al, 2020 (Fig 2G)[17]. Doing so provides context for discussing our putative DA neuron subtypes with regards to prior literature. For example, cluster Sox6^{Tafa1} was found to be equivalent to a population we previously defined by high expression of *Anxa1*, which holds notably distinct functional and anatomical properties[25]. Several of the subtypes that emerged in our new dataset also appear to be novel divisions within clusters that previously showed evidence of internal heterogeneity based on their cluster stability metrics[25]. For instance, a *Vip*+ cluster (Calb1^{Gipr}) emerged from within a highly variable cluster (defined as #10 in Azcorra & Gaertner et al.[25]), and additional subdivisions emerged from previously described unstable Sox6+ clusters (most notably #3 in [25]) (Fig 2G). This emergence of unique populations, such as Calb1^{Gipr}, is likely due to the increased yield and sample quality of our new dataset which has enabled better detection of small populations and gene markers with lower expression levels. This is further supported by the high cluster stability of the new Calb1^{Gipr} (Fig S2A,B), resolving the instability of the

equivalent cluster in Azcorra & Gaertner et al[25]. We re-clustered our snRNA-seq data into the clusters defined in Azcorra & Gaertner et al. and found our clusters aligned with or further subdivided many of the previously defined clusters (Fig S3A). Lastly, to further corroborate our cluster identities, we used an independent bioinformatic platform, Scanpy, to re-cluster our nuclei (Fig S3B). Comparing the Seurat and Scanpy clusters using a Sankey diagram (Fig S3C) shows a high correlation between the Seurat and Scanpy clusters.

Mapping the spatial location of midbrain DA neurons using MERFISH

Having identified multiple clusters of midbrain DA neurons, we next resolved to determine the spatial distribution of these populations using MERFISH[44], a form of spatial transcriptomics that offers subcellular resolution imaging-based RNA quantification. We assembled a 500 gene panel which included 120 of the top DEGs identified within the snRNA-seq dataset, as well as 380 general markers of neuronal and non-neuronal cell populations (Supp Table 1). Seven coronal slices were collected from adult mouse midbrains and processed for MERFISH (Fig 3A). The coronal slices were selected containing ventral midbrain regions between -2.9mm to -3.9mm from bregma. Cells were segmented using the CellPose algorithm based on DAPI and Poly-T staining and passed through quality control filters to remove cells with too large of a volume (potential doublets) or too few transcripts detected (false cells; Fig S4A-C). This resulted in 429,713 identified cells (Fig 3B). Cells from the seven sections displayed significantly correlated gene expression and minimal batch effects (Fig S4D). Cells were integrated in Seurat using reciprocal principal component analysis (RPCA) algorithm and then clustered, resulting in 40 final clusters comprising distinct classes of neurons, oligodendrocytes, oligodendrocyte precursor cells (OPCs), astrocytes, microglia and meningeal cells (Fig 3B). These clusters could be attributed to 25 functional groups based on the expression of cell type markers from Supp. Table 1 and their spatial distribution. Glia-associated genes were mostly contained within their expected cell types: oligodendrocytes (*Mog*, *Cnp*), OPC (*Pgfra*, *Olig2*), astrocytes (*Aqp4*, *Gfap*), and microglia (*Aif1*, *Tmem119*). However, we noted the unexpected distribution of the oligodendrocyte-specific transcript *Mbp* in non-glial cells which we believe might indicate the presence of this mRNA in glial-processes that overlap non-glial cells[45]. Within the neuronal class we identified 16 excitatory clusters (presence of *Slc17a6* or *Slc17a7*), 7 inhibitory clusters (presence of *Slc32a1*, *Gad1* or *Gad2*), one serotonin neuron cluster and one DA neuron cluster (Fig 3B-D, S4G). Although our gene panel was designed to identify neuronal heterogeneity of the midbrain, it had sufficient resolution to delineate 7 excitatory neuron classes present in distinct layers of the cortex and 6 cortical interneuron classes (Fig S4E,F). Cortical excitatory neurons were absent from subcortical structures such as the thalamus,

hypothalamus, midbrain and hindbrain. This fundamental dichotomy between cortical and subcortical neurons in the adult mouse brain is supported by recent spatial transcriptomic studies[46-48].

A population of DA neurons was identified among our 40 clusters based on the expression *Th*, *Slc6a3* and *Slc18a2* (Fig 3C). The location of the dopaminergic cluster (blue in Fig 3D) considerably overlaps with *Th* expression (Fig 3E). This cluster comprised 4,532 cells, was entirely located in the midbrain, and was also enriched for *Ddc*, *Dlk1*, and *Drd2* (not shown). Within this population, we detected expression of genes associated with glutamatergic and GABAergic neurotransmission, in line with previous reports of multi-lingual characteristics of DA neurons[49-54], for instance the glutamatergic marker *Slc17a6* (Vglut2) and GABAergic markers *Gad2* and *Slc32a1* (Vgat) (Fig S4G). Markers for cholinergic (*Chat*) and serotonergic (*Tph2*) neurotransmission were not detected in putative DA neurons and the glutamatergic transporters *Slc17a7* (Vglut1) and *Slc17a8* (Vglut3) were also absent. We next wanted to characterize the neuroanatomical location of this DA neuron cluster. We manually delineated neuroanatomical boundaries on each of the slices for seven midbrain regions known to contain most dopamine neurons including SNc, VTA, retrorubral area (RR), caudal linear nucleus (CLi), interfascicular nucleus (IF), rostral linear nucleus (RLi), and periaqueductal grey (PAG). These boundaries were delineated based on cell density (dapi), white matter landmarks, and *Th* transcript distribution using the Allen Brain Atlas as a reference (Fig S5A). We found that cells in our DA neuron cluster were located in the following locations: 35.2% in the SNc, 19.8% in the VTA, 7.8% in the CLi, 15.3% in the RR, 2.0% in the IF, 0.6% in the RLi, 4% in the PAG and 15.3% located elsewhere (Fig S5B). Finally, to explore the diversity of DA neurons identified by MERFISH, we subclustered the 4,532 cells to resolve 12 distinct subgroups (Fig S5C-F). For clarity, we designated the clusters identified solely with MERFISH transcript levels as MER-0 to 11. Many of the clusters were characterized by increased expression of known subtype markers such as *Calb1* (MER-6, MER-8 and MER-10), *Aldh1a1* (MER-5), *Ndnf* (MER-0), *Gad2* (MER-7 and MER-9) and *Slc17a6* (MER-2).

Cell label transfer allows the mapping of snRNA-seq clusters

To provide the spatial distribution of the clusters identified using snRNA-seq, we sought to project the MERFISH dataset onto the snRNA-seq UMAP space. We first compared the expression level of the 500 genes utilized in our panel to select features for integration with comparable expression levels. We selected 281 genes to be included based on similarity in the average counts per cell between the datasets. Some excluded genes potentially reflected faulty probes or improper assignment of RNA transcripts to segmented cells. For instance, *Sox6* transcripts were not detected across the entire brain (data not shown). We successfully projected the MERFISH dataset onto the snRNA-seq UMAP space and selected for 2,297

dopamine neurons with a cell similarity score > 0.5 (mean value, 0.538), increasing confidence in the correct assignment to snRNA-seq defined clusters. Almost all clusters were represented by more than 20 cells in our MERFISH dataset (Fig. 4A and S6A). Clusters Calb1^{Ccdc192} and Calb1^{Gipr} were not resolved in this label transfer either due to their relatively smaller number or location in an under-sampled region. Indeed, the *Vip*-expressing Calb1^{Gipr} has been shown to be located in the caudal midbrain (PAG/DR)[38], a region not well covered in our MERFISH experiment. Comparing snRNA-seq clusters with the ones obtained by clustering the MERFISH dataset alone showed a strong correspondence (Fig S6B). Half of the snRNA-seq clusters were predominantly comprised of single MERFISH clusters (10/20) and 25% (5/20) were comprised of 2 MERFISH clusters, demonstrating that the integration with the snRNA-seq dataset allowed us to further refine our MERFISH-based classification (Fig S6B). We then imputed expression data from the snRNA-seq dataset onto the MERFISH cells to better visualize spatial expression levels. The imputed gene expression is extrapolated from anchors established from pairwise correspondences of cell expression levels between MERFISH and snRNA-Seq datasets. The imputed data across the genes used for data integration correlated strongly with MERFISH transcripts, particularly for *Calb1*, *Gad2* and *Aldh1a1* genes (Fig S6C-F and S7). Altogether, the dataset integration allowed us to map the neuroanatomical location of 18 of the clusters identified by snRNA-seq as well as impute expression of the whole transcriptome.

Locating DA neuron clusters within the midbrain

Clusters were sorted into Gad2, Sox6 and Calb1 families based on our hierarchical dendrogram (Fig 4A; Gad2 n=124 cells, Sox6 n=887 cells, Calb1 n=590 cells). Imputed *Gad2*, *Sox6* and *Calb1* gene expression within DA neurons showed a distinct medio-lateral and rostro-caudal distribution (Fig 4C-E). The spatial distribution of imputed *Sox6* and *Calb1* transcripts within DA neurons correlated well with previous reports[38,43,55] whereas the distribution of the Gad2 family is predominantly observed in the rostral linear/posterior hypothalamus region and CLi, and to a lesser extent in the VTA (Fig 4B, S8). Each family was overrepresented in a neuroanatomically-defined area; for instance, the SNc is composed of 78.4% of neurons from the Sox6 family (Fig S8A), consistent with previous reports[43]. However, each family was also found across multiple neuroanatomical regions (Fig S8A). We also observed spatial heterogeneity within families. For example, one branch of the Calb1 family dendrogram (Fig 4, label in shades of yellow-orange defined by higher *Lepr* expression) was enriched in the dorsal VTA, whereas the other Calb1 branch (Fig 4, label in shades of red defined by higher *Ntm* expression) populated mostly the ventromedial VTA. Similar observations could be made for the Sox6 family, where subtypes displayed biased distributions. This led us

to explore the distribution of each individual cluster and representative markers toward the goal of generating a granular map of the midbrain DA system.

The hierarchical clustering of the Sox6 family reveals two major branches (Fig 2A and 4A): 1) Branch 1 defined by higher levels of *Slc44a5* and composed of subtypes Sox6^{Tmem132d} and Sox6^{March3} and 2) Branch 2 defined by *Cntnap5* and composed of subtypes Sox6^{Tafa1}, Sox6^{Arhgap28}, Sox6^{Kcnmb2} and Sox6^{Vcan}. We observed an important discrepancy between these two branches, with Branch 1 being located more dorsal and medial than Branch 2 (Fig 5A). Interestingly, this dorsoventral organization breaks down in the caudal SNc and RR where neurons from both branches are mostly intermixed. Only minor differences in location were observed between clusters of Branch 1, maybe reflecting their molecular similarities. Of note, whereas Sox6^{Tmem132d} is predominant in the rostral SNc and VTA for this branch, neurons of Sox6^{March3} are more numerous in the RR (Fig 5B, Fig S8B). Within Branch 2, the molecular identity correlated with the medio-lateral distribution. For instance, neurons of cluster Sox6^{Arhgap28} were found to be more lateral in the SNc compared to neurons of Sox6^{Tafa1}, Sox6^{Kcnmb2} or Sox6^{Vcan}. This medio-lateral distribution of neurons of the SNc ventral tier might reflect the differential topographical distribution of nigrostriatal projections previously reported[26, 27.]

The Calb1 family was the most molecularly diverse which was also reflected in its localization (Fig 6). Two major branches of Calb1 DA neurons were identified: 1) Branch 1 (Fig 6A-B) containing subtypes Calb1^{Ptprt}, Calb1^{Sulf1}, and Calb1^{Sox6} and 2) Branch 2 (Fig 6C-D) containing subtypes Calb1^{Chrm2}, Calb1^{Kctd8}, Calb1^{Stac}, Calb1^{Pde11a} and Calb1^{Lpar1}. Broadly speaking, Calb1 subtypes were predominantly localized to the VTA and CLi with some neurons showing localization to the dorsal SNc, SNpl, or RR (Fig S8). Within the VTA, Branch 1 Calb1 neurons tended to localize to the dorsal VTA (Fig 6A-B) while Branch 2 Calb1 neurons showed distinct ventromedial localization (Fig 6C-D). In more caudal sections the difference was even more pronounced, although intermingling was always observed. Within Branch 1, subtypes Calb1^{Ptprt} and Calb1^{Sox6} were largely absent from the SNc dorsal tier and RR, whereas subtype Calb1^{Sulf1} was present. Some notable distinctions were observed in both groups, specifically subtypes Calb1^{Sulf1} and Calb1^{Lpar1} which were each observed in significant numbers in the dorsolateral SNc/SNpl region. Of these, only Calb1^{Lpar1} expressed significant *Slc17a6* (data not shown), making it a likely candidate for the SNpl Vglut2+ DA neuron population found to innervate the tail of the striatum[27]. Indeed, these laterally located neurons have been genetically targeted with a Slc17a6-Cre driver, and display deceleration-correlated responses, and robust responses to physically aversive stimuli[25, 27].

Within the Gad2 family we were able to resolve two subtypes. Subtypes Gad2^{Egfr} and Gad2^{Ebf2} had very similar spatial distributions and were prominent in the posterior IF, RLi, and CLi (Fig 7, Fig S8). These regions have been associated with small DA neurons of lower *Th* and *Slc6a3* levels. Subtype Gad2^{Egfr} neurons were more plentiful than cluster Gad2^{Ebf2}, with the latter being absent in more caudal sections. All Gad2 subtypes expressed *Slc32a1*, opening the possibility that these neurons co-release GABA. Further, they also expressed *Slc17a6*, revealing potentially multilingual DA neuron subtypes.

LRRK2^{G2019S} KI results in pan-dopaminergic gene expression changes consistent with deficits observed in this model, without altering DA subtype proportions

With cluster identities and locations defined, we next sought to explore the downstream gene expression characteristics of DA neurons in Lrrk2 mutants. The role of Lrrk2 in PD pathophysiology remains unclear[56], and Lrrk2 mutations are often proposed as being an indirect source of dysfunction for the DA system, such as through its high expression in glial or striatal cells[34, 57-60]. However, previous studies have shown alterations in DA signaling or gene expression even when the mutant LRRK2 was expressed specifically in DA neurons[61, 62]. Thus, as a first step in demonstrating the possibility that mutant Lrrk2 in DA neurons themselves can contribute to dysfunction, we mapped the expression of Lrrk2 across all DA neuron subtypes (Fig 8A). We found detectable expression across most subtypes, with the notable exception of Gad2^{Egfr} neurons.

Given the distinctive properties of different DA neuron subtypes, we next sought to test if phenotypic changes in Lrrk2 mutant mice are driven by changes in the relative proportions of said subtypes. While previous studies have shown no overt DA neuron loss, it is possible that there was a selective reduction of a subtype(s) that would be undetectable by previous methods. We first graphed the proportions of each subtype within our control and mutant datasets (Fig 8B), and found remarkable similarity across these samples, consistent with previous reports showing no DA neuron loss in these mice[36]. To further assess for changes in cell state among clusters across conditions, we next compared the similarity of each cluster to all other clusters across conditions (Fig 8C). MetaNeighbor analysis revealed that the most similar population to each wildtype cluster was its corresponding subtype in the Lrrk2 mutant samples. This suggests that the general subtype organization is largely impervious to LRRK2^{G2019S} perturbation, and that our clustering scheme is likely more reflective of cell type rather than cell state.

To assess pan-dopaminergic gene expression changes as a function of Lrrk2 genotype, we next looked at globally differentially expressed genes across our mutant and control samples. Calculating

differentially expressed genes across all clusters revealed a large number of significantly enriched or diminished genes as a function of Lrrk2 mutation (1646 genes in total with BH corrected p-value <0.05, Wilcoxon rank sum test). The results of this differential expression analysis are plotted in Fig S9A, with full results and statistics available in Extended Data. Notably, most of these changes appear to be small in magnitude; among FDR-significant genes, only 51 showed increased expression and 7 showed decreased expression with log2 fold changes greater than 0.5 (Fig S9A). Two genes with high enrichment in Lrrk2 mutants, Mir124a-1hg and AC149090.1 were both of particular interest due to their potential relevance to PD pathogenesis. AC149090.1 is orthologous to the human PISD gene, which is involved in autophagy and implicated in mitochondrial dysfunction[63-65]. This gene has been proposed as a sensitive marker of biological aging in the brain[63, 66], an interesting finding given that age is the single largest risk factor for PD. Also of note, Mir124a-1hg is a host gene (though not the exclusive gene) for miR-124, a microRNA with purported neuroprotective effects that has been found to be downregulated in PD patients or some PD mouse models[67-70]. While the opposite effect was seen here (i.e., an upregulation of this gene in our model), one might speculate that miR-124 is initially increased during toxic metabolic insults as a compensatory response, given its proposed neuroprotective effects. In our prodromal model without observable degeneration, this could thus represent an early sign of cell stress. Conversely, in PD patients or overtly degenerative models, a lack of compensatory miR-124 or fulminant cell death among vulnerable cells could result in an observed decrease in miR-124 expression.

We next performed pathway enrichment analyses using Gene Set Enrichment Analysis (GSEA) and gene ontology (GO) to compare our Lrrk2 mutant samples to our controls. Using GSEA, we found 18 gene sets with significant differential expression (FDR adjusted p-values < 0.05). The top 13 of these enriched pathways (those with NES scores > 1.8) are shown in Fig 8D. While a broad range of biological domains are found among these results, many of the top enriched pathways were related to energy production via mitochondrial oxidative respiration, remarkably similar to prior comparisons between vulnerable and resistant DA neuron populations[43]. Mitochondrial energy production pathways, and specifically the extraordinary bioenergetic burdens of DA neurons, have been heavily implicated in PD vulnerability[71, 72]. Thus, the enrichment of these pathways in Lrrk2 mutants is particularly intriguing as it may reflect that the structural and functional alterations reported with LRRK2 mutations in patients[73] and preclinical models[74, 75] further predispose these cells to dysfunction and ultimately degeneration. Utilizing GO, we again found many pathways displaying significant differential expression between Lrrk2 mutants and controls. Among these, the top pathways are all associated with synaptic organization and functions (Fig 8E). This was particularly intriguing given that pre-synaptic dysfunction in Lrrk2 mutants such as endocytosis[76-78] or

axonal cargo trafficking impairments[79] could underpin the deficits in DA signaling observed in these mice[80, 81]. Of note, decreased endocytosis has been observed in DA neurons but not cortical or hippocampal LRRK2^{G2019S} primary neurons[82]. These results, although *in vitro*, suggest cell type specific synaptic effects on DA neurons, consistent with their vulnerability in LRRK2-related PD.

Comparing individual subtypes across conditions allows insights into subtype-specific dysfunction

Given that global gene expression differences across our conditions appear to hold relevance for putative mechanisms of PD pathogenesis, we next sought to address the question of whether gene expression within any given subtype is specifically altered in our PD model. Although Lrrk2 mRNA seemed uniform across subtypes, downstream effects could be subtype-specific secondary to differential Lrrk2 kinase activity, distinct kinase targets, or superimposed differences in other properties of each subtype that culminate in distinct downstream transcriptional alterations. Given the proposed vulnerability of Sox6+ DA neurons and relative resilience of Calb1+ neurons in PD[3, 41, 43], we first compared these two cluster families across Lrrk2 mutants versus control mice to see if any changes in DA neurons are specific to vulnerable or resistant cell types. We found 729 DEGs among the Sox6 family, and 679 among the Calb1 family (BH corrected P value <0.05). Among those genes, we found 327 DEGs were shared between the Sox6 and Calb1 families (Fig S9B). The full list of DEGs and statistics is available in Extended Data. Notably, AC149090.1 was highly significant in both cohorts, implying that the pan-DA differential expression of this gene was not driven by a particular subset of DA neurons. To further compare changes within cluster families, we once again performed GSEA and GO but on groups of clusters in isolation (Fig 9A-B). In line with our previous results, GO revealed pathways that largely corresponded to synaptic and axonal function, but with higher significance of results in the Sox6 family than in Calb1 or Gad2 families (Fig 9A). Within the Sox6 family, GSEA analysis showed an upregulation of energy production and metabolism pathways similar to those observed at the global level, as well as a downregulation of intercellular communication pathways (Fig 9B, top 5 upregulated and downregulated pathways shown for each family). Finally, to extend the granularity of our analyses, we applied the same GO pipeline to two individual clusters, Sox6^{Tafa1} and Calb1^{Stac}, which have the highest expression of *Anxa1* in the SNc and VTA respectively (Fig S10A). The Sox6^{Tafa1} SNc subtype was of particular interest to us given recent results showing that *Anxa1*+ ventral tier SNc neuronal activity is selectively correlated with acceleration in mice running on a treadmill, leading to the hypothesis that degeneration of these neurons may contribute to motor deficits seen in PD[25]. The Calb1^{Stac} VTA subtype was chosen as a comparator due to its similar expression profile, including key markers *Anxa1* and *Aldh1a1*, but within the relatively resistant VTA. Similar pathways were

observed in each subtype, including several pathways once again corresponding to synaptic function, however with substantially higher significance values observed in the Sox6^{Tafa1} cluster. Of note, while this enrichment is intriguing, the associations to such pathways cannot be interpreted as unique to this population, as technical limitations confound direct comparisons of DEGs among individual clusters due to variable cluster sizes and internal heterogeneity (see Methods for more details).

Next, we compared the synaptic compartment and biological functions of DEGs (BH adjusted p values < 0.05) in Lrrk2^{G2019S} mutants versus controls using SynGO, a database for systematic annotation of synaptic genes[83]. We focused primarily on the Sox6 and Calb1 families, but also on cluster Sox6^{Tafa1} given that this population (SNc neurons with highest expression of *Anxa1*) is specifically linked to movement[25]. Of the DEGs in the Sox6 family, Calb1 family, and Sox6^{Tafa1} cluster, 24.82%, 20.19% and 29.41% of these genes had synaptic localizations, respectively, with more significant synaptic associations among populations localized to ventral tier SNc (Fig 9C-E, Fig S9C). Further compartment categorization revealed that among these synaptic DEGs annotated by SynGO, 45.83% of Sox6 family (11.38% of total DEGs), 64.0% of Sox6^{Tafa1} (18.82% of total DEGs) and 48.82% of Calb1 family (9.86% of total DEGs) DEGs have presynaptic localizations, raising the possibility of enhanced disruption of presynaptic function in acceleration-correlated DA neurons. From these presynaptic DEGs, 20/77 in the Sox6 family, 5/16 in Sox6^{Tafa1}, and 16/62 in the Calb1 family have specific annotations for active zones, the primary sites of evoked DA release. Annotations based on the biological functions showed that 20.83%, 28.0% and 15.74% of Sox6, Sox6^{Tafa1} and Calb1 DEGs, respectively, are associated with processes in the presynapse (Fig S9D). Overall, greater proportions of DEGs are associated with presynaptic locations in cells from vulnerable DA neurons (Sox6 family, and in particular, Sox6^{Tafa1}), compared to less vulnerable ones (Calb1 family).

Given the differential effects of Lrrk2 mutation across subtypes, we also sought to map the association of gene expression in our dataset with co-expression of PD-associated risk loci identified by GWAS[84] using a polygenic enrichment score method, scDRS[85]. We first calculated a PD GWAS risk score for each individual cell, which revealed clear bias for higher risk scores among clusters that localize to the ventral SNc (Fig 9F). We next sought to extend the cell-level calculations of scDRS to explore the relative PD risk among different subtypes. To do so, we calculated the mean PD GWAS risk score for each cluster and utilized bootstrapping to generate a corresponding 95% confidence intervals for these mean values, allowing us to determine if the observed means were significantly different from the null. We confirmed significant risk associations in several clusters, particularly those in the Sox6 family (Fig 9G). Interestingly, the Gad2 family of clusters all showed significant negative associations with PD risk loci (Fig 9G). Of note,

the highest scDRS scores were in clusters Sox6^{Tafa1} and Sox6^{Vcan}, the only two SNc clusters that express *Anxa1*. By calculating the average scores for each cluster family as a whole, as well as these two Anxa1-expressing SNc clusters, we found that only the Sox6 family (and Anxa1-expressing SNc clusters within it) show significant associations with PD risk loci at a population level (Fig 9H).

Finally, due to the large number of potential comparisons to be made across clusters, cluster families, and genotypes, we have also developed an online tool that allows for exploration of our datasets, which we have termed Dopabase (URL: dopabase.org). Utilizing this tool, users have the ability to access many analysis and plotting functions to continue to explore populations of interest.

Discussion

Our work provides five key advances. 1. We provide a large snRNAseq dataset that provides high granularity to our taxonomic schemes. 2. We provide a user-friendly portal to query this dataset. 3. We plot the spatial location of most identified DA neuron subtypes, showing heterogeneity even within sub-domains in traditional anatomical areas. 4. We demonstrate that cluster proportions do not change in LRRK2^{G2019S} mice, although gene expression changes are observed, most notably in mitochondrial bioenergetics and synapse organization pathways in vulnerable DA subtypes. 5. PD GWAS risk is most prominent in Anxa1 SNc neurons.

We demonstrate the presence of 20 molecularly distinct clusters of midbrain DA neurons. These are hierarchically organized, the dendrogram being split into three levels, which we refer to as family, branch and subtype. Family (Level 1) distinctions are driven by the expression of *Gad2/Kctd16* vs *Slc6a3/Ddc* high expressing cells. *Slc6a3/Ddc* high cells are further divided by *Sox6/Col25a1* and *Calb1/Ndst3* (Level 2) among other genes. It is likely that these fundamental divisions are, at least in part, set early in development in the mesodiencephalic floor plate, where *Sox6* demarcates progenitors with a substantially higher probability of ventral SNc and lateral VTA fate, and conversely, *Sox6*- progenitors have a higher probability of a *Calb1*+ fate[43, 55]. The Sox6 family is split into two branches with six subtypes, the Calb1 family is split into two branches with ten subtypes, whereas the Gad2 family can be further split into two subtypes. This scheme represents a more granular taxonomy than previous studies[19-21, 23, 25, 38, 41], and bears some resemblance to a recent study[86]. Key congruencies between these studies include: 1) parsing midbrain DA system into broad groups of subtypes based on expression of *Sox6* and *Calb1*, and to a lesser extent *Gad2*, populations, 2) greater diversity observed in *Calb1*+ clusters, 3) the identification of some

consensus and homogenous subtypes. Key discrepancies include the extent to which each cluster is further subdivided and the final number and molecular signature of DA neuron subtypes. In the latter study[86], they refer to levels as territories and neighborhoods. For simplicity of communication, we refer to the lowest level of our dendrogram of molecularly defined DA neuron clusters, as subtypes. However, we acknowledge that because of the common developmental origin from the floor plate[87-90], their high relatedness, the blurriness of boundaries between clusters, and the polythetic nature of clustering, referring to these groups as neighborhoods is an alternative and reasonable possibility. Notwithstanding these considerations, defining markers of some of these clusters and generating Cre/Flp lines, may be useful to at very least enrich for DA neuron populations with distinctive properties as has recently been shown for the *Anxa1*+ SNc neurons, which at a population level are selectively correlated with acceleration, while reward responses are almost entirely absent[25].

Our spatial MERFISH analysis provides the most granular representation of DA subtype locations to date. This resolution was obtained by the integration of snRNA-Seq and MERFISH datasets. Our data complement current spatial transcriptomic datasets[46-48]. For instance, using MERFISH, Zishen et al. identified 10 clusters in the VTA (A10), 9 clusters in the SNc (A9), and 2 clusters in the RR although little detail on the molecular identity or location of these subtypes is provided. In addition, Langlieb et al [48], provide a molecular atlas of the adult mouse brain in which they identified 13 clusters of midbrain DA neurons and located these using Slide-Seq. By contrast, Salmani et al.[86] used established markers to locate seven DA neuron territories and 16 neighborhoods. They report a distribution of *Gad2*+ cells in the midline including CLi, RLi and IF, which bears similarity with our Gad2 family. They also *Pdia5+/Calb1+* DA neurons in the pars lateralis, likely representing an overlapping population with our Calb1^{Postn} and Calb1^{Sulf1}. Within the SNc, our data fits also well with the location of DA neurons that could be labeled with Anxa1-Cre, although here we show that in our Seurat based clustering, Anxa1 expression is detected in Sox6^{Tafa1} and to a lesser extent in Sox6^{Vcan} cells. Taken together, with our increasing appreciation of DA neuron heterogeneity, a consensus is bound to arise on both the identity and location of DA neuron subtypes. Integration of all these emerging datasets will be necessary to achieve a comprehensive DA neuron taxonomy that will ultimately encompass all morphological, connectional, physiological, molecular, and neuroanatomical parameters.

A key finding of this study is the heterogeneity observed even within sub-domains of the traditional anatomical clusters. An example of this is the lateral SNc/SNpl. Current models depict the SN heterogeneity as a mediolateral gradient[91]. Our data substantially refines this by showing that even within

the lateral SNc/SNpl region, there are several distinct subtypes. One interesting finding is that two of the lateral subtypes from distinct families (Sox6 and Calb1), show some common gene expression signatures. For example, *Asic2* or *Ankfn1* are highly expressed in Calb1^{Lpar1} and Sox6^{Arhgap28} subtype, both of which are laterally located. One interpretation of this is that secondary gene expression programs may have been superimposed upon the early developmental sub-divisions, for instance during circuit assembly. A second example of heterogeneity within sub-domains is in the dorsolateral VTA. Again, we find enormous heterogeneity in this region, which is unaccounted for in previous models. It is likely that neurons in this region will have vastly distinct anatomical and functional properties, as exemplified by the complex intermixing of neurons with diverging axonal projections[7]. Careful intersectional genetic interrogation methods will need to be developed to distinguish these populations.

A strength of our study is that it utilizes advantages of each transcriptomic approach, the deep molecular profiling of individual cells using snRNA-Seq and the spatial resolution of MERFISH. For instance, we relied on gene expression imputation to ascribe expression level to genes not covered/detected in our MERFISH probe panel. Gene imputation as described by Stuart et al.[92] has been used in several recent studies integrating spatial and transcriptomic data[46, 47]. It relies on identifying anchors that enable projection of MERFISH data onto the UMAP space of a snRNA-Seq dataset and then uses neighboring cells to extrapolate the expression of genes not included in our probe panel. This approach was used to impute Sox6 expression, which accurately reflects what has been reported in prior immunofluorescence and in situ hybridization studies[11, 27, 38, 43, 55]. Moreover, imputed gene expression levels correlated strongly with MERFISH detected transcript for most genes further supporting our approach (Fig S6 and S7). Nevertheless, dataset integration has limitations that should be considered. First, imputed gene expression relies on the ability to identify reliable anchors linking the snRNA-Seq and MERFISH datasets. These anchors are determined in part by the choice of genes included on probe panels and thus could indirectly influence the reliability of imputed gene expression. Secondly, gene counts per cell in MERFISH are determined via segmentation of images, which is susceptible to artifacts and bias from centrally versus peripherally localized gene transcripts. In summary, although limitations are present in multi-modal transcriptomic analyses, merging these two approaches provided a molecular and spatial map of the DA system that could not have been resolved by either method alone.

We also attempted to address whether our clustering schemes are more reflective of cell types or cell states. We show that the proportions of subtypes are largely consistent between controls and mutants, and that mutant subtypes are most closely related to their control counterpart. This suggests that our

taxonomic scheme is agnostic to a mild perturbation such as LRRK2^{G2019S}, suggesting that our clusters are reflective of cell types, rather than cell states. It is possible that with more severe perturbations, such as a toxin lesion, more substantial alterations of taxonomic schemes are observed[86, 93]. However, we expect that for mild insults, day to day behavioral changes, or pharmacological paradigms, our clusters will be resistant to changes, although individual gene levels may vary. Nonetheless, we cannot definitively confirm that a given DA neuron cannot convert from one subtype to another. Ultimately, alternative approaches such as detailed fate mapping of clusters or RNAseq-based trajectory analyses with greater numbers of sampled cells could be used to resolve this question.

Our work describes a rich heterogeneity of DA neurons in the murine SNc that might serve as a reference for understanding patterns of degeneration in PD. The majority of SNc neurons belong to the Sox6 family, but some are from the Calb1 family. Several studies on post-mortem human PD brain have demonstrated a relative resilience of Calb1+ DA neurons[43, 94, 95]. Calb1+ neurons in mice SNc and SNpl project most densely to the dorsomedial striatum (DMS) and tail of striatum (TS) respectively[3, 27]. This could match the patterns of axon preservation observed in post-mortem PD brain, which show relative sparing of medial, ventral and caudal regions of the caudate-putamen[96, 97]. Indeed it is plausible that patterns of degeneration could be associated with distinct clinical presentations of PD – a recent study suggests that patients with relatively spared caudate projections have a higher probability of tremor[98]. Additionally, we postulate that spared DA neurons projections may be a cellular substrate mediating the side-effects of levodopa particularly at high doses. In line with this, high DA release in the TS is associated with hallucination like behaviors in mice[99].

Given the relevance of LRRK2 to genetic and idiopathic PD, we examined the effects of LRRK2 on DA neuron gene expression to start appreciating the molecular pathophysiology of relevant subpopulations in PD. We relied on a KI LRRK2 model, expressing the G2019S pathogenic mutation due to the highest prevalence of this mutation over other mutations. Although these mice do not exhibit DA neuron loss, they do have nigrostriatal synaptic alterations within a physiological range[35]. Our transcriptomic interrogation has three main takeaways. First, the *Lrrk2* gene is expressed in most DA neurons in mice; in humans *LRRK2* transcript is enriched in the SOX6 population[41]. While previous work has shown LRRK2 protein expression in midbrain DA neurons[60, 100], its relative expression across DA neuron subtypes in mice requires further investigation. Thus, transcriptomic alterations observed within DA neurons are potentially cell-autonomous effects, albeit indirect, since LRRK2 functions as a protein kinase with phosphoregulation primarily affected. Second, while individual genes showed only modest

differences, pathway analysis showed a strong dysregulation of synaptic pathways. This is conceptually aligned with prior work describing that LRRK2 phosphorylates various key presynaptic targets (not necessarily in DA axons) associated with SV cycle[101]. Impairments of these processes at the nigrostriatal terminals may contribute to the DA release deficits reported in LRRK2 mutants[35-37]. Third, alterations in pathways associated with oxidative phosphorylation and energy production are in line with the described LRRK2 impact on mitochondrial structure and mitophagy in the LRRK2^{G2019S} KI mice. As DA transmission changes and mitochondrial abnormalities are the two main and consistent phenotypes with the KI mouse model, our transcriptomic findings indicate that functional changes are at least partly linked with gene expression changes.

Taken together, our work complements perfectly the recent taxonomy of mouse[46-48, 86] and human brain cell types[102]. Focusing our efforts on the DA system allowed us to reveal a comprehensive transcriptomic signature of midbrain DA neurons as well as ascribe with precision their spatial location. Moreover, we provide a web resource to explore this dataset through multiple angles. We also uncovered cell type specific transcriptomic changes in a prodromal model of PD revealing the molecular consequences of higher LRRK2 activity. Disentangling the molecular diversity of DA neurons, now defining 20 transcriptomic subtypes, raises important questions. For instance, what are the functional implications of this diversity and are these transcriptomic differences reflected in other cellular features? Decades of experiments might be needed to answer such questions, however, by defining genetic anchor points unique to each subtype, we have provided an avenue to target these populations in the mouse.

Methods

snRNASeq

Animals

All animals used in this study were maintained and cared following protocols approved by the Northwestern Animal Care and Use Committee. Cre mouse lines were maintained heterozygous by breeding to wild-type C57BL6 mice (RRID:IMSR_CRL:027), with the exception of DAT-IRES-Cre (RRID:IMSR_JAX:027178, referred to as DAT-Cre in Fig 1), CAG-Sun1/sfGFP (RRID:IMSR_JAX:021039, referred to as RC-LSL-Sun1/GFP in Fig 1), which was maintained homozygous and crossed to Lrrk2^{G2019S} mice (RRID:IMSR_JAX:030961) only for generation of experimental mice for nuclei isolation. Both males and females were used in equal number for all RNAseq experiments.

Sample Preparation

To isolate nuclei for snRNA-seq library generation, n=8 6-month old DAT-IRES-CRE (RRID:IMSR_JAX:027178), CAG-Sun1/sfGFP (RRID:IMSR_JAX:021039), Lrrk2$^{G2019S+/wt}$ or littermate Lrrk2$^{wt/wt}$ mice (2 female control, 2 male control, 2 female mutant, 2 male mutant) were sacrificed and rapidly decapitated for extraction of brain tissue as previously described for isolation of GFP+ (i.e. dopaminergic) nuclei. A total of 4 independent samples (2 control, 2 mutants; equal sexes per sample) were isolated using n=2 pooled mice in each. Each sample was processed in its own GEM well to produce 4 distinct libraries that were subsequently sequenced and analyzed as described below. A 2mm thick block of ventral midbrain tissue was dissected out from each mouse and collected for isolation. Tissue was dounce homogenized in a nuclear extraction buffer (10mM Tris, 146mM NaCl, 1mM $CaCl_2$, 21mM $MgCl_2$, 0.1% NP-40, 40u/mL Protector RNAse inhibitor (Roche 3335399001). Dounce homogenizer was washed with 4mL of a washing buffer (10mM Tris, 146mM NaCl, 1mM $CaCl_2$, 21mM MgCl2, 0.01% BSA, 40U/mL Protector RNAse inhibitor) and filtered through a 30uM cell strainer. After three rounds of washing by centrifugation at 500g for 5 minutes, nuclei pellets were resuspended resuspension buffer (10mM Tris, 146mM NaCl, 1mM CaCl2, 21mM $MgCl_2$, 2% BSA, 0.02% Tween-20) and filtered through a 20uM strainer. This nuclei suspension was loaded onto a MACSQuant Tyto HS chip and diluted with 1x PBS. Nuclei were sorted using gates set for isolation of GFP+ singlet nuclei. Sorted nuclei were subsequently used for preparation of four 10X Genomics Chromium libraries. Protocol can be found at dx.doi.org/10.17504/protocols.io.14egn6wryl5d/v1 .

Library preparations were performed by the Northwestern University NUSeq Core Facility. Nuclei number and viability were first analyzed using Nexcelom Cellometer Auto2000 with AOPI fluorescent staining method. Sixteen thousand nuclei were loaded into the Chromium Controller (10X Genomics, PN-120223) on a Chromium Next GEM Chip G (10X Genomics, PN-1000120), and processed to generate single nucleus gel beads in the emulsion (GEM) according to the manufacturer's protocol. The cDNA and library were generated using the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (10X Genomics, PN-1000286) and Dual Index Kit TT Set A (10X Genomics, PN-1000215) according to the manufacturer's manual with following modification: PCR cycle used for cDNA generation was 16 and the resulting PCR products was size-selected using 0.8X SPRI beads instead of 0.6X SPRI beads as stated in protocol. Quality control for constructed library was performed by Agilent Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, 5067-4626) and Qubit DNA HS assay kit for qualitative and quantitative analysis, respectively.

The multiplexed libraries were pooled and sequenced on an Illumina Novaseq6000 sequencer with paired-end 50 kits using the following read length: 28 bp Read1 for cell barcode and UMI and 91 bp Read2 for transcript. Raw sequence reads were then demultiplexed and transcript reads were aligned to mm10 genome using CellRanger (v7.0.1 – RRID:SCR_017344 – https://www.10xgenomics.com/support/software/cell-ranger/latest). snRNA datasets can be found at GEO accession GSE271781.

Data curation and analysis

Analysis was performed using R (RRID:SCR_001905). Outputs from CellRanger were read into Seurat (version 5.0.1; RRID:SCR_007322) using the Read10X command for each sample. Numbers of UMIs, features and ribosomal reads and mitochondrial reads were plotted for each dataset and used to determine cutoffs for quality control pre-filtering of each sample. The following QC filtering commands were used in R: control1 <- subset(control1, subset = nFeature_RNA > 1200 & nFeature_RNA < 7800 & percent.mt < 0.5 & nCount_RNA < 29000 & percent.ribo < 0.5); lrrk1 <- subset(lrrk1, subset = nFeature_RNA > 1200 & nFeature_RNA < 6500 & percent.mt < 0.5 & nCount_RNA < 26000 & percent.ribo < 0.5); control2 <- subset(control2, subset = nFeature_RNA > 800 & nFeature_RNA < 5000 & percent.mt < 0.5 & nCount_RNA < 15000 & percent.ribo < 0.5); lrrk2 <- subset(lrrk2, subset = nFeature_RNA > 1000 & nFeature_RNA < 7500 & percent.mt < 0.5 & nCount_RNA < 20000 & percent.ribo < 0.5) (post-filtering violin plots shown in Fig S1A). The male and female datasets were then normalized and integrated using the SCTransform V2 method[103] in Seurat V5[104]. Cells were initially clustered using 35 principal components at a resolution of 0.8. Plotting clusters on UMAP showed one small cluster exceedingly distant from all other clusters, which was suspected to represent EW nucleus cells and was subsequently removed. Afterwards, the integration and clustering pipeline was repeated without these cells to remove their impact on PCA. Final clustering was achieved using the FindNeighbors() command with the following parameters: dims = 1:32, reduction = "cca", n.trees = 500, k.param = 40, and with the FindClusters() command with the following parameters: resolution = 0.8, algorithm = 1, group.singletons = TRUE, graph.name = "SCT_snn". UMAP reduction was calculated using the RunUMAP command with the following parameters: reduction = "cca", dims = 1:32, reduction.name = "umap.cca", n.epochs = 500, min.dist = .2, n.neighbors = 1000. In total, the integration resulted in a final dataset of 28532 nuclei, with a median UMI count of 7750.5 and median of 3056 features. All clusters were represented in both male and female samples (Fig S1B).

Cluster dendrogram was produced using the Seurat BuildClusterTree() function with the same parameters used for clustering; dims = 1:32, reduction = "cca" . Differential expression between branches of the

dendrogram were calculated using FindMarkers() with the following parameters to maximize detection of DEGs with binary differences across branches, rather than genes expressed in both branches at different levels; min.diff.pct = 0.25, logfc.threshold = 1, min.pct = 0.2. The top positively and negatively enriched genes were then plotted for manual curation of branch point defining markers that best distinguished each branch, with preference given to genes previously described in DA neuron subtype literature.

Sex of original mice was inferred for each cell by calculating the expression ratio of genes *Uty* and *Eif* relative to *Tsix* and *Xist* for each cell, revealing roughly equal proportions of male and female cells in the final dataset (Fig S1C). All clusters were represented in each individual sample (Fig S1D), and with no overt differences in proportions across conditions (Fig S1E, Fig 8B). Clusters 21 and 16 were removed from downstream analyses due to high expression of glial markers Atp1a2 and Mbp, respectively (Fig S1F). All differential expression calculations were also performed in Seurat using the FindMarkers() command; individual parameters for differential expression at each step of our analysis are provided in the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal). Of note, the default Seurat differential expression calculation method is known to overestimate significance of differential gene expression due to treating individual cells as independent samples. While pseudo bulk differential gene expression is preferable given these limitations, the number of independent 10X library preparations (2 per group) does not provide sufficient sample numbers for the recommended minimums for pseudo bulk differential gene expression approaches, and thus the default Seurat methodology was used; caution should be used when interpreting p values of individual DEGs such as those shown in the volcano plot of Fig S9.

ShinyCell (RRID:SCR_022756) web browser was created using the SCT assay of the Seurat Object described above[105]. The code can be found at the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal).

Mapping data query sets were done using TranserData() seurat function. The dataset and clusters from Azcorra & Gaertner et al. (Dataset can be found in the Gene Expression Omnibus GSE222558) [25] was used as the reference, and the query dataset was the one described in Figure 1. Any cells clustered with a prediction.max.score <.5 were removed from the data set (2.4% of cells removed). The exact workflow can be found at the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal).

Sankey plots were made using the Network3D package in R and the sankeyNetwork function. The 'source' and 'target' nodes are the corresponding cluster labels. ’Value’ is set as the number of cells clustered in

both the source and target nodes. The exact workflow can be found at the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal).

Cluster stability and homogeneity analyses

To quantify how homogenous each cluster is and uncover potential further subdivisions that may exist within our clusters, we applied two approaches highly similar to those used for our previously described dataset[25]. First, we generated a measure of stability for each cluster through random downsampling and reclustering of the dataset using the same custom R scripts as previously described[25]. Doing so provided cluster stability metrics (Fig S2A) representing the propensity for cells within a cluster to continue to co-cluster when data is removed.

To understand potential sources of lower cluster stability values (e.g. two clusters being grouped together as one, or additional small clusters being divided into two adjacent clusters), we utilized the ClusTree (v0.5.1; RRID:SCR_016293 – https://github.com/lazappi/clustree) R package to plot clustering at ten incremental resolutions of 0.1 through 1.1 (Fig S2B).

Scanpy

Analysis done using Python (v 3.10.13[106] - RRID:SCR_008394). Cells were clustered using the scanpy toolkit (Version #1.10.1 RRID: SCR_018139 https://scanpy.readthedocs.io/en/stable/). Outputs from CellRanger were read into Scanpy using scanpy.read_10x_mtx(). Cells were filtered with the same parameters as described above in the Seurat workflow. Datasets were individually clustered using the Leiden clustering algorithm. The four datasets were then integrated using scanpy.tl.ingest.(). Each dataset was mapped onto Control 1. The exact workflow can be found at the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal) .

Cluster Similarity across Control and Lrrk2 Samples

To assess the molecular signature similarity between the clusters from Control and Lrrk2^{G2019S} samples, we first corrected for dropouts the expression matrix of each sample using Adaptively-thresholded Low Rank Approximation (ALRA)[107]. Then, using MetaNeighbor standard pipeline[108], we picked the intersect of variable genes across all but the top decile of expression bins for Control and Lrrk2 samples. On the 5282 selected genes, MetaNeighbor was performed to assess an AUROC score of clusters similarity. The exact workflow can be found at the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal).

Pathway Enrichment Analyses

To evaluate pathways and gene sets enriched in subtypes or across conditions, we utilized the fgsea (v3.19; https://bioconductor.org/packages/release/bioc/html/fgsea.html; SCR_020938) and clusterProfiler (v4.8.3; RRID:SCR_016884 – https://doi.org/doi:10.18129/B9.bioc.clusterProfiler) R packages. Pathways for GSEA analysis (mouse canonical pathways gene sets) were obtained from MSigDB (v2023.2.Hs, 2023.2.Mm; https://www.gsea-msigdb.org/gsea/msigdb; SCR_016863). Custom R scripts were created for visualization of GSEA results and are available in the deposited code for this paper. fgsea was run using the following parameters: nperm=1000, stats=ranks, minSize = 25, maxSize = 500. Differential expression calculations for fgsea package input were performed using Seurat's FindMarkers command to generate a ranked list of genes, with filtering out of low-expression genes based on a minimum of 10% detection in either population being compared.

For gene ontology, the biological processes gene sets were obtained using the org.Mm.eg.db R package (v3.19; https://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html; SCR_002774), and GO (part of clusterProfiler package) was run using default settings with BH adjustment of p-values and utilizing FDR-adjusted significant ($p < 0.05$) DEGs for the respective comparisons. A custom genetic background was set for every individual comparison calculated to ensure hypergeometric testing was not simply enriched for cell type specific pathways; for comparisons of subsets within the dataset (i.e. a family or cluster) across conditions, a minimum detection level of 10% of cells was used to define the genetic background. These same thresholds were applied to filter the DEG lists used as input for GO. Results were visualized using the enrichplot R package (v3.19; https://www.bioconductor.org/packages/release/bioc/html/enrichplot.html; SCR_006442).

scDRS Cell-level Scoring of GWAS Risk Associations

The scDRS score represents the relative association for each individual cell's expression profile (among all other cells in the dataset) with PD risk loci by utilizing the underlying SNPs and associations described in GWAS summary statistics. Gene expression data ('corrected' UMI counts) was extracted from the SCT assay. Expression matrices, cell metadata, and feature metadata were loaded using Pandas (v2.2.2; https://pandas.pydata.org/; SCR_018214) in Jupyter (Jupyter: v7.2.1; https://jupyter.org/install ; SCR_018315) notebook. The data were converted into an AnnData object using the anndata package (AnnData v0.10.8; https://anndata.readthedocs.io/en/latest/; SCR_018209) and stored in HDF5 format. Normalization and scaling were performed using the Scanpy package (v1.8.1) (PMID: 29409532). Total counts for each cell were normalized to 10,000 reads using the sc.pp.normalize_total function and log-transformed using sc.pp.log1p. The sc.pp.highly_variable_genes function (parameters:

min_mean=0.0125, max_mean=3, and min_disp=0.5) was used to identify highly variable genes. Finally, the anndata object was scaled using the sc.pp.scale function, with a max_value parameter of 10.

GWAS summary statistics for Parkinson's Disease risk were obtained from EBI GWAS Catalog Study ID: GCST009325 (PMID: 31701892) and processed using the MAGMA tool on FUMA's online platform (v1.5.2; https://fuma.ctglab.nl/snp2gene; SCR_017521; RRID:SCR_005757) to generate a scored list (z-scores) of significant genes. Default settings were used with a 10x10 gene window. Scored gene list was processed using the munge-gs function of the scDRS package (v1.0.2; https://martinjzhang.github.io/scDRS/) to create .gs gene sets (PMID: 34431100) and loaded into the Jupyter Notebook using the scdrs.util.load_gs function. To align with the anndata object data, the gene set's human genes were mapped to mouse orthologs and intersected with genes present in the dataset. Preprocessing was completed using the scdrs.preprocess function, binning genes by mean and variance with parameters n_mean_bin=20 and n_var_bin=20. Scoring for Parkinson's disease was performed using the scdrs.score_cell function (parameters: ctrl_match_key=mean_var, n_ctrl=1000, weight_opt=vs, return_ctrl_raw_score=False, and return_ctrl_norm_score=True).

Since scDRS does not include a native method for population-level p-values, we calculated a mean scDRS score for each cluster or family by averaging its component cells and assessed for significance by creating a bootstrapped confidence interval for these means. A score of 0 represents the null of no association between gene expression and PD risk loci, and thus if the 95% confidence interval does not overlap 0, the mean scDRS score for a given group can be regarded as significant as there is a less than 5% chance of the true group mean containing the null. Bootstrapped 95% confidence intervals for mean score within clusters or cluster families were calculated in R using 10,000 permutations for each group and plotted as mean scDRS score per group with CI represented as error bars.

Parkinson's Disease MAGMA Analysis Ranked Genes dataset can be found at 10.5281/zenodo.13076447

SynGO Analyses

SynGO analysis protocol is described in detail at dx.doi.org/10.17504/protocols.io.x54v92r94l3e/v1.

MERFISH

Animals

Adult C57Bl/6 male mice (RRID:IMSR_CRL:027) aged 2-3 months were used for the MERFISH experiments. Animals were housed at the Montreal Neurological Institute (MNI) on a 12h-12h light-dark cycle (light cycle 7:00 to 19:00) with ad libitum access to food and water. Experiments were approved by the MNI Animal Care Committee and conducted according to guidelines and regulations from the Canadian Council on Animal Care.

Gene selection and panel assembly for MERFISH

Hybridization probes were generated by Vizgen using their custom gene portal. DA neuron specific genes were selected based on differential gene expression analysis from the 20 clusters identified in the single-nuclei sequencing data. Also included were: 1) general markers of glutamatergic, GABAergic, cholinergic, serotonergic neuron populations, 2) markers of non-neuronal cell types, 3) neuropeptide precursor genes and neuropeptide receptors and 4) general ion channels. The list of all probes on the panel can be found in Supplemental Table 1.

Tissue preparation and tissue imaging for MERFISH

Mice were perfused with 1x PBS followed by 4% paraformaldehyde (PFA). Brains were extracted and post-fixed in PFA for 16 hours, embedded in O.C.T. (Tissue-Tek O.C.T.; 25608-930, VWR), and stored at −80 °C until sectioning. Frozen brains were sectioned at −18 °C on a cryostat (Leica CM3050S). Within 3 animals we collected 10um-thick sequential coronal brain sections every 200µm on specialized glass slides (approximately 5-6 sections per mouse). Seven sections were chosen with the best morphology across the midbrain for processing. Sample preparation was completed using Vizgen gene imaging kits provided according to manufacturer. All steps were performed under RNAse free conditions. Sections were first washed with 1x PBS three times before incubation in 70% Ethanol overnight at 4°C to permeabilize cell membranes. The following day each section was washed with sample wash buffer followed by an incubation in formamide wash buffer at 37 °C for 30 minutes. The gene panel mix containing hybridization probes was then placed on top of the tissue and placed in a humidified 37 °C chamber for 36-48h. Afterwards the tissue was incubated with formamide buffer for 30m at 47 °C twice before washing with sample wash buffer. The hybridized tissue was then embedded in a 0.05% ammonium persulfate gel and cleared overnight in a solution of Proteinase K (NEB Cat# P8107S) at 37 °C. The following morning the sample was incubated in a solution containing DAPI and PolyT stain to aid in cell segmentation. Once processed, brain sections were prepped for imaging using the commercial Merscope system developed by

Vizgen. 500-gene panel cartridges were thawed in a 37°C water bath for 1h before the start of imaging. Mouse RNAse inhibitor (NEB Cat# M0314) was added to imaging activation buffer and added to the cartridge fluidic system before loading into the Merscope to prime the fluidics chamber. Slides were removed from wash buffer and placed in the imaging capsule, connected to the fluidics chamber. Sections were imaged according to the manufacturer (Vizgen). The exact protocol for *tissue preparation and tissue imaging for MERFISH* can be found at dx.doi.org/10.17504/protocols.io.14egn6opyl5d/v1 .

Preprocessing and quality control of MERFISH data

Cell by gene matrices were processed in R using the Seurat V5 package[104]. To exclude cells that underwent poor cell segmentation we removed cells that had >500µm^3 and <4000µm^3 volumes. To remove cells that show low transcript identity or falsely identified cells we removed any cell that had <40 total RNA transcripts detected. The remaining cells then underwent SCT normalization[103] before running RPCA integration to integrate datasets from individual sections. Shared nearest neighbour clustering was completed using the first 37 principle components using a resolution of k=0.7. Genetic markers of each cluster were determined using a roc test and was used to annotate our dataset. High-level annotations for glia and neuron types were confirmed using the SCType package in R (v3.0; https://sctype.app/). The putative DA cluster was subsetted, re-normalized and subclustered using the first 19 principle components and a resolution of k=0.65. Gene markers of each subcluster was determined using a ROC test and the top three genes from each cluster were identified based on the high positive log-fold change with a low p-values (adjusted p-value<10^{-50}). Code can be found on the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal) .

Integration of MERFISH data with scRNA-seq data

To integrate MERFISH data with snRNA-seq data we utilized the weighted-nearest neighbour algorithm defined in Stuart, Butler, et al [92]. Briefly, anchor points were established between the two datasets using the FindTransferAnchors and MapQuery functions in Seurat. This allowed for a projection of the MERFISH data into the scRNA-seq space and a transfer of cluster identity between the two. Importantly, a label transfer prediction score was assigned to each MERFISH cell and cells with scores greater than 0.5 were classified as high confidence cells. Whole genome expression was imputed onto MERFISH cells using snRNA-seq data using the TransferData function in Seurat using the snRNA-seq normalized data as a reference. Cluster identifications were compared between transfer data and MERFISH subclustering using a Sankeyplot in R. Briefly, the number of cells that transferred between two pairs of clusters were

calculated and expressed as a proportion within each MERFISH cluster. Only flows >10% were included in the visualization. Spatial visualizations of DA subtypes were generated using the ImageDimPlot function in Seurat and overlaid on the Allen Brain Atlas mouse anatomical template or the Paxinos mouse brain coronal axis using affinity designer. Gene expression values were obtained using the ImageFeaturePlot function in Seurat. Code can be found on the Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal) .

MERFISH anatomical localization

We utilized the Vizualizer software (v2.2, Vizgen Inc. https://portal.vizgen.com/resources/software) to delineate mouse midbrain regions into 11 distinct regions: the caudal linear nucleus (CLi), interfascicular nucleus (IF), interpeduncular nucleus (IPN), periaqueductal grey region (PAG), posterior hypothalamus (PH), rostral linear nucleus (RLi), retrorubral field (RR), substantia nigra pars compacta (SNc), Substantia nigra pars reticulata (SNr), Superior Colliculus (SC) and the ventral tegmental area (VTA). Brain regions were manually sectioned based on the intensity of DAPI imaging as well as the presence of *Th* transcripts compared to the Allen Brain Atlas. Cells within each ROI were counted to generate the localization of each family and cluster by counts and proportions.

MERFISH data can be found at DOI #: 10.5281/zenodo.12636328.

Code Availability Statement

Code is available at Awatramani Lab Github page (github.com/AwatramaniLab/Gaertner_Oram_etal) and DOI # 10.5281/zenodo.13820931.

Acknowledgements

ZG and CO performed snRNAseq and MERFISH experiments and data analysis. CB, ES, AS and CC performed data analysis and shinycell app construction. LP and DD provided scientific input on DA classification and LRRK2 alterations. ZG, CO, LP, JFP, RA wrote the manuscript. We would like to acknowledge Dr. Kevin Petrecca, Dr. Phuong Le and Michael Luo for generously providing access to Merscope and technical advice. RA, DD, AS and LP were funded by Aligning Science Across Parkinson's (ASAP-020600). For the purpose of open access, the authors have applied a CC-BY 4.0 public copyright license to this manuscript. RA was also funded by 1R01NS119690-01, P50 DA044121-01A1. LP was funded by R01 NS097901. ZG was

funded by NINDS 1F31NS115524-01A1. JFP was funded by CIHR (PJT-183760), HBHL, Parkinson Canada. CB was funded by Parkinson Canada, FRQS and CIHR.

Figure Legends:

Figure 1. Generation of a large RNAseq dataset of midbrain dopamine neurons. A) Schematic of single nucleus isolation pipeline using microfluidic chip-based sorting from n=8 DAT-IRES-Cre (referred to as DAT-Cre), CAG-Sun1/sfGFP (referred to as RC-LSL-Sun1/GFP) mice. B) UMAP representation of clusters. Of note, putative clusters 16 and 21 are shown in grey, and are removed from analyses due to co-expression of glial markers. C) Expression of the defining midbrain dopaminergic neuron markers vesicular monoamine transporter 2 (*Vmat2*; *Slc18a2*), DOPA decarboxylase (*Ddc*), tyrosine hydroxylase (*Th*), and dopamine transporter (*Dat; Slc6a3*). Robust expression is observed throughout all clusters. D) Dot plot showing expression of several markers of previously described dopamine neuron subtypes, consistent with finding additional heterogeneity within established subtypes. E) Expression patterns of midbrain floorplate markers. All clusters show robust expression with exception of 0 and 17, which indicate these populations are not midbrain DA neurons but might be located in nearby hypothalamic nuclei.

Figure 2. Mapping subtype identities, marker genes, and relationships to previously described populations. A) Dendrogram of DA neuron subtypes in our dataset. Distance between branch points represent approximations of relatedness between subsequent nodes, and genes labeled at each branch point represent the top differentially expressed genes distinguishing the downstream groups. Subtypes fell largely into three "families," with expression patterns generally defined by *Gad2*, *Sox6*, or *Calb1*. Notably, clusters 1, 11, and 13 are shown with an asterisk as these do not display differential expression in line with the defining genes of their cluster families (i.e. are not significantly enriched for their eponymous family genes). B) Heatmap of top differentially expressed genes for each cluster. The top 3 genes for each cluster are shown. C-F) Co-expression of genes at specified branch points denoted in 2A. G) Table describing putative clusters and relation to previous literature. Top row: cluster number. Second row: Equivalent cluster in Azcorra & Gaertner et al[25]. Third row: stepwise genetic signatures based on following sequential branch points of cluster dendrogram. Bottom row: Proposed shorthand name for subtypes, based on cluster family and top defining genes for that cluster within each family.

Figure 3. Identification of dopamine neurons in MERFISH. A) Schematic showing the processing and imaging of brain tissue with MERFISH. B) Clustering and annotation of all cells identified by MERFISH. C) Relative expression of *Th*, *Slc6a3* and *Slc18a2* shows the presence of a single cluster (blue cluster in B) that expresses all three genes. D) Spatial location of neuronal clusters for whole brain MERFISH along three rostral-caudal delineations associated with -3.0, -3.2 and -3.6mm Bregma. E) Cellular expression of *Th* in the sections shown in D.

Figure 4. Mapping of MERFISH data with snRNA-seq clustering: A) Grouping and recoloring of snRNA-seq clusters into distinct families represented by Gad2, Sox6 and Calb1 (Left). Relative proportion of dopamine neurons comprising each cluster from the snRNA-seq (SN, orange) and MERFISH label transfer (MER, grey) data (Right). B) Spatial representation of predicted clustering across the rostral-caudal axis, zoom of one representative hemisphere shown below. C-E) Imputed cellular expression of *Gad2*, *Calb1* and *Sox6* across the rostral-caudal axis.

Figure 5. Spatial localization of Sox6 family of DA neurons. A) Distribution of Sox6 family DA neuron subtypes relative to all other DA neurons along the rostral-caudal axis. DA cells belonging to the Calb1 and Gad2 family are colored in light grey. B) Location of individual subtype within the Sox6 family (left two panels) with other cells in the Sox6 family shown in dark grey. Relative cellular expression of genetic markers associated with each subfamily shown in a UMAP and a midbrain section.

Figure 6. Spatial localization of Calb1 family DA neurons. A) Calb1 family of neurons were divided into two distinct branches based on hierarchical clustering. Distribution of Branch 1 of Calb1 family DA neurons along the rostral-caudal axis with all other DA cells colored in grey. B) Location of individual clusters within Branch 1 of the Calb1 family. Only cells in the Branch 1 subset are shown. Relative cellular expression of genetic markers associated with each subtype shown in a UMAP and brain section. C) and D) Equivalent plots as A-B, but for Branch 2 of the Calb1 family of DA neurons. Relative cellular expression of genetic markers associated with each subfamily shown in a UMAP and a midbrain section.

Figure 7. Spatial location of Gad2 family DA neurons. A) Distribution of Gad2 family DA neurons along the rostral-caudal axis with other DA cells colored in grey. B) Location of individual subtypes within the Gad2 family with only cells of the Gad2 family shown in grey.

Figure 8. Global effects of Lrrk2^{G2019S} mutation are observed across all clusters. A) Expression pattern of Lrrk2 RNA, which is present in all clusters, though notably low in Gad2^{Egfr} (cluster 15). B) Cluster proportions in control and mutant samples. Relative proportions of each subtype within the dataset are relatively unchanged. C) MetaNeighbor-generated cluster similarity heatmap shows that each cluster from control samples is more similar to its corresponding cluster in the mutant samples, suggesting there is no large-scale change in cell type across samples. D) Gene set enrichment analysis (GSEA) results comparing all clusters across conditions. Several top enriched pathways in Lrrk2 mutants are related to mitochondrial energy production/oxidative phosphorylation. E) Gene ontology (GO) results comparing all clusters across conditions. Top enriched pathways are related to synapse organization and function, consistent with previously described synaptic dysfunction in Lrrk2^{G2019S} mutant mice.

Figure 9. Changes in gene expression are observed in PD-implicated DA neuron subtypes in Lrrk2^{G2019S} mice. A) GO results comparing each cluster family across conditions (top 5 enriched pathways shown). Several pathways are highly relevant to locomotor behavior and previously described dysfunction (such as synaptic organization pathways) in Lrrk2^{G2019S} mutant mice. B) GSEA results comparing Sox6 cluster family across conditions. While twenty pathways were positively enriched in mutant Sox6 family clusters (BH-corrected p-values < 0.05), several additional pathways were enriched in either direction using a less conservative cutoff of $p < 0.1$. The top five positively and negatively enriched pathways with this less conservative cutoff are shown. Pathway enrichment was similar to global changes, but more pronounced in the Sox6 family of clusters. C-E) Results of SynGO analyses for Sox6 family (9C), Calb1 family (9D), or Sox6^{Tafa1} clusters (9E). SynGO-annotated (i.e. synaptic) genes were enriched among DEGs for ventral SNc populations, particularly Sox6^{Tafa1}. Of note, among these synaptic genes, Sox6^{Tafa1} cluster showed much heavier enrichment for those localized to the presynapse, which may suggest subtype-specific presynaptic function in acceleration-corrected subtypes. F) Feature plot of each cell's association with PD risk loci proposed by GWAS as calculated by scDRS scores. Clear differences are observed across clusters. G) Mean and 95% confidence intervals for PD GWAS risk (scDRS scores) for each cluster. Sox6 family of clusters are

particularly elevated. H) Average risk scores with 95% confidence intervals among cluster families, as well as the combination of clusters Sox6^{Tafa1} and Sox6^{Vcan}, the two *Anxa1*-expressing SNc clusters, which showed the highest average risk scores.

Supplementary Figure Legends

Figure S1. Quality control for dataset generation. A) Violin plots for each of our 4 samples (post-filtering) for each of the four QC metrics used for filtering our datasets. B) Expression patterns of sex-specific genes. All clusters are represented by both sexes. C) Histogram of number of cells plotted by male to female gene ratios (scored continuously along the x-axis). Three discrete peaks emerge, representing cells likely originating from either sex or those with indeterminate ratios due to technical drop-off in RNAseq reads. Similar numbers of male and female cells were observed, as represented by the sums of adjacent bins in each of the three marked regions on the x-axis. D) Cluster representation from each individual RNAseq library. All clusters were represented in all samples. E) Cluster representation from pooled control and Lrrk2 mutant samples. Distributions are roughly equivalent between conditions, suggesting no overt change in subtype composition as a function of genotype. F) Dotplot of expression for glial marker genes *Mbp* and *Atp1a2*, showing high expression in clusters 16 and 21, respectively, indicating likely doublets of DA neurons and glia.

Figure S2. Representations of cluster heterogeneity. A) Cluster stability metrics shown as a box plot for each cluster. Clusters with lower stability contain cells that more easily collapse into other clusters. Outliers (defined as more than 1.5 times the interquartile range (IQR) above the third quartile or below the first quartile) are shown as circles. B) Cluster tree displaying evolution of clusters when calculated at different resolutions. As resolution increases, new levels of heterogeneity emerge, but ultimately become largely stable at higher resolutions. Resolution used for our clustering scheme is highlighted by dotted line.

Figure S3. Validating snRNA-seq clusters with Python and published DA clusters. A) Comparison of UMAP from Azcorra & Gaertner et al., 2023[25] and UMAP from Figure 1. A Sankey diagram was created by mapping nuclei collected in Fig 1 onto the clusters from Azcorra*,Gaertner* et al., 2023[25]. The sankey diagrams only include cell transfers that make up at least 5% of the cluster. B) UMAP representation of clusters using

Scanpy package on Python and expression of *Anxa1* gene. C) Sankey diagram comparing the clustering of dataset from Figure 1 using Scanpy and Seurat. Note, cells in cluster 19 of the Scanpy plot did not transfer to one cluster in Seurat with >5% of its total cell count.

Figure S4. Quality control metrics for MERFISH experiments. A-C) Volume, number of features (nFeatures) and number of total transcripts counts (nCounts) metrics for each cell identified in MERFISH experiments after quality control. D) Example correlation coefficient between total cellular transcripts detected for each gene across two experiments and a matrix showing experimental similarity. Spatial location of cortical excitatory (E) and inhibitory (F) neurons show distinct layering patterns by MERFISH imaging. G) Relative expression levels of groups of genes associated with synthesis and secretion of glutamate (*Slc17a6, Slc17a7, Slc17a8*), GABA (*Slc32a1, Gad1, Gad2*), dopamine (*Slc6a3, Th*) and serotonin (*Tph2, Slc6a4*) showcase the multilingual nature of the dopaminergic and serotonergic systems.

Figure S5. Clustering of DA neurons identified by MERFISH. A) Schematic showing the manual outlining of anatomical regions of the ventral midbrain. B) Relative proportions of total DA neurons identified by MERFISH distributed throughout the following ventral midbrain structures: substantia nigra pars compacta (SNc), ventral tegmental area (VTA), interfascicular nucleus (IF), rostral linear nucleus (RLi), caudal linear nucleus (CLi), periaqueductal grey (PAG), and retrorubral region (RR). C) Subclustering of DA neurons from MERFISH dataset reveals 12 distinct subclusters. D) Heatmap showing relative expression of genes associated with dopaminergic character and the top three differentially expressed genes within each cluster identified in c). E) Relative expression of genes associated with dopaminergic character (*Slc6a3*, *Th*, *Slc18a2*) distributed across DA clusters. F) Midbrain location of clusters identified by MERFISH using the manual outlining described in A).

Figure S6. Integrating MERFISH with snRNA-seq: A) Overview of MERFISH integration with snRNA-seq dataset (top). snRNA-seq UMAP space and cluster identification were transferred to MERFISH cells by identifying mutual nearest neighbors (MNN) and assigning a score based on the overlap of neighbor cells that share the same labelling (middle). High confidence MERFISH cells (bottom) were identified as those that showed a cell similarity score >0.5 (2,298 of 4,399 original DA neurons). B) Sankey plot showing the proportion of cells from MERFISH clustering identities that make up the snRNA-seq clustering identities.

Flow chart only includes cell transfers that make up at least 10% of the snRNA-Seq cluster. C) Heatmap showing the distribution of correlation coefficients for each gene in our MERFISH panel. Correlation coefficients were found by calculating the correlation between MERFISH transcript counts vs. Predicted (Imputed) transcript counts for the 500 genes used in our panel. The heatmap is made up of 500 columns with each column colored to associate with that genes r-value, so it is not a linear plot. The labels point to the column representing genes associated with major family subdivisions. The minimum correlation value is -0.189 for Sox6. D-F) Comparison of the transcript detection in snRNA-seq dataset, MERFISH dataset and imputed gene expression for *Gad2*, *Calb1* and *Aldh1a1*. Spatial location of normalized transcript counts and imputed data show similar distributions (middle). Example representations of individual transcripts detected by MERFISH in the VTA for each gene (right and inset).

Figure S7. Correlation of MERFISH vs imputed transcript counts for markers of the Sox6 (A) and Calb1 (B) families. Expression plots were generated comparing the expression of snRNA-seq (far left), MERFISH (second from left) and imputed (second from right) transcripts for genes associated with each cluster family. Only high confidence cells from the MERFISH experiments were used in these plots figure. Correlation plot of MERFISH vs imputed transcript counts (far right) show the strength of imputation. Correlation coefficients were calculated using only cells with non-zero transcript counts in MERFISH and imputed datasets. Each family shows two genes that correlate strongly and two genes that correlate weakly, the latter of which may be due to technical limitations.

Figure S8. Quantifications of subtypes by anatomic localization. A) Anatomical distribution of DA neurons colored by cluster family. B) Anatomical distribution of DA neurons colored by individual cluster identity.

Figure S9. Analysis of differentially expressed genes and their synaptic localizations and functional roles using SynGO. A) Volcano plot of differentially expressed genes across conditions. Each point is colored according to its significance and magnitude of fold change. B) Venn diagram of differentially expressed genes (DEGs) between Sox6 and Calb1 families. C) Mapping of pre- and post-synaptic genes found among significant DEGs in Lrrk2^{G2019S} mutants compared to controls for Sox6 family, Sox6^{Tafa1}, and Calb1 family of clusters. Bars are colored according to relative association with each cellular compartment. D) Functional roles of synaptic DEGs in each population in Lrrk2^{G2019S} mutants compared to controls.

Figure S10. Comparison of pathway analysis results between Anxa1+ clusters. A) Gene ontology results for clusters 4 ($Sox6^{Tafa1)}$ and 11 ($Calb1^{Stac}$) showing increased significance of pathway changes in Anxa1+ SNc neurons compared to an Anxa1+ VTA cluster.

Extended Data:

Extended data files contain raw differential expression analysis results in Excel format. Tabs within the file are provided for various comparisons referenced/used throughout the text.

References

1. Björklund A, Dunnett SB. Dopamine neuron systems in the brain: an update. Trends Neurosci. 2007;30(5):194-202. Epub 20070403. doi: 10.1016/j.tins.2007.03.006. PubMed PMID: 17408759.
2. Fu Y, Yuan Y, Halliday G, Rusznák Z, Watson C, Paxinos G. A cytoarchitectonic and chemoarchitectonic analysis of the dopamine cell groups in the substantia nigra, ventral tegmental area, and retrorubral field in the mouse. Brain Struct Funct. 2012;217(2):591-612. Epub 20110921. doi: 10.1007/s00429-011-0349-2. PubMed PMID: 21935672.
3. Gaertner Z, Azcorra M, Dombeck DA, Awatramani R. Molecular heterogeneity in the substantia nigra: A roadmap for understanding PD motor pathophysiology. Neurobiol Dis. 2022;175:105925. Epub 20221111. doi: 10.1016/j.nbd.2022.105925. PubMed PMID: 36372290.
4. Evans RC, Zhu M, Khaliq ZM. Dopamine Inhibition Differentially Controls Excitability of Substantia Nigra Dopamine Neuron Subpopulations through T-Type Calcium Channels. J Neurosci. 2017;37(13):3704-20. Epub 20170306. doi: 10.1523/jneurosci.0117-17.2017. PubMed PMID: 28264982; PMCID: PMC5373143.
5. Menegas W, Akiti K, Amo R, Uchida N, Watabe-Uchida M. Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. Nat Neurosci. 2018;21(10):1421-30. Epub 20180903. doi: 10.1038/s41593-018-0222-1. PubMed PMID: 30177795; PMCID: PMC6160326.
6. Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw ND, Tank DW, Witten IB. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature. 2019;570(7762):509-13. Epub 20190529. doi: 10.1038/s41586-019-1261-9. PubMed PMID: 31142844; PMCID: PMC7147811.
7. Lammel S, Hetzel A, Häckel O, Jones I, Liss B, Roeper J. Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron. 2008;57(5):760-73. doi: 10.1016/j.neuron.2008.01.022. PubMed PMID: 18341995.
8. da Silva JA, Tecuapetla F, Paixão V, Costa RM. Dopamine neuron activity before action initiation gates and invigorates future movements. Nature. 2018;554(7691):244-8. Epub 20180131. doi: 10.1038/nature25457. PubMed PMID: 29420469.
9. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature. 2009;459(7248):837-41. Epub 20090517. doi: 10.1038/nature08028. PubMed PMID: 19448610; PMCID: PMC2739096.
10. Beier KT, Steinberg EE, DeLoach KE, Xie S, Miyamichi K, Schwarz L, Gao XJ, Kremer EJ, Malenka RC, Luo L. Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. Cell. 2015;162(3):622-34. doi: 10.1016/j.cell.2015.07.015. PubMed PMID: 26232228; PMCID: PMC4522312.
11. Avvisati R, Kaufmann AK, Young CJ, Portlock GE, Cancemi S, Costa RP, Magill PJ, Dodson PD. Distributional coding of associative learning in discrete populations of midbrain dopamine neurons. Cell Rep. 2024;43(4):114080. Epub 20240404. doi: 10.1016/j.celrep.2024.114080. PubMed PMID: 38581677.
12. Brischoux F, Chakraborty S, Brierley DI, Ungless MA. Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. Proc Natl Acad Sci U S A. 2009;106(12):4894-9. Epub 20090304. doi: 10.1073/pnas.0811507106. PubMed PMID: 19261850; PMCID: PMC2660746.
13. Heymann G, Jo YS, Reichard KL, McFarland N, Chavkin C, Palmiter RD, Soden ME, Zweifel LS. Synergy of Distinct Dopamine Projection Populations in Behavioral Reinforcement. Neuron. 2020;105(5):909-20.e5. Epub 20191223. doi: 10.1016/j.neuron.2019.11.024. PubMed PMID: 31879163; PMCID: PMC7060117.
14. Howe MW, Dombeck DA. Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature. 2016;535(7613):505-10. Epub 20160711. doi: 10.1038/nature18942. PubMed PMID: 27398617; PMCID: PMC4970879.

15. Lerner TN, Shilyansky C, Davidson TJ, Evans KE, Beier KT, Zalocusky KA, Crow AK, Malenka RC, Luo L, Tomer R, Deisseroth K. Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. Cell. 2015;162(3):635-47. doi: 10.1016/j.cell.2015.07.014. PubMed PMID: 26232229; PMCID: PMC4790813.
16. Phillips RA, 3rd, Tuscher JJ, Black SL, Andraka E, Fitzgerald ND, Ianov L, Day JJ. An atlas of transcriptionally defined cell populations in the rat ventral tegmental area. Cell Rep. 2022;39(1):110616. doi: 10.1016/j.celrep.2022.110616. PubMed PMID: 35385745; PMCID: PMC10888206.
17. Poulin JF, Gaertner Z, Moreno-Ramos OA, Awatramani R. Classification of Midbrain Dopamine Neurons Using Single-Cell Gene Expression Profiling Approaches. Trends Neurosci. 2020;43(3):155-69. Epub 20200211. doi: 10.1016/j.tins.2020.01.004. PubMed PMID: 32101709; PMCID: PMC7285906.
18. Garritsen O, van Battum EY, Grossouw LM, Pasterkamp RJ. Development, wiring and function of dopamine neuron subtypes. Nat Rev Neurosci. 2023;24(3):134-52. Epub 20230118. doi: 10.1038/s41583-022-00669-3. PubMed PMID: 36653531.
19. La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, Borm LE, Stott SRW, Toledo EM, Villaescusa JC, Lönnerberg P, Ryge J, Barker RA, Arenas E, Linnarsson S. Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell. 2016;167(2):566-80.e19. doi: 10.1016/j.cell.2016.09.027. PubMed PMID: 27716510; PMCID: PMC5055122.
20. Kramer DJ, Risso D, Kosillo P, Ngai J, Bateup HS. Combinatorial Expression of Grp and Neurod6 Defines Dopamine Neuron Populations with Distinct Projection Patterns and Disease Vulnerability. eNeuro. 2018;5(3). Epub 20180613. doi: 10.1523/eneuro.0152-18.2018. PubMed PMID: 30135866; PMCID: PMC6104179.
21. Hook PW, McClymont SA, Cannon GH, Law WD, Morton AJ, Goff LA, McCallion AS. Single-Cell RNA-Seq of Mouse Dopaminergic Neurons Informs Candidate Gene Selection for Sporadic Parkinson Disease. Am J Hum Genet. 2018;102(3):427-46. doi: 10.1016/j.ajhg.2018.02.001. PubMed PMID: 29499164; PMCID: PMC5985341.
22. Tiklová K, Björklund Å K, Lahti L, Fiorenzano A, Nolbrant S, Gillberg L, Volakakis N, Yokota C, Hilscher MM, Hauling T, Holmström F, Joodmardi E, Nilsson M, Parmar M, Perlmann T. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. Nat Commun. 2019;10(1):581. Epub 20190204. doi: 10.1038/s41467-019-08453-1. PubMed PMID: 30718509; PMCID: PMC6362095.
23. Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, de Rivera H, Bien E, Baum M, Bortolin L, Wang S, Goeva A, Nemesh J, Kamitaki N, Brumbaugh S, Kulp D, McCarroll SA. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell. 2018;174(4):1015-30.e16. doi: 10.1016/j.cell.2018.07.028. PubMed PMID: 30096299; PMCID: PMC6447408.
24. Yaghmaeian Salmani B, Lahti L, Gillberg L, Jacobsen JK, Mantas I, Svenningsson P, Perlmann T. Transcriptomic atlas of midbrain dopamine neurons uncovers differential vulnerability in a Parkinsonism lesion model. Cold Spring Harbor Laboratory; 2023.
25. Azcorra M, Gaertner Z, Davidson C, He Q, Kim H, Nagappan S, Hayes CK, Ramakrishnan C, Fenno L, Kim YS, Deisseroth K, Longnecker R, Awatramani R, Dombeck DA. Unique functional responses differentially map onto genetic subtypes of dopamine neurons. Nat Neurosci. 2023;26(10):1762-74. Epub 20230803. doi: 10.1038/s41593-023-01401-9. PubMed PMID: 37537242; PMCID: PMC10545540.
26. Wu J, Kung J, Dong J, Chang L, Xie C, Habib A, Hawes S, Yang N, Chen V, Liu Z, Evans R, Liang B, Sun L, Ding J, Yu J, Saez-Atienzar S, Tang B, Khaliq Z, Lin DT, Le W, Cai H. Distinct Connectivity and Functionality of Aldehyde Dehydrogenase 1a1-Positive Nigrostriatal Dopaminergic Neurons in Motor Learning. Cell Rep. 2019;28(5):1167-81.e7. doi: 10.1016/j.celrep.2019.06.095. PubMed PMID: 31365862; PMCID: PMC6741442.
27. Poulin JF, Caronia G, Hofer C, Cui Q, Helm B, Ramakrishnan C, Chan CS, Dombeck DA, Deisseroth K, Awatramani R. Mapping projections of molecularly defined dopamine neuron subtypes using

intersectional genetic approaches. Nat Neurosci. 2018;21(9):1260-71. Epub 20180813. doi: 10.1038/s41593-018-0203-4. PubMed PMID: 30104732; PMCID: PMC6342021.
28. Greggio E, Jain S, Kingsbury A, Bandopadhyay R, Lewis P, Kaganovich A, van der Brug MP, Beilina A, Blackinton J, Thomas KJ, Ahmad R, Miller DW, Kesavapany S, Singleton A, Lees A, Harvey RJ, Harvey K, Cookson MR. Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. Neurobiol Dis. 2006;23(2):329-41. Epub 20060605. doi: 10.1016/j.nbd.2006.04.001. PubMed PMID: 16750377.
29. Tokars V, Chen C, Parisiadou L. Closing the structure-to-function gap for LRRK2. Trends Biochem Sci. 2022;47(3):187-8. Epub 20211028. doi: 10.1016/j.tibs.2021.10.003. PubMed PMID: 34756665.
30. Taymans JM, Fell M, Greenamyre T, Hirst WD, Mamais A, Padmanabhan S, Peter I, Rideout H, Thaler A. Perspective on the current state of the LRRK2 field. NPJ Parkinsons Dis. 2023;9(1):104. Epub 20230701. doi: 10.1038/s41531-023-00544-7. PubMed PMID: 37393318; PMCID: PMC10314919.
31. Di Maio R, Hoffman EK, Rocha EM, Keeney MT, Sanders LH, De Miranda BR, Zharikov A, Van Laar A, Stepan AF, Lanz TA, Kofler JK, Burton EA, Alessi DR, Hastings TG, Greenamyre JT. LRRK2 activation in idiopathic Parkinson's disease. Sci Transl Med. 2018;10(451). doi: 10.1126/scitranslmed.aar5429. PubMed PMID: 30045977; PMCID: PMC6344941.
32. Khan SS, Jaimon E, Lin Y-E, Nikoloff J, Tonelli F, Alessi DR, Pfeffer SR. Loss of primary cilia and dopaminergic neuroprotection in pathogenic LRRK2-driven and idiopathic Parkinson’s disease. bioRxiv. 2024:2024.01.15.575737. doi: 10.1101/2024.01.15.575737.
33. Matikainen-Ankney BA, Kezunovic N, Mesias RE, Tian Y, Williams FM, Huntley GW, Benson DL. Altered Development of Synapse Structure and Function in Striatum Caused by Parkinson's Disease-Linked LRRK2-G2019S Mutation. J Neurosci. 2016;36(27):7128-41. doi: 10.1523/jneurosci.3314-15.2016. PubMed PMID: 27383589; PMCID: PMC4938860.
34. Chen C, Soto G, Dumrongprechachan V, Bannon N, Kang S, Kozorovitskiy Y, Parisiadou L. Pathway-specific dysregulation of striatal excitatory synapses by LRRK2 mutations. Elife. 2020;9. Epub 20201002. doi: 10.7554/eLife.58997. PubMed PMID: 33006315; PMCID: PMC7609054.
35. Xenias HS, Chen C, Kang S, Cherian S, Situ X, Shanmugasundaram B, Liu G, Scesa G, Savio Chan C, Parisiadou L. R1441C and G2019S LRRK2 knockin mice have distinct striatal molecular, physiological, and behavioral alterations. Commun Biol. 2022;5(1):1211. Epub 20221110. doi: 10.1038/s42003-022-04136-8. PubMed PMID: 36357506; PMCID: PMC9649688.
36. Yue M, Hinkle KM, Davies P, Trushina E, Fiesel FC, Christenson TA, Schroeder AS, Zhang L, Bowles E, Behrouz B, Lincoln SJ, Beevers JE, Milnerwood AJ, Kurti A, McLean PJ, Fryer JD, Springer W, Dickson DW, Farrer MJ, Melrose HL. Progressive dopaminergic alterations and mitochondrial abnormalities in LRRK2 G2019S knock-in mice. Neurobiol Dis. 2015;78:172-95. Epub 20150331. doi: 10.1016/j.nbd.2015.02.031. PubMed PMID: 25836420; PMCID: PMC4526103.
37. Tozzi A, Tantucci M, Marchi S, Mazzocchetti P, Morari M, Pinton P, Mancini A, Calabresi P. Dopamine D2 receptor-mediated neuroprotection in a G2019S Lrrk2 genetic model of Parkinson's disease. Cell Death Dis. 2018;9(2):204. Epub 20180212. doi: 10.1038/s41419-017-0221-2. PubMed PMID: 29434188; PMCID: PMC5833812.
38. Poulin JF, Zou J, Drouin-Ouellet J, Kim KY, Cicchetti F, Awatramani RB. Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. Cell Rep. 2014;9(3):930-43. Epub 20141030. doi: 10.1016/j.celrep.2014.10.008. PubMed PMID: 25437550; PMCID: PMC4251558.
39. Soden ME, Miller SM, Burgeno LM, Phillips PEM, Hnasko TS, Zweifel LS. Genetic Isolation of Hypothalamic Neurons that Regulate Context-Specific Male Social Behavior. Cell Rep. 2016;16(2):304-13. Epub 20160623. doi: 10.1016/j.celrep.2016.05.067. PubMed PMID: 27346361; PMCID: PMC4945459.
40. Nouri N, Awatramani R. A novel floor plate boundary defined by adjacent En1 and Dbx1 microdomains distinguishes midbrain dopamine and hypothalamic neurons. Development. 2017;144(5):916-27. Epub 20170207. doi: 10.1242/dev.144949. PubMed PMID: 28174244; PMCID: PMC5374351.

41. Kamath T, Abdulraouf A, Burris SJ, Langlieb J, Gazestani V, Nadaf NM, Balderrama K, Vanderburg C, Macosko EZ. Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. Nat Neurosci. 2022;25(5):588-95. Epub 20220505. doi: 10.1038/s41593-022-01061-1. PubMed PMID: 35513515; PMCID: PMC9076534.
42. Anderegg A, Poulin JF, Awatramani R. Molecular heterogeneity of midbrain dopaminergic neurons--Moving toward single cell resolution. FEBS Lett. 2015;589(24 Pt A):3714-26. Epub 20151023. doi: 10.1016/j.febslet.2015.10.022. PubMed PMID: 26505674; PMCID: PMC4679573.
43. Pereira Luppi M, Azcorra M, Caronia-Brown G, Poulin JF, Gaertner Z, Gatica S, Moreno-Ramos OA, Nouri N, Dubois M, Ma YC, Ramakrishnan C, Fenno L, Kim YS, Deisseroth K, Cicchetti F, Dombeck DA, Awatramani R. Sox6 expression distinguishes dorsally and ventrally biased dopamine neurons in the substantia nigra with distinctive properties and embryonic origins. Cell Rep. 2021;37(6):109975. doi: 10.1016/j.celrep.2021.109975. PubMed PMID: 34758317; PMCID: PMC8607753.
44. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015;348(6233):aaa6090. Epub 20150409. doi: 10.1126/science.aaa6090. PubMed PMID: 25858977; PMCID: PMC4662681.
45. Bradl M, Lassmann H. Oligodendrocytes: biology and pathology. Acta Neuropathol. 2010;119(1):37-53. Epub 20091022. doi: 10.1007/s00401-009-0601-5. PubMed PMID: 19847447; PMCID: PMC2799635.
46. Zhang M, Pan X, Jung W, Halpern AR, Eichhorn SW, Lei Z, Cohen L, Smith KA, Tasic B, Yao Z, Zeng H, Zhuang X. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. Nature. 2023;624(7991):343-54. Epub 20231213. doi: 10.1038/s41586-023-06808-9. PubMed PMID: 38092912; PMCID: PMC10719103.
47. Yao Z, van Velthoven CTJ, Kunst M, Zhang M, McMillen D, Lee C, Jung W, Goldy J, Abdelhak A, Aitken M, Baker K, Baker P, Barkan E, Bertagnolli D, Bhandiwad A, Bielstein C, Bishwakarma P, Campos J, Carey D, Casper T, Chakka AB, Chakrabarty R, Chavan S, Chen M, Clark M, Close J, Crichton K, Daniel S, DiValentin P, Dolbeare T, Ellingwood L, Fiabane E, Fliss T, Gee J, Gerstenberger J, Glandon A, Gloe J, Gould J, Gray J, Guilford N, Guzman J, Hirschstein D, Ho W, Hooper M, Huang M, Hupp M, Jin K, Kroll M, Lathia K, Leon A, Li S, Long B, Madigan Z, Malloy J, Malone J, Maltzer Z, Martin N, McCue R, McGinty R, Mei N, Melchor J, Meyerdierks E, Mollenkopf T, Moonsman S, Nguyen TN, Otto S, Pham T, Rimorin C, Ruiz A, Sanchez R, Sawyer L, Shapovalova N, Shepard N, Slaughterbeck C, Sulc J, Tieu M, Torkelson A, Tung H, Valera Cuevas N, Vance S, Wadhwani K, Ward K, Levi B, Farrell C, Young R, Staats B, Wang MM, Thompson CL, Mufti S, Pagan CM, Kruse L, Dee N, Sunkin SM, Esposito L, Hawrylycz MJ, Waters J, Ng L, Smith K, Tasic B, Zhuang X, Zeng H. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature. 2023;624(7991):317-32. Epub 20231213. doi: 10.1038/s41586-023-06812-z. PubMed PMID: 38092916; PMCID: PMC10719114.
48. Langlieb J, Sachdev NS, Balderrama KS, Nadaf NM, Raj M, Murray E, Webber JT, Vanderburg C, Gazestani V, Tward D, Mezias C, Li X, Flowers K, Cable DM, Norton T, Mitra P, Chen F, Macosko EZ. The molecular cytoarchitecture of the adult mouse brain. Nature. 2023;624(7991):333-42. Epub 20231213. doi: 10.1038/s41586-023-06818-7. PubMed PMID: 38092915; PMCID: PMC10719111.
49. Trudeau LE, Hnasko TS, Wallén-Mackenzie A, Morales M, Rayport S, Sulzer D. The multilingual nature of dopamine neurons. Prog Brain Res. 2014;211:141-64. doi: 10.1016/b978-0-444-63425-2.00006-4. PubMed PMID: 24968779; PMCID: PMC4565795.
50. Descarries L, Bérubé-Carrière N, Riad M, Bo GD, Mendez JA, Trudeau LE. Glutamate in dopamine neurons: synaptic versus diffuse transmission. Brain Res Rev. 2008;58(2):290-302. Epub 20071026. doi: 10.1016/j.brainresrev.2007.10.005. PubMed PMID: 18042492.
51. Morales M, Root DH. Glutamate neurons within the midbrain dopamine regions. Neuroscience. 2014;282:60-8. Epub 20140527. doi: 10.1016/j.neuroscience.2014.05.032. PubMed PMID: 24875175; PMCID: PMC4397110.

52. Morales M, Margolis EB. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. Nat Rev Neurosci. 2017;18(2):73-85. Epub 20170105. doi: 10.1038/nrn.2016.165. PubMed PMID: 28053327.
53. Tritsch NX, Granger AJ, Sabatini BL. Mechanisms and functions of GABA co-release. Nat Rev Neurosci. 2016;17(3):139-45. Epub 20160211. doi: 10.1038/nrn.2015.21. PubMed PMID: 26865019; PMCID: PMC6980171.
54. Conrad WS, Oriol L, Faget L, Hnasko TS. Proportion and distribution of neurotransmitter-defined cell types in the ventral tegmental area and substantia nigra pars compacta. bioRxiv. 2024:2024.02.28.582356. doi: 10.1101/2024.02.28.582356.
55. Panman L, Papathanou M, Laguna A, Oosterveen T, Volakakis N, Acampora D, Kurtsdotter I, Yoshitake T, Kehr J, Joodmardi E, Muhr J, Simeone A, Ericson J, Perlmann T. Sox6 and Otx2 control the specification of substantia nigra and ventral tegmental area dopamine neurons. Cell Rep. 2014;8(4):1018-25. Epub 20140807. doi: 10.1016/j.celrep.2014.07.016. PubMed PMID: 25127144.
56. Singh A, Zhi L, Zhang H. LRRK2 and mitochondria: Recent advances and current views. Brain Research. 2019;1702:96-104. doi: https://doi.org/10.1016/j.brainres.2018.06.010.
57. Choi I, Kim B, Byun JW, Baik SH, Huh YH, Kim JH, Mook-Jung I, Song WK, Shin JH, Seo H, Suh YH, Jou I, Park SM, Kang HC, Joe EH. LRRK2 G2019S mutation attenuates microglial motility by inhibiting focal adhesion kinase. Nat Commun. 2015;6:8255. Epub 20150914. doi: 10.1038/ncomms9255. PubMed PMID: 26365310; PMCID: PMC4647842.
58. Parisiadou L, Yu J, Sgobio C, Xie C, Liu G, Sun L, Gu XL, Lin X, Crowley NA, Lovinger DM, Cai H. LRRK2 regulates synaptogenesis and dopamine receptor activation through modulation of PKA activity. Nat Neurosci. 2014;17(3):367-76. Epub 20140126. doi: 10.1038/nn.3636. PubMed PMID: 24464040; PMCID: PMC3989289.
59. Cook DA, Kannarkat GT, Cintron AF, Butkovich LM, Fraser KB, Chang J, Grigoryan N, Factor SA, West AB, Boss JM, Tansey MG. LRRK2 levels in immune cells are increased in Parkinson's disease. NPJ Parkinsons Dis. 2017;3:11. Epub 20170328. doi: 10.1038/s41531-017-0010-8. PubMed PMID: 28649611; PMCID: PMC5459798.
60. West AB, Cowell RM, Daher JP, Moehle MS, Hinkle KM, Melrose HL, Standaert DG, Volpicelli-Daley LA. Differential LRRK2 expression in the cortex, striatum, and substantia nigra in transgenic and nontransgenic rodents. J Comp Neurol. 2014;522(11):2465-80. Epub 20140412. doi: 10.1002/cne.23583. PubMed PMID: 24633735; PMCID: PMC4076169.
61. Pallos J, Jeng S, McWeeney S, Martin I. Dopamine neuron-specific LRRK2 G2019S effects on gene expression revealed by translatome profiling. Neurobiol Dis. 2021;155:105390. Epub 20210511. doi: 10.1016/j.nbd.2021.105390. PubMed PMID: 33984508; PMCID: PMC8223504.
62. Liu G, Sgobio C, Gu X, Sun L, Lin X, Yu J, Parisiadou L, Xie C, Sastry N, Ding J, Lohr KM, Miller GW, Mateo Y, Lovinger DM, Cai H. Selective expression of Parkinson's disease-related Leucine-rich repeat kinase 2 G2019S missense mutation in midbrain dopaminergic neurons impairs dopamine release and dopaminergic gene expression. Hum Mol Genet. 2015;24(18):5299-312. Epub 20150629. doi: 10.1093/hmg/ddv249. PubMed PMID: 26123485; PMCID: PMC4550828.
63. Buckley MT, Sun ED, George BM, Liu L, Schaum N, Xu L, Reyes JM, Goodell MA, Weissman IL, Wyss-Coray T, Rando TA, Brunet A. Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. Nat Aging. 2023;3(1):121-37. Epub 20221219. doi: 10.1038/s43587-022-00335-4. PubMed PMID: 37118510; PMCID: PMC10154228.
64. Zhao T, Goedhart CM, Sam PN, Sabouny R, Lingrell S, Cornish AJ, Lamont RE, Bernier FP, Sinasac D, Parboosingh JS, Vance JE, Claypool SM, Innes AM, Shutt TE. PISD is a mitochondrial disease gene causing skeletal dysplasia, cataracts, and white matter changes. Life Sci Alliance. 2019;2(2). Epub 20190311. doi: 10.26508/lsa.201900353. PubMed PMID: 30858161; PMCID: PMC6412922.

65. Thomas HE, Zhang Y, Stefely JA, Veiga SR, Thomas G, Kozma SC, Mercer CA. Mitochondrial Complex I Activity Is Required for Maximal Autophagy. Cell Rep. 2018;24(9):2404-17.e8. doi: 10.1016/j.celrep.2018.07.101. PubMed PMID: 30157433; PMCID: PMC6298213.
66. Jin K, Yao Z, van Velthoven CTJ, Kaplan ES, Glattfelder K, Barlow ST, Boyer G, Carey D, Casper T, Chakka AB, Chakrabarty R, Clark M, Departee M, Desierto M, Gary A, Gloe J, Goldy J, Guilford N, Guzman J, Hirschstein D, Lee C, Liang E, Pham T, Reding M, Ronellenfitch K, Ruiz A, Sevigny J, Shapovalova N, Shulga L, Sulc J, Torkelson A, Tung H, Levi B, Sunkin SM, Dee N, Esposito L, Smith K, Tasic B, Zeng H. Cell-type specific molecular signatures of aging revealed in a brain-wide transcriptomic cell-type atlas. bioRxiv. 2023. Epub 20230727. doi: 10.1101/2023.07.26.550355. PubMed PMID: 38168182; PMCID: PMC10760145.
67. Zhang F, Yao Y, Miao N, Wang N, Xu X, Yang C. Neuroprotective effects of microRNA 124 in Parkinson's disease mice. Archives of Gerontology and Geriatrics. 2022;99:104588. doi: https://doi.org/10.1016/j.archger.2021.104588.
68. 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. J Neuroinflammation. 2018;15(1):13. Epub 20180112. doi: 10.1186/s12974-018-1053-4. PubMed PMID: 29329581; PMCID: PMC5767033.
69. Han D, Dong X, Zheng D, Nao J. MiR-124 and the Underlying Therapeutic Promise of Neurodegenerative Disorders. Front Pharmacol. 2019;10:1555. Epub 20200117. doi: 10.3389/fphar.2019.01555. PubMed PMID: 32009959; PMCID: PMC6978711.
70. Angelopoulou E, Paudel YN, Piperi C. miR-124 and Parkinson's disease: A biomarker with therapeutic potential. Pharmacol Res. 2019;150:104515. Epub 20191107. doi: 10.1016/j.phrs.2019.104515. PubMed PMID: 31707035.
71. González-Rodríguez P, Zampese E, Stout KA, Guzman JN, Ilijic E, Yang B, Tkatch T, Stavarache MA, Wokosin DL, Gao L, Kaplitt MG, López-Barneo J, Schumacker PT, Surmeier DJ. Disruption of mitochondrial complex I induces progressive parkinsonism. Nature. 2021;599(7886):650-6. Epub 20211103. doi: 10.1038/s41586-021-04059-0. PubMed PMID: 34732887; PMCID: PMC9189968.
72. Surmeier DJ, Obeso JA, Halliday GM. Selective neuronal vulnerability in Parkinson disease. Nat Rev Neurosci. 2017;18(2):101-13. doi: 10.1038/nrn.2016.178. PubMed PMID: 28104909; PMCID: PMC5564322.
73. Mortiboys H, Johansen KK, Aasly JO, Bandmann O. Mitochondrial impairment in patients with Parkinson disease with the G2019S mutation in LRRK2. Neurology. 2010;75(22):2017-20. doi: 10.1212/WNL.0b013e3181ff9685. PubMed PMID: 21115957.
74. Singh F, Prescott AR, Rosewell P, Ball G, Reith AD, Ganley IG. Pharmacological rescue of impaired mitophagy in Parkinson's disease-related LRRK2 G2019S knock-in mice. Elife. 2021;10. Epub 20210803. doi: 10.7554/eLife.67604. PubMed PMID: 34340748; PMCID: PMC8331189.
75. Liu H, Ho PW, Leung CT, Pang SY, Chang EES, Choi ZY, Kung MH, Ramsden DB, Ho SL. Aberrant mitochondrial morphology and function associated with impaired mitophagy and DNM1L-MAPK/ERK signaling are found in aged mutant Parkinsonian LRRK2(R1441G) mice. Autophagy. 2021;17(10):3196-220. Epub 20201210. doi: 10.1080/15548627.2020.1850008. PubMed PMID: 33300446; PMCID: PMC8526027.
76. Nguyen M, Krainc D. LRRK2 phosphorylation of auxilin mediates synaptic defects in dopaminergic neurons from patients with Parkinson's disease. Proc Natl Acad Sci U S A. 2018;115(21):5576-81. Epub 20180507. doi: 10.1073/pnas.1717590115. PubMed PMID: 29735704; PMCID: PMC6003526.
77. Soukup SF, Kuenen S, Vanhauwaert R, Manetsberger J, Hernández-Díaz S, Swerts J, Schoovaerts N, Vilain S, Gounko NV, Vints K, Geens A, De Strooper B, Verstreken P. A LRRK2-Dependent EndophilinA Phosphoswitch Is Critical for Macroautophagy at Presynaptic Terminals. Neuron. 2016;92(4):829-44. Epub 20161006. doi: 10.1016/j.neuron.2016.09.037. PubMed PMID: 27720484.

78. Islam MS, Nolte H, Jacob W, Ziegler AB, Pütz S, Grosjean Y, Szczepanowska K, Trifunovic A, Braun T, Heumann H, Heumann R, Hovemann B, Moore DJ, Krüger M. Human R1441C LRRK2 regulates the synaptic vesicle proteome and phosphoproteome in a Drosophila model of Parkinson's disease. Hum Mol Genet. 2016;25(24):5365-82. doi: 10.1093/hmg/ddw352. PubMed PMID: 27794539; PMCID: PMC6078604.
79. Boecker CA, Goldsmith J, Dou D, Cajka GG, Holzbaur ELF. Increased LRRK2 kinase activity alters neuronal autophagy by disrupting the axonal transport of autophagosomes. Curr Biol. 2021;31(10):2140-54.e6. Epub 20210324. doi: 10.1016/j.cub.2021.02.061. PubMed PMID: 33765413; PMCID: PMC8154747.
80. Skelton PD, Tokars V, Parisiadou L. LRRK2 at Striatal Synapses: Cell-Type Specificity and Mechanistic Insights. Cells. 2022;11(1). Epub 20220105. doi: 10.3390/cells11010169. PubMed PMID: 35011731; PMCID: PMC8750662.
81. Pischedda F, Piccoli G. LRRK2 at the pre-synaptic site: A 16-years perspective. Journal of Neurochemistry. 2021;157(2):297-311. doi: https://doi.org/10.1111/jnc.15240.
82. Pan PY, Li X, Wang J, Powell J, Wang Q, Zhang Y, Chen Z, Wicinski B, Hof P, Ryan TA, Yue Z. Parkinson's Disease-Associated LRRK2 Hyperactive Kinase Mutant Disrupts Synaptic Vesicle Trafficking in Ventral Midbrain Neurons. J Neurosci. 2017;37(47):11366-76. Epub 20171020. doi: 10.1523/jneurosci.0964-17.2017. PubMed PMID: 29054882; PMCID: PMC5700420.
83. Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijsouw T, Coba MP, Cornelisse LN, Farrell RJ, Goldschmidt HL, Howrigan DP, Hussain NK, Imig C, de Jong APH, Jung H, Kohansalnodehi M, Kramarz B, Lipstein N, Lovering RC, MacGillavry H, Mariano V, Mi H, Ninov M, Osumi-Sutherland D, Pielot R, Smalla KH, Tang H, Tashman K, Toonen RFG, Verpelli C, Reig-Viader R, Watanabe K, van Weering J, Achsel T, Ashrafi G, Asi N, Brown TC, De Camilli P, Feuermann M, Foulger RE, Gaudet P, Joglekar A, Kanellopoulos A, Malenka R, Nicoll RA, Pulido C, de Juan-Sanz J, Sheng M, Südhof TC, Tilgner HU, Bagni C, Bayés À, Biederer T, Brose N, Chua JJE, Dieterich DC, Gundelfinger ED, Hoogenraad C, Huganir RL, Jahn R, Kaeser PS, Kim E, Kreutz MR, McPherson PS, Neale BM, O'Connor V, Posthuma D, Ryan TA, Sala C, Feng G, Hyman SE, Thomas PD, Smit AB, Verhage M. SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. Neuron. 2019;103(2):217-34.e4. Epub 20190603. doi: 10.1016/j.neuron.2019.05.002. PubMed PMID: 31171447; PMCID: PMC6764089.
84. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, Bras J, Young E, von Coelln R, Simon-Sanchez J, Schulte C, Sharma M, Krohn L, Pihlstrom L, Siitonen A, Iwaki H, Leonard H, Faghri F, Gibbs JR, Hernandez DG, Scholz SW, Botia JA, Martinez M, Corvol JC, Lesage S, Jankovic J, Shulman LM, Sutherland M, Tienari P, Majamaa K, Toft M, Andreassen OA, Bangale T, Brice A, Yang J, Gan-Or Z, Gasser T, Heutink P, Shulman JM, Wood NW, Hinds DA, Hardy JA, Morris HR, Gratten J, Visscher PM, Graham RR, Singleton AB, andMe Research T, System Genomics of Parkinson's Disease C, International Parkinson's Disease Genomics C. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019;18(12):1091-102. doi: 10.1016/S1474-4422(19)30320-5. PubMed PMID: 31701892; PMCID: PMC8422160.
85. Zhang MJ, Hou K, Dey KK, Sakaue S, Jagadeesh KA, Weinand K, Taychameekiatchai A, Rao P, Pisco AO, Zou J, Wang B, Gandal M, Raychaudhuri S, Pasaniuc B, Price AL. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. Nature Genetics. 2022;54(10):1572-80. doi: 10.1038/s41588-022-01167-z.
86. Yaghmaeian Salmani B, Lahti L, Gillberg L, Jacobsen JK, Mantas I, Svenningsson P, Perlmann T. Transcriptomic atlas of midbrain dopamine neurons uncovers differential vulnerability in a Parkinsonism lesion model. Elife. 2024;12. Epub 20240408. doi: 10.7554/eLife.89482. PubMed PMID: 38587883; PMCID: PMC11001297.

87. Blaess S, Bodea GO, Kabanova A, Chanet S, Mugniery E, Derouiche A, Stephen D, Joyner AL. Temporal-spatial changes in Sonic Hedgehog expression and signaling reveal different potentials of ventral mesencephalic progenitors to populate distinct ventral midbrain nuclei. Neural Dev. 2011;6:29. Epub 20110620. doi: 10.1186/1749-8104-6-29. PubMed PMID: 21689430; PMCID: PMC3135491.
88. Joksimovic M, Anderegg A, Roy A, Campochiaro L, Yun B, Kittappa R, McKay R, Awatramani R. Spatiotemporally separable Shh domains in the midbrain define distinct dopaminergic progenitor pools. Proc Natl Acad Sci U S A. 2009;106(45):19185-90. Epub 20091022. doi: 10.1073/pnas.0904285106. PubMed PMID: 19850875; PMCID: PMC2776440.
89. Yan CH, Levesque M, Claxton S, Johnson RL, Ang SL. Lmx1a and lmx1b function cooperatively to regulate proliferation, specification, and differentiation of midbrain dopaminergic progenitors. J Neurosci. 2011;31(35):12413-25. doi: 10.1523/jneurosci.1077-11.2011. PubMed PMID: 21880902; PMCID: PMC6703256.
90. Andersson E, Tryggvason U, Deng Q, Friling S, Alekseenko Z, Robert B, Perlmann T, Ericson J. Identification of intrinsic determinants of midbrain dopamine neurons. Cell. 2006;124(2):393-405. doi: 10.1016/j.cell.2005.10.037. PubMed PMID: 16439212.
91. Cox J, Witten IB. Striatal circuits for reward learning and decision-making. Nat Rev Neurosci. 2019;20(8):482-94. doi: 10.1038/s41583-019-0189-2. PubMed PMID: 31171839; PMCID: PMC7231228.
92. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-902.e21. Epub 20190606. doi: 10.1016/j.cell.2019.05.031. PubMed PMID: 31178118; PMCID: PMC6687398.
93. Tang L, Xu N, Huang M, Yi W, Sang X, Shao M, Li Y, Hao ZZ, Liu R, Shen Y, Yue F, Liu X, Xu C, Liu S. A primate nigrostriatal atlas of neuronal vulnerability and resilience in a model of Parkinson's disease. Nat Commun. 2023;14(1):7497. Epub 20231118. doi: 10.1038/s41467-023-43213-2. PubMed PMID: 37980356; PMCID: PMC10657376.
94. Yamada T, McGeer PL, Baimbridge KG, McGeer EG. Relative sparing in Parkinson's disease of substantia nigra dopamine neurons containing calbindin-D28K. Brain Res. 1990;526(2):303-7. doi: 10.1016/0006-8993(90)91236-a. PubMed PMID: 2257487.
95. Damier P, Hirsch EC, Agid Y, Graybiel AM. The substantia nigra of the human brain. I. Nigrosomes and the nigral matrix, a compartmental organization based on calbindin D(28K) immunohistochemistry. Brain. 1999;122 (Pt 8):1421-36. doi: 10.1093/brain/122.8.1421. PubMed PMID: 10430829.
96. Kish SJ, Shannak K, Hornykiewicz O. Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications. N Engl J Med. 1988;318(14):876-80. doi: 10.1056/nejm198804073181402. PubMed PMID: 3352672.
97. Kordower JH, Olanow CW, Dodiya HB, Chu Y, Beach TG, Adler CH, Halliday GM, Bartus RT. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. Brain. 2013;136(Pt 8):2419-31. doi: 10.1093/brain/awt192. PubMed PMID: 23884810; PMCID: PMC3722357.
98. Mendonça MD, Ferreira PC, Oliveira F, Barbosa R, Meira B, Costa DC, Oliveira-Maia AJ, da Silva JA. Relative sparing of dopaminergic terminals in the caudate nucleus is a feature of rest tremor in Parkinson's disease. NPJ Parkinsons Dis. 2024;10(1):209. Epub 20241118. doi: 10.1038/s41531-024-00818-8. PubMed PMID: 39557871; PMCID: PMC11574046.
99. Schmack K, Bosc M, Ott T, Sturgill JF, Kepecs A. Striatal dopamine mediates hallucination-like perception in mice. Science. 2021;372(6537). doi: 10.1126/science.abf4740. PubMed PMID: 33795430.
100. Mandemakers W, Snellinx A, O'Neill MJ, de Strooper B. LRRK2 expression is enriched in the striosomal compartment of mouse striatum. Neurobiol Dis. 2012;48(3):582-93. Epub 20120729. doi: 10.1016/j.nbd.2012.07.017. PubMed PMID: 22850484.
101. Pischedda F, Piccoli G. LRRK2 at the pre-synaptic site: A 16-years perspective. J Neurochem. 2021;157(2):297-311. Epub 20210205. doi: 10.1111/jnc.15240. PubMed PMID: 33206398.

102. Siletti K, Hodge R, Mossi Albiach A, Lee KW, Ding SL, Hu L, Lönnerberg P, Bakken T, Casper T, Clark M, Dee N, Gloe J, Hirschstein D, Shapovalova NV, Keene CD, Nyhus J, Tung H, Yanny AM, Arenas E, Lein ES, Linnarsson S. Transcriptomic diversity of cell types across the adult human brain. Science. 2023;382(6667):eadd7046. Epub 20231013. doi: 10.1126/science.add7046. PubMed PMID: 37824663.
103. Choudhary S, Satija R. Comparison and evaluation of statistical error models for scRNA-seq. Genome Biol. 2022;23(1):27. Epub 20220118. doi: 10.1186/s13059-021-02584-9. PubMed PMID: 35042561; PMCID: PMC8764781.
104. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology. 2024;42(2):293-304. doi: 10.1038/s41587-023-01767-y.
105. Ouyang JF, Kamaraj US, Cao EY, Rackham OJL. ShinyCell: simple and sharable visualization of single-cell gene expression data. Bioinformatics. 2021;37(19):3374-6. doi: 10.1093/bioinformatics/btab209. PubMed PMID: 33774659.
106. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. Epub 20180206. doi: 10.1186/s13059-017-1382-0. PubMed PMID: 29409532; PMCID: PMC5802054.
107. Linderman GC, Zhao J, Roulis M, Bielecki P, Flavell RA, Nadler B, Kluger Y. Zero-preserving imputation of single-cell RNA-seq data. Nat Commun. 2022;13(1):192. Epub 20220111. doi: 10.1038/s41467-021-27729-z. PubMed PMID: 35017482; PMCID: PMC8752663.
108. Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. Nat Commun. 2018;9(1):884. Epub 20180228. doi: 10.1038/s41467-018-03282-0. PubMed PMID: 29491377; PMCID: PMC5830442.

Figure 1

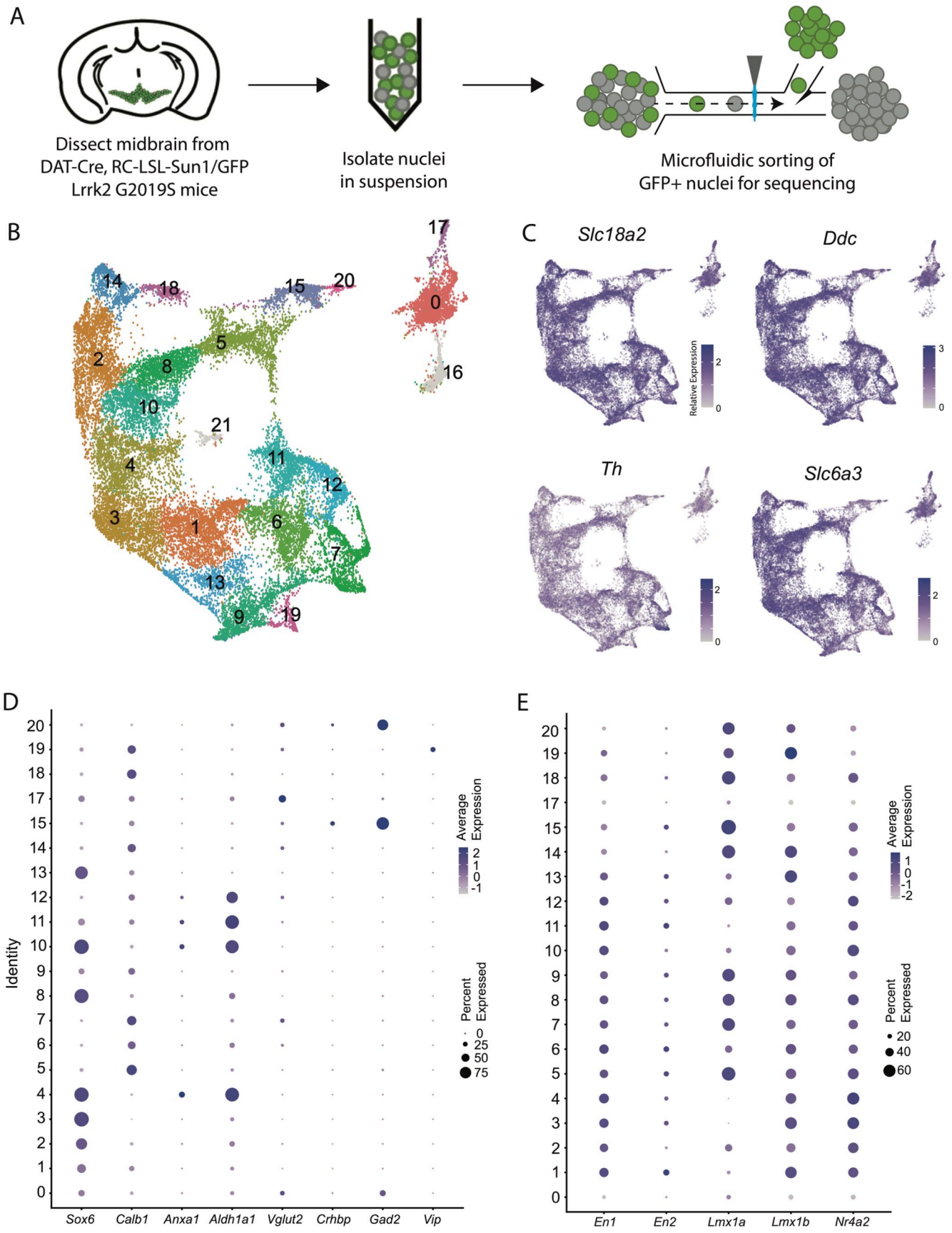
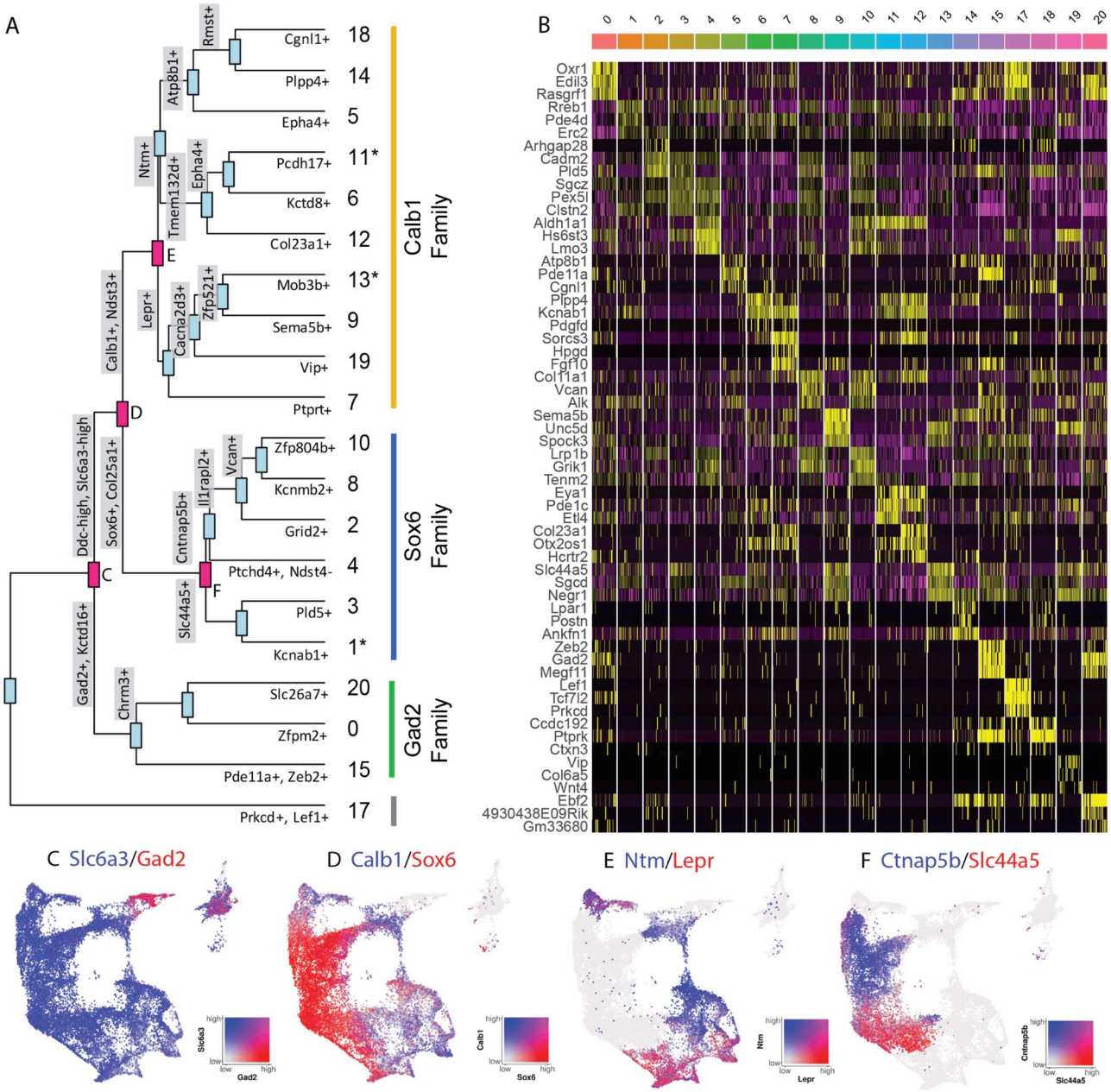

A
Dissect midbrain from
DAT-Cre, RC-LSL-Sun1/GFP
Lrrk2 G2019S mice
Isolate nuclei
in suspension
Microfluidic sorting of
GFP+ nuclei for sequencing
B
C
Slc18a2
Ddc
Th
Slc6a3
Relative Expression
D
Identity
Average Expression
Percent Expressed
Sox6
Calb1
Anxa1
Aldh1a1
Vglut2
Crhbp
Gad2
Vip
E
Average Expression
Percent Expressed
En1
En2
Lmx1a
Lmx1b
Nr4a2

Figure 2



G

Cluster Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	20
Azcorra & Gaertner et al., Nat. Neuro. 2023	12	3	2	3	4	11	6, 7	5	1	10	1	6	6	3	9	8	15	9	10	8
Poulin et al., TiNS 2020	N/A	Sox6+, Aldh1a1-	Sox6+, Aldh1a1-	Sox6+, Aldh1a1-	Aldh1a1+, Sox6+	Poorly Defined	Aldh1a1+, Otx2+	Vglut2+ (VTA)	Aldh1a1+, Sox6+	Vglut2+ (VTA)	Aldh1a1+, Sox6+	Aldh1a1+, Otx2+	Aldh1a1+, Otx2+	Sox6+, Aldh1a1-	Vglut2+ (SNc)	Vgat+	N/A	Vglut2+ (SNc)	Vip+	Vgat+
Unique Genetic Signature	Gad2, Chrm3, Zfpm2	Sox6, Slc44a5, Kcnab1	Sox6, Cntnap5b, Il1rapl2, Grid2	Sox6, Slc44a5, Pld5	Sox6, Cntnap5b, Ptchd4	Calb1, Ntm, Atp8b1, Epha4	Calb1, Ntm, Tmem132d, Epha4, Kctd8	Calb1, Lepr, Ptprt	Sox6, Cntnap5b, Il1rapl2, Vcan, Kcnmb2	Calb1, Lepr, Cacna2d3, Zfp521, Sema5b	Sox6, Cntnap5b, Il1rapl2, Vcan, Zfp804b	Calb1, Ntm, Tmem132d, Epha4, Pcdh17	Calb1, Ntm, Tmem132d, Col23a1	Calb1, Lepr, Cacna2d3, Zfp521, Mob3b	Calb1, Ntm, Atp8b1, Rmst, Plpp4	Gad2, Pde11a, Zeb2	Lef1, Prkcd	Calb1, Ntm, Atp8b1, Rmst, Cgnl1	Calb1, Lepr, Cacna2d3, Vip	Gad2, Chrm3, Slc26a7
Proposed Shorthand Identity	Gad2: Syndig1	Sox6: Tmem132d	Sox6: Arhgap28	Sox6: March3	Sox6: Tafa1	Calb1: Pde11a	Calb1: Kctd8	Calb1: Ptprt	Sox6: Kcnmb2	Calb1: Sulf1	Sox6: Vcan	Calb1: Stac	Calb1: Chrm2	Calb1: Sox6	Calb1: Lpar1	Gad2: Egfr	Lef1	Calb1: Ccdc192	Calb1: Gipr	Gad2: Ebf2

Figure 3

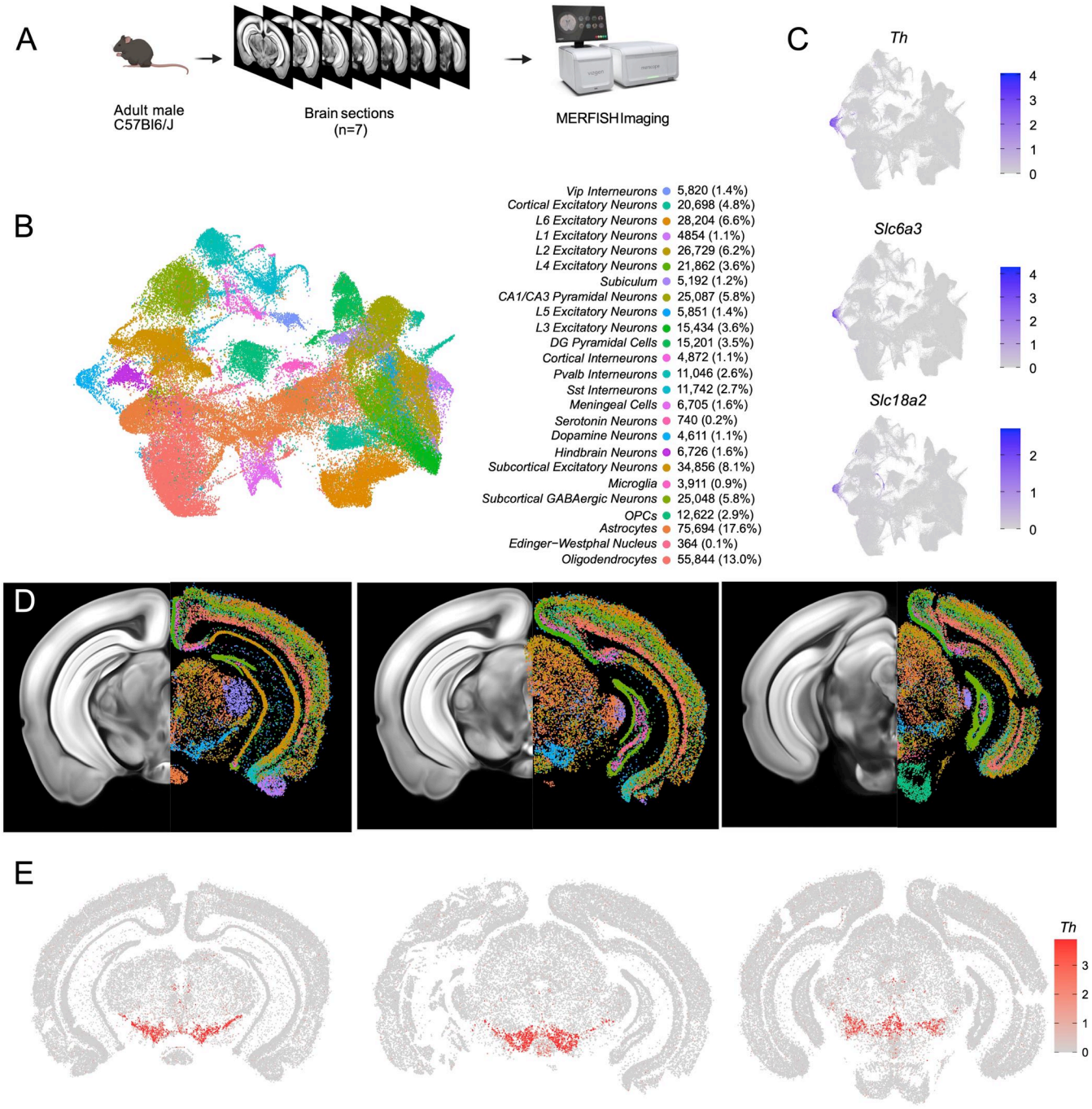

A
Adult male
C57Bl6/J
Brain sections
(n=7)
MERFISH Imaging
B
Vip Interneurons 5,820 (1.4%)
Cortical Excitatory Neurons 20,698 (4.8%)
L6 Excitatory Neurons 28,204 (6.6%)
L1 Excitatory Neurons 4854 (1.1%)
L2 Excitatory Neurons 26,729 (6.2%)
L4 Excitatory Neurons 21,862 (3.6%)
Subiculum 5,192 (1.2%)
CA1/CA3 Pyramidal Neurons 25,087 (5.8%)
L5 Excitatory Neurons 5,851 (1.4%)
L3 Excitatory Neurons 15,434 (3.6%)
DG Pyramidal Cells 15,201 (3.5%)
Cortical Interneurons 4,872 (1.1%)
Pvalb Interneurons 11,046 (2.6%)
Sst Interneurons 11,742 (2.7%)
Meningeal Cells 6,705 (1.6%)
Serotonin Neurons 740 (0.2%)
Dopamine Neurons 4,611 (1.1%)
Hindbrain Neurons 6,726 (1.6%)
Subcortical Excitatory Neurons 34,856 (8.1%)
Microglia 3,911 (0.9%)
Subcortical GABAergic Neurons 25,048 (5.8%)
OPCs 12,622 (2.9%)
Astrocytes 75,694 (17.6%)
Edinger−Westphal Nucleus 364 (0.1%)
Oligodendrocytes 55,844 (13.0%)
C
Th
4
3
2
1
0
Slc6a3
4
3
2
1
0
Slc18a2
2
1
0
D
E
Th
3
2
1
0

Figure 4

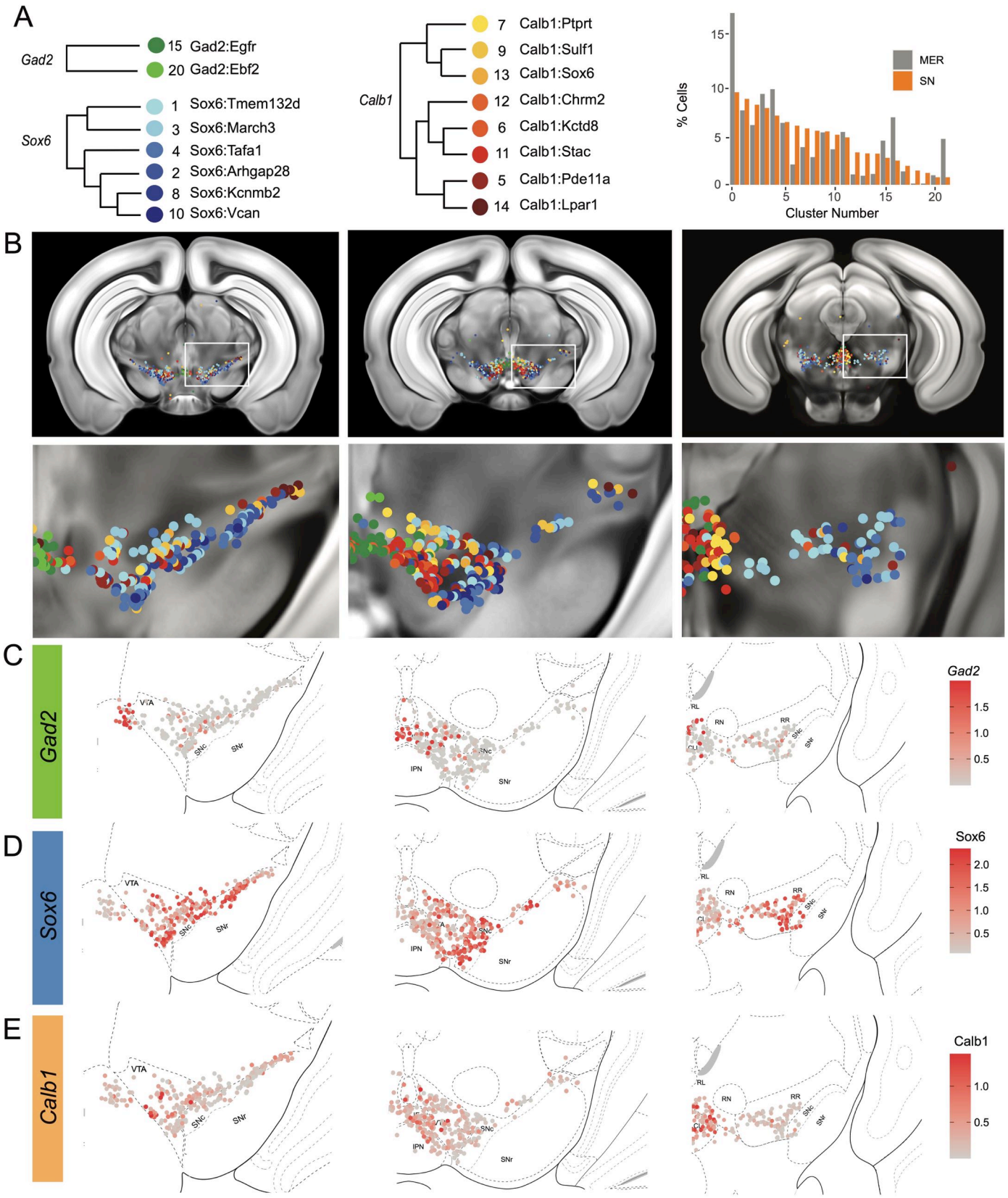

A
Gad2
15 Gad2:Egfr
20 Gad2:Ebf2
Sox6
1 Sox6:Tmem132d
3 Sox6:March3
4 Sox6:Tafa1
2 Sox6:Arhgap28
8 Sox6:Kcnmb2
10 Sox6:Vcan
Calb1
7 Calb1:Ptprt
9 Calb1:Sulf1
13 Calb1:Sox6
12 Calb1:Chrm2
6 Calb1:Kctd8
11 Calb1:Stac
5 Calb1:Pde11a
14 Calb1:Lpar1
% Cells
Cluster Number
MER
SN
B
C
Gad2
D
Sox6
E
Calb1
VTA
SNc
SNr
IPN
RL
RN
RR
CLi

Figure 5

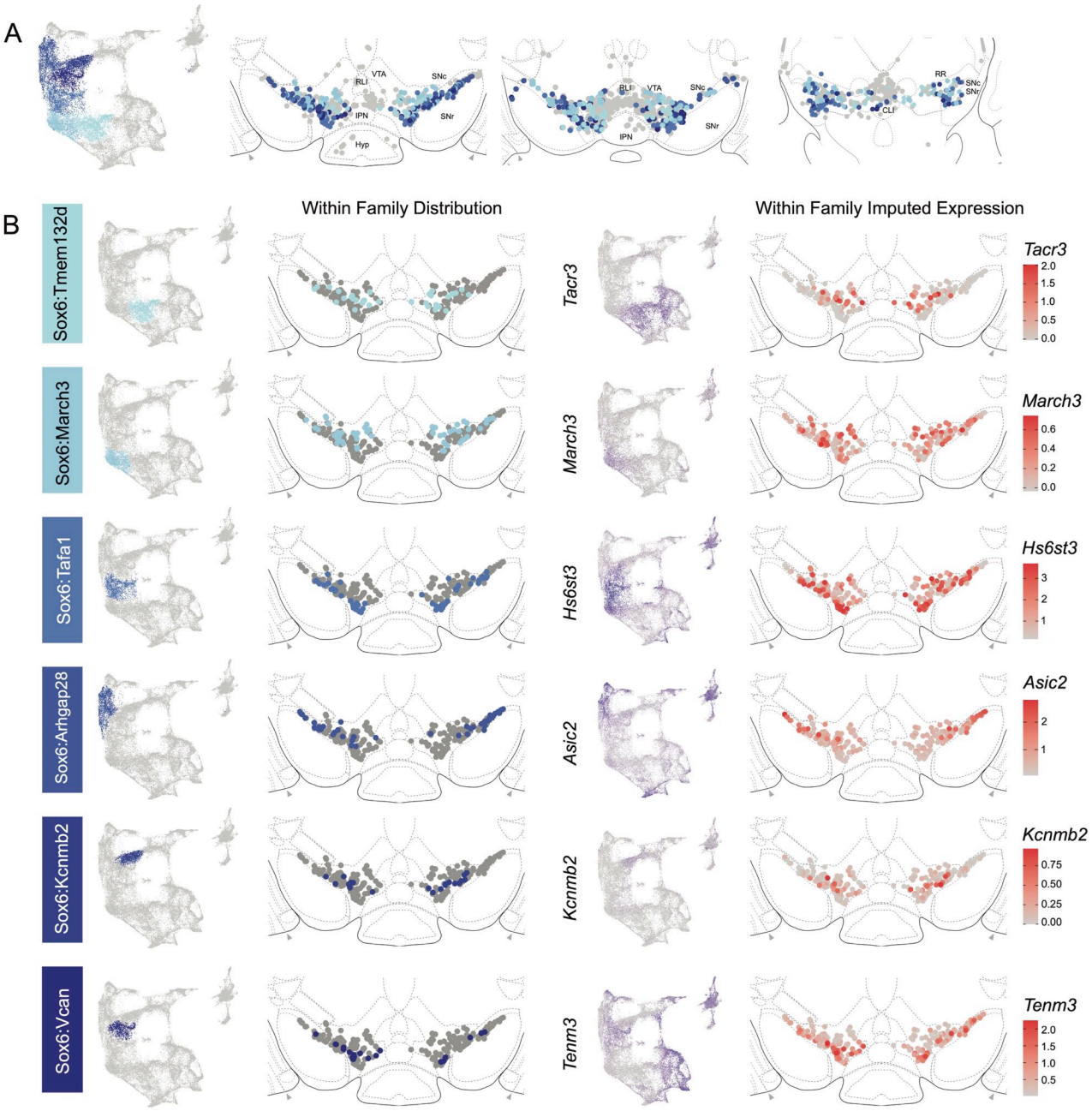

A
VTA
RLi
SNc
IPN
SNr
Hyp
RLi
VTA
SNc
IPN
SNr
RR
SNc
SNr
CLi
B
Within Family Distribution
Within Family Imputed Expression
Sox6:Tmem132d
Sox6:March3
Sox6:Tafa1
Sox6:Arhgap28
Sox6:Kcnmb2
Sox6:Vcan
Tacr3
March3
Hs6st3
Asic2
Kcnmb2
Tenm3
Tacr3
2.0
1.5
1.0
0.5
0.0
March3
0.6
0.4
0.2
0.0
Hs6st3
3
2
1
Asic2
2
1
Kcnmb2
0.75
0.50
0.25
0.00
Tenm3
2.0
1.5
1.0
0.5

Figure 6

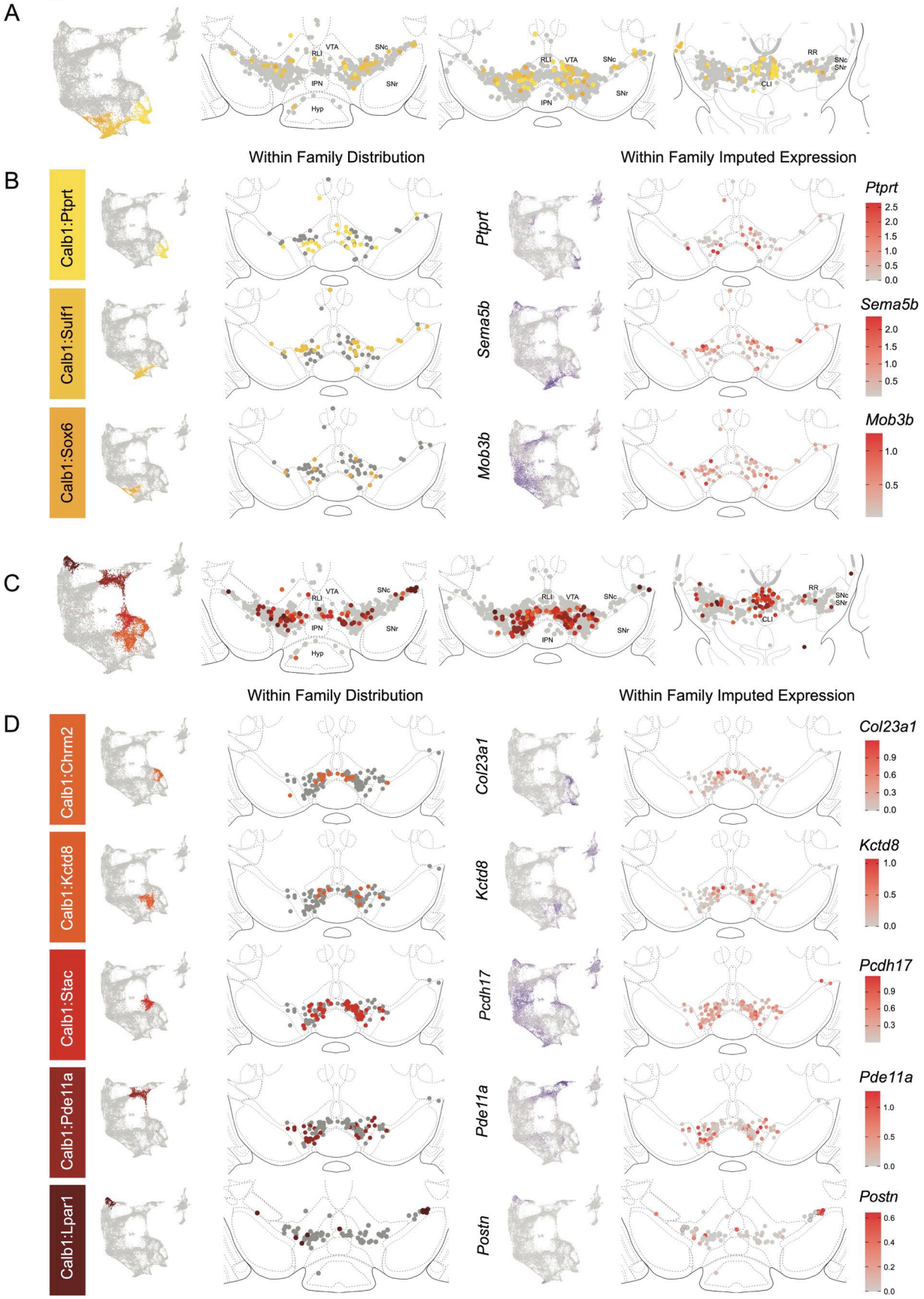

A
VTA
RLi
SNc
IPN
SNr
Hyp
RR
CLi
Within Family Distribution
Within Family Imputed Expression
B
Calb1:Ptprt
Calb1:Sulf1
Calb1:Sox6
Ptprt
Sema5b
Mob3b
C
D
Calb1:Chrm2
Calb1:Kctd8
Calb1:Stac
Calb1:Pde11a
Calb1:Lpar1
Col23a1
Kctd8
Pcdh17
Pde11a
Postn

Figure 7

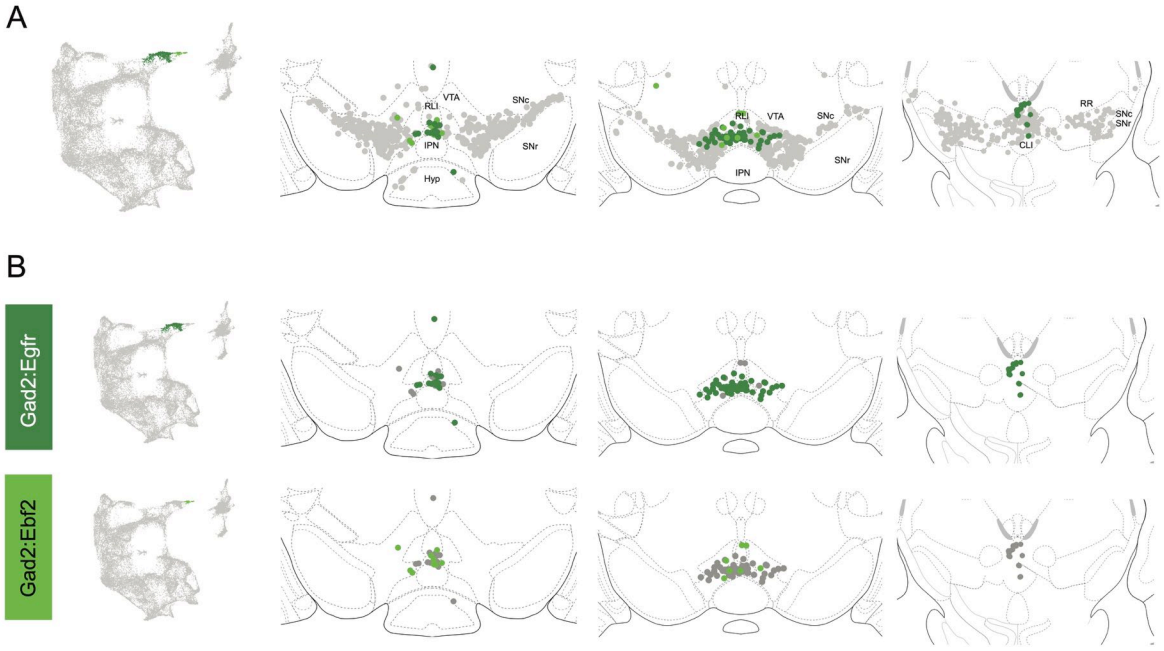

A
VTA
RLi
SNc
IPN
SNr
Hyp
RLi
VTA
SNc
IPN
SNr
RR
SNc
SNr
CLi
B
Gad2:Egfr
Gad2:Ebf2

Figure 8

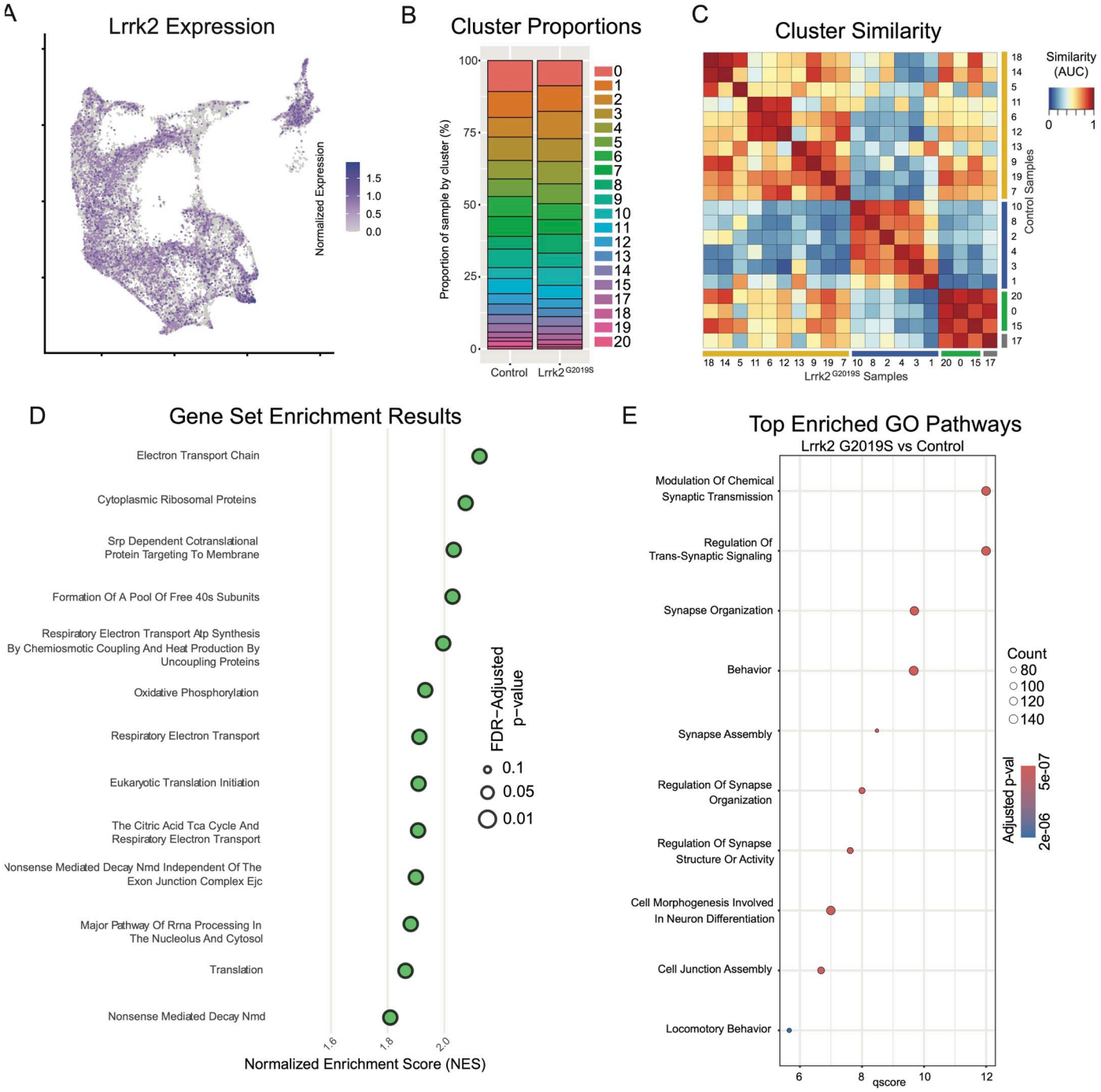

A
Lrrk2 Expression
Normalized Expression
1.5
1.0
0.5
0.0
B
Cluster Proportions
Proportion of sample by cluster (%)
100
75
50
25
0
Control
Lrrk2 G2019S
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17 18 19 20
C
Cluster Similarity
Similarity (AUC)
0
1
Control Samples
18 14 5 11 6 12 13 9 19 7 10 8 2 4 3 1 20 0 15 17
Lrrk2 G2019S Samples
D
Gene Set Enrichment Results
Electron Transport Chain
Cytoplasmic Ribosomal Proteins
Srp Dependent Cotranslational Protein Targeting To Membrane
Formation Of A Pool Of Free 40s Subunits
Respiratory Electron Transport Atp Synthesis By Chemiosmotic Coupling And Heat Production By Uncoupling Proteins
Oxidative Phosphorylation
Respiratory Electron Transport
Eukaryotic Translation Initiation
The Citric Acid Tca Cycle And Respiratory Electron Transport
Nonsense Mediated Decay Nmd Independent Of The Exon Junction Complex Ejc
Major Pathway Of Rrna Processing In The Nucleolus And Cytosol
Translation
Nonsense Mediated Decay Nmd
1.6
1.8
2.0
Normalized Enrichment Score (NES)
FDR-Adjusted p-value
0.1
0.05
0.01
E
Top Enriched GO Pathways
Lrrk2 G2019S vs Control
Modulation Of Chemical Synaptic Transmission
Regulation Of Trans-Synaptic Signaling
Synapse Organization
Behavior
Synapse Assembly
Regulation Of Synapse Organization
Regulation Of Synapse Structure Or Activity
Cell Morphogenesis Involved In Neuron Differentiation
Cell Junction Assembly
Locomotory Behavior
6
8
10
12
qscore
Count
80
100
120
140
Adjusted p-val
5e-07
2e-06

Figure 9

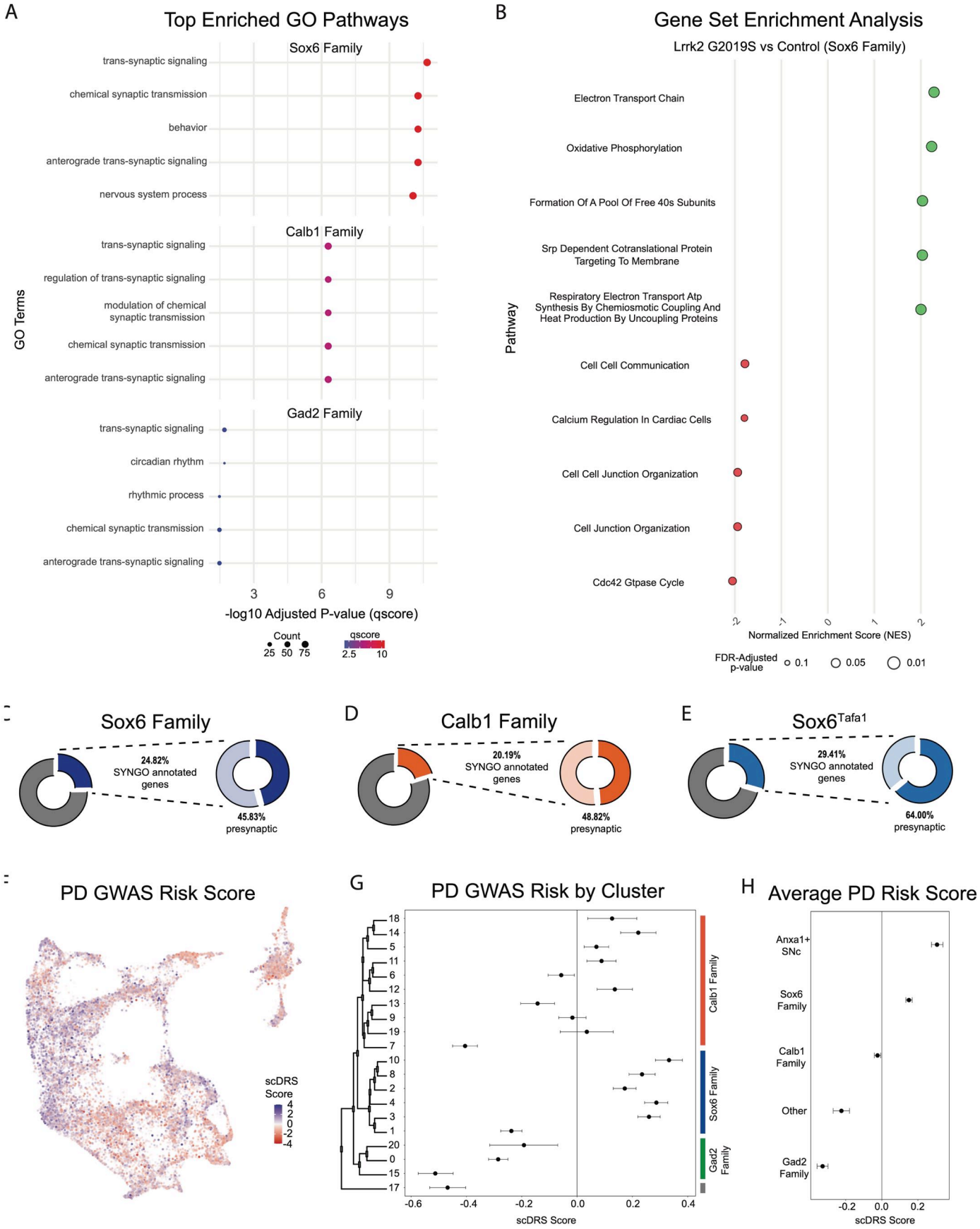

A
Top Enriched GO Pathways
Sox6 Family
trans-synaptic signaling
chemical synaptic transmission
behavior
anterograde trans-synaptic signaling
nervous system process
Calb1 Family
trans-synaptic signaling
regulation of trans-synaptic signaling
modulation of chemical synaptic transmission
chemical synaptic transmission
anterograde trans-synaptic signaling
Gad2 Family
trans-synaptic signaling
circadian rhythm
rhythmic process
chemical synaptic transmission
anterograde trans-synaptic signaling
GO Terms
3
6
9
-log10 Adjusted P-value (qscore)
Count
25 50 75
qscore
2.5 10
B
Gene Set Enrichment Analysis
Lrrk2 G2019S vs Control (Sox6 Family)
Electron Transport Chain
Oxidative Phosphorylation
Formation Of A Pool Of Free 40s Subunits
Srp Dependent Cotranslational Protein Targeting To Membrane
Respiratory Electron Transport Atp Synthesis By Chemiosmotic Coupling And Heat Production By Uncoupling Proteins
Cell Cell Communication
Calcium Regulation In Cardiac Cells
Cell Cell Junction Organization
Cell Junction Organization
Cdc42 Gtpase Cycle
Pathway
-2
-1
0
1
2
Normalized Enrichment Score (NES)
FDR-Adjusted p-value
0.1
0.05
0.01
Sox6 Family
24.82%
SYNGO annotated genes
45.83%
presynaptic
D
Calb1 Family
20.19%
SYNGO annotated genes
48.82%
presynaptic
E
Sox6Tafa1
29.41%
SYNGO annotated genes
64.00%
presynaptic
PD GWAS Risk Score
scDRS Score
4
2
0
-2
-4
G
PD GWAS Risk by Cluster
18
14
5
11
6
12
13
9
19
7
10
8
2
4
3
1
20
0
15
17
Calb1 Family
Sox6 Family
Gad2 Family
-0.6
-0.4
-0.2
0.0
0.2
0.4
scDRS Score
H
Average PD Risk Score
Anxa1+ SNc
Sox6 Family
Calb1 Family
Other
Gad2 Family
-0.2
0.0
0.2
scDRS Score

Figure S1

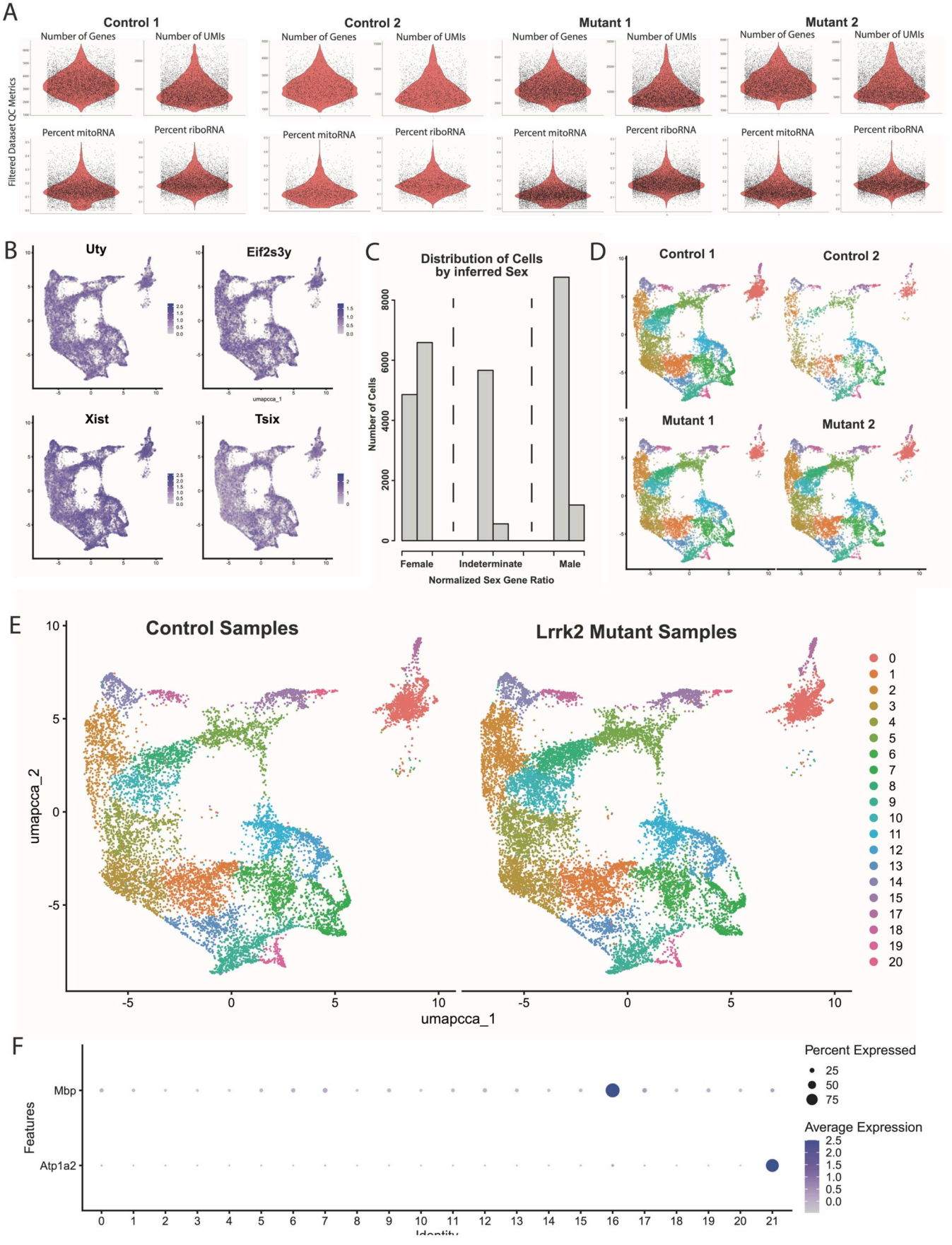

A
Control 1
Control 2
Mutant 1
Mutant 2
Number of Genes
Number of UMIs
Percent mitoRNA
Percent riboRNA
Filtered Dataset QC Metrics
B
Uty
Eif2s3y
Xist
Tsix
umapcca_1
C
Distribution of Cells by inferred Sex
Number of Cells
Female
Indeterminate
Male
Normalized Sex Gene Ratio
D
Control 1
Control 2
Mutant 1
Mutant 2
E
Control Samples
Lrrk2 Mutant Samples
umapcca_2
umapcca_1
F
Mbp
Atp1a2
Features
Identity
Percent Expressed
Average Expression

Figure S2

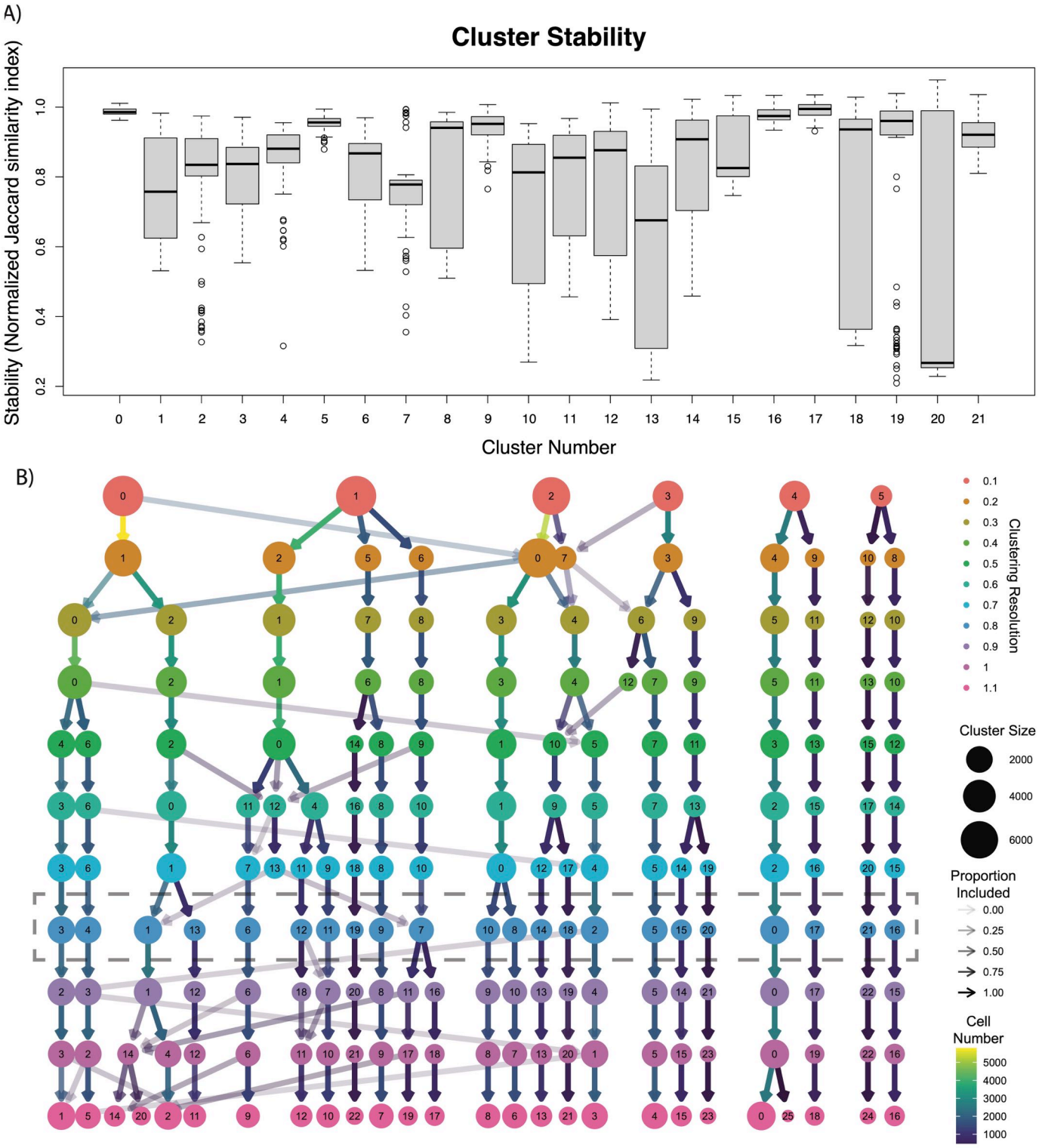
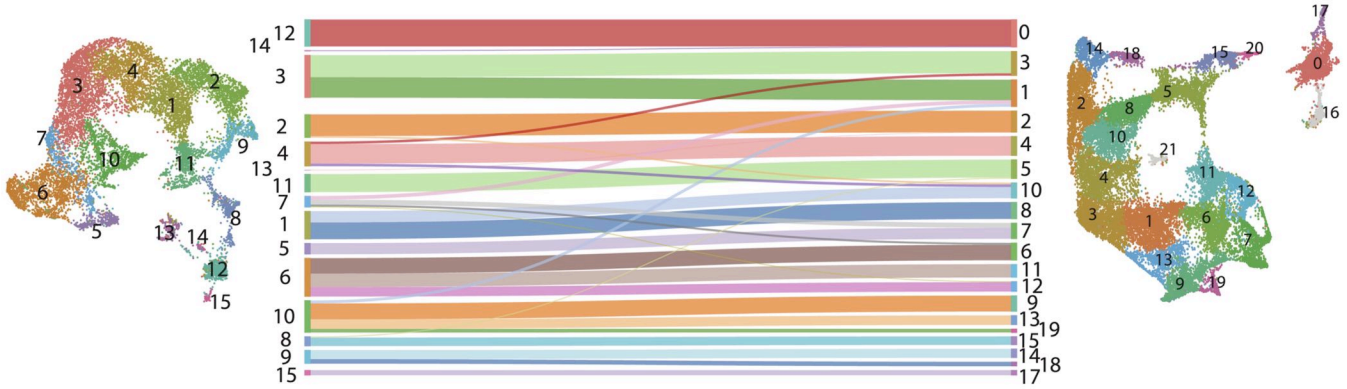
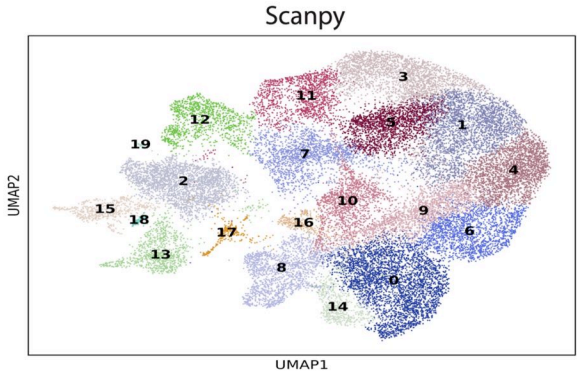

A)
Cluster Stability
Stability (Normalized Jaccard similarity index)
Cluster Number
B)
Clustering Resolution
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
Cluster Size
2000
4000
6000
Proportion Included
0.00
0.25
0.50
0.75
1.00
Cell Number
5000
4000
3000
2000
1000

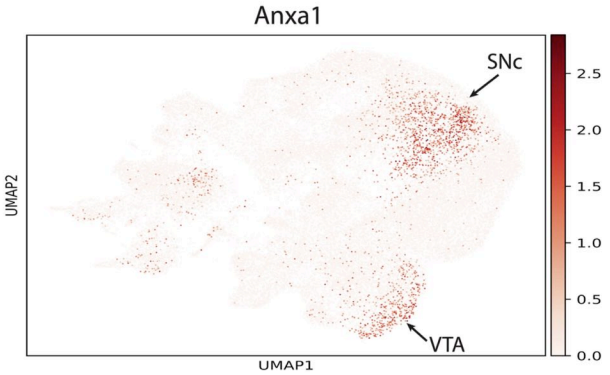
Figure S3

A



B





C

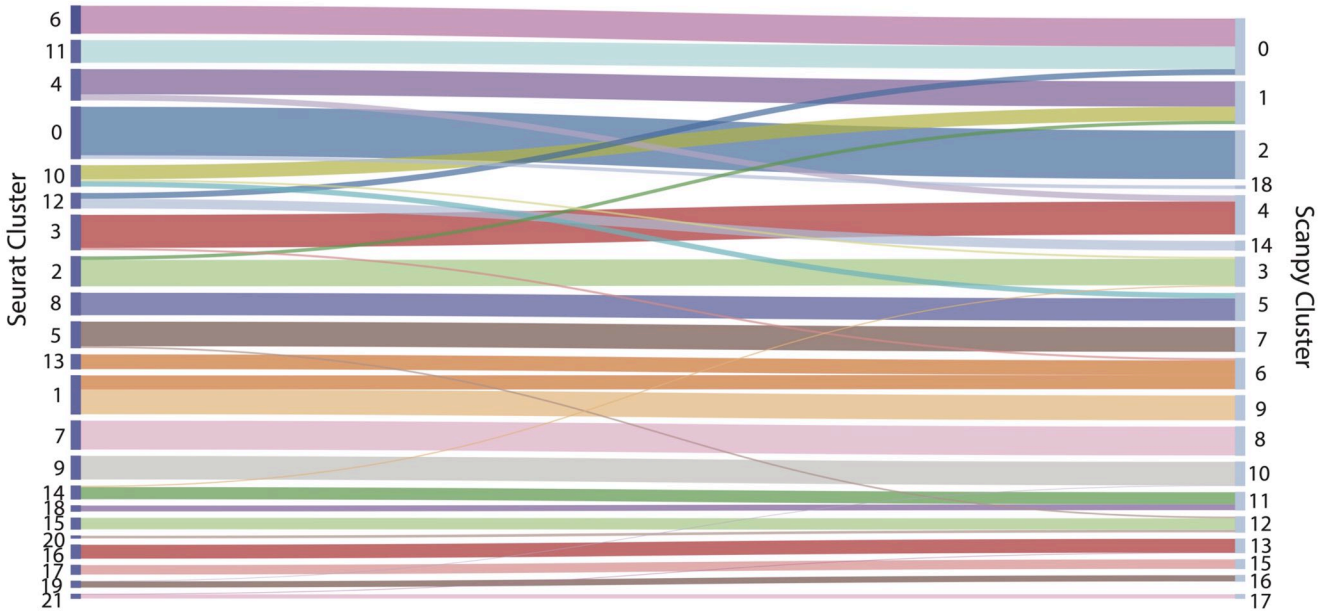


Figure S4

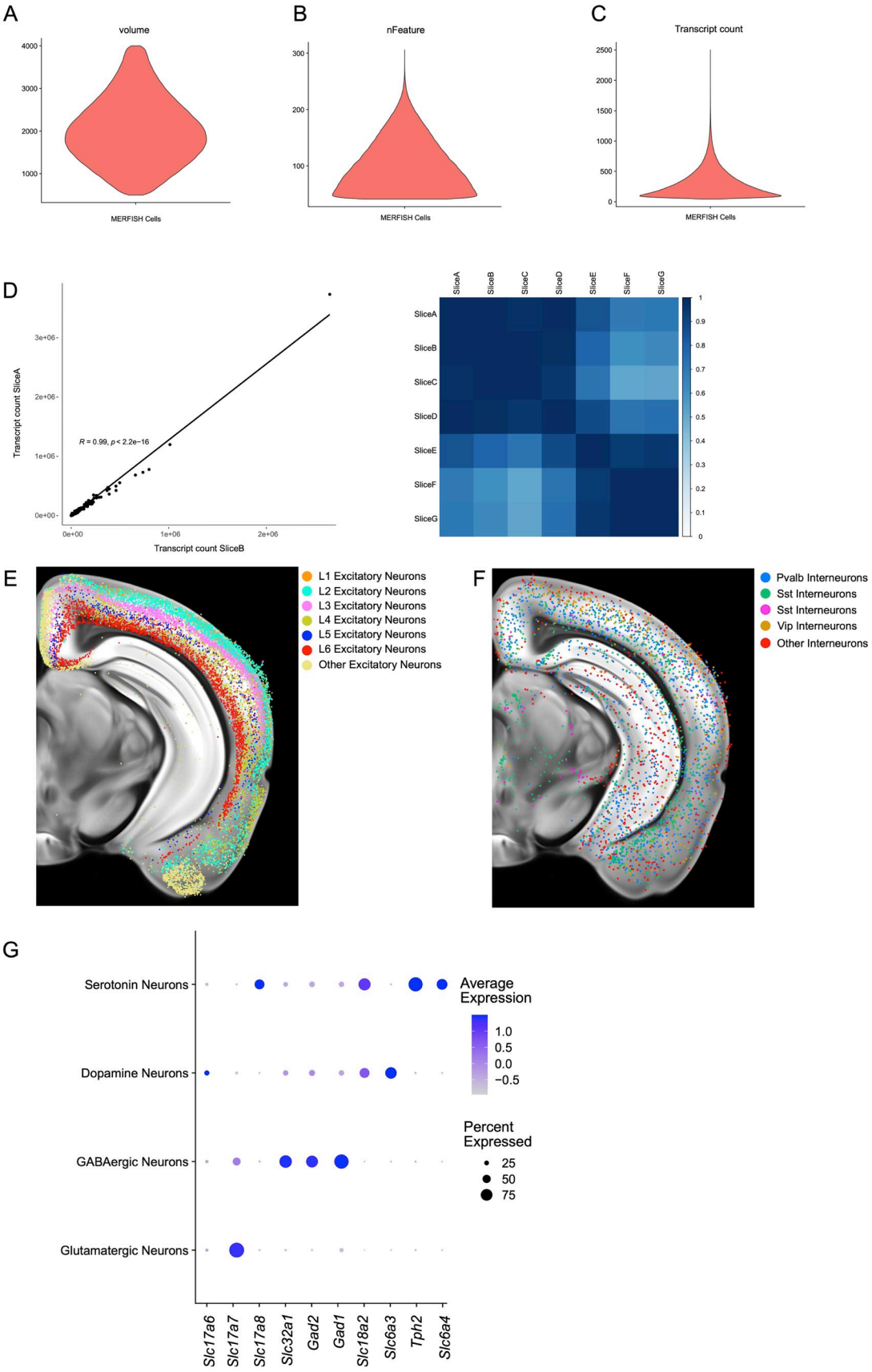

A
volume
4000
3000
2000
1000
MERFISH Cells
B
nFeature
300
200
100
MERFISH Cells
C
Transcript count
2500
2000
1500
1000
500
0
MERFISH Cells
D
Transcript count SliceA
Transcript count SliceB
R = 0.99, p < 2.2e−16
0e+00
1e+06
2e+06
3e+06
SliceA
SliceB
SliceC
SliceD
SliceE
SliceF
SliceG
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
E
L1 Excitatory Neurons
L2 Excitatory Neurons
L3 Excitatory Neurons
L4 Excitatory Neurons
L5 Excitatory Neurons
L6 Excitatory Neurons
Other Excitatory Neurons
F
Pvalb Interneurons
Sst Interneurons
Sst Interneurons
Vip Interneurons
Other Interneurons
G
Serotonin Neurons
Dopamine Neurons
GABAergic Neurons
Glutamatergic Neurons
Slc17a6
Slc17a7
Slc17a8
Slc32a1
Gad2
Gad1
Slc18a2
Slc6a3
Tph2
Slc6a4
Average Expression
1.0
0.5
0.0
−0.5
Percent Expressed
25
50
75

Figure S5

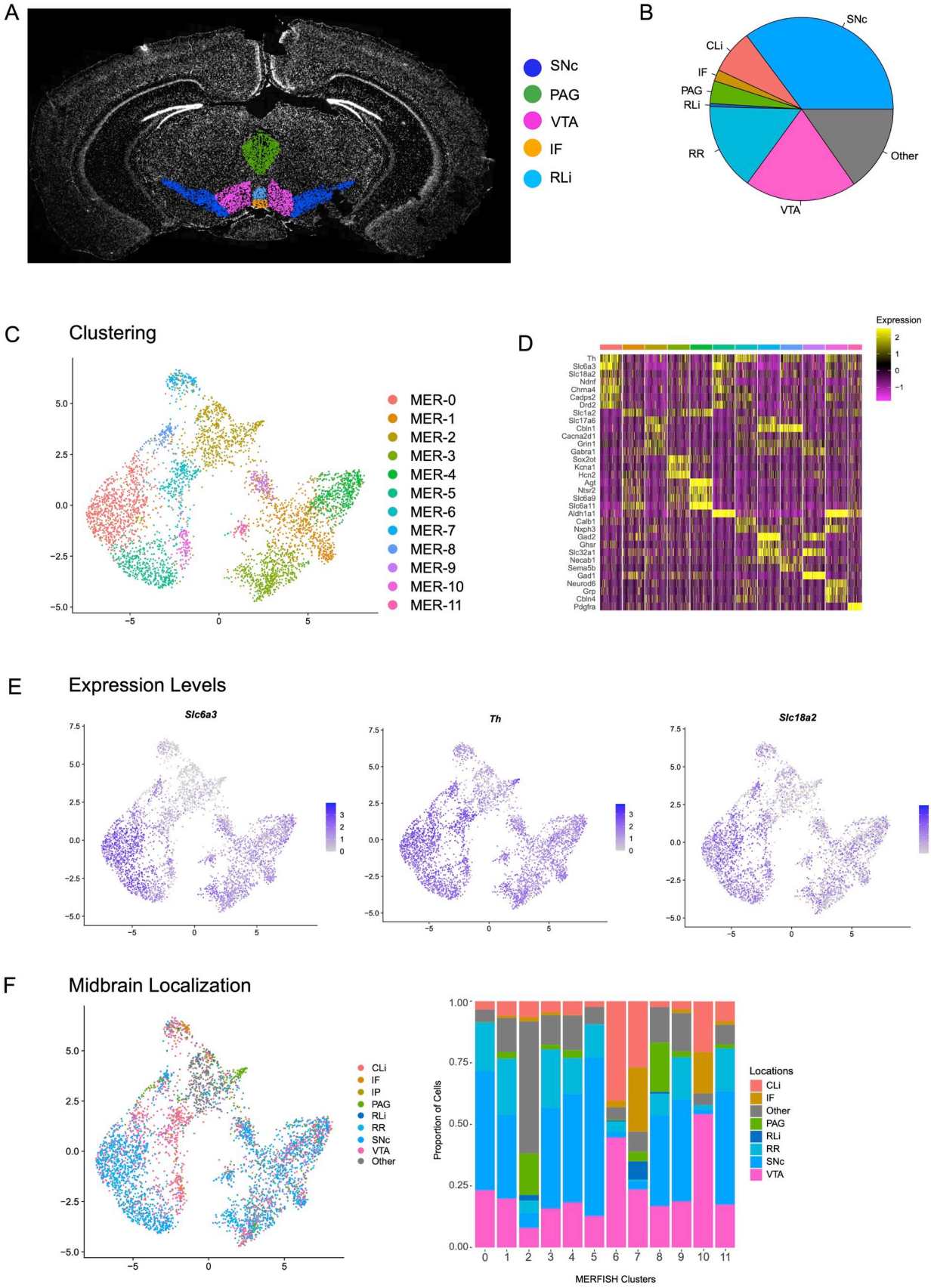

A
SNc
PAG
VTA
IF
RLi
B
SNc
CLi
IF
PAG
RLi
RR
VTA
Other
C
Clustering
MER-0
MER-1
MER-2
MER-3
MER-4
MER-5
MER-6
MER-7
MER-8
MER-9
MER-10
MER-11
D
Expression
Th
Slc6a3
Slc18a2
Ndnf
Chma4
Cadps2
Drd2
Slc1a2
Slc17a6
Cbln1
Cacna2d1
Grin1
Gabra1
Sox2ot
Kcna1
Hcn2
Agt
Ntsr2
Slc6a9
Slc6a11
Aldh1a1
Calb1
Nxph3
Gad2
Ghsr
Slc32a1
Necab1
Sema5b
Gad1
Neurod6
Grp
Cbln4
Pdgfra
E
Expression Levels
Slc6a3
Th
Slc18a2
F
Midbrain Localization
CLi
IF
IP
PAG
RLi
RR
SNc
VTA
Other
Proportion of Cells
MERFISH Clusters
Locations
CLi
IF
Other
PAG
RLi
RR
SNc
VTA

Figure S6

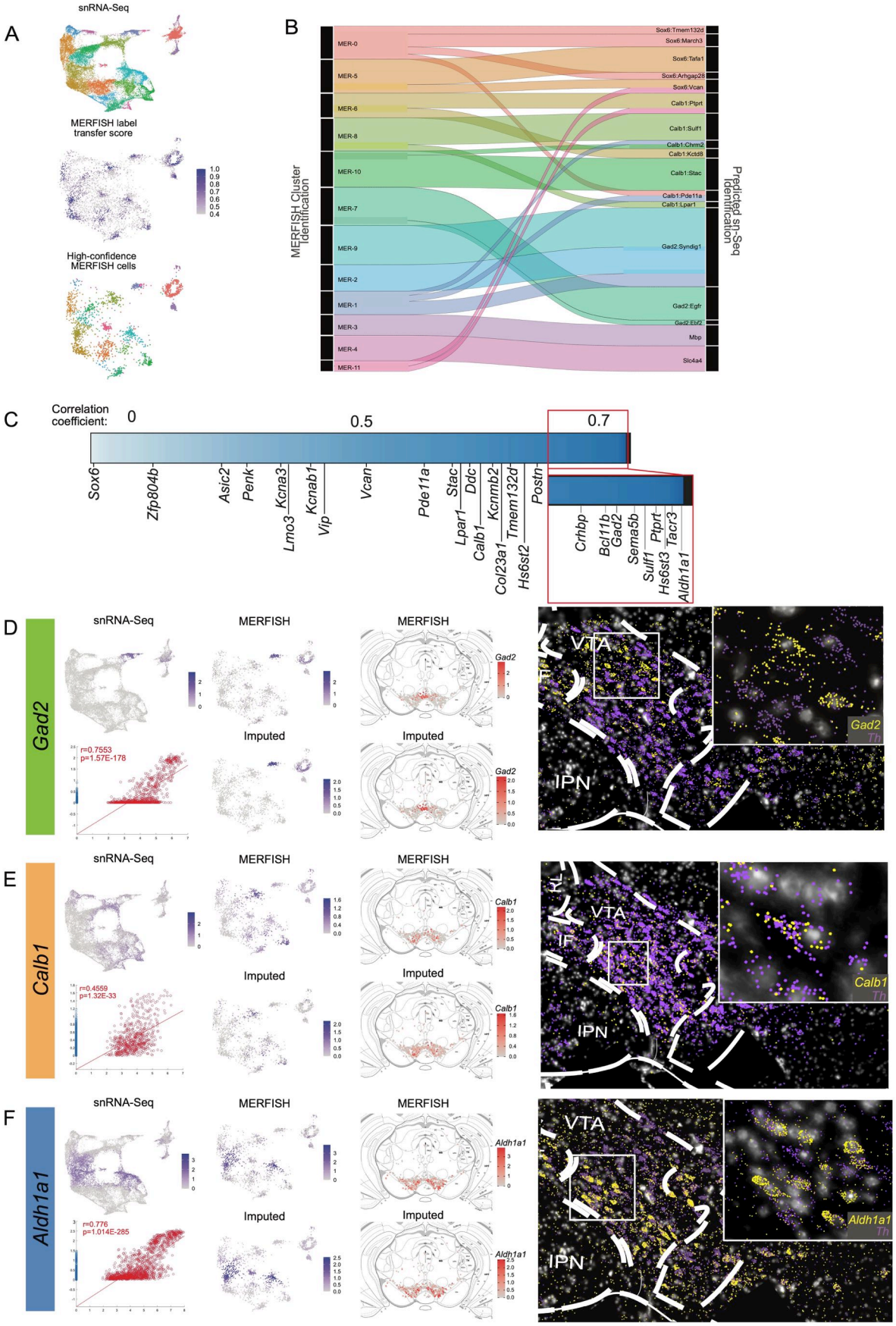

A
snRNA-Seq
MERFISH label transfer score
High-confidence MERFISH cells
B
MERFISH Cluster Identification
Predicted sn-Seq Identification
MER-0
MER-5
MER-6
MER-8
MER-10
MER-7
MER-9
MER-2
MER-1
MER-3
MER-4
MER-11
Sox6:Tmem132d
Sox6:March3
Sox6:Tafa1
Sox6:Arhgap28
Sox6:Vcan
Calb1:Ptprt
Calb1:Sulf1
Calb1:Chrm2
Calb1:Kcld8
Calb1:Stac
Calb1:Pde11a
Calb1:Lpar1
Gad2:Syndig1
Gad2:Egfr
Gad2:Ebf2
Mbp
Slc4a4
C
Correlation coefficient:
0
0.5
0.7
Sox6
Zfp804b
Asic2
Penk
Kcna3
Lmo3
Kcnab1
Vip
Vcan
Pde11a
Stac
Ddc
Lpar1
Calb1
Kcnmb2
Col23a1
Tmem132d
Hs6st2
Postn
Crhbp
Bcl11b
Gad2
Sema5b
Sulf1
Ptprt
Hs6st3
Tacr3
Aldh1a1
D
Gad2
snRNA-Seq
MERFISH
Imputed
r=0.7553
p=1.57E-178
VTA
IPN
E
Calb1
r=0.4559
p=1.32E-33
HL
IF
F
Aldh1a1
r=0.776
p=1.014E-285
Th

Figure S7

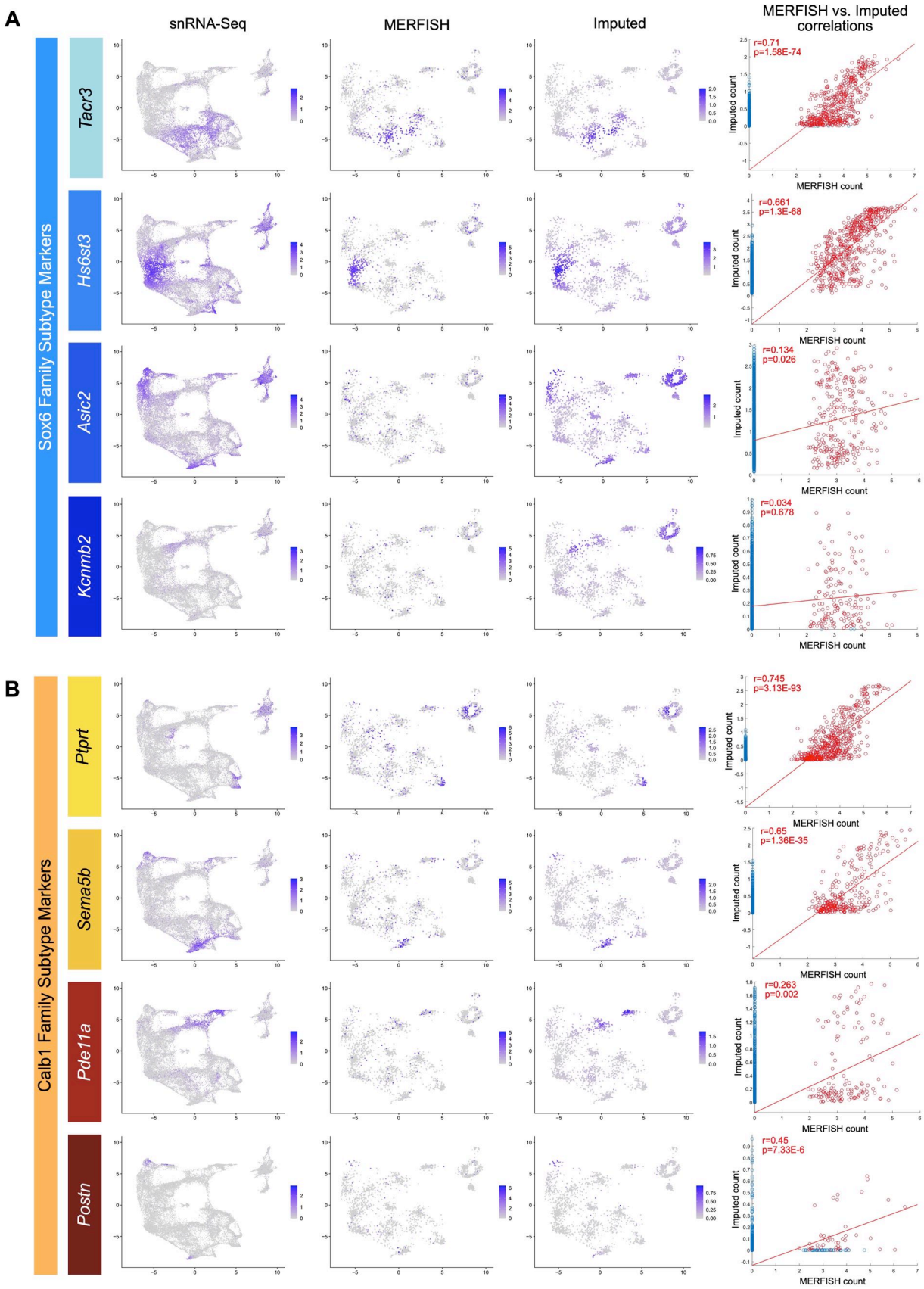

A
snRNA-Seq
MERFISH
Imputed
MERFISH vs. Imputed correlations
Sox6 Family Subtype Markers
Tacr3
r=0.71
p=1.58E-74
Hs6st3
r=0.661
p=1.3E-68
Asic2
r=0.134
p=0.026
Kcnmb2
r=0.034
p=0.678
Imputed count
MERFISH count
B
Calb1 Family Subtype Markers
Ptprt
r=0.745
p=3.13E-93
Sema5b
r=0.65
p=1.36E-35
Pde11a
r=0.263
p=0.002
Postn
r=0.45
p=7.33E-6

Figure S8

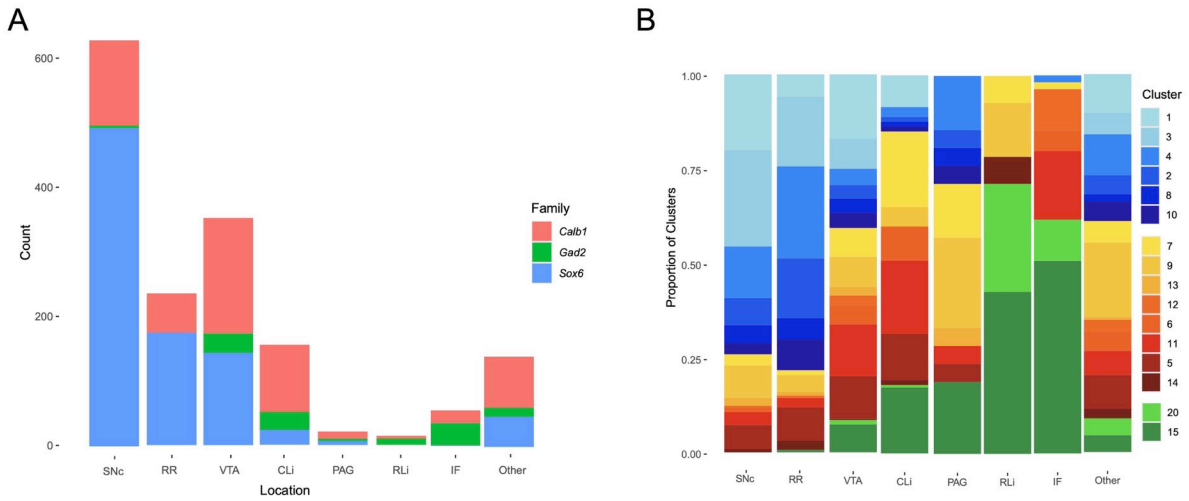
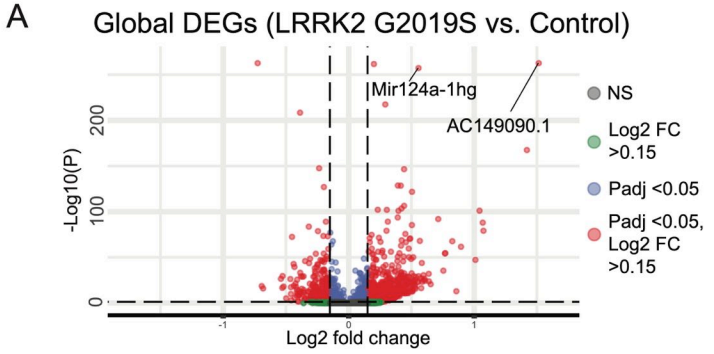
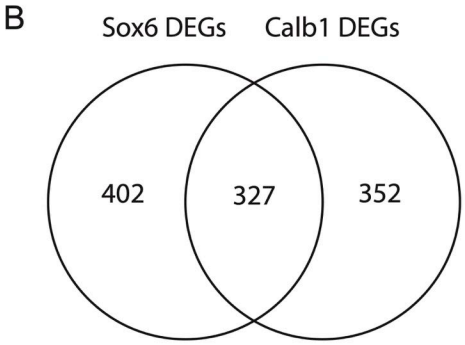
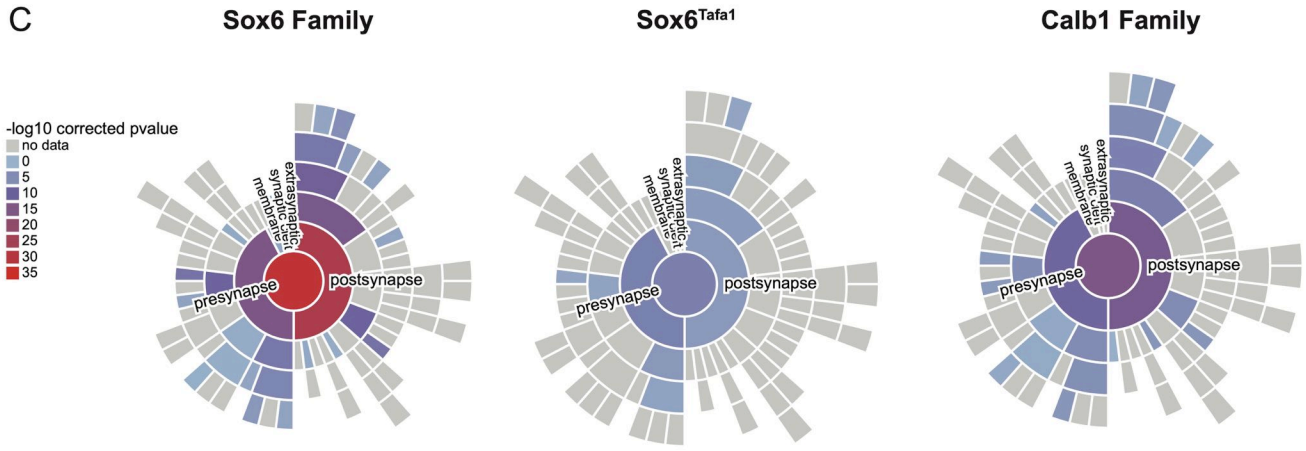

A
600
400
200
0
Count
SNc
RR
VTA
CLi
PAG
RLi
IF
Other
Location
Family
Calb1
Gad2
Sox6
B
1.00
0.75
0.50
0.25
0.00
Proportion of Clusters
SNc
RR
VTA
CLi
PAG
RLi
IF
Other
Cluster
1
3
4
2
8
10
7
9
13
12
6
11
5
14
20
15

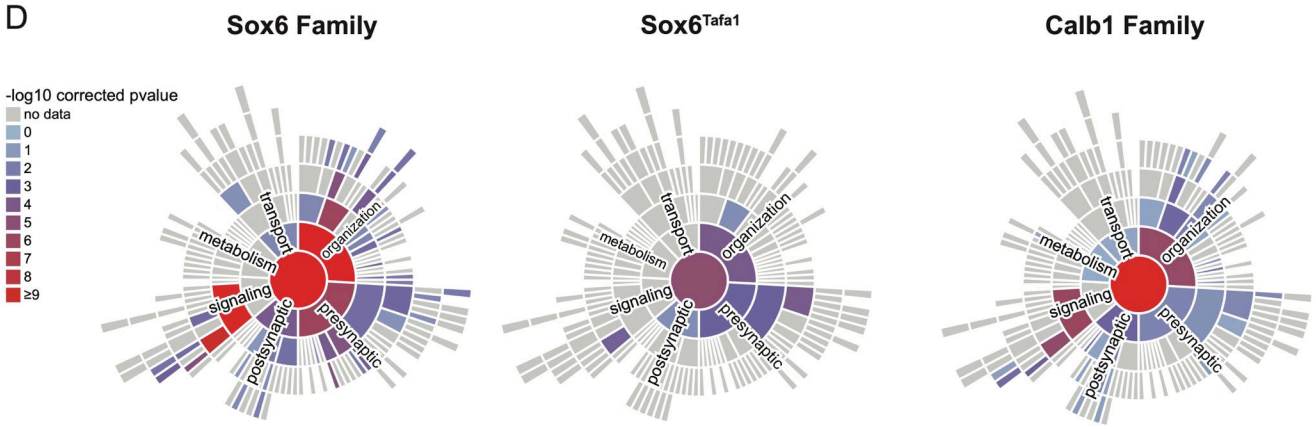
Figure S9







Localization of Differentially Expressed Genes



Functional roles of Differentially Expressed Genes

Figure 310

A
Top Enriched GO Pathways
Anxa1 SNc (Sox6^Tafa1)
trans-synaptic signaling
chemical synaptic transmission
behavior
anterograde trans-synaptic signaling
cell junction organization
Anxa1 VTA (Calb1^Stac)
regulation of synapse structure or activity
regulation of synapse organization
cell junction organization
cell junction assembly
intermediate filament-based process
1.5
2.0
2.5
-log10 Adjusted P-value (qscore)
qscore
1.6
2.8
Count
2.5
5.0
7.5
10.0
12.5