

Reviewed Preprint

v2 • March 5, 2025

Revised by authors

Reviewed Preprint

v1 • November 21, 2024

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

Zachary Gaertner, Cameron Oram, Amanda Schneeweis, Elan Schonfeld, Cyril Bolduc, Chuyu Chen, Daniel Dombeck, Loukia Parisiadou, Jean-Francois Poulin ✉, Rajeshwar Awatramani ✉

Northwestern University Feinberg School of Medicine, Dept of Neurology, Chicago, United States • Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, United States • McGill University (Montreal Neurological Institute), Faculty of Medicine and Health Sciences, Dept of Neurology and Neurosurgery, Montreal, Canada • Northwestern University Feinberg School of Medicine, Dept of Pharmacology, Chicago, United States • Northwestern University, Dept of Neurobiology, Evanston, United States

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

This **important** study combines single nucleus transcriptional profiling with spatial transcriptomics to identify and map heterogeneity among dopamine neurons in the mouse ventral midbrain. The **compelling** results separate dopamine neurons into three broad families that have unique (yet overlapping) spatial distribution within the ventral tegmental area and substantia nigra, and also identify population-specific changes in a LRRK2 mouse model of Parkinson's Disease. The creation of a public-facing app where the snRNA-seq data can be investigated by anyone is a major strength.

<https://doi.org/10.7554/eLife.101035.2.sa3>

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.

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

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 bias²⁵. 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 dataset²⁵ 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 premammillary 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

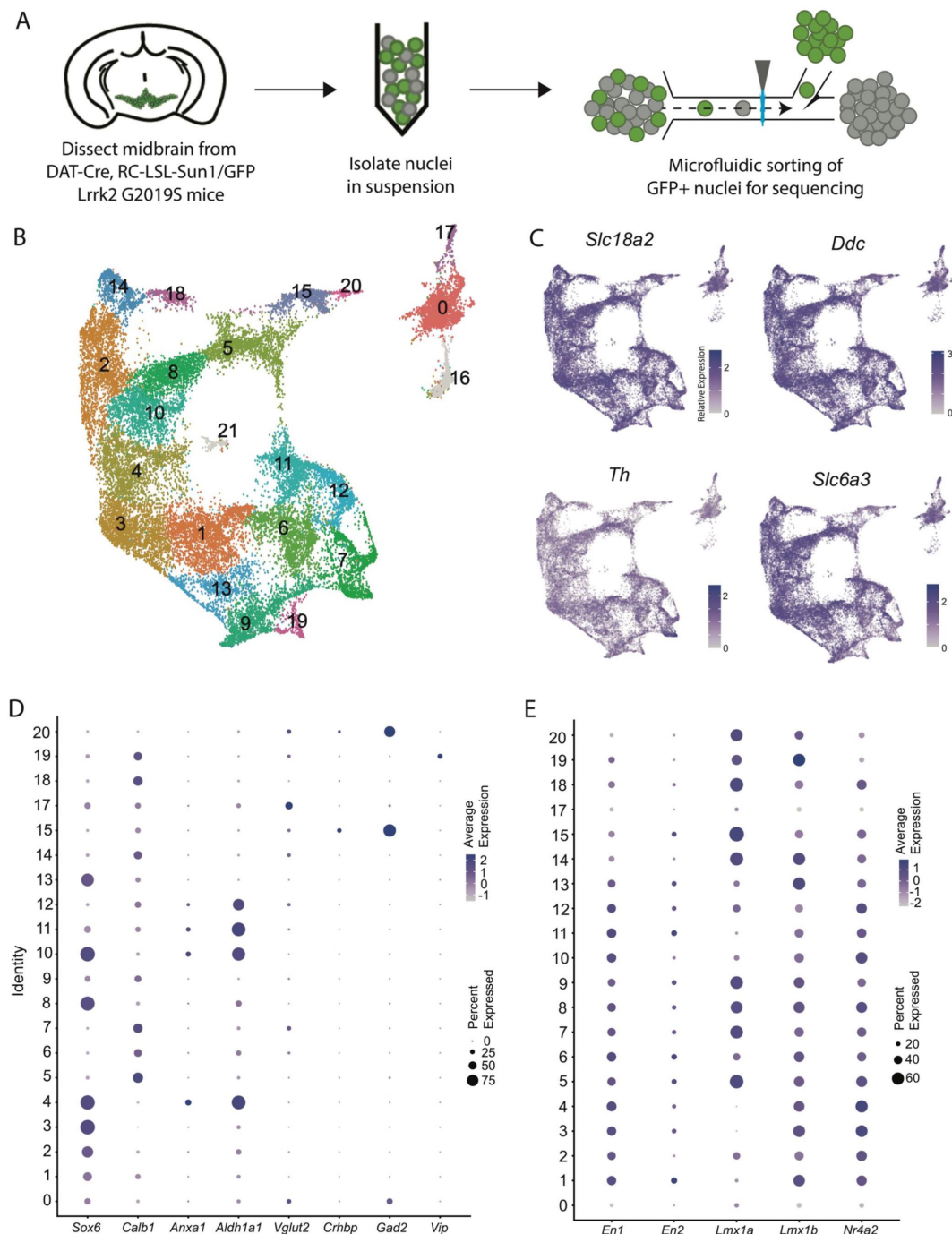


Figure 1.

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.

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 work³⁸, 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 humans⁴¹. 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 study²⁵, 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 properties²⁵. 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 metrics²⁵. 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 MERFISH⁴⁴, 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 cells⁴⁵. 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 studies⁴⁶⁻⁴⁸.

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 neurons⁴⁹⁻⁵⁴, 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

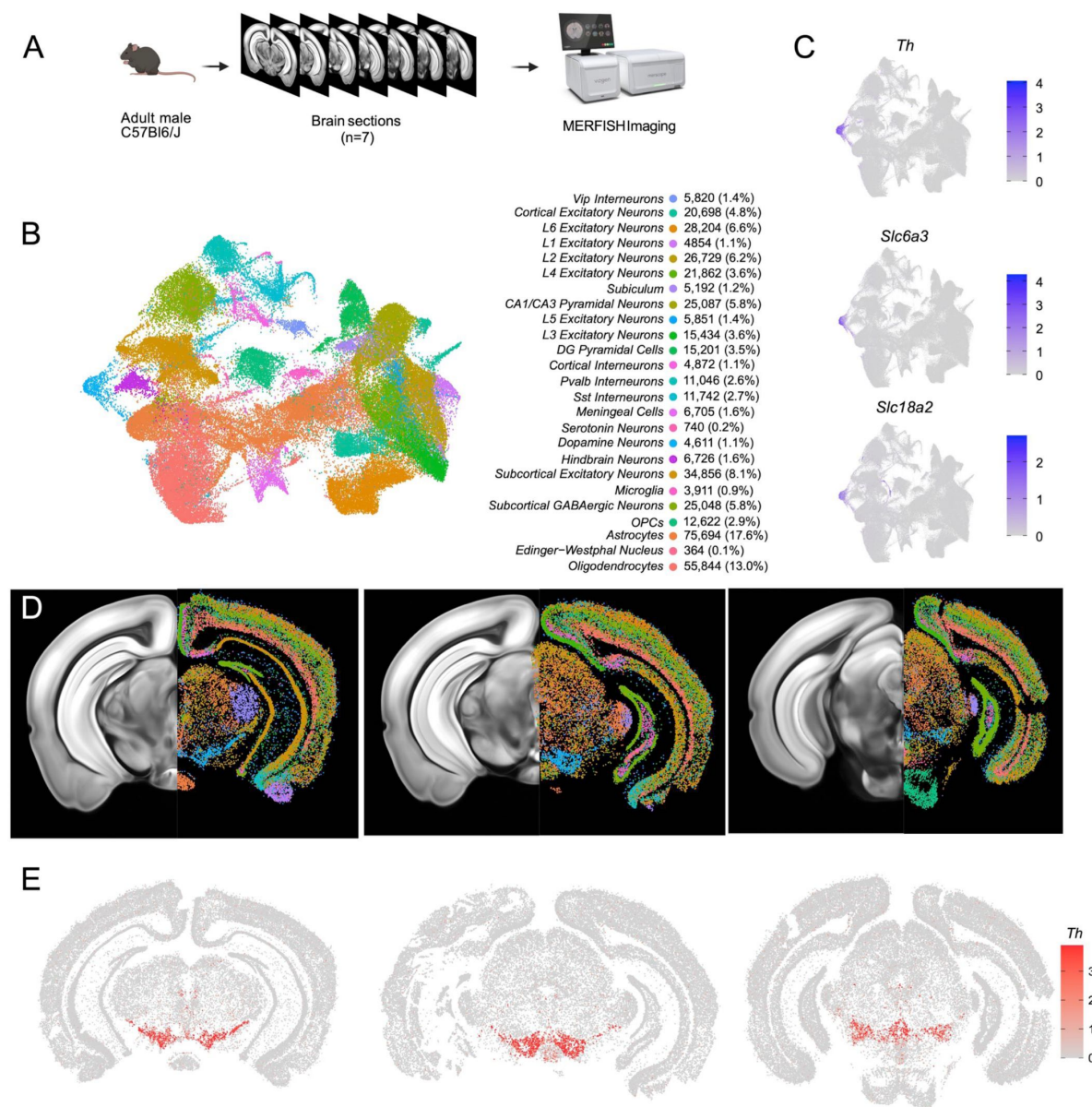


Figure 3.

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.

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 reports⁴³. 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, S8A),

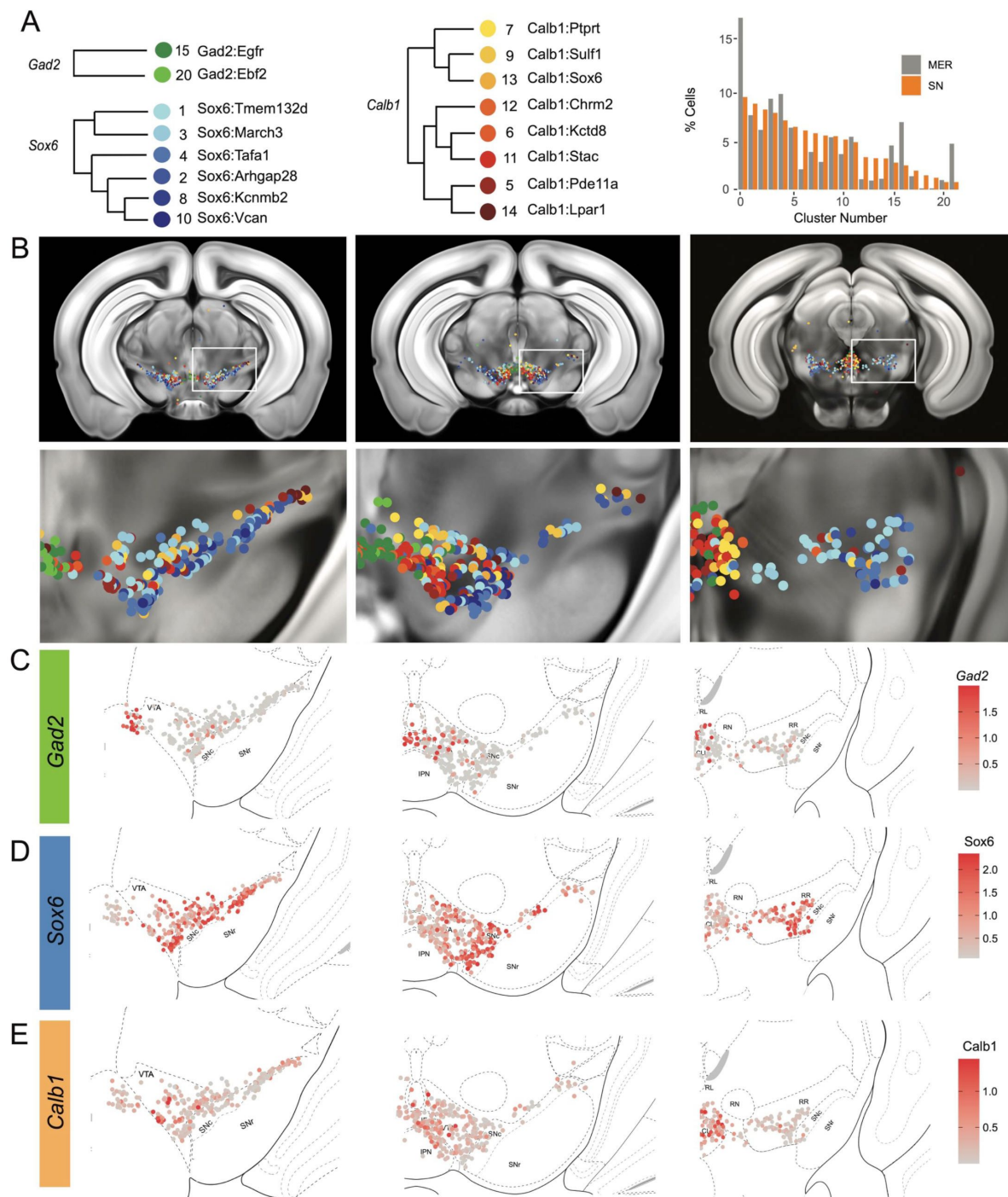


Figure 4.

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.

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

- defined by higher levels of *Slc44a5* and composed of subtypes Sox6^{Tmem132d} and Sox6^{March3} and 2) Branch
- 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^{Ptppt}, 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^{Ptppt} 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 striatum²⁷. 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.

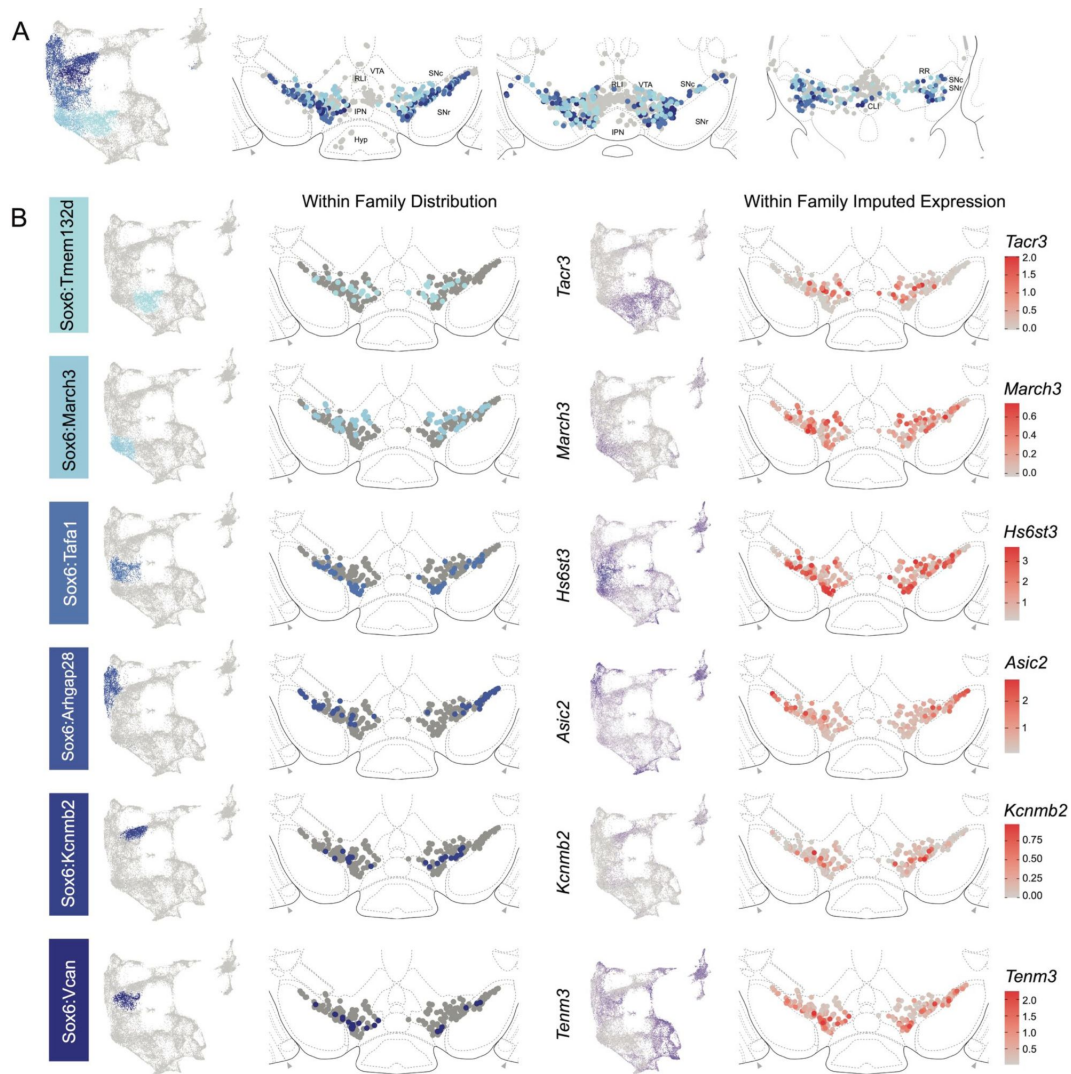


Figure 5.

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.

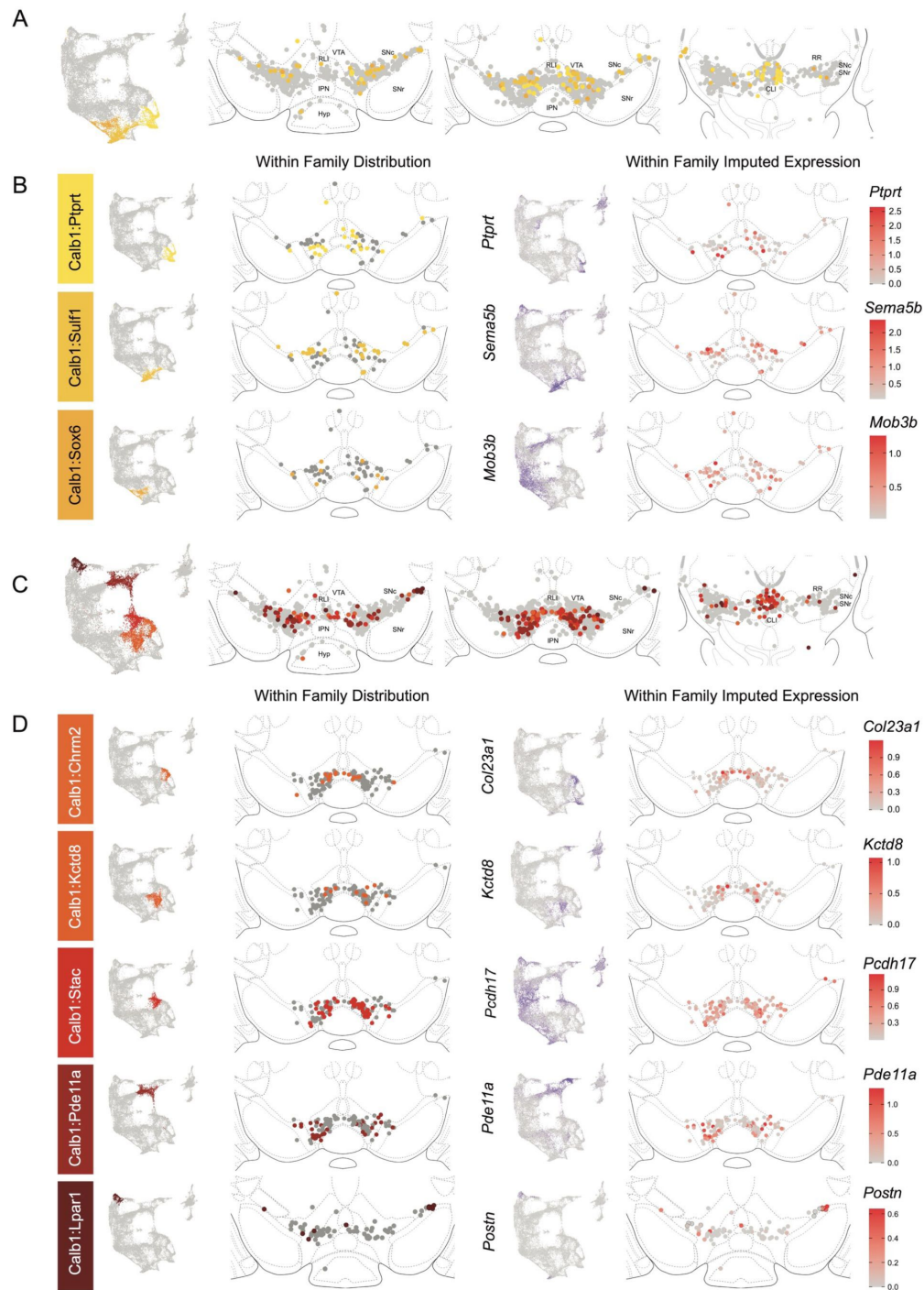


Figure 6.

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.

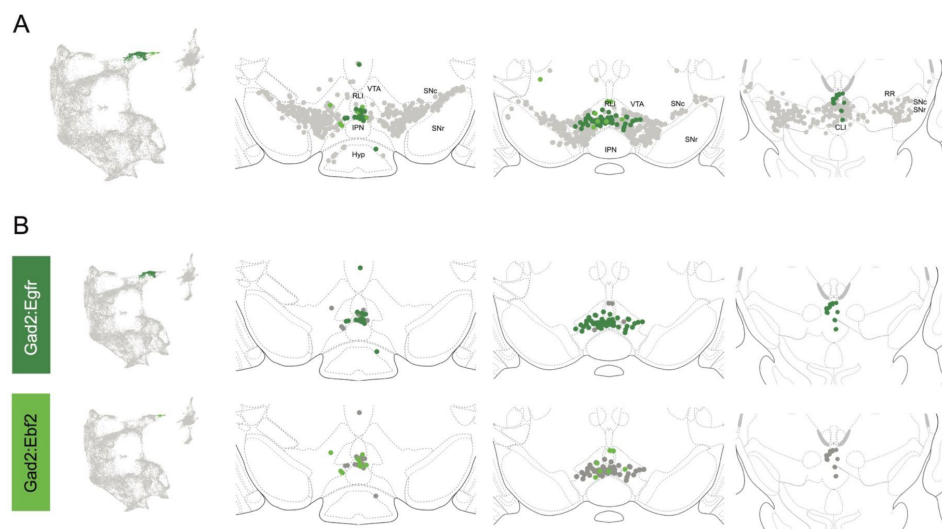


Figure 7.

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.

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 unclear⁵⁶, 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 cells³⁴, ⁵⁷–⁶⁰. However, previous studies have shown alterations in DA signaling or gene expression even when the mutant LRRK2 was expressed specifically in DA neurons⁶¹, 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 mice³⁶. 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 dysfunction⁶³–⁶⁵. This gene has been proposed as a sensitive marker of biological aging in the brain⁶³, 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 models⁶⁷–⁷⁰. 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.

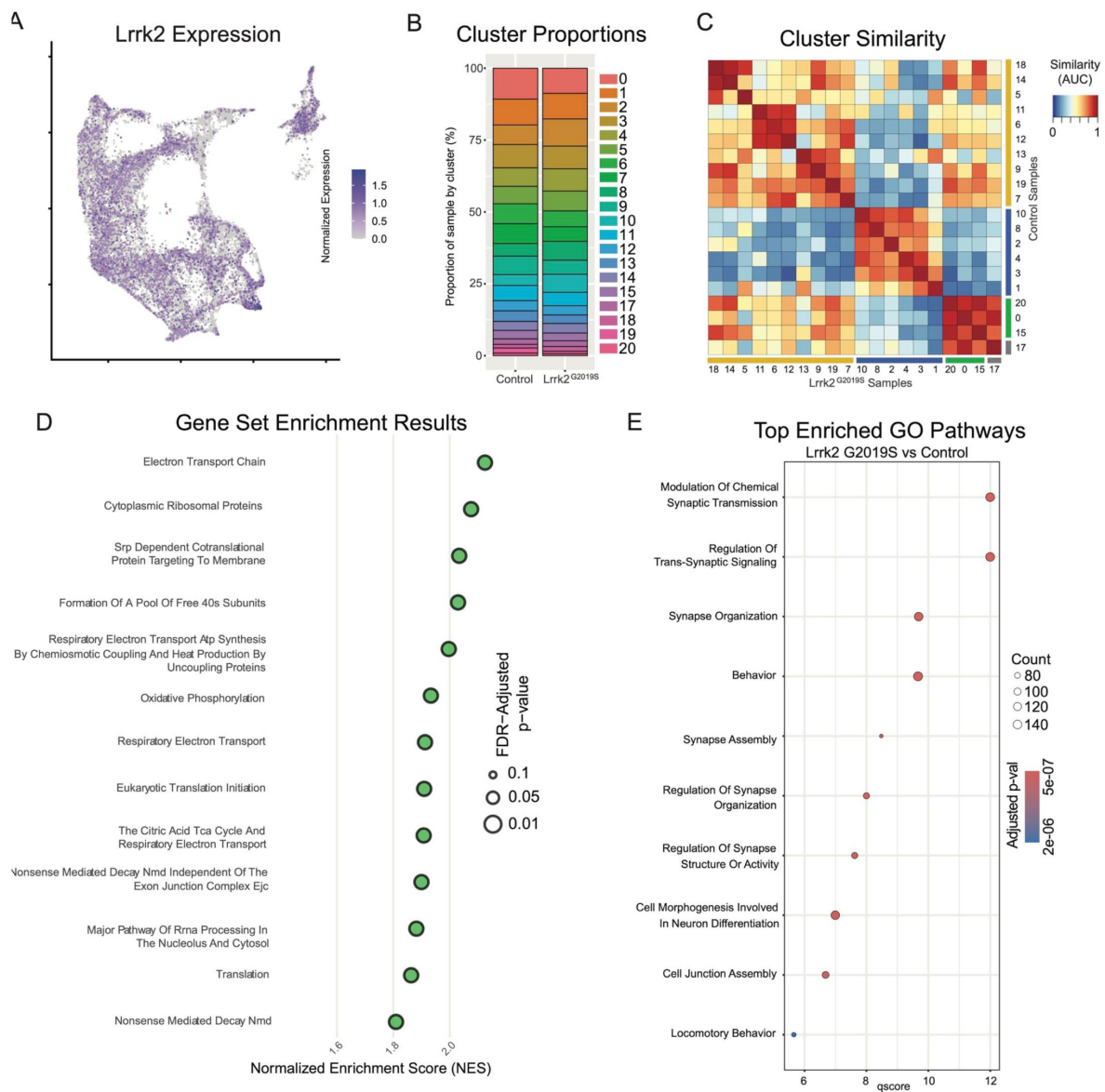


Figure 8.

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.

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 populations⁴³. 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 patients⁷³ 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 impairments⁷⁹ 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 neurons⁸². 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 PD²⁵. 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 genes⁸³. 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 movement²⁵. 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 GWAS⁸⁴ using a polygenic enrichment score method, scDRS⁸⁵. 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

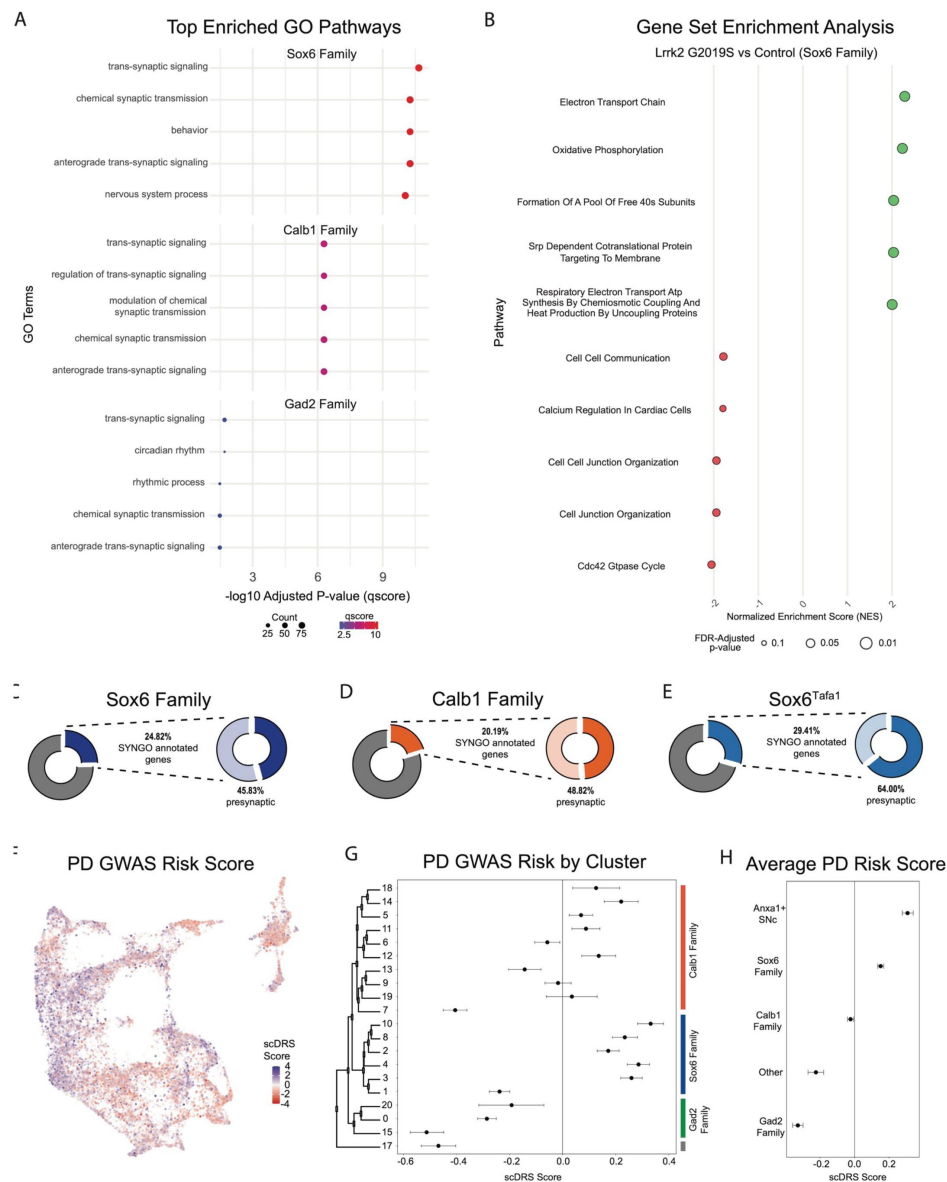


Figure 9.

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.

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 study⁸⁶. 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 study⁸⁶, 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 absent²⁵.

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 gradient⁹¹. 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 projections⁷. 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 tremor⁹⁸. 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 mice⁹⁹.

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 range³⁵. 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 population⁴¹. 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 cycle¹⁰¹. 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 types¹⁰². 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 CaCl₂, 21mM MgCl₂, 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 CaCl₂, 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 MgCl₂, 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](https://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 method¹⁰³ in Seurat V5¹⁰⁴. 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 above¹⁰⁵. 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 dataset²⁵. 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 described²⁵. 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 3.10.13¹⁰⁶ - 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 pipeline¹⁰⁸, 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_002 774), 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 10×10 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](https://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 10µm-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](https://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 package¹⁰⁴. To exclude cells that underwent poor cell segmentation we removed cells that had $>500\mu\text{m}^3$ and $<4000\mu\text{m}^3$ 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 normalization¹⁰³ 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 al⁹². 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.

Supplementary Figure Legends

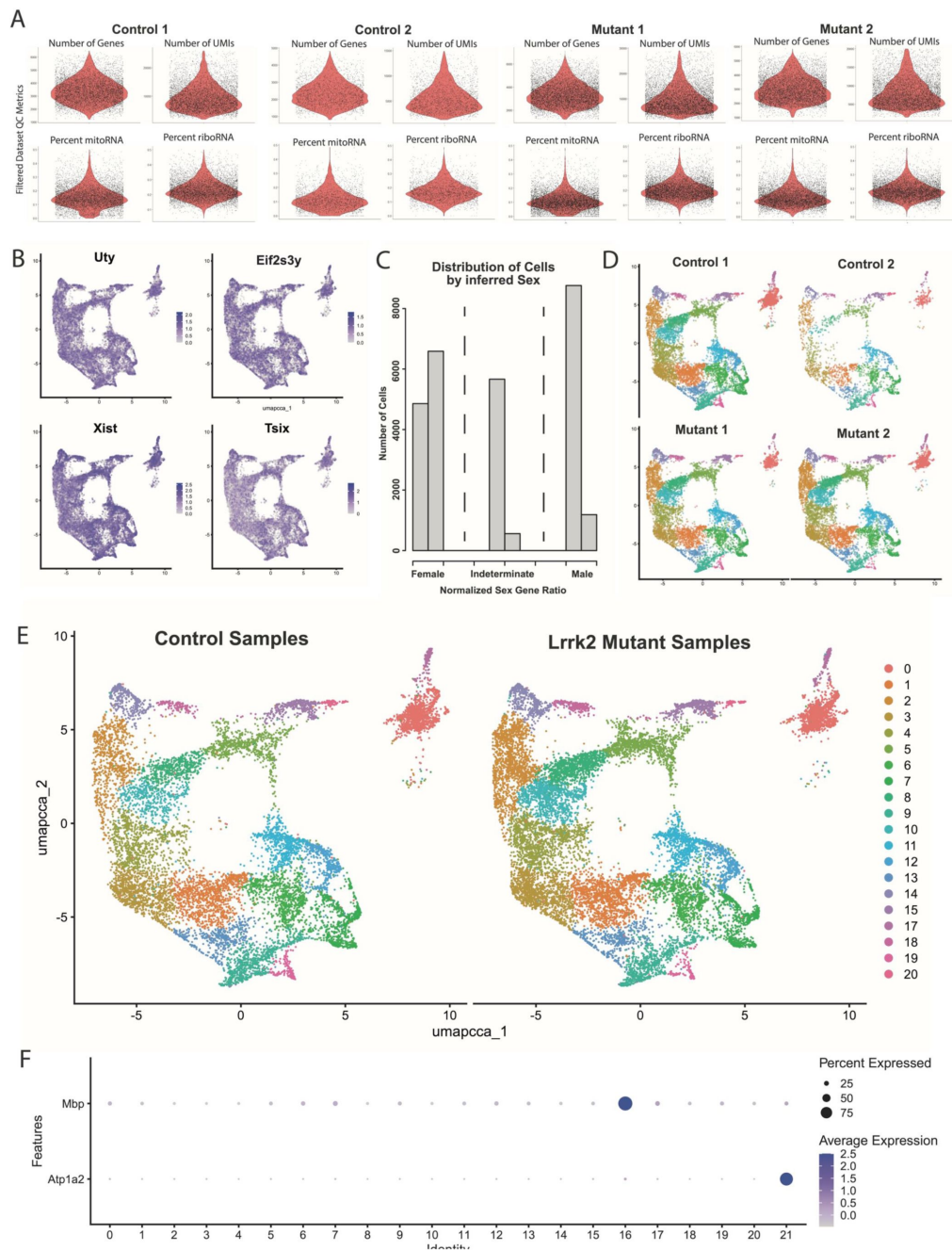


Figure S1.

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.

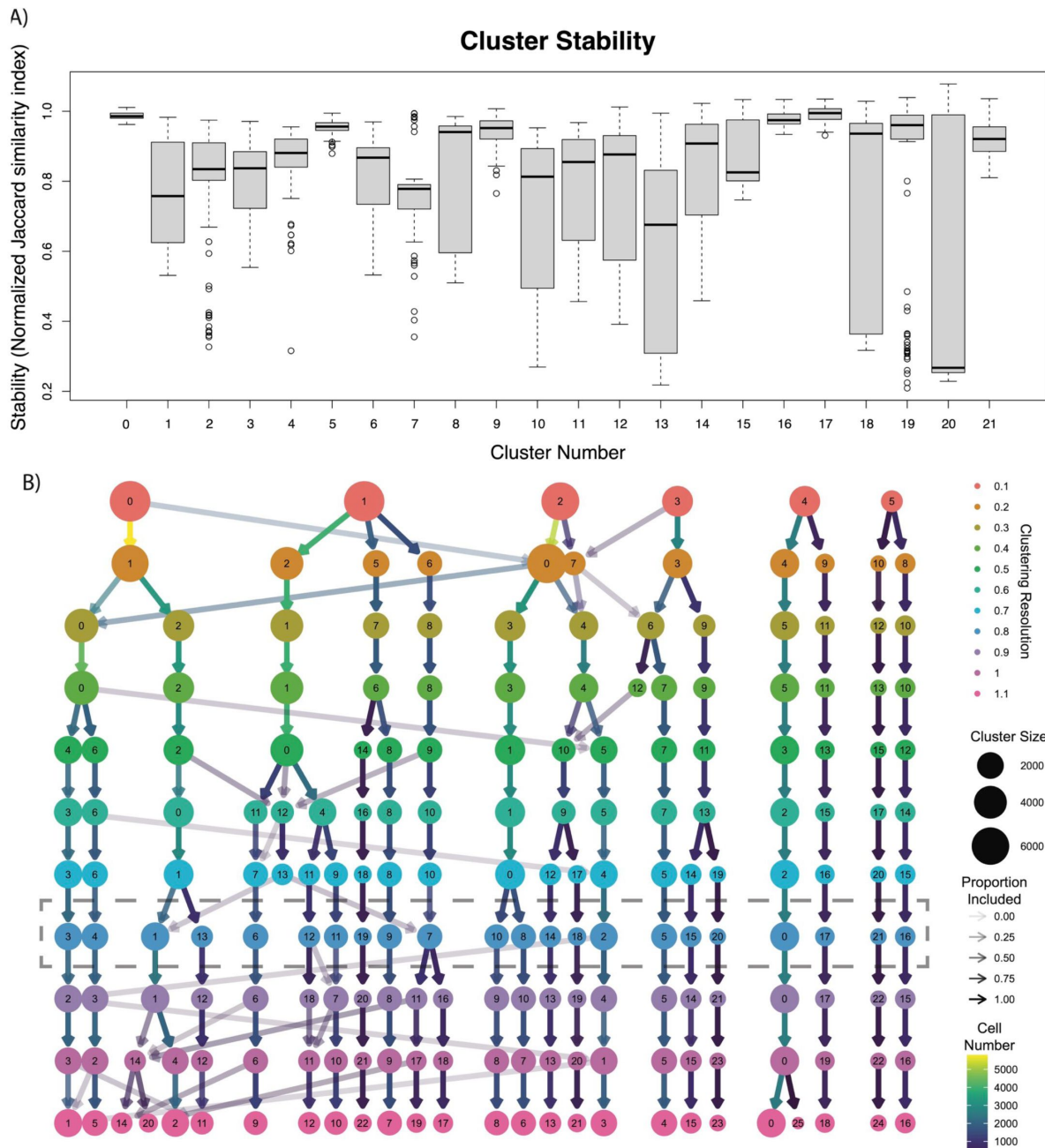


Figure S2.

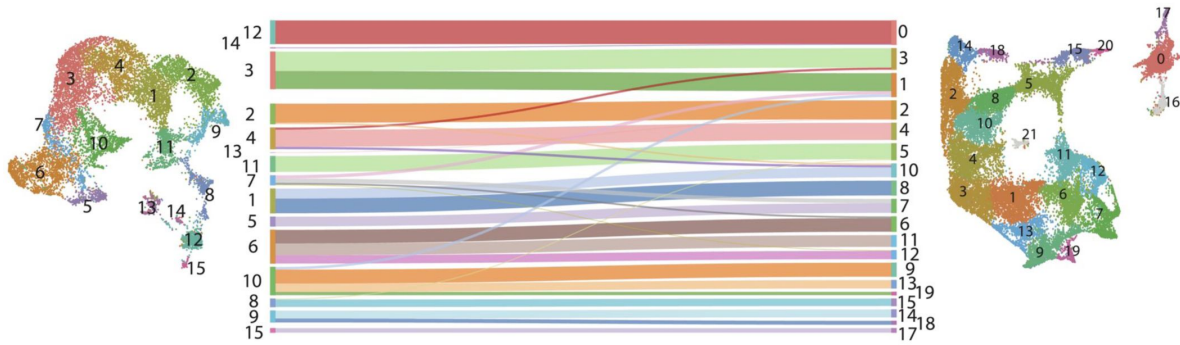
Representations of cluster heterogeneity.

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.

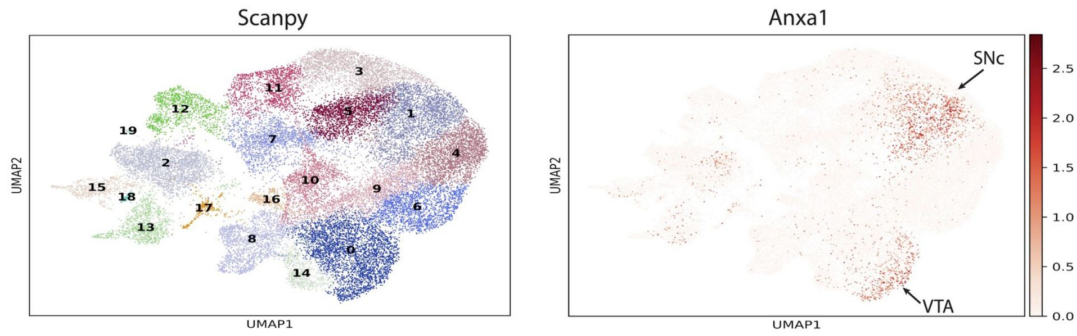
A

Azcorra*, Gaertner* et al., 2023

Gaertner*, Oram* et al., 2024



B



C

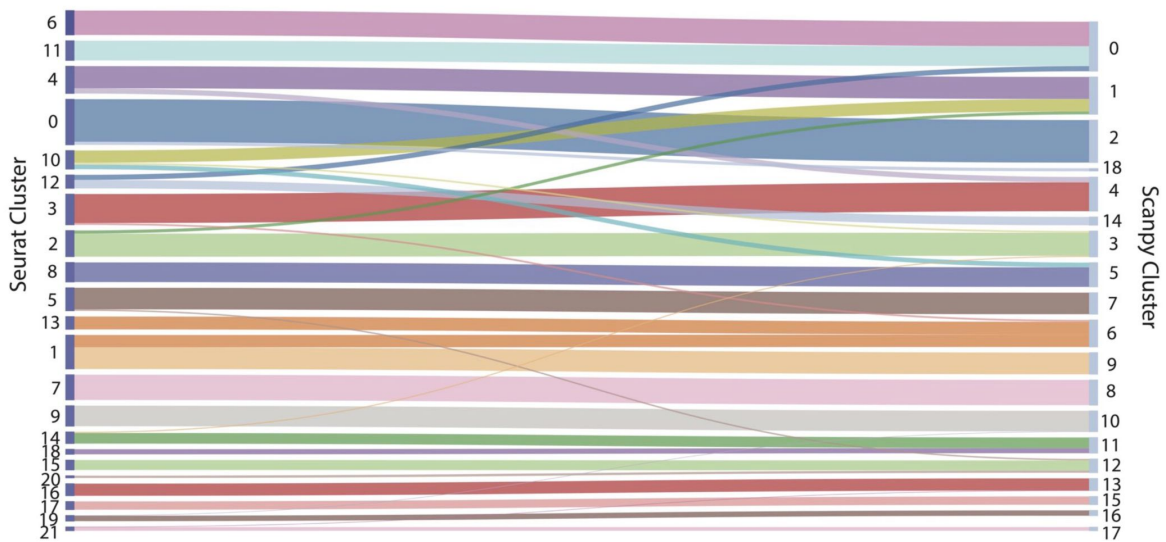


Figure S3.

Validating snRNA-seq clusters with Python and published DA clusters.

A) Comparison of UMAP from Azcorra & Gaertner et al., 2023²⁵ and UMAP from **Figure 1**. A Sankey diagram was created by mapping nuclei collected in **Figure 1** onto the clusters from Azcorra*, Gaertner* et al., 2023²⁵. 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.

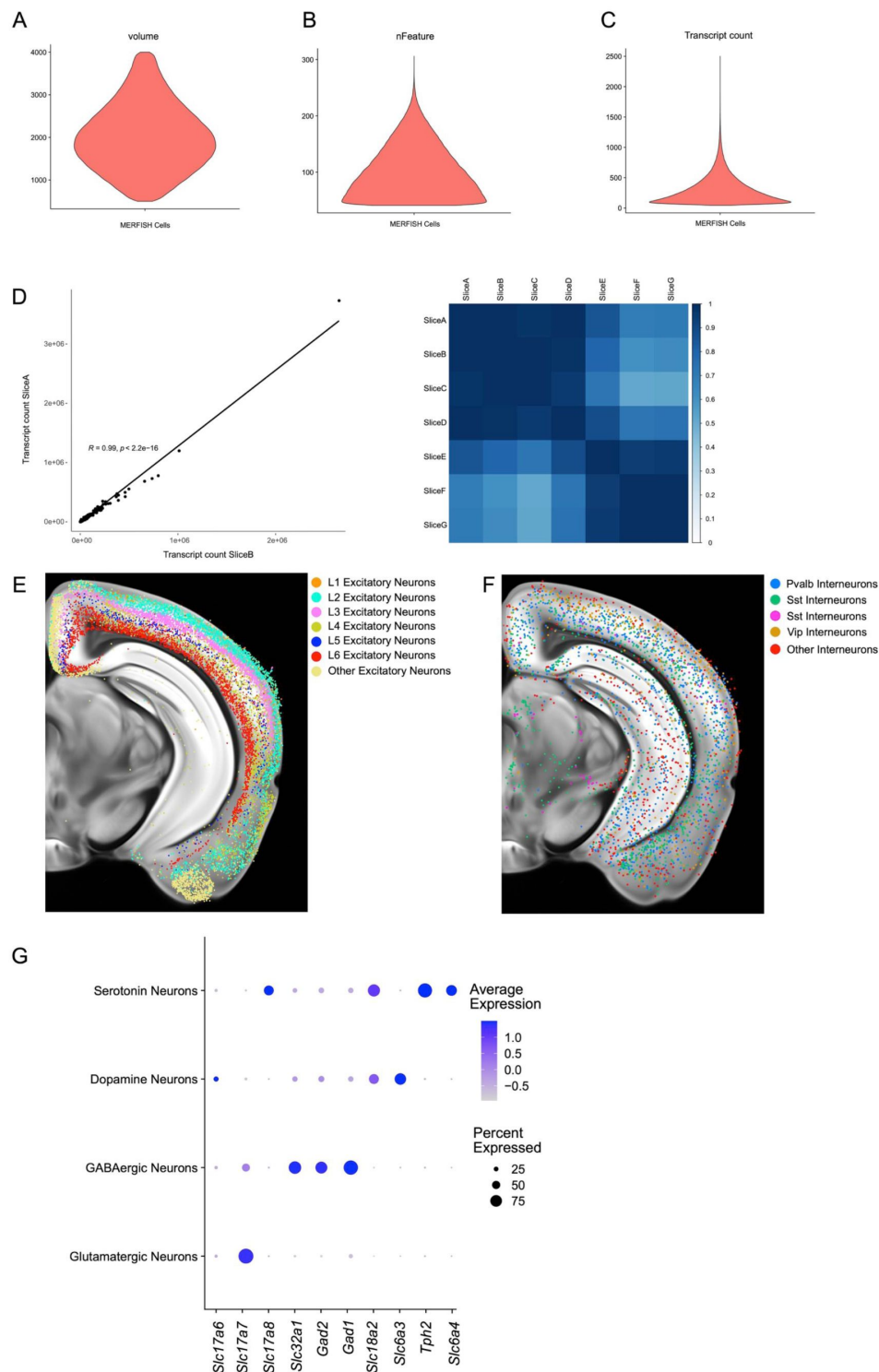


Figure S4.

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.

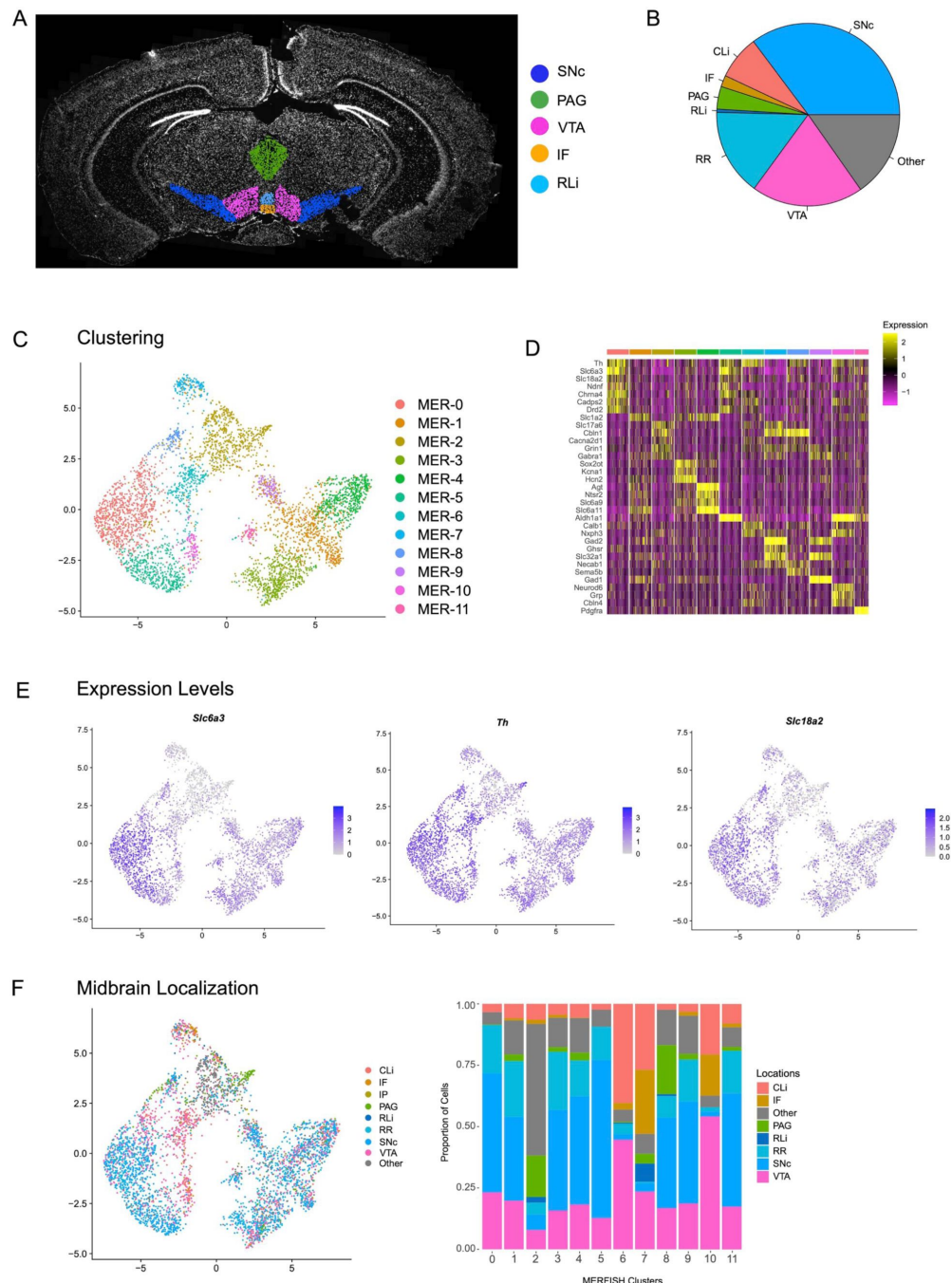


Figure S5.

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).

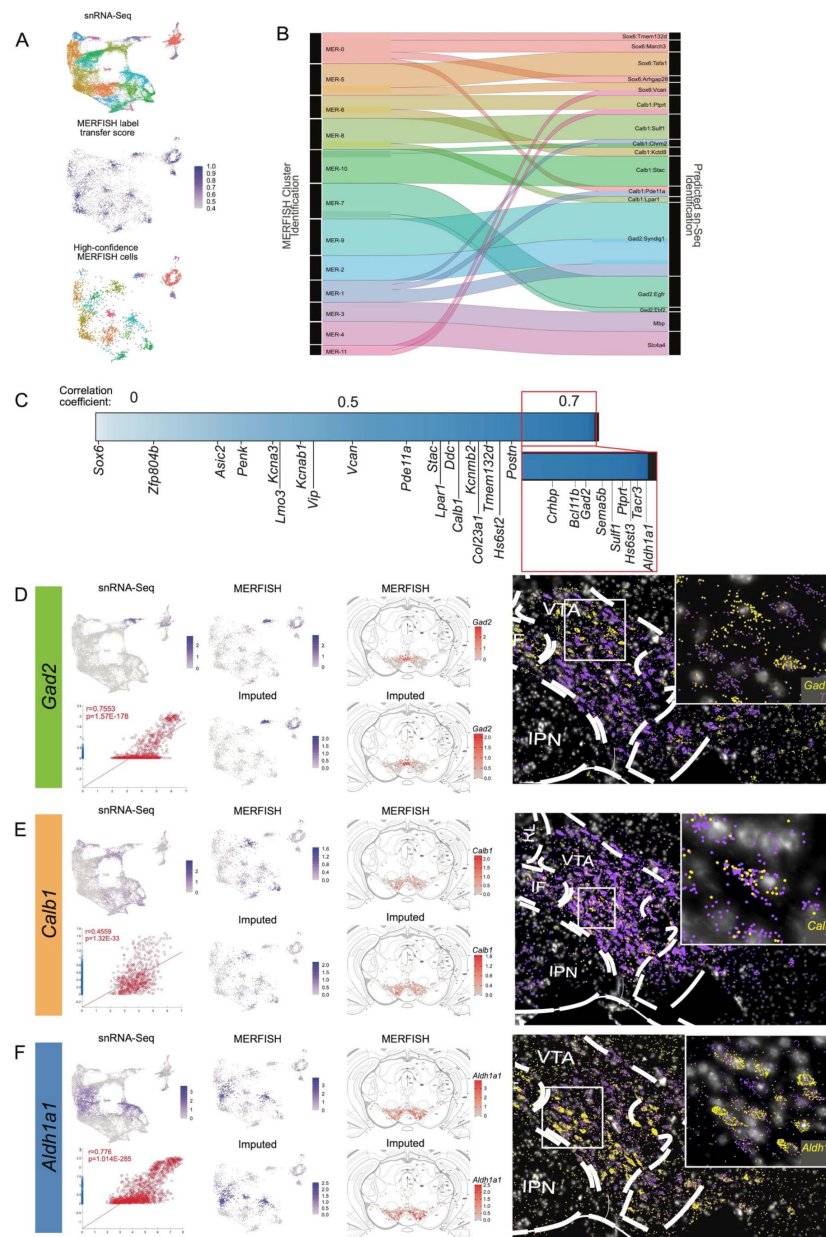


Figure S6.

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).

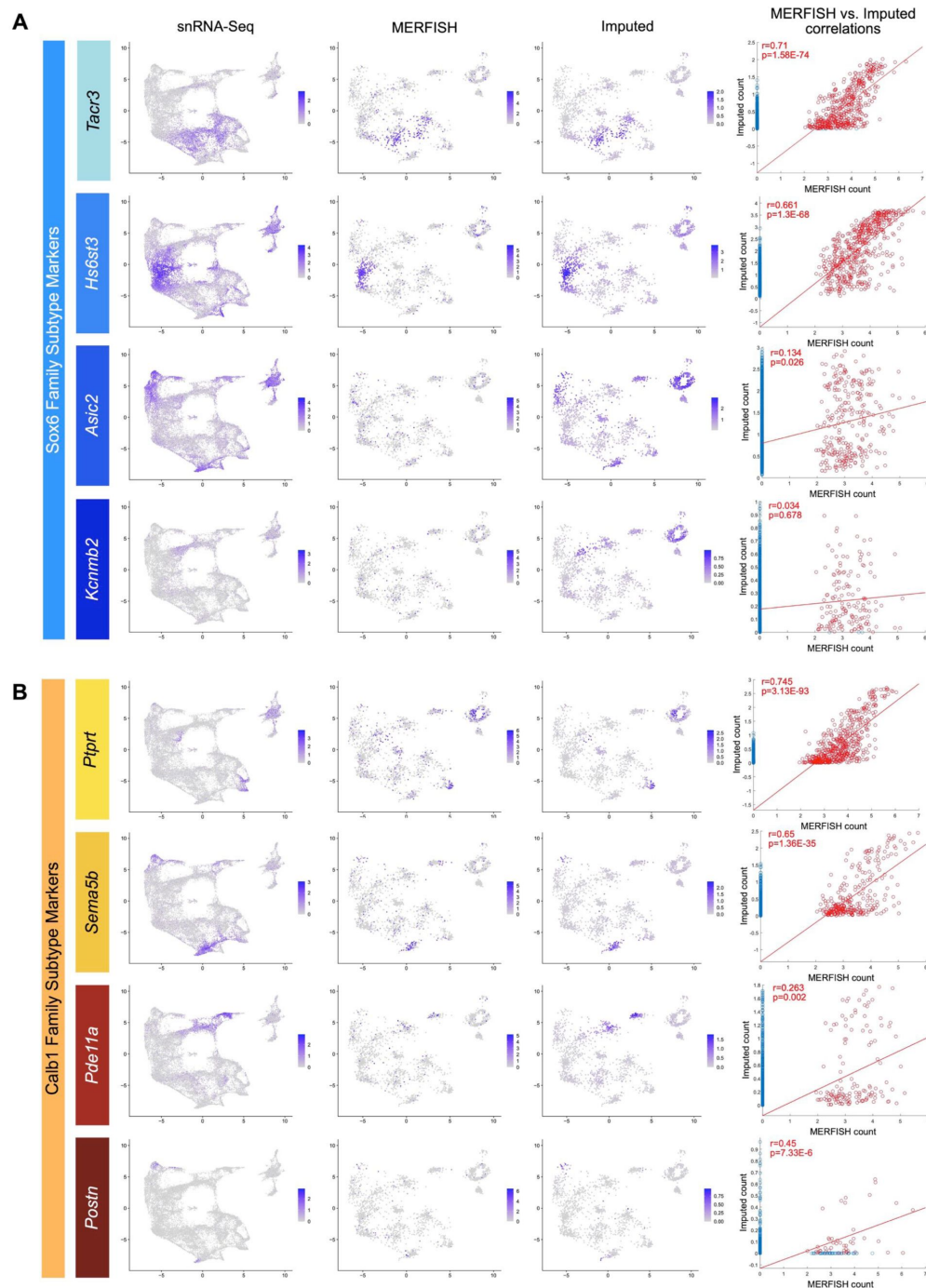


Figure S7.

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.

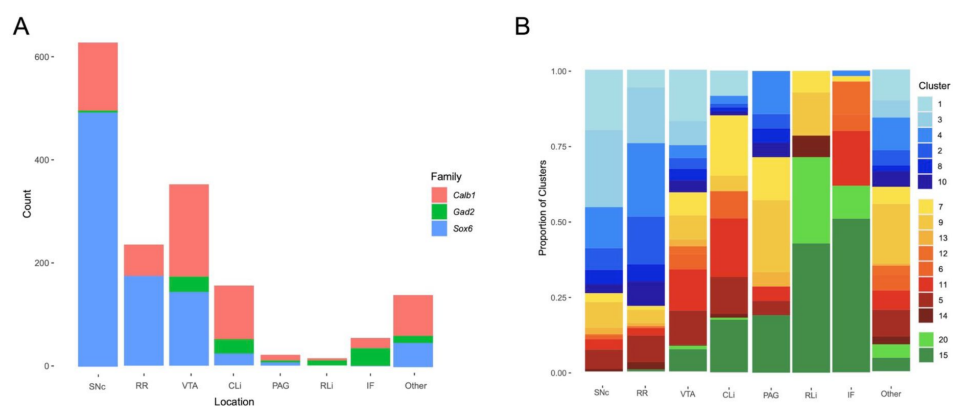


Figure S8.

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.

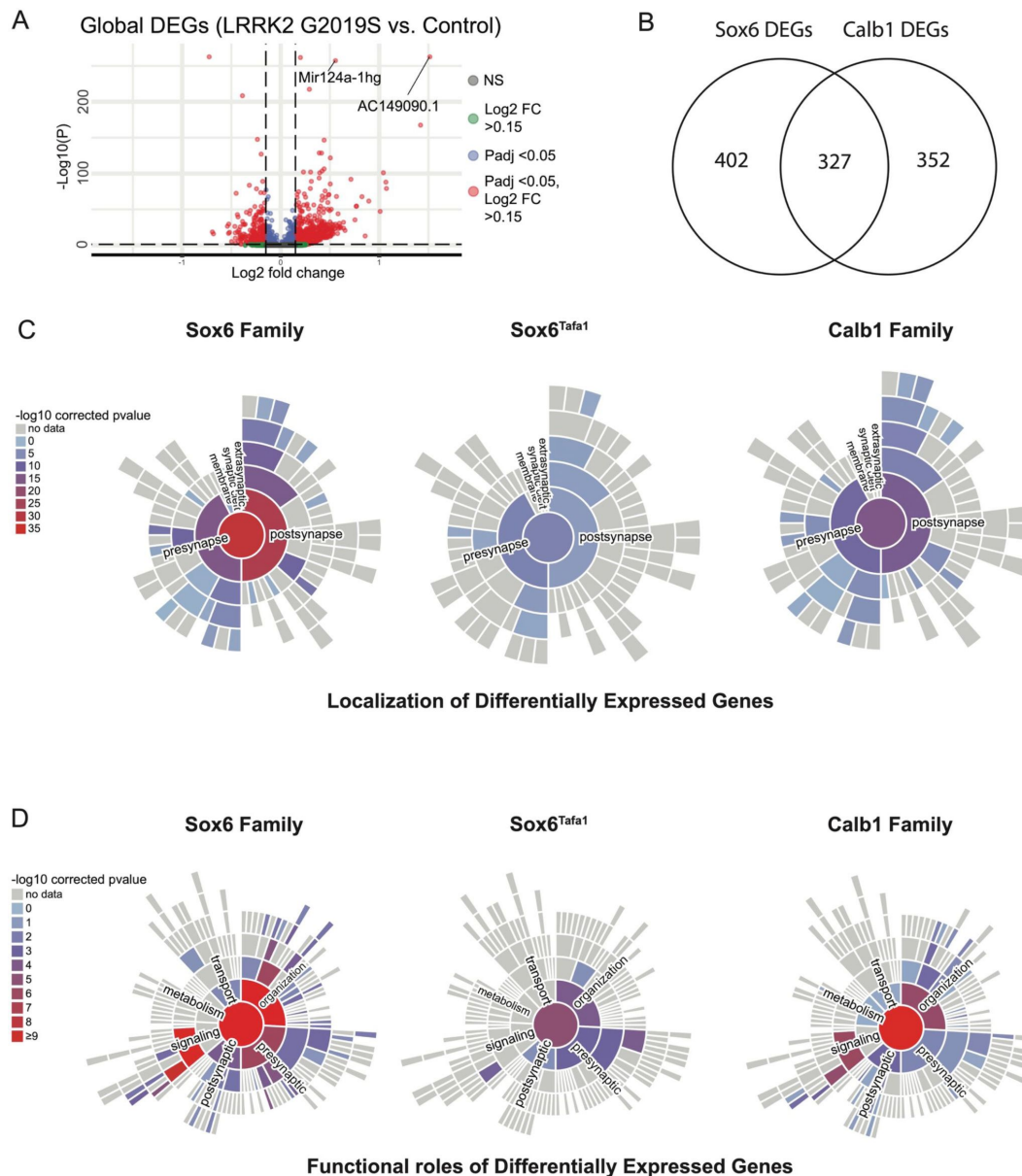


Figure S9.

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.

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.

Additional information

Code Availability Statement

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

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.

Additional files

Supplemental Table 1 [↗](#)

Extended Data [↗](#)

References

1. Björklund A, Dunnett SB (2007) **Dopamine neuron systems in the brain: an update** *Trends Neurosci* **30**:194–202 <https://doi.org/10.1016/j.tins.2007.03.006>
2. Fu Y, Yuan Y, Halliday G, Rusznák Z, Watson C, Paxinos G (2012) **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* **217**:591–612 <https://doi.org/10.1007/s00429-011-0349-2>
3. Gaertner Z, Azcorra M, Dombeck DA, Awatramani R (2022) **Molecular heterogeneity in the substantia nigra: A roadmap for understanding PD motor pathophysiology** *Neurobiol Dis* **175**:105925 <https://doi.org/10.1016/j.nbd.2022.105925>
4. Evans RC, Zhu M, Khaliq ZM (2017) **Dopamine Inhibition Differentially Controls Excitability of Substantia Nigra Dopamine Neuron Subpopulations through T-Type Calcium Channels** *J Neurosci* **37**:3704–20 <https://doi.org/10.1523/jneurosci.0117-17.2017>
5. Menegas W, Akiti K, Amo R, Uchida N, Watabe-Uchida M (2018) **Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli** *Nat Neurosci* **21**:1421–30 <https://doi.org/10.1038/s41593-018-0222-1>
6. Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw ND, Tank DW, Witten IB (2019) **Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons** *Nature* **570**:509–13 <https://doi.org/10.1038/s41586-019-1261-9>
7. Lammel S, Hetzel A, Häckel O, Jones I, Liss B, Roeper J (2008) **Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system** *Neuron* **57**:760–73 <https://doi.org/10.1016/j.neuron.2008.01.022>
8. da Silva JA, Tecuapetla F, Paixão V, Costa RM (2018) **Dopamine neuron activity before action initiation gates and invigorates future movements** *Nature* **554**:244–8 <https://doi.org/10.1038/nature25457>
9. Matsumoto M, Hikosaka O (2009) **Two types of dopamine neuron distinctly convey positive and negative motivational signals** *Nature* **459**:837–41 <https://doi.org/10.1038/nature08028>
10. Beier KT, Steinberg EE, DeLoach KE, Xie S, Miyamichi K, Schwarz L, Gao XJ, Kremer EJ, Malenka RC, Luo L (2015) **Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping** *Cell* **162**:622–34 <https://doi.org/10.1016/j.cell.2015.07.015>
11. Avvisati R, Kaufmann AK, Young CJ, Portlock GE, Cancemi S, Costa RP, Magill PJ, Dodson PD (2024) **Distributional coding of associative learning in discrete populations of midbrain dopamine neurons** *Cell Rep* **43**:114080 <https://doi.org/10.1016/j.celrep.2024.114080>
12. Brischoux F, Chakraborty S, Brierley DI, Ungless MA (2009) **Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli** *Proc Natl Acad Sci U S A* **106**:4894–9 <https://doi.org/10.1073/pnas.0811507106>

13. Heymann G, Jo YS, Reichard KL, McFarland N, Chavkin C, Palmiter RD, Soden ME, Zweifel LS (2020) **Synergy of Distinct Dopamine Projection Populations in Behavioral Reinforcement** *Neuron* **105**:909–20 <https://doi.org/10.1016/j.neuron.2019.11.024>
14. Howe MW, Dombeck DA (2016) **Rapid signalling in distinct dopaminergic axons during locomotion and reward** *Nature* **535**:505–10 <https://doi.org/10.1038/nature18942>
15. Lerner TN, Shilyansky C, Davidson TJ, Evans KE, Beier KT, Zalocusky KA, Crow AK, Malenka RC, Luo L, Tomer R, Deisseroth K (2015) **Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits** *Cell* **162**:635–47 <https://doi.org/10.1016/j.cell.2015.07.014>
16. Phillips RA, Tuscher JJ, Black SL, Andraka E, Fitzgerald ND, Ianov L, Day JJ (2022) **An atlas of transcriptionally defined cell populations in the rat ventral tegmental area** *Cell Rep* **39**:110616 <https://doi.org/10.1016/j.celrep.2022.110616>
17. Poulin JF, Gaertner Z, Moreno-Ramos OA, Awatramani R (2020) **Classification of Midbrain Dopamine Neurons Using Single-Cell Gene Expression Profiling Approaches** *Trends Neurosci* **43**:155–69 <https://doi.org/10.1016/j.tins.2020.01.004>
18. Garritsen O, van Battum EY, Grossouw LM, Pasterkamp RJ (2023) **Development, wiring and function of dopamine neuron subtypes** *Nat Rev Neurosci* **24**:134–52 <https://doi.org/10.1038/s41583-022-00669-3>
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 (2016) **Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells** *Cell* **167**:566–80 <https://doi.org/10.1016/j.cell.2016.09.027>
20. Kramer DJ, Risso D, Kosillo P, Ngai J, Bateup HS (2018) **Combinatorial Expression of Grp and Neurod6 Defines Dopamine Neuron Populations with Distinct Projection Patterns and Disease Vulnerability** *eNeuro* **5** <https://doi.org/10.1523/eneuro.0152-18.2018>
21. Hook PW, McClymont SA, Cannon GH, Law WD, Morton AJ, Goff LA, McCallion AS (2018) **Single-Cell RNA-Seq of Mouse Dopaminergic Neurons Informs Candidate Gene Selection for Sporadic Parkinson Disease** *Am J Hum Genet* **102**:427–46 <https://doi.org/10.1016/j.ajhg.2018.02.001>
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 (2019) **Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development** *Nat Commun* **10**:581 <https://doi.org/10.1038/s41467-019-08453-1>
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 (2018) **Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain** *Cell* **174**:1015–30 <https://doi.org/10.1016/j.cell.2018.07.028>
24. Yaghmaeian Salmani B, Lahti L, Gillberg L, Jacobsen JK, Mantas I, Svenningsson P, Perlmann T (2023) **Transcriptomic atlas of midbrain dopamine neurons uncovers differential vulnerability in a Parkinsonism lesion model** *Cold Spring Harbor Laboratory*

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 2023) **Unique functional responses differentially map onto genetic subtypes of dopamine neurons** *Nat Neurosci* **26**:1762–74 <https://doi.org/10.1038/s41593-023-01401-9>
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 2019) **Distinct Connectivity and Functionality of Aldehyde Dehydrogenase 1a1-Positive Nigrostriatal Dopaminergic Neurons in Motor Learning** *Cell Rep* **28**:1167–81 <https://doi.org/10.1016/j.celrep.2019.06.095>
27. Poulin JF, Caronia G, Hofer C, Cui Q, Helm B, Ramakrishnan C, Chan CS, Dombeck DA, Deisseroth K, Awatramani R 2018) **Mapping projections of molecularly defined dopamine neuron subtypes using intersectional genetic approaches** *Nat Neurosci* **21**:1260–71 <https://doi.org/10.1038/s41593-018-0203-4>
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 2006) **Kinase activity is required for the toxic effects of mutant LRRK2/dardarin** *Neurobiol Dis* **23**:329–41 <https://doi.org/10.1016/j.nbd.2006.04.001>
29. Tokars V, Chen C, Parisiadou L 2022) **Closing the structure-to-function gap for LRRK2** *Trends Biochem Sci* **47**:187–8 <https://doi.org/10.1016/j.tibs.2021.10.003>
30. Taymans JM, Fell M, Greenamyre T, Hirst WD, Mamais A, Padmanabhan S, Peter I, Rideout H, Thaler A 2023) **Perspective on the current state of the LRRK2 field** *NPJ Parkinsons Dis* **9**:104 <https://doi.org/10.1038/s41531-023-00544-7>
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 2018) **LRRK2 activation in idiopathic Parkinson's disease** *Sci Transl Med* **10** <https://doi.org/10.1126/scitranslmed.aar5429>
32. Khan SS, Jaimon E, Lin Y-E, Nikoloff J, Tonelli F, Alessi DR, Pfeffer SR 2024) **Loss of primary cilia and dopaminergic neuroprotection in pathogenic LRRK2-driven and idiopathic Parkinson's disease** *bioRxiv* <https://doi.org/10.1101/2024.01.15.575737>
33. Matikainen-Ankney BA, Kezunovic N, Mesias RE, Tian Y, Williams FM, Huntley GW, Benson DL 2016) **Altered Development of Synapse Structure and Function in Striatum Caused by Parkinson's Disease-Linked LRRK2-G2019S Mutation** *J Neurosci* **36**:7128–41 <https://doi.org/10.1523/jneurosci.3314-15.2016>
34. Chen C, Soto G, Dumrongprechachan V, Bannon N, Kang S, Kozorovitskiy Y, Parisiadou L 2020) **Pathway-specific dysregulation of striatal excitatory synapses by LRRK2 mutations** *eLife* **9** <https://doi.org/10.7554/eLife.58997>
35. Xenias HS, Chen C, Kang S, Cherian S, Situ X, Shanmugasundaram B, Liu G, Scesa G, Savio Chan C, Parisiadou L 2022) **R1441C and G2019S LRRK2 knockin mice have distinct striatal molecular, physiological, and behavioral alterations** *Commun Biol* **5**:1211 <https://doi.org/10.1038/s42003-022-04136-8>
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 2015) **Progressive dopaminergic alterations**

and mitochondrial abnormalities in LRRK2 G2019S knock-in mice *Neurobiol Dis* **78**:172–95 <https://doi.org/10.1016/j.nbd.2015.02.031>

37. Tozzi A, Tantucci M, Marchi S, Mazzocchetti P, Morari M, Pinton P, Mancini A, Calabresi P (2018) **Dopamine D2 receptor-mediated neuroprotection in a G2019S Lrrk2 genetic model of Parkinson's disease** *Cell Death Dis* **9**:204 <https://doi.org/10.1038/s41419-017-0221-2>
38. Poulin JF, Zou J, Drouin-Ouellet J, Kim KY, Cicchetti F, Awatramani RB (2014) **Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling** *Cell Rep* **9**:930–43 <https://doi.org/10.1016/j.celrep.2014.10.008>
39. Soden ME, Miller SM, Burgeno LM, Phillips PEM, Hnasko TS, Zweifel LS (2016) **Genetic Isolation of Hypothalamic Neurons that Regulate Context-Specific Male Social Behavior** *Cell Rep* **16**:304–13 <https://doi.org/10.1016/j.celrep.2016.05.067>
40. Nouri N, Awatramani R (2017) **A novel floor plate boundary defined by adjacent En1 and Dbx1 microdomains distinguishes midbrain dopamine and hypothalamic neurons** *Development* **144**:916–27 <https://doi.org/10.1242/dev.144949>
41. Kamath T, Abdulraouf A, Burris SJ, Langlieb J, Gazestani V, Nadaf NM, Balderrama K, Vanderburg C, Macosko EZ (2022) **Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease** *Nat Neurosci* **25**:588–95 <https://doi.org/10.1038/s41593-022-01061-1>
42. Anderegg A, Poulin JF, Awatramani R (2015) **Molecular heterogeneity of midbrain dopaminergic neurons--Moving toward single cell resolution** *FEBS Lett* **589**:3714–26 <https://doi.org/10.1016/j.febslet.2015.10.022>
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 (2021) **Sox6 expression distinguishes dorsally and ventrally biased dopamine neurons in the substantia nigra with distinctive properties and embryonic origins** *Cell Rep* **37**:109975 <https://doi.org/10.1016/j.celrep.2021.109975>
44. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X (2015) **RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells** *Science* **348**:aaa6090 <https://doi.org/10.1126/science.aaa6090>
45. Bradl M, Lassmann H (2010) **Oligodendrocytes: biology and pathology** *Acta Neuropathol* **119**:37–53 <https://doi.org/10.1007/s00401-009-0601-5>
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 (2023) **Molecularly defined and spatially resolved cell atlas of the whole mouse brain** *Nature* **624**:343–54 <https://doi.org/10.1038/s41586-023-06808-9>
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, Moonman 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 2023) **A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain** *Nature* **624**:317–32 <https://doi.org/10.1038/s41586-023-06812-z>
48. Langlieb J, Sachdev NS, Balderrama KS, Nadaf NM, Raj M, Murray E, Webber JT, Vanderburg C, Gazestani V, Tward D, Mezas C, Li X, Flowers K, Cable DM, Norton T, Mitra P, Chen F, Macosko EZ 2023) **The molecular cytoarchitecture of the adult mouse brain** *Nature* **624**:333–42 <https://doi.org/10.1038/s41586-023-06818-7>
 49. Trudeau LE, Hnasko TS, Wallén-Mackenzie A, Morales M, Rayport S, Sulzer D 2014) **The multilingual nature of dopamine neurons** *Prog Brain Res* **211**:141–64 <https://doi.org/10.1016/b978-0-444-63425-2.00006-4>
 50. Descarries L, Bérubé-Carrière N, Riad M, Bo GD, Mendez JA, Trudeau LE 2008) **Glutamate in dopamine neurons: synaptic versus diffuse transmission** *Brain Res Rev* **58**:290–302 <https://doi.org/10.1016/j.brainresrev.2007.10.005>
 51. Morales M, Root DH 2014) **Glutamate neurons within the midbrain dopamine regions** *Neuroscience* **282**:60–8 <https://doi.org/10.1016/j.neuroscience.2014.05.032>
 52. Morales M, Margolis EB 2017) **Ventral tegmental area: cellular heterogeneity, connectivity and behaviour** *Nat Rev Neurosci* **18**:73–85 <https://doi.org/10.1038/nrn.2016.165>
 53. Tritsch NX, Granger AJ, Sabatini BL 2016) **Mechanisms and functions of GABA co-release** *Nat Rev Neurosci* **17**:139–45 <https://doi.org/10.1038/nrn.2015.21>
 54. Conrad WS, Oriol L, Faget L, Hnasko TS. 2024) **Proportion and distribution of neurotransmitter-defined cell types in the ventral tegmental area and substantia nigra pars compacta** *bioRxiv* <https://doi.org/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 2014) **Sox6 and Otx2 control the specification of substantia nigra and ventral tegmental area dopamine neurons** *Cell Rep* **8**:1018–25 <https://doi.org/10.1016/j.celrep.2014.07.016>
 56. Singh A, Zhi L, Zhang H 2019) **LRRK2 and mitochondria: Recent advances and current views** *Brain Research* **1702**:96–104 <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 2015) **LRRK2 G2019S mutation attenuates microglial motility by inhibiting focal adhesion kinase** *Nat Commun* **6**:8255 <https://doi.org/10.1038/ncomms9255>
 58. Parisiadou L, Yu J, Sgobio C, Xie C, Liu G, Sun L, Gu XL, Lin X, Crowley NA, Lovinger DM, Cai H 2014) **LRRK2 regulates synaptogenesis and dopamine receptor activation through modulation of PKA activity** *Nat Neurosci* **17**:367–76 <https://doi.org/10.1038/nn.3636>
 59. Cook DA, Kannarkat GT, Cintron AF, Butkovich LM, Fraser KB, Chang J, Grigoryan N, Factor SA, West AB, Boss JM, Tansey MG 2017) **LRRK2 levels in immune cells are increased in Parkinson's disease** *NPJ Parkinsons Dis* **3**:11 <https://doi.org/10.1038/s41531-017-0010-8>

60. West AB, Cowell RM, Daher JP, Moehle MS, Hinkle KM, Melrose HL, Standaert DG, Volpicelli-Daley LA (2014) **Differential LRRK2 expression in the cortex, striatum, and substantia nigra in transgenic and nontransgenic rodents** *J Comp Neurol* **522**:2465–80 <https://doi.org/10.1002/cne.23583>
61. Pallos J, Jeng S, McWeeney S, Martin I (2021) **Dopamine neuron-specific LRRK2 G2019S effects on gene expression revealed by transcriptome profiling** *Neurobiol Dis* **155**:105390 <https://doi.org/10.1016/j.nbd.2021.105390>
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 (2015) **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* **24**:5299–312 <https://doi.org/10.1093/hmg/ddv249>
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 (2023) **Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain** *Nat Aging* **3**:121–37 <https://doi.org/10.1038/s43587-022-00335-4>
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 (2019) **PISD is a mitochondrial disease gene causing skeletal dysplasia, cataracts, and white matter changes** *Life Sci Alliance* **2** <https://doi.org/10.26508/lsa.201900353>
65. Thomas HE, Zhang Y, Stefely JA, Veiga SR, Thomas G, Kozma SC, Mercer CA (2018) **Mitochondrial Complex I Activity Is Required for Maximal Autophagy** *Cell Rep* **24**:2404–17 <https://doi.org/10.1016/j.celrep.2018.07.101>
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 (2023) **Cell-type specific molecular signatures of aging revealed in a brain-wide transcriptomic cell-type atlas** *bioRxiv* <https://doi.org/10.1101/2023.07.26.550355>
67. Zhang F, Yao Y, Miao N, Wang N, Xu X, Yang C (2022) **Neuroprotective effects of microRNA 124 in Parkinson's disease mice** *Archives of Gerontology and Geriatrics* **99**:104588 <https://doi.org/10.1016/j.archger.2021.104588>
68. Yao L, Ye Y, Mao H, Lu F, He X, Lu G, Zhang S (2018) **MicroRNA-124 regulates the expression of MEK3 in the inflammatory pathogenesis of Parkinson's disease** *J Neuroinflammation* **15**:13 <https://doi.org/10.1186/s12974-018-1053-4>
69. Han D, Dong X, Zheng D, Nao J (2019) **MiR-124 and the Underlying Therapeutic Promise of Neurodegenerative Disorders** *Front Pharmacol* **10**:1555 <https://doi.org/10.3389/fphar.2019.01555>
70. Angelopoulou E, Paudel YN, Piperi C (2019) **miR-124 and Parkinson's disease: A biomarker with therapeutic potential** *Pharmacol Res* **150**:104515 <https://doi.org/10.1016/j.phrs.2019.104515>
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 (2021)

Disruption of mitochondrial complex I induces progressive parkinsonism *Nature* **599**:650–6 <https://doi.org/10.1038/s41586-021-04059-0>

72. Surmeier DJ, Obeso JA, Halliday GM (2017) **Selective neuronal vulnerability in Parkinson disease** *Nat Rev Neurosci* **18**:101–13 <https://doi.org/10.1038/nrn.2016.178>
73. Mortiboys H, Johansen KK, Aasly JO, Bandmann O (2010) **Mitochondrial impairment in patients with Parkinson disease with the G2019S mutation in LRRK2** *Neurology* **75**:2017–20 <https://doi.org/10.1212/WNL.0b013e3181ff9685>
74. Singh F, Prescott AR, Rosewell P, Ball G, Reith AD, Ganley IG (2021) **Pharmacological rescue of impaired mitophagy in Parkinson's disease-related LRRK2 G2019S knock-in mice** *eLife* **10** <https://doi.org/10.7554/eLife.67604>
75. Liu H, Ho PW, Leung CT, Pang SY, Chang EES, Choi ZY, Kung MH, Ramsden DB, Ho SL (2021) **Aberrant mitochondrial morphology and function associated with impaired mitophagy and DNMT1-MAPK/ERK signaling are found in aged mutant Parkinsonian LRRK2(R1441G) mice** *Autophagy* **17**:3196–220 <https://doi.org/10.1080/15548627.2020.1850008>
76. Nguyen M, Krainc D (2018) **LRRK2 phosphorylation of auxilin mediates synaptic defects in dopaminergic neurons from patients with Parkinson's disease** *Proc Natl Acad Sci U S A* **115**:5576–81 <https://doi.org/10.1073/pnas.1717590115>
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 (2016) **A LRRK2-Dependent EndophilinA Phosphoswitch Is Critical for Macroautophagy at Presynaptic Terminals** *Neuron* **92**:829–44 <https://doi.org/10.1016/j.neuron.2016.09.037>
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 (2016) **Human R1441C LRRK2 regulates the synaptic vesicle proteome and phosphoproteome in a Drosophila model of Parkinson's disease** *Hum Mol Genet* **25**:5365–82 <https://doi.org/10.1093/hmg/ddw352>
79. Boecker CA, Goldsmith J, Dou D, Cajka GG, Holzbaur ELF (2021) **Increased LRRK2 kinase activity alters neuronal autophagy by disrupting the axonal transport of autophagosomes** *Curr Biol* **31**:2140–54 <https://doi.org/10.1016/j.cub.2021.02.061>
80. Skelton PD, Tokars V, Parisiadou L (2022) **LRRK2 at Striatal Synapses: Cell-Type Specificity and Mechanistic Insights** *Cells* **11** <https://doi.org/10.3390/cells11010169>
81. Pischedda F, Piccoli G (2021) **LRRK2 at the pre-synaptic site: A 16-years perspective** *Journal of Neurochemistry* **157**:297–311 <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 (2017) **Parkinson's Disease-Associated LRRK2 Hyperactive Kinase Mutant Disrupts Synaptic Vesicle Trafficking in Ventral Midbrain Neurons** *J Neurosci* **37**:11366–76 <https://doi.org/10.1523/jneurosci.0964-17.2017>
83. Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijssouw 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, Hugarir 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 (2019) **SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse** *Neuron* **103**:217–34 <https://doi.org/10.1016/j.neuron.2019.05.002>

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, Me Research T, System Genomics of Parkinson's Disease C, International Parkinson's Disease Genomics C (2019) **Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies** *Lancet Neurol* **18**:1091–102 [https://doi.org/10.1016/S1474-4422\(19\)30320-5](https://doi.org/10.1016/S1474-4422(19)30320-5)
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 (2022) **Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data** *Nature Genetics* **54**:1572–80 <https://doi.org/10.1038/s41588-022-01167-z>
86. Yaghmaeian Salmani B, Lahti L, Gillberg L, Jacobsen JK, Mantas I, Svenningsson P, Perlmann T (2024) **Transcriptomic atlas of midbrain dopamine neurons uncovers differential vulnerability in a Parkinsonism lesion model** *eLife* **12** <https://doi.org/10.7554/eLife.89482>
87. Blaess S, Bodea GO, Kabanova A, Chanet S, Mugniery E, Derouiche A, Stephen D, Joyner AL (2011) **Temporal-spatial changes in Sonic Hedgehog expression and signaling reveal different potentials of ventral mesencephalic progenitors to populate distinct ventral midbrain nuclei** *Neural Dev* **6**:29 <https://doi.org/10.1186/1749-8104-6-29>
88. Joksimovic M, Anderegg A, Roy A, Campochiaro L, Yun B, Kittappa R, McKay R, Awatramani R (2009) **Spatiotemporally separable Shh domains in the midbrain define distinct dopaminergic progenitor pools** *Proc Natl Acad Sci U S A* **106**:19185–90 <https://doi.org/10.1073/pnas.0904285106>
89. Yan CH, Levesque M, Claxton S, Johnson RL, Ang SL (2011) **Lmx1a and Lmx1b function cooperatively to regulate proliferation, specification, and differentiation of midbrain dopaminergic progenitors** *J Neurosci* **31**:12413–25 <https://doi.org/10.1523/jneurosci.1077-11.2011>
90. Andersson E, Tryggvason U, Deng Q, Friling S, Alekseenko Z, Robert B, Perlmann T, Ericson J (2006) **Identification of intrinsic determinants of midbrain dopamine neurons** *Cell* **124**:393–405 <https://doi.org/10.1016/j.cell.2005.10.037>
91. Cox J, Witten IB (2019) **Striatal circuits for reward learning and decision-making** *Nat Rev Neurosci* **20**:482–94 <https://doi.org/10.1038/s41583-019-0189-2>

92. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R 2019) **Comprehensive Integration of Single-Cell Data** *Cell* **177**:1888–902 <https://doi.org/10.1016/j.cell.2019.05.031>
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 2023) **A primate nigrostriatal atlas of neuronal vulnerability and resilience in a model of Parkinson's disease** *Nat Commun* **14**:7497 <https://doi.org/10.1038/s41467-023-43213-2>
94. Yamada T, McGeer PL, Baimbridge KG, McGeer EG 1990) **Relative sparing in Parkinson's disease of substantia nigra dopamine neurons containing calbindin-D28K** *Brain Res* **526**:303–7 [https://doi.org/10.1016/0006-8993\(90\)91236-a](https://doi.org/10.1016/0006-8993(90)91236-a)
95. Damier P, Hirsch EC, Agid Y, Graybiel AM 1999) **The substantia nigra of the human brain. I. Nigrosomes and the nigral matrix, a compartmental organization based on calbindin D(28K) immunohistochemistry** *Brain* **122**:1421–36 <https://doi.org/10.1093/brain/122.8.1421>
96. Kish SJ, Shannak K, Hornykiewicz O 1988) **Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications** *N Engl J Med* **318**:876–80 <https://doi.org/10.1056/nejm198804073181402>
97. Kordower JH, Olanow CW, Dodiya HB, Chu Y, Beach TG, Adler CH, Halliday GM, Bartus RT 2013) **Disease duration and the integrity of the nigrostriatal system in Parkinson's disease** *Brain* **136**:2419–31 <https://doi.org/10.1093/brain/awt192>
98. Mendonça MD, Ferreira PC, Oliveira F, Barbosa R, Meira B, Costa DC, Oliveira-Maia AJ, da Silva JA 2024) **Relative sparing of dopaminergic terminals in the caudate nucleus is a feature of rest tremor in Parkinson's disease** *NPJ Parkinsons Dis* **10**:209 <https://doi.org/10.1038/s41531-024-00818-8>
99. Schmack K, Bosc M, Ott T, Sturgill JF, Kepecs A 2021) **Striatal dopamine mediates hallucination-like perception in mice** *Science* **372** <https://doi.org/10.1126/science.abf4740>
100. Mandemakers W, Snellinx A, O'Neill MJ, de Strooper B 2012) **LRRK2 expression is enriched in the striosomal compartment of mouse striatum** *Neurobiol Dis* **48**:582–93 <https://doi.org/10.1016/j.nbd.2012.07.017>
101. Pischedda F, Piccoli G 2021) **LRRK2 at the pre-synaptic site: A 16-years perspective** *J Neurochem* **157**:297–311 <https://doi.org/10.1111/jnc.15240>
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 2023) **Transcriptomic diversity of cell types across the adult human brain** *Science* **382**:eadd7046 <https://doi.org/10.1126/science.add7046>
103. Choudhary S, Satija R 2022) **Comparison and evaluation of statistical error models for scRNA-seq** *Genome Biol* **23**:27 <https://doi.org/10.1186/s13059-021-02584-9>
104. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R 2024) **Dictionary learning for integrative, multimodal and scalable single-cell analysis** *Nature Biotechnology* **42**:293–304 <https://doi.org/10.1038/s41587-023-01767-y>
105. Ouyang JF, Kamaraj US, Cao EY, Rackham OJL 2021) **ShinyCell: simple and sharable visualization of single-cell gene expression data** *Bioinformatics* **37**:3374–6 <https://doi.org/10.1093/bioinformatics/btab192>

.1093/bioinformatics/btab209

106. Wolf FA, Angerer P, Theis FJ (2018) **SCANPY: large-scale single-cell gene expression data analysis** *Genome Biol* **19**:15 <https://doi.org/10.1186/s13059-017-1382-0>
107. Linderman GC, Zhao J, Roulis M, Bielecki P, Flavell RA, Nadler B, Kluger Y (2022) **Zero-preserving imputation of single-cell RNA-seq data** *Nat Commun* **13**:192 <https://doi.org/10.1038/s41467-021-27729-z>
108. Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J (2018) **Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor** *Nat Commun* **9**:884 <https://doi.org/10.1038/s41467-018-03282-0>

Author information

Zachary Gaertner[#]

Northwestern University Feinberg School of Medicine, Dept of Neurology, Chicago, United States, Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, United States

ORCID iD: [0000-0002-1760-6549](https://orcid.org/0000-0002-1760-6549)

[#]equal contribution

Cameron Oram[#]

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

ORCID iD: [0009-0003-2778-6031](https://orcid.org/0009-0003-2778-6031)

[#]equal contribution

Amanda Schneeweis

Northwestern University Feinberg School of Medicine, Dept of Neurology, Chicago, United States, Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, United States

ORCID iD: [0000-0003-4141-6064](https://orcid.org/0000-0003-4141-6064)

Elan Schonfeld

Northwestern University Feinberg School of Medicine, Dept of Neurology, Chicago, United States

ORCID iD: [0000-0001-7368-1562](https://orcid.org/0000-0001-7368-1562)

Cyril Bolduc

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

ORCID iD: [0009-0003-6104-8462](https://orcid.org/0009-0003-6104-8462)

Chuyu Chen

Northwestern University Feinberg School of Medicine, Dept of Pharmacology, Chicago, United States, Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, United States

ORCID iD: [0000-0001-5666-5173](https://orcid.org/0000-0001-5666-5173)

Daniel Dombeck

Northwestern University, Dept of Neurobiology, Evanston, United States, Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, United States

ORCID iD: [0000-0003-2576-5918](https://orcid.org/0000-0003-2576-5918)

Loukia Parisiadou

Northwestern University Feinberg School of Medicine, Dept of Pharmacology, Chicago, United States, Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, United States

ORCID iD: [0000-0002-2569-4200](https://orcid.org/0000-0002-2569-4200)

Jean-Francois Poulin

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

ORCID iD: [0000-0002-1039-4985](https://orcid.org/0000-0002-1039-4985)

For correspondence: j-francois.poulin@mcgill.ca

Rajeshwar Awatramani

Northwestern University Feinberg School of Medicine, Dept of Neurology, Chicago, United States, Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, United States

ORCID iD: [0000-0002-0713-2140](https://orcid.org/0000-0002-0713-2140)

For correspondence: r-awatramani@northwestern.edu

Editors

Reviewing Editor

Andrew West

Duke University, Durham, United States of America

Senior Editor

Kate Wassum

University of California, Los Angeles, Los Angeles, United States of America

Reviewer #1 (Public review):

Summary:

Dopamine neurons contribute to motivated and motor behaviors in many ways, and ample recent evidence has suggested that distinct dopamine neuron subclasses support discrete behavioral and circuit functions. Prior studies have subdivided dopamine neurons by spatial localization, gene expression patterns, and physiological properties. However, many of these studies were bound by previous technical limitations that made comprehensive subclassification efforts difficult or impossible. The main goal of this manuscript was to

characterize and further define dopamine neuron heterogeneity in the ventral midbrain. The study uses cutting-edge single nucleus RNA-seq (on the 10X Genomics platform) and spatial transcriptomics (on the MERFISH platform) to define dopamine neuron heterogeneity with unprecedented resolution. The result is a convincing and comprehensive subclassification of dopamine neurons into three main families, each with major branches and subtypes. In addition, the study reports comparisons between wild type mice and mice that harbor a G2019S mutation in the *Lrrk2* gene, which models a common cause of autosomally dominant Parkinson's Disease in humans. These results, while less robust due to the nature of the group comparisons, nevertheless identify vulnerability within specific dopamine neuron subpopulations. This vulnerability may contribute unique risk to dopamine neuron loss in the context of Parkinson's disease. Overall, the study is careful and rigorous and provides a critical resource for the rapidly evolving knowledge of dopamine neuron subtypes.

Strengths:

- The creation of a public-facing app where the snRNA-seq data can be investigated by anyone is a major strength.
- The manuscript includes careful comparisons to prior datasets that have sought to explore dopamine neuron heterogeneity. The result is a useful synthesis of new findings with previously published work, which is helpful for moving the field forward in this area.
- The integration of snRNA-seq with MERFISH results is particularly strong, and enables insight not only into subclassification, but also into how this relates to spatial localization. The careful neuroanatomy reveals important distinctions between Sox6, Calb1, and Gad2 positive dopamine neuron families, with some degree of spatial overlap.

<https://doi.org/10.7554/eLife.101035.2.sa2>

Reviewer #2 (Public review):

Gaertner and colleagues present a study examining the transcriptomic diversity and spatial location of dopaminergic neurons from mice and examine the changes in gene expression resulting from knock in of the Parkinson's LRRK G2019S risk variant. Overall, I found the manuscript presented their study very clearly, well written with very clear figures for the most part. I am not an expert on mouse neuroanatomy but found their classification reasonably well justified and spatial orientation of dopaminergic neurons within the mouse brain informative and clear. While trends were clear and well presented, the apparent spatial heterogeneity suggests that knowledge of the functional connections and roles of these neurons will be required to better interpret the results presented but nonetheless their findings exposed significant detail that is required for further understanding.

The study of the transcriptional effects of the LRRK2 KI was also informative and clearly framed in terms of a focused analyses on the effects of the KI only on dopaminergic neurons.

I thank the authors for addressing my previous concerns and comments, and feel they have done so well. I agree that as GSEA only includes ranked genes from the specific study, the gene set is already limited to the relevant background.

<https://doi.org/10.7554/eLife.101035.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Reviewer #1 (Public review):

Weaknesses:

(1) Important details about the nature of DEG comparisons between the wild type and the Lrrk2 G2019S model are missing.

Please see the recommendations section below for specific responses to individual comments from Reviewer #1.

(2) Some aspects of the integration between snRNA-seq and MERFISH data are not clear, and many MERFISH-identified cells do not appear to have a high-confidence cluster transfer into the snRNA-seq data space. Imputation is used to overcome some issues with the MERFISH dataset, but it is not clear that this is appropriate.

Please see the recommendations section below for specific responses to individual comments from Reviewer #1.

Reviewer #2 (Public review):

(1) In the GO pathway analyses (both GSEA and DEG GO), I did not see a correction applied to the gene background considered. The study focusses on dopaminergic neurons and thus the gene background should be restricted to genes expressed in dopaminergic neurons, rather than all genes in the mouse genome. The problem arises that if we randomly sample genes from dopaminergic neurons instead of the whole genome, we are predisposed to sampling genes enriched in relevant cell-type-specific roles (and their relevant GO terms) and correspondingly depleted in genes enriched in functions not associated with this cell type. Thus, I am unsure whether the results presented in Figures 8 and 9 may be more likely to be obtained just by randomly sampling genes from a dopaminergic neuron. The background should be limited and these functional analyses rerun.

Thank you for pointing out this important concern. We agree that overrepresentation analyses (ORAs) are vulnerable to selecting cell-type specific markers as significantly differentially expressed and thus inflating detection of cell-type associated gene sets rather than those truly altered as a function of experimental condition. We have thus re-run the GO analyses in our study with the genetic background being adjusted for each individual comparison. For dataset-level GO in Fig 8, genetic background was defined as genes with expression detected in at least 5% of all cells (to approximate the inclusion of cluster-specific genes). 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. Interestingly, this correction appears to have filtered out or lowered the significance of some of the more generic brain-associated pathways that we initially presented, such as axonogenesis or learning and memory, and we feel even more confident in our original interpretation.

Functional class scoring methods like GSEA, however, are unlike ORAs in that they do utilize a hypergeometric test to calculate overrepresentation as no distinction is made between significant and non-significant differential gene expression (nor is a genetic background provided as input to this tool). GSEA takes as input the full DE results, ranking genes according to their association with either group. Thus, genes simply enriched in DA neurons should be present towards both extremes of the rank list, rather than uniformly skewed toward one extreme. Per the GSEA authors' user manual and original source paper, the

entirety of DE testing should be provided as input for GSEA (barring genes with detection levels so low that their differential expression and/or ranking is likely to be artifactual):

“The GSEA algorithm does not filter the expression dataset and generally does not benefit from your filtering of the expression dataset. During the analysis, genes that are poorly expressed or that have low variance across the dataset populate the middle of the ranked gene list and the use of a weighted statistic ensures that they do not contribute to a positive enrichment score. By removing such genes from your dataset, you may actually reduce the power of the statistic and processing time is rarely a factor as GSEA can easily analyze 22,000 genes with even modest processing power. However, an exception exists for RNA-seq datasets where GSEA may benefit from the removal of extremely low count genes (i.e., genes with artifactual levels of expression such that they are likely not actually expressed in any of the samples in the dataset).” [<https://www.gsea-msigdb.org/gsea/doc/GSEAUserGuideFrame.html>]

In our study, this filtering of very low expression genes (to account for artifactually inflated fold changes or a large number of ties in the rank list that are subsequently ordered at random) occurred at the level of DE testing using the Seurat FindMarkers command, in which differential expression calculations were only performed for genes that were detected in a minimum of 10% of cells in the dataset.

(2) In the scDRS results, I am unsure what is significant and what isn't. The authors refer to relative measures in the text ("highest") but I do not know whether these differences are significant nor whether any associations are significantly unexpected. Can the x-axis of scDRS results presented in Figure 9 H and I be replaced with a corrected p-value instead of the scDRS score?

An important distinction should be made here between scDRS and similar approaches that utilize overrepresentation analyses to assess for associations of DEGs with putative risk genes, similar to the GO analyses performed in our paper. 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 (see Methods or Zhang et al., Nat Genetics 2022 for more details). While scDRS can be used to generate a p value for each individual cell in the dataset, scDRS does not have a native method for defining group-level p values, nor have we attempted to calculate group-level p values here. In order to compare cluster-level mean scDRS scores and determine their significance, we created bootstrapped 95% confidence intervals for the mean scDRS score of each cluster or family (shown by the error bars in forest plots 9G, 9H). A score of 0 represents the null hypothesis 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. Similarly, groups can be compared to each other in the same way to determine if the group-level mean scDRS score is significantly different across a given pair. However, this overlap of confidence intervals should be interpreted cautiously, as there are a large number of potential comparisons that can be made, creating the potential for Type I error. We have added language to clarify what the scDRS score represents, and to ensure it is not conflated with approaches such as GO or GSEA.

(3) The results discussed at the bottom of page 13 [page 14 of new version] state that 48.82% of the proteins encoded by the Calb1 DEGs have pre-synaptic localisations as opposed to 45.83% of the SOX6 DEGs, which does not support the statement that "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)".

Thank you for pointing this out; the error here lies in the wording of the results. The percentages mentioned above describe the percentages within the synaptic localized genes rather than the total DEG lists. We have rephrased this section for clarity to include both the percentages within this category as well as the total (the results of which are in line with our original statement).

(4) While an interest in the Sox6^{Tafa1} subtype is explained through their expression of Anxa1 denoting a previously identified subtype associated with locomotory behaviours, it was unclear to me how to interpret the functional associations made to DEGs in this subtype taken out of context of other subtypes. Given all the other subtypes, it is not possible to ascertain how specific and thus how interesting these results are unless other subtypes are analysed in the same way and this Sox6^{Tafa1} subtype is demonstrated as unusual given results from other subtypes.

In our study, we chose to specifically focus on this population given its unique acceleration-locked functional activity pattern observed in Azcorra & Gaertner et al, Nat Neuro 2023, as there are technical limitations that warrant cautious application of the above approach. We agree that the associations of this population to the described DEGs cannot be interpreted as unique to this population given the data presented and have added language to this effect within the text. There are two major challenges to analyzing all other subtypes to provide a comparison. Firstly, given the number of subtypes involved and number of downstream analyses, it is computationally intensive to carry out this analysis. More importantly however, the results cannot be easily compared across different populations due to the variability in both cluster size and internal heterogeneity of each cluster, as the statistical power in calculating DEGs will be inherently different across these populations (i.e. smaller or more heterogeneous clusters would be expected to show a lower number of DEGs reaching significance). While pseudo bulk testing is effective for mitigating these factors, our limited sample number (n=2 independently generated datasets per group) dramatically underpowers differential expression testing using pseudo bulk analysis. One solution is to uniformly limit each cluster size to the minimally observed cluster size through random down-sampling. While this allows the 'n' in DE calculations to be uniform, this potentially worsens the problem of internal heterogeneity, which would remain roughly constant but in the setting of a lower 'n', increasing the variability in results for larger clusters. To provide a comparator for the population of interest we focused on, we have performed this down sampling approach in order to compare Sox6^{Tafa1} to another cluster within the VTA, Calb1^{Stac}, that also expresses high levels of *Anxa1* and *Aldh1a1* given the broad interest in these markers as proxies for vulnerability. The results of this comparison are now shown in Figure S10.

(5) On p12, the authors highlight Mir124a-1hg that encodes miR-124. This is upregulated in Figure 8D but the authors note this has been to be downregulated in PD patients and some PD mouse models. Can the authors comment on the directional difference?

We have adjusted the text to reflect this discrepancy and speculate on why this may be observed. In short, one hypothesis is that miR-124, given its proposed neuroprotective effects, is increased in DA neurons facing toxic metabolic insults as a compensatory response. In our prodromal model without observable degeneration, this could represent an early sign of cell stress. While speculative, in PD patients or overtly degenerative models, lack of compensatory miR-124 or fulminant cell death among vulnerable cells could result in an observed decrease in miR-124 expression.

(6) Lastly, can the authors comment on the selection of a LogFC cut-off of 0.15 for their DEG selection? I couldn't see this explained (apologies if I missed it).

The 0.15 cutoff was selected arbitrarily based on the observed range of fold changes seen among our differentially expressed genes. However, importantly, this cutoff was not used for defining DEGs for downstream analyses such as GSEA or GO, nor for defining significance of differential expression, which was done purely based on FDR-adjusted p values <0.05. The selection of 0.15 affects only the coloring seen in the volcano plot, which we have decided to move to supplemental figures given the uniformly small effect size seen in individual genes and a separate reviewer comment regarding concern in the field over differential expression testing methods in single-cell datasets. Instead, this figure now focuses on highlighting pathway- and gene-set level comparisons that can provide easier interpretation of small, but concordant changes across swaths of genes.

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

(1) In the MERFISH dataset, only around half of the DAergic cells (2,297 of 4,532) were successfully projected into the snRNA-seq UMAP space, based on a similarity score > 0.5. Additionally, key transcripts that were used to define the snRNA-seq clusters (such as Sox6) were not identified at all in the MERFISH dataset. This raises some questions about the ability to integrate and compare these datasets directly, which are not fully considered in the manuscript. These discrepancies are smoothed over using imputation, which allows specific class-defining genes such as Sox6 to be plotted on spatial coordinates in Figure 4D. However, imputation is not without caveats, and the appropriateness of the imputation is not well considered in the text.

We fully agree with the reviewer that the use of an imputation approach needs to be clarified and justified thoroughly. We added a sentence to better clarify the process of imputation on Page 9 “The imputed gene expression is extrapolated from anchors established from pairwise correspondences of cell expression levels between MERFISH and snRNA-Seq datasets.” This pair-wise cell correspondence as defined by anchors can be assessed using Seurat confidence score. We acknowledge the fact that only about 50% of cells could confidently be transferred onto the snRNA-Seq data. This is the result of using a stringent confidence level of 0.5 (similar to previous publications, PMID: 38092916 & 38092912). We preferred mapping fewer high-confidence cells than potentially misrepresenting the spatial location of some of these clusters.

It is also important to demonstrate the reliability of gene imputation. Indeed as pointed out by the reviewer, some probes such as Sox6 were not detected in the MERFISH dataset. To strengthen our data integration and as already mentioned in the manuscript, we excluded 219 genes based on the deviation of average counts per cell between the datasets. The fact that the imputed expression of Sox6 perfectly reflects its well-characterized distribution (PMIDs: 25127144, 30104732, 25437550, 34758317) strengthened our confidence in our imputation pipeline. We also looked at the correlation of imputed gene expression with the detected transcripts in our MERFISH experiments. We added a new supplemental figure (S7) highlighting the correlations between MERFISH and imputed gene expression of 8 genes (4 for each Sox6 and Calb1 family). Together Fig S6 and S7 show the range of correlations between imputed and actual MERFISH transcript. Altogether, we can observe relatively high correlation between the number of detected transcripts per gene in snRNA-Seq and MERFISH datasets

In addition, we added a paragraph discussing limitations of gene expression imputation on page 17: “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

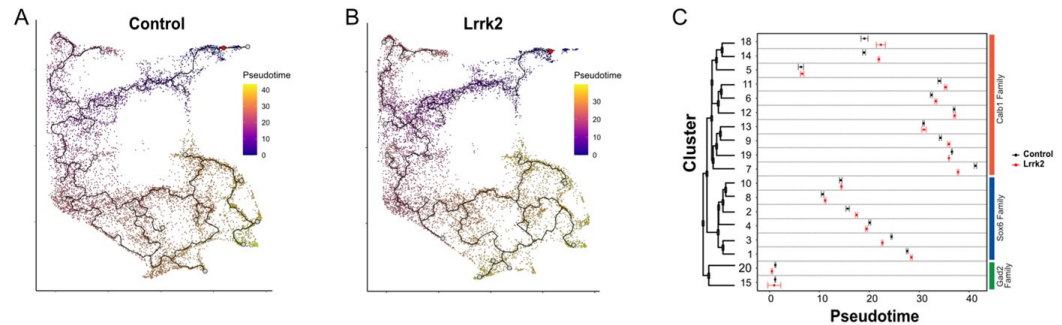
as described by Stuart et al.(92) has been used in several recent studies integrating spatial and transcriptomic data(46, 47). It relies on identifying anchors that enable projection of MERFISH data onto the UMAP space of a snRNA-Seq dataset and then uses neighboring cells to extrapolate the expression 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.”

(2) In the discussion, the authors argue that the cellular classifications identified here for DA neurons are more likely to reflect discrete cell types than cell states. The rationale for this conclusion is largely based on the absence of subtype differences between wild-type and LRRK2 G2019S transgenic mice. I do not find this argument to be convincing, because it is still possible that certain subdivisions simply reflect dynamic cell states that are also not grossly altered in the mutant mouse. A stronger argument for this claim would be to include trajectory-based analyses that do not show predicted transition points between nearby or related clusters.

We thank the reviewer for pointing out this particular limitation as differentiating “cell type” and “cell states” been debated in the field for years with no consensus emerging how to address the issue. As suggested, we performed a trajectory analysis using Monocle3 on both control and Lrrk2 samples. We’ve built the trajectory map, taking cluster 20 as the starting node. To avoid potential biased trajectories induced by different cell coverage, we’ve down sampled the Lrrk2 condition to match the number of cells of wildtype. As expected, since most of the DA clusters are not segregated in the UMAP space, the trajectory analysis showed predicted transitions between clusters (see Author response image 1A and 1B). Even though some clusters’ pseudotime score were statistically different between the wildtype and Lrrk2 samples, they overall remained similar (Author response image 1C). This analysis suggests that the LRRK2G2019S mutation induces a mild transcriptional perturbation but does not result in a major cell state drift. Indeed, we believe changes in the observed trajectory path would disappear as the number of cells analyzed increases. Because of this bias introduced by cell coverage, we prefer not to include this trajectory analysis in the manuscript to avoid misleading readers. Thus, as suggested by the reviewer, we softened our claim to “This suggests that our taxonomic scheme is agnostic to a mild perturbation such as LRRK2G2019S, 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.”.

Author response image 1.

A) Trajectory analysis of wildtype and B) LRRK2^{G2019S} samples. C) Pseudotime scores for each cluster across wildtype and Lrrk2 conditions. Error bars represent the confidence of error for false positives discovery rate of 5%.



(3) The relationship between individual samples, GEMwell, and sequenced library should be clarified. If independent samples were combined into one GEMwell, this should be explicitly stated for clarity.

We have revised the text to better clarify the methodology. In brief, each of our 4 independent samples (2 control, 2 mutants; equal sexes per sample) were isolated from n=2 pooled mice (for a total n=8 mice across the 4 samples). Each sample was processed in its own GEM well to produce 4 distinct libraries that were subsequently sequenced and analyzed as described.

(4) Please include more details on DEG testing in the manuscript, this is key for interpreting the robustness of certain findings. Ideally, pseudobulk comparisons would be used here (given concerns in the field that DEG testing where N = number of cells artificially inflates the statistical power, violates assumptions of independence, and results in false positive DEGs).

While we agree that pseudobulk analysis would be ideal for reducing false positives, our study, while exceptionally large in total numbers of DA cells profiled, was generated from 4 total 10X libraries as described above, without any mechanism to definitively demultiplex to the original n=8 source mice. Thus, pseudobulk comparisons would be performed using only n=2 per group, which is below the recommended sample size for these methods. Given this concern, we have moved the volcano plot from Figure 8D to the supplementals and added language to the methods and relevant figure legend acknowledging the limitation in Seurat's default differential expression analysis methodology.

<https://doi.org/10.7554/eLife.101035.2.sa0>