Transcriptomic atlas of midbrain dopamine neurons uncovers differential vulnerability in a Parkinsonism lesion model

Behzad Yaghmaeian Salmani[1], Laura Lahti[1], Linda Gillberg[1], Jesper Kjaer Jacobsen[1, 2], Ioannis Mantas[3], Per Svenningsson[3] and Thomas Perlmann[1, *]

[1]Department of Cell and Molecular Biology, Karolinska Institutet, SE-171 77, Stockholm, Sweden

[2]Department of Neurology, Karolinska University Hospital, Stockholm, Sweden

[3]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[*]Corresponding author: Thomas Perlmann, (thomas.perlmann@ki.se)

Abstract

Midbrain dopamine (mDA) neurons comprise diverse cells with unique innervation targets and functions. This is illustrated by the selective sensitivity of mDA neurons of the substantia nigra compacta (SNc) in patients with Parkinson's disease, while those in the ventral tegmental area (VTA) are relatively spared. Here we used single nuclei RNA sequencing (snRNA-seq) of approximately 70,000 mouse midbrain cells to build a high-resolution atlas of mouse mDA neuron diversity at the molecular level. The results showed that differences between mDA neuron groups could best be understood as a continuum without sharp differences between subtypes. Thus, we assigned mDA neurons to several "territories" and "neighborhoods" within a shifting gene expression landscape where boundaries are gradual rather than discrete. Based on the enriched gene expression patterns of these territories and neighborhoods, we were able to localize them in the adult mouse midbrain. Moreover, because the underlying mechanisms for the variable sensitivities of diverse mDA neurons to pathological insults are not well understood, we analyzed surviving neurons after partial 6-hydroxydopamine (6-OHDA) lesions to unravel gene expression patterns that correlate with mDA neuron vulnerability and resilience. Together, this atlas provides a basis for further studies on the neurophysiological role of mDA neurons in health and disease.

Introduction

Dopamine (DA) neurons localized in the ventral midbrain (mDA neurons) comprise the brain's largest and most extensive source of DA neurotransmission (Björklund and Dunnett, 2007). These neurons are known to perform a wide variety of functions, and distinct mDA neuron assemblies are known to play diverse functional roles. For example, mDA neurons of the substantia nigra pars compacta (SNc) are mainly associated with controlling motor functions. In contrast, those in the ventral tegmental area (VTA) and retrorubral field (RRF) have been associated more strongly with motivation, learning, and cognition (Björklund and Dunnett, 2007). Notably, mDA neurons are also characterized by their varying impact on disease. For example, the hallmark motor symptoms of Parkinson's disease (PD) result from the degeneration of SNc neurons. On the other hand, mDA neurons in the VTA have been linked to neural pathways that are affected in addiction, schizophrenia, attention deficit hyperactivity disorder, and other psychiatric conditions (Björklund and Dunnett, 2007; Garritsen et al., 2023; Poulin et al., 2020).

The emergence of single-cell RNA sequencing and other methods that resolve differences in, e.g., chromatin state between single cells, have increased our understanding of mDA neuron diversity (Garritsen et al., 2023; Poulin et al., 2020). This diversity is of great significance in normal brain functions and disease, e.g., in forming multiple DA-controlled circuits enabling versatile DA neurotransmission and explaining different vulnerabilities of distinct groups of mDA neurons. However, methods dependent on capturing certain cell types, such as neurons sensitive to cell dissociation, can result in biased datasets where the most sensitive neuron groups are underrepresented. In addition, as the scRNAseq studies of mDA neurons published so far have analyzed a relatively low number of mDA neurons, the resolution of these datasets is still quite limited. Single-nucleus RNA sequencing (snRNA-seq) is a more efficient and reliable sampling alternative for individual neurons

(Krishnaswami et al., 2016). Indeed, snRNAseq of human mDA neurons, enriched by sorting for NR4A2-labelled nuclei, yielded a 180-fold increase in absolute neuronal numbers compared to similar datasets based on scRNAseq (Kamath et al., 2022).

Our present study combined fluorescent sorting and snRNA-seq to explore the diversity of a large number of mouse mDA neurons, resulting in a comprehensive atlas of mDA neuron populations. By additionally analyzing nuclei from a partial toxin-induced mDA neuron degeneration model, we were able to investigate questions of selective and variable vulnerability among these distinct populations. Our findings provide a foundation for future investigations into the projections and functions of molecularly distinct mDA neuron groups.

Results

Single nuclei RNA sequencing of the ventral midbrain

Perhaps partly because of the challenges in dissociating and isolating whole mDA neurons, previous studies have only analyzed a relatively small fraction of all mDA neurons from each dissected brain (Dougalis et al., 2012; Hook et al., 2018; Manno et al., 2016; Poulin et al., 2014; Tiklova et al., 2019). This raises concerns about the potential for sampling bias in those studies. To overcome this limitation, we used single nuclei RNA sequencing (snRNA-seq) for mice expressing nuclear mCherry under the control of the *Slc6a3* (*DAT*, dopamine transporter) promoter (Figure 1A-C)(Hupe et al., 2014). Nuclei isolated by fluorescent sorting were sequenced using the 10X Genomics Chromium platform (10Xv3). Using this strategy, more than 20% of all mDA neurons were consistently collected and sequenced from each brain, an improvement of over 50X from our previous study (Tiklova et al., 2019).

In addition to assessing the diversity among the mDA neuron population, we also aimed to analyze how diverse mDA neuron groups were affected by cellular stress. Accordingly, mice (n=6) were unilaterally injected with 6-hydroxydopamine (6-OHDA, 0.7µg or 1.5µg) into the medial forebrain bundle, which leads to partial cell death of mDA neurons on one hemisphere of the brain (Figure 1A). To determine differential survival among diverse mDA neurons, nuclei from intact and lesioned hemispheres were collected six weeks after lesioning when the acute phase of cell death had already occurred. To avoid omitting low red fluorescent protein (RFP)-expressing nuclei from analysis, a relatively relaxed gating in fluorescently activated nuclear sorting (FANS) was used, resulting in an approximately equal number of mCherry^{+} and mCherry^{-} nuclei from each sample per brain (Figure 1–figure supplement 1). After snRNAseq and quality control, 36,051 nuclei from the intact and 32,863 from the lesioned hemisphere were retained for bioinformatic analysis (see Methods, Figure 1 - figure supplement 2).

Sequencing data were subjected to normalization, variance stabilization, and dimensional reduction and then visualized by uniform manifold approximation (UMAP; Figure 1D). Louvain clustering was adopted, dividing all nuclei into 26 clusters (Methods, Figure 1–figure supplement 3). A hierarchical clustering dendrogram illustrating the relationship between clusters was also constructed, and cluster-specific markers were identified and used to annotate previously known cell types (Figure 1D, E; Figure 1–figure supplement 3, Supplementary file 1). This analysis revealed the presence of different cell types, including oligodendrocytes; oligodendrocyte precursor cells; astrocytes; and various types of neurons known to be localized in the midbrain but also with some inclusion of cell types known to be localized both in the forebrain and hindbrain. Neuronal clusters were mainly comprised of different types of glutamatergic (GLU), gamma-aminobutyric acid (GABA), and dopamine neurons (Figure 1–figure supplement 3). This dataset is accessible via: http://193.10.16.198:5005/.

Dopamine neuron diversity

Plotting known markers for mDA neurons in the UMAP revealed that a substantial number of all analyzed nuclei, clusters 1, 4, 7, and 10, were derived from mDA neurons (Figure 1–figure supplement 3). To focus the analysis on dopamine neurons, 33,052 nuclei expressing above threshold levels of either *Th* or *Slc6a3* (Th^{+}/DAT^{+}) were selected for further computational analysis. Most of these filtered nuclei, visualized in a *de novo* UMAP, were positive for several genes known to be expressed in mDA neurons (Figure 2A, B). This dataset is accessible via: http://193.10.16.198:5006/.

K-means clustering, followed by hierarchical clustering, was used to divide Th^{+}/DAT^{+} nuclei into 71 unique clusters whose relationship was also visualized in a dendrogram (Figure 2C). Although the analysis of gene expression in a dot-plot confirmed that most nuclei were

dopaminergic, a smaller number of nuclei distributed in clusters expressing low levels of mDA neuron markers showed characteristics of GABAergic, glutamatergic neurons, or maturing myelinating oligodendrocytes (mODC). A few clusters remained unassigned (Figure 2A, C). Almost all clusters expressing genes required for synthesis and handling DA (dopaminergic) also expressed mDA neuron markers, including *En1, Nr4a2,* and *Pitx3*, the latter at a lower abundance, except one that expressed hypothalamic dopaminergic markers, including*, Satb2, Prlr,* and *Gad1* (Figure 2A, C; referred to as HyDA neurons). The mDA neuron clusters form a major separate branch in the dendrogram, while HyDA neurons constitute a more distant branch (Figure 2C). Two closely related clusters (#12 and #26) form a distinctly separate group, predominantly (~95%) composed of nuclei from the lesioned hemisphere. Upon closer inspection, they share the same parent node with mDA clusters marked by high abundance and expression level of *Th.* However*,* the expression of *Th* and other dopaminergic markers in these two lesion-specific clusters is relatively low (Figure 2C). The lesion-specific clusters and the consequences of lesioning will be further described below.

mDA neuron territories and neighborhoods

Illustration of diversity among different sub-categories of cells, such as mDA neurons, using a dendrogram can give a somewhat misleading impression that individual branching endpoints (referred to as "leaf nodes") would represent transcriptionally distinct cellular subtypes. However, it may be more accurate to consider differences between closely related mDA neuron groups as a gradually shifting continuum within a gene expression landscape. Thus, we found using a hierarchical definition of the relationship more accurate when using the dendrogram to describe diversity. By this definition, mDA neurons are distinct and constitute a neuron type, clearly different from, for example, HyDA and other types of GABAergic and

glutamatergic neurons. Within individual neuron types, such as the closely related mDA neurons, we define the next hierarchical level as "territories," while "neighborhoods" is the most highly resolved level of diversity that, based on the bioinformatic analysis, can be defined meaningfully. According to this classification, each one of the territories and neighborhoods should constitute specified adjacent leaf nodes of the cluster tree with a common ancestor node in the hierarchical dendrogram. Notably, leaf nodes of a neighborhood or territory need not be sibling nodes.

Furthermore, they should be distinguishable by clearly defined gene expression signatures and occupy a relatively well-defined space, i.e., not scattered, in the UMAP. This classification subdivided mDA neurons into 7 territories and 16 neighborhoods (Figure 3A, C; Figure 3–figure supplement 1). The territories were named according to selected distinguishing genes (*Sox6, Gad2*, *Fbn2*, *Pcsk6*, *Pdia5*, *Ebf1,* and *Otx*2) and can be identified by additional markers as indicated in Figure 3 (see also Figure 3–figure supplement 1). It should be noted that these markers were typically not unique for a particular territory but, in combination, defined gene signatures that distinguished territories from each other.

The transcription factor Sox6 is known to be expressed in mDA neurons of the SNc and in some neurons of the VTA (Luppi et al., 2021; Panman et al., 2014; Poulin et al., 2014). The *Sox6*-territory included approximately one-third of all sequenced mDA neuron nuclei (Figure 3A). This territory expressed other previously defined SNc markers, including *Kcnj6 (Girk2)* (Figure 3–figure supplement 1) and additional genes representing a gene signature that included *Col25a1* and *Tigar*, as well as *Nwd2, Pex5l, Cdh11,* and *Serpine2* (Figure 3C; Figure 3–figure supplement 1). In contrast, expression of *Calb1*, encoding the known VTA marker Calbindin 1, was either absent or was expressed at low levels in *Sox6*-territory nuclei (Figure 3C). This almost non-intersecting, mutually exclusive expression of *Sox6* and *Calb1* defined a

striking major division among all analyzed mDA neuron nuclei, following the neuroanatomical division of mDA neurons into SNc and VTA (Figure 3B).

The remaining six territories consisting of approximately 15,000 mDA nuclei were more diverse and, based on the prominent expression of *Calb1* and absence of high *Sox6* expression, presumably represented mDA neurons localized outside SNc. Indeed, these territories expressed additional VTA markers, including *Otx2* (Figure 3C). Expression of mDA neuron markers such as *Th, Ddc,* and *Slc6a3* was particularly weak in the *Gad2*-territory, which displayed typical features of GABAergic neurons, including *Gad2* and *Slc32a1*.

Our next focus was to describe mDA neuron diversity within territories by dividing them into more highly resolved neighborhoods (NHs, Figure 3C, Figure 3–figure supplement 1). In rare cases, the smallest hierarchical domain for a neighborhood can be one dendrogram leaf node. Still, most consist of at least two leaf nodes. Four neighborhoods were defined within the *Sox6*-territory, two of which expressed high levels of *Aldh1a1*, a previously recognized marker for some SNc neurons (NH1 and NH4). *Sox6*$^+$ neighborhoods were also distinguished by, *Vcan* and *Col11a* (NH1 and NH2), *Anxa1* and *Fam19a4* (NH1 and NH4), *Grin2c, Ndnf* and *Vill* (NH3 and NH4), and *Aldh1a7* (NH4), among other markers.

The other six territories were each divided into two neighborhoods (Figure 3C, Figure 3-figure supplement 1). *Gad2*_NH1 expressed *Chrm2* and *Zfp536*, whereas *Gad2*_NH2 expressed *Megf11, Zeb2*, and *Met*. *Fbn2*_NH1 was specified by *Rxfp1* and *Hs3st2* while *Fbn2*_NH2 by *Col23a1, Dsg2* and *Ism1*. *Pcsk6*_NH1 was positive for *Cald1, Pde11a and Cpne2*. *Pcsk6*_NH2 contained *Tacr3*$^+$ *Aldh1a1*$^-$ cells, which also expressed *Igf1*. There were also *Tacr3*$^+$ *Aldh1a1*$^+$ cells, but they belonged to *Otx2*_NH1, which also expressed *Grp, Eya1, and Plekhg1*. *Otx2*_NH2, which was low for *Slc6a3*, was identified by *Csf2rb2, Nfib,* and *Nfia*. The *Pdia5*_NH1 expressed *Jph1 Cpne5,* and *Pdia5*_NH2 was positive for *Npy1r,*

Postn, and *Wnt7b*. The *Ebf1*_NH1 expressed *Col24a1,* whereas *Ebf1*_NH2 was positive for *Vip*.

Anatomical distribution of diverse mDA neurons

Next, we investigated the anatomical distribution of the seven mDA neuron territories and their neighborhoods in the adult mouse midbrain. We based our spatial mapping on Allen Mouse Brain Atlas (mouse.brain-map.org) and on our own *in situ* hybridization and immunohistochemical analyses. *In situ* hybridization experiments used single probes on individual sections combined with immunohistochemistry. We also used an *in situ* sequencing strategy that included 49 selected probes, hybridized simultaneously on seven sections across the entire midbrain (Figure 4B-E, Figure 4–figure supplements 1-3, Supplementary files 2 and 3)(Gyllborg et al., 2020; Krzywkowski et al., 2017; Qian et al., 2020). We also used coronal *in situ* Allen Mouse Brain Atlas data (atlas.brain-map.org) to guide when relevant data were available. For the anatomical localization of territories and neighborhoods, we mainly followed Paxinos and colleagues (Fu et al., 2011) and the coronal reference atlas from Allen Mouse Brain Atlas. Together these analyses provided the basis of the schematic in Figure 4F, showing the distribution of different territories in rostrocaudal levels of the midbrain.

Consistent with previously published results (Panman et al., 2014; Poulin et al., 2014), ***Sox6* territory** neurons were confined to the SNc and its subdomains (Figure 4F, Figure 4–figure supplement 1, See also Supplementary Results). Cells of the ***Ebf1* territory** were scattered within the periaqueductal grey (PAG), parabrachial pigmented nucleus (PBP), rostral linear nucleus (RLi), and para nigral nucleus (PN) (Figure 4F, Figure 4–figure supplement 2), consistent with previous single-cell studies identifying *Vip*-expressing mDA neurons within the PAG (Dougalis et al., 2012; Hook et al., 2018; Manno et al., 2016; Poulin et al., 2014; Tiklova et al., 2019). Cells of the ***Pdia5* territory** were mapped to the dorsal

VTA and the most lateral SN tip (Figure 4F, Figure 4–figure supplement 2). Consistent with previous analyses (Hook et al., 2018; Manno et al., 2016; Poulin et al., 2014; Saunders et al., 2018; Tiklova et al., 2019), ***Otx2* territory** was localized in the ventral tier of VTA, including PBP, PN, and interfascicular nucleus (IF), but some cells were also found in RLi and caudal linear nucleus (CLi) (Figure. 4, Figure 4–figure supplement 2). ***Gad2* territory** was primarily localized in the most medial VTA, especially to IF, RLi, and CLi (Figure 4A, F, Figure 4–figure supplement 3), supporting findings from previous studies (Garritsen et al., 2023; Poulin et al., 2020). A few scattered neurons matching the smallest territory, ***Fbn2***, were detected in the anteromedial PBP (Figure 4A, F, Figure 4–figure supplement 3). Finally, neurons of the ***Pcsk6* territor**y were localized broadly in VTA, although being most numerous in PBP (Figure 4F, Figure 4–figure supplement 3).

The analysis of the anatomical localization of neighborhoods is described further in Supplementary Results.

Selective vulnerability to 6-OHDA lesioning

As mentioned, nuclei for snRNAseq were sampled from animals that received a unilateral injection of 6-OHDA in the medial forebrain bundle. To assess the extent of mDA neuron fiber loss in the lesioned hemisphere, striatal tissue was also isolated, sectioned, and analyzed by DAT-binding autoradiography. Animals with partial lesions were chosen for sampling and RNA sequencing of nuclei (Methods, Figure 5–figure supplement 1). The nuclei that expressed either *Th* or *Slc6a3* (mDA nuclei dataset, in total 33,052 nuclei) were visualized by UMAP and split by condition to indicate their origin from either lesioned or intact hemispheres (Figure 5A, B). Unsurprisingly, fewer nuclei (8,243) were recovered from the lesioned compared to the intact hemisphere (24,809).

Two conclusions were immediately apparent from the UMAP visualization: First, many nuclei from the lesioned hemisphere were distributed outside of the mDA neuron territories (Figure 5A-B, F, see Figures 2 and 3), and, as expected, the numbers of nuclei in these territories were less affected by 6-OHDA. Second, most mDA neuron nuclei remaining from the lesioned hemispheres were evenly dispersed together with nuclei from the non-lesioned hemispheres within mDA neuron territories and neighborhoods, indicating that surviving cells were relatively similar to non-lesioned cells at the gene expression level (Figure 5A, B). Next, normalized cell loss per territory and neighborhood was calculated based on the sub-clusters cellular composition per condition, accounting for the global FANS yield per condition and sub-cluster size (see methods, Supplementary file 4). The results demonstrated a pervasive loss in the *Sox6* territory (Figure 5C), which is consistent with previous studies showing that *Sox6*-expressing cells of the SNc are more vulnerable to pathological stress in rodents and in PD (Luppi et al., 2021; Panman et al., 2014). Cell loss was also notably high in the *Pcsk6* territory. The *Ebf1* territory showed the lowest level of cell loss, while the other territories were affected at intermediate levels (Figure 5C, D). Analyses of normalized cell loss were also calculated for neighborhoods. Within the *Sox6* territory, a consistent and high level of cell loss was seen in all neighborhoods except *Sox6*_NH3, which showed a somewhat milder but still significant decrease (Fig. 5E). Notably, different levels of cell loss were often seen within individual territories e.g., *Gad2*_NH1, and *Otx2*_NH1 as compared to *Gad2*_NH2 and *Otx2*_NH2 (Figure 5E). As described above, the corresponding mDA neurons were mostly confined to regions within the VTA, demonstrating that high vulnerability to 6-OHDA was not exclusive to neurons distributed within the SNc.

Identification of genes associated with vulnerability or resilience

Next, we set out to identify genes that might predict either mDA neuron vulnerability or resilience to 6-OHDA. Because the goal was to identify genes with predictive value expressed in normal mDA neurons, only gene expression in nuclei from the non-lesioned hemisphere was analyzed. Thereby direct effects of the toxin were excluded from the analysis. Accordingly, genes that are commonly and significantly enriched in intact nuclei from clusters with either >90% (vulnerability genes) or <50% (resilience genes) normalized cell loss were identified (Figure 6–figure supplement 1, Supplementary file 4). 39 genes fulfilled the stringent cutoff criteria for vulnerability genes. In contrast, only ten genes fulfilled the cutoff criteria for resilience genes (Methods). The selected genes were ranked based on average Log2 fold change, and vulnerability and resilience gene modules were constructed, consisting of 20 and 8 most significantly enriched genes, respectively (Figure 6F, G, Supplementary file 5). The composite average expression of the gene sets (modules) was visualized in violin plots for neighborhoods and territories (Figure 6A-D, Supplementary File 5). As expected, the vulnerability and resilience gene module expression in territories and neighborhoods corresponded to the distribution of the clusters used as input for identifying the gene modules (Figure 6–figure supplement 1). However, linear regression analysis also demonstrated a highly significant overall correlation between cell loss and the vulnerability module scores (p-value = 3.89e-10), even when the input clusters were excluded from analysis (Figure 6–figure supplement 1C). Thus, vulnerability and presumably resilience gene modules can predict cell loss in the data set. This suggests that the identified module genes may be functionally associated with either sensitivity to or protection against 6-OHDA-induced cell stress. One of the vulnerability genes, *Slc6a3*, encodes the dopamine transporter used to import 6-OHDA to the cytosol. Thus, it is unsurprising that neurons of the Gad2 and Ebf1 territories expressing low levels of *Slc6a3* are less vulnerable. However, when regression analysis was repeated using a vulnerability gene module in which *Slc6a3* had been

excluded, correlation to cell loss remained highly significant (Figure 6–figure supplement 1D, p-value = 1.73e-09). Therefore, the other genes of the vulnerability module may influence sensitivity to toxic stress.

The *Th*$^{+}$/*DAT*$^{+}$ dataset of 33,052 nuclei was re-analyzed by the Seurat integration method treating nuclei from lesioned and intact hemispheres as two different experimental conditions (Stuart et al., 2019). The main goal of this approach was to identify common transcriptomic states in nuclei from the different conditions via a learned joint structure (see Methods). We used this integrated dataset to validate further the predictive value of the identified vulnerability and resilience gene sets. Vulnerability and resilience modules were plotted across the integrated clusters and were color-coded by territory identities (Figure 6–figure supplement 2). In the integrated dataset, the clusters are completely independent of the merged data-based annotations assigned to integrated territories and neighborhoods. Therefore, the distribution of the two modules' scores across the integrated clusters is a good indicator of how predictive they can be, given the territorial identities of the integrated clusters. Of note, the module scores corresponded well to the clusters, as integrated clusters with the *Sox6* territory identity and most of the clusters with the *Pcsk6* territory identity showed considerably high and low values of the vulnerability and resilience modules, respectively.

Conversely, integrated clusters with *Gad2* and *Ebf1* territory identities showed high resilience module values and low vulnerability module values (Figure 6–figure supplement 2). In addition, most of the pairwise comparisons of module scores across the mDA-integrated clusters were significant (Figure 6–figure supplement 2, Supplementary file 6). Non-mDA clusters were excluded from this statistical analysis. Thus, a strong predictive correlation between the expression of vulnerability and resilience modules is apparent across the snRNA-seq data.

Long-lasting gene expression changes in response to 6-OHDA

As mentioned above, cells from lesioned and intact hemispheres were relatively evenly distributed within mDA neuron territories in the UMAP diagram, indicating that most of the surviving cells from the lesioned hemisphere were transcriptionally similar to corresponding cells on the intact hemisphere (Figure 5A, B). However, two clusters consisted almost exclusively of nuclei from the lesioned hemisphere. To distinguish them from territories harboring nuclei characteristic of the normal non-lesioned /intact hemisphere of the brain, they were named ***mostly lesion*** **(ML)** clusters (Figure 7A, B). Although expression of typical mDA neuron markers such as *Th* and *Slc6a3* was low, the ML-clusters were part of the mDA neuron dendrogram branch and shared a parent node with *Sox6*-territory (Figure 3A, C). Thus, these nuclei were likely derived from surviving mDA neurons that have become transcriptionally altered due to cellular stress. We took advantage of the integrated dataset described above to analyze this further. Using original territory labels, visualizing the integrated data by UMAP was consistent with data from non-integrated original analysis and demonstrated that the mDA neuron nuclei clustered together (Figure 7C, Figure 6-figure supplement 2). Next, the integrated data territories were annotated based on the cluster composition of inherited merged (original) territory identities and the enriched markers for the original territories (see Figure 3C). Notably, the nuclei forming the ML-clusters were now entirely integrated within the mDA neuron territories, supporting the hypothesis that they originated from healthy mDA neurons. In addition to ML clusters being predominantly made up of lesioned nuclei, *Sox6* expression level and abundance were also clearly higher than *Calb1* (Figure 7D, E). Both visualizations of ML-clusters nuclei in the integrated UMAP and quantifications demonstrated that most (>80%) of all ML nuclei were confined to two territories, *Sox6* and *Pcsk6*. It thus seems plausible that the ML nuclei originated from altered

neurons in these two territories. Notably, these territories were also the most vulnerable when exposed to 6-OHDA (Figure 7F-H, also see Figure 6A, C).

Differential gene expression analysis by condition was performed in two different ways. First, gene expression in all nuclei from the lesioned hemisphere was compared to all nuclei from the intact hemisphere. Consistent with the similar distribution of most of the lesioned and intact nuclei in the UMAP, relatively few (242 in total) significantly differentially expressed genes were identified (Figure 7–figure supplement 1, Supplementary file 7). Second, nuclei from the ML-clusters were compared to nuclei from all other territories. Consistent with a more dramatic and long-lasting influence from lesioning in the ML-clusters, 1577 genes were found to be differentially expressed in this comparison. Several of the most strongly upregulated genes, including *Atf3, Creb5, Xirp2, Clic4, Sprr1a*, and *Mmp12*, have in previous studies been associated with cellular stress, cell death, and cell signaling (Figure 7I, Figure 7-figure supplement 1, Supplementary file 7). *In situ* hybridization analysis combined with immunohistochemistry shows apparent upregulation of ML-enriched genes in the lesioned tissue compared to control (Figure 7J-M).

Although 6-OHDA was injected unilaterally, a potential influence on cells in the intact hemisphere could interfere with some of our conclusions. Therefore, to assess whether the stereotactic injection had any effect on the transcriptional profile of the nuclei from the intact hemisphere, we created and analyzed a separate dataset including all Th^{+}/DAT^{+} nuclei from the intact hemisphere of the brain, merged with a new dataset of 6,001 mCherry^{+} nuclei isolated from the ventral midbrain of untreated mice. UMAP visualization of this combined data set indicated an almost complete overlap between the two groups of nuclei. The dataset was annotated similar to the procedure above and all major cell types were also identifiable in it (Figure 7–figure supplement 2). Moreover, comparative differential gene expression analysis also supported the conclusion that any differences between the two groups were

attributable to undefined technical parameters in experimental conditions rather than to the effects of the toxin (Figure 7–figure supplement 2, Supplementary file 7).

Relationship between dopamine neuron groups in the mouse and human brain

Understanding how the emerging knowledge of mouse mDA neuron diversity compares to the molecular landscape in the human brain is of major interest. We, therefore, integrated and reanalyzed our mouse data with human snRNAseq data from a recently published paper (Kamath et al., 2022). Canonical correlation analysis (CCA) was used to integrate 25,003 nuclei from human control and diseased (PD and Lewy body disease) donors with our lesioned and intact nuclei from the Th^{+}/DAT^{+} dataset (see methods). UMAP visualization revealed a relatively large overlap between nuclei from the two species, despite some regions being predominated by one specific species or condition. Detection of mouse-specific UMAP regions is not surprising, given the comprehensiveness of the mDA-enriched mouse dataset. However, human specific UMAP regions were also observed (Figure 8). Next, we used the mouse intact nuclei from the mDA territories (see Figure 3) to build a reference UMAP structure onto which the human control nuclei (query dataset) were projected (see methods). The results demonstrated that almost 75% of the human control nuclei are mapped to mouse *Sox6* territory (Figure 8-figure supplement 1A-D). Upon using the mouse intact nuclei, only from *Sox6* territory as a reference, and the human control nuclei that were mapped to *Sox6* territory in the previous step as query, most of this human query subset were mapped to *Sox6*-NH2, while almost no nuclei mapped to *Sox6*-NH4 (Figure 8-figure supplement 1E-H). The findings that most human nuclei bear the closest resemblance to mouse *Sox6* territory nuclei were expected, as the human nuclei were derived from SNc.

Discussion

When histochemical methods for mapping catecholaminergic cells and projections were developed (Falck et al., 1962), it became clear that mDA neurons consist of heterogeneous groups of cells with distinct anatomical locations and innervation targets. However, we now know that a substantial diversity exists beyond the anatomically defined cell groups, supported by studies focused on mDA neurons using rapidly emerging single-cell methods to characterize cellular diversity (Garritsen et al., 2023; Hook et al., 2018; Manno et al., 2016; Poulin et al., 2020, 2014; Tiklova et al., 2019). Here the goal was to create a comprehensive gene expression atlas over the mouse mDA neuron gene expression landscape by using snRNA-seq for the analyses. In addition, we wished to address the differential vulnerability to neuronal stress within the diverse groups of mDA neurons.

Unsurprisingly, global gene expression patterns differ substantially between cell types such as astrocytes, neurons, or endothelial cells. In addition, different neuron types with distinct developmental origins usually differ substantially in scRNA sequencing analyses. Thus, DA neurons from the hypothalamus, which have a distinct developmental origin from mDA neurons, differ from mDA neurons by many differentially expressed genes. Our study was primarily motivated by the more challenging need to describe differences across the population of related mDA neurons, which have a common developmental origin and a high degree of molecular similarities among them. The challenge in the analysis is evident from UMAP visualizations, as presented here, where mDA neurons are distributed in a shifting gene expression landscape without clearly distinct borders between groups of mDA neurons. Thus, defining distinctive sub-types within such a nuanced gene expression landscape may be misleading, as noted in several previous single-cell studies (Kozareva et al., 2021; Saunders et al., 2018; Tasic et al., 2018; Yao et al., 2021). Therefore, we chose to use a nomenclature that better reflects continuity between mDA neuron groups by referring to categories of mDA

neurons as belonging to "territories" and "neighborhoods." We do not propose that the taxonomy presented here should be viewed as universal or a definitive endpoint, but rather should be seen as high-resolution gene expression analysis that will help to guide further studies that use strategies to understand connectivity, neurophysiology, and behavior in investigations of relevant neuronal sub-circuits influenced by mDA neurons.

The mDA neuron atlas presented here is based on the sequencing of a large number of mDA neurons (40,000 in total), which was achieved by using snRNA-seq, which we found to be superior to scRNA-seq for reducing sampling bias. Furthermore, in support of a robust sampling, a large proportion of mDA neurons population was analyzed from each dissected brain (consistently above 20%), a 50-100 times improvement from our previous single-cell study (Tiklova et al., 2019). Indeed, the neurons belonging to the SNc (*Sox6* territory) corresponded to roughly one-third of all mDA neurons, a proportion that conforms to previous stereological cell counting of mouse mDA neurons (Björklund and Dunnett, 2007; Zaborszky and Vadasz, 2001). Thus, we assume that the cellular stoichiometry, as presented in the present snRNA-seq atlas, roughly corresponds to the distribution of mDA neurons in the adult mouse brain.

Previous scRNA-seq studies have described mDA neuron groups characterized by gene expression patterns that resemble those in different territories and neighborhoods. Comparisons to those studies are described and referenced in Supplementary Results and show that distinguishing gene expression features seen in six of the seven territories described here are similar to those found in previous single-cell studies (Hook et al., 2018; Manno et al., 2016; Poulin et al., 2014; Saunders et al., 2018; Tiklova et al., 2019). However, the presented mDA neuron atlas illustrates previously unknown diversity within these groups. Thus, the *Sox6* territory was divided into four neighborhoods with partly unique anatomical localizations within the SNc. In addition, the gene expression signature of the *Pcsk6* territory,

comprising a relatively large number of cells, did not correspond well to previously described mDA neuron groups. Thus, the *Pcsk6* territory neurons apparently correspond to an mDA neuron group previously not recognized.

Moreover, while *Vip*$^+$ mDA neurons were identified in several previous single cell studies, we show here that the *Ebf1* territory can be divided into two related neighborhoods encompassing *Vip*$^+$ and *Col24a1*$^+$ mDA neurons, respectively. Although these cells were most enriched in PAG, they were also found scattered in other regions such as PBP and CLi. *Col24a1*$^+$ likely corresponds to a specific population of *Npw*$^+$ mDA neurons, localized in PAG, VTA, and the dorsal raphe, which project to the extended central amygdala and modulate mouse behavior under stress (Motoike et al., 2016).

We also extended our analysis to neurons exposed to neuronal stress by sampling nuclei from the lesioned after partial unilateral exposure to the neurotoxin 6-OHDA. The mechanisms of neuronal stress leading to the degeneration of mDA neurons in PD differ from events following exposure to neurotoxins (Blesa and Przedborski, 2014). However, it is known that most insults to mDA neurons, including those resulting from 6-OHDA exposure, affect SNc more severely than neurons of the VTA, indicating an inherent higher vulnerability of these neurons. This is evident in our analysis, where normalized cell loss was most severe within the *Sox6* territory. However, a notable high vulnerability was also seen outside the *Sox6* territory, e.g., within the *Pdia5* and *Pcsk6* territories. Indeed, the *Pcsk6*_NH1, apparently localized within the PBP, is even more vulnerable than the *Sox6*_NH3. Within the VTA territories, resilient neighborhoods included several that express Slc17a6 encoding the glutamate transporter VGLUT2, a protein linked to increased and decreased dopamine neuron vulnerability (Kouwenhoven et al., 2020; Steinkellner et al., 2018).

The information on normalized cell death was used to identify genes enriched in highly vulnerable or resilient clusters. Regression analysis showed a highly significant

correlation between the vulnerability module scores and normalized cell loss in neuron groups, i.e., clusters that were not used as input to identify the genes used in calculating the module. We note that several genes in these modules may be functionally linked to either vulnerability or resilience. For example, *Kcnj6* is a known marker for SNc neurons and has, in previous studies, been shown to be associated with the vulnerability of these neurons (Liss et al., 2005). In addition, the protein encoded by the *Vsnl1* gene is a calcium sensor that is known to be increased in cerebrospinal fluid in Alzheimer's patients and may be linked to disease progression (Liao et al., 2022; Zetterberg, 2017). Therefore, the analysis of 6-OHDA vulnerability complements the atlas in ways that will be valuable for understanding how the molecular gene expression landscape influences sensitivity to pathological stress. It will be important in future studies to extend these studies to mouse neurons exposed to pathology that more resembles that seen in PD.

To avoid extensive influence on gene expression from the acute toxin effects in dying cells, brain tissue was collected for nuclei isolation several weeks after 6-OHDA exposure. Most nuclei derived from the lesioned hemispheres of dissected brains were scattered in UMAP analysis with nuclei from the intact hemisphere, suggesting that the toxin had not caused any severe gene expression alterations. However, nuclei in the ML clusters were an exception, as they were derived almost exclusively from neurons dissected from the lesioned hemisphere. These toxin-exposed neurons displayed more profound and long-lasting alterations in gene expression. We addressed the origin of these cells in an integrated data set, and it is plausible from this analysis that most of these altered neurons were derived from neurons belonging to the two most vulnerable *Sox6* and *Pcsk6* territories. Gene expression changes in ML-clusters' included severe down-regulation of typical mDA neuron gene markers such as *Th* and *Slc6a3*. In addition, many enriched genes have previously been associated with neurodegeneration or response to neuron stress. For example, *Atf3, Lifr,*

Sprr1a have been reported to be part of a regeneration-associated gene program in peripheral nerves (Kanaan et al., 2015; Seijffers et al., 2007). *Atf3* is also known to be upregulated as a result of activation of the unfolded protein response pathway via PERK (Jiang et al., 2004). *Ngf*, encoding the nerve growth factor, is known to be neurotrophic to mDA neurons and is dysregulated in PD (Mogi et al., 1999). Thus, part of this stress response is likely to be neuroprotective and likely reflects a cellular survival response in affected cells.

In conclusion, the work presented here provides a comprehensive atlas over mDA neuron diversity in mice. Combined with the included anatomical analysis and assessment of neuron vulnerability after toxin stress, we anticipate that this will be a useful resource in studies to understand the intricate functions of neurocircuitries involving dopamine neurotransmission functionally.

Materials and Methods

Mouse lines and genotyping

Dat^{Cre} (Ekstrand et al., 2007), *Pitx3*^{GFP} (Zhao et al., 2004) *Rpl10a-mCherry* (*TrapC*^{fl}) (Hupe et al., 2014) and their genotyping have been described before. *Dat*^{Cre} and *TrapC*^{fl} lines were intercrossed to generate *Dat*$^{Cre/+}$*; TrapC*$^{fl/fl}$ mice. All experimental procedures followed the guidelines and recommendations of Swedish animal protection legislation and were approved by Stockholm North Animal Ethics board.

Partial lesions with 6-OHDA

C57BL6 female mice were pretreated (30 min before surgery) with 25 mg/kg desipramine (i.p.; Sigma–Aldrich) and 5 mg/kg pargyline (i.p.; Sigma–Aldrich), anesthetized with 2-2.5% isoflurane, placed in a stereotaxic frame, and unilaterally injected, over 2 min, with 0.7 or 1.5 µg of 6-OHDA in 0.01% ascorbate (Sigma–Aldrich) into the medial forebrain bundle (MFB) of the right hemisphere. The coordinates for injection were AP, −1.1 mm; ML, −1.1 mm; and DV, −4.75 mm relative to bregma and the dural surface (Paxinos and Franklin, 2001). The tissues were collected six weeks after 6-OHDA administration.

DAT binding autoradiography

Fresh frozen thaw-mounted striatal sections (12 µm thick) were initially incubated in 50 mM Tris–HCl/120 mM NaCl (pH 7.5) for 20 min. Then the same sections were incubated in binding buffer (50 mM Tris– HCl/120 mM NaCl, pH 7.5/1 µM fluoxetine) containing 50 pM of the DAT ligand (^{125}I) RTI55 (PerkinElmer Life Sciences, Boston, USA) for 60 min. To detect nonspecific binding, 100 µM nomifensine was added to the assay. The sections were

washed 2 × 10 s in ice-cold binding buffer, rapidly dipped in deionized water, air-dried, and exposed to autoradiographic films in X-ray cassettes at room temperature for 2 days.

Fluorescent RNA in situ hybridization and immunohistochemistry

Mice were deeply anesthetized, and the brains collected after intracardial perfusion with +37C PBS, followed by 4% PFA in PBS. The midbrain piece was coronally cut out using a brain matrix and post-fixed for 2-3 days in PFA at RT, changing the PFA solution daily. The tissues were dehydrated and processed into Histosec wax (Merck) using an automated tissue processor (Leica TP1020), and coronal 6 µm sections were collected at Superfrost Plus slides (Thermo-Scientific) using an automated waterfall microtome (Epredia HM355S). Immunohistochemistry and fluorescent RNA in situ hybridization were performed essentially as described previously (Tiklova et al., 2019), except for Pdia5 antibody Triton X-100 in the washing buffers and blocking solutions was replaced by 0.1% Tween-20. For the analysis of *Tacr3*, parallel slides were stained with an antibody for Aldh1a1. 2-3 animals were used for each experiment. Each ML-territory probe was tested on lesioned (n=3) and control Dat^{Cre}; $TrapC^{fl/fl}$ mice.

The ML-clusters specific probes *Clic4*, *Creb5*, *Mmp12* and *Sprr1a* were cloned fragments from the respective protein coding sequences of mouse cDNA. Size of fragments: *Clic4* 722 bp, *Creb5* 591 bp, *Mmp12* 759 bp and *Sprr1a* 598 bp. The probe used for *Vip* was cloned with primers from Allen Mouse Brain Atlas (experiment 77371835). The *Gad2* probe was a cloned 869 bp fragment from the protein coding sequence of mouse *Gad2* cDNA. *Tacr3* probe was a 992 bp Genestrings fragment (Invitrogen) from the protein coding sequence of mouse *Tacr3* cDNA. All cDNA fragments were cloned into pCR-Blunt II TOPO vector (Invitrogen 450245) and DIG-UTP labelled cRNA probes synthetized from linearized plasmids using either SP6 or T7 RNA polymerases (Roche) according to manufacturer's

instructions. The list of Allen Mouse Brain Atlas experiments of which images were shown in Fig 4 and in Figure 1-figure supplement 3, Figure 4-figure supplements 1-3, as well as the list of primary and secondary antibodies used, are in Supplementary Materials (Supplementary file 2).

Image capture and processing

Images of fluorescent in situ hybridizations and immunostainings were taken using a Zeiss confocal microscope LSM700. Tiling of the individual images was done with Fiji (ImageJ) Pairwise stitching plug-in (Preibisch et al., 2009). Brightness and contrast were adjusted after tiling, if needed.

In situ RNA sequencing

The *in situ* RNA sequencing experiments were performed as described (Gyllborg et al., 2020). In brief, fresh-frozen 10 µm thick mouse brain sections were sectioned on a cryostat and stored at -80 °C until fixation. After fixation in 3% (w/v) paraformaldehyde (Sigma) in DEPC-treated PBS (DEPC-PBS) for 10 min at RT, tissues were washed in DEPC-PBS and reverse transcription was performed at 37 °C for 16 hrs. Tissues were then treated with RNaseH at +37°C for 45 min to make single-stranded cDNAs, followed by padlock probe (IDT Corelville, Iowa; 4nmole, standard desalting, 5 prime phosphorylated; Supplementary file 3) hybridization and ligation at 45 °C for 60 min. Rolling circle amplification was performed at 30 °C for 16 hrs. Rolling circle products were detected by using fluorescently labeled detection oligo (Supplementary file 3, IDT Corelville, Iowa).

Imaging and data processing of ISS

Images were acquired with Leica DMi8 epifluorescence microscope equipped with an external LED light source (Lumencor® SPECTRA X light engine), automatic multi-slide

stage (LMT200-HS), sCMOS camera (Leica DFC9000 GTC), and objective (HCX PL APO 40X/1.10 W CORR). A series of images (10% overlap between neighboring images) were obtained, and maximum intensity projected to two-dimensional images. These images were aligned between cycles and stitched together using Ashlar algorithm. Stitching was followed by retiling to get smaller (6000x6000 pixel) images. Those images were used for decoding using PerRoundMaximumChannel Decode Spots Algorithm. The resulting spots were filtered based on minimum quality. The preprocessing and decoding pipeline can be found on the Moldia GitHub page (https://github.com/Moldia).

Tissue processing for FANS

Dat^{Cre} *TrapC*$^{fl/fl}$ (n=6 for lesion cohort, n=6 for control cohort) mice were euthanized with CO_2, and brains were rapidly removed and transferred into cold PBS. The midbrain was dissected and snap-frozen on dry ice. For lesioned brains, the lesioned hemisphere was separated from the intact hemisphere before being snap-frozen. Tissue processing for FANS was performed as described previously (Toskas et al., 2022), with the addition of adding 0.1% Triton and 5mM $CaCl_2$ to the lysis buffer. For the control cohort, resuspended and filtered nuclei were stained with NeuN-647 conjugated (1:1000, MAB377 Millipore Corp) antibody for a few hours before being sorted. Nuclei were sorted using a FACSAria Fusion or FACSAriaIII cell sorter with the FACSDiva software (BD Biosciences). The nuclei were identified by forward- and side-scatter gating, a 405-nm laser with a 450/50 or 450/40 filter, a 561-nm laser with a 610/20 filter, and a 633-nm laser with a 660/20 filter, quantifying DNA content per event to assure that only singlets were collected. A 100 μm nozzle and an acquisition rate of ~3000 events per second was used. Nuclei were stained with DAPI (1:200, Merck) 5 min before sorting and preserved in 2% BSA until downstream procedures.

Library preparations and sequencing

Ventral midbrains tissues from 6 mice (6 lesioned, 6 non-lesioned (intact) hemispheres) were used to obtain 70464 nuclei (36789 un-lesioned (intact) and 33675 lesioned). Ventral midbrain tissue from 6 control mice were used to obtain 8311 nuclei. Chromium Single Cell 3′ v3, dual indexing (10x Genomics) was used to make single-nucleus libraries, in accordance with the manufacturer's protocols. Libraries were sequenced on a NovaSeq 6000 system (ESCG, SciLifeLab, Karolinska Institute). Library preparations and sequencing for the nuclei from the intact and lesioned ventral midbrain hemispheres were done together but separately from the control nuclei.

snRNAseq analysis

Sequenced reads were demultiplexed and aligned to a CellRanger customized reference package "transcriptome: mm10-3.0.0_premrna", with pre-mRNA added to ensure both premature and mature mRNAs are mapped accurately, using CellRanger Pipeline version 3.1.0 from 10x Genomics (Zheng et al., 2017). A count matrix (filtered_feature_bc_matrix) per sample was generated using CellRanger Pipeline version 3.1.0. Seurat package version 4 was used for downstream analyses of the data (Hao et al., 2021). Seurat Read10x() was used to load the sparse count matrices of 10X genomics CellRanger output and CreateSeuratObject() function was used to make one Seurat Object per sample. All 12 samples (6 lesioned, 6 un-lesioned) were merged to generate one Seurat object for the entire data. Quality control and filtering of data were done in the following steps. First, the percentage of transcripts that map to mitochondrial genes and ribosomal genes were calculated using Seurat function PercentageFeatureSet and the corresponding values added to the created percent_mito% and percent_ribo% metadata columns, respectively. Then, CellCycleScoring function was used to score cell cycle phases, and phase metadata was

calculated and added based on the expression of G2-M and S phase canonical markers. In the next step, the ratio of a gene's unique molecular identifier (UMI) counts to the sum of all UMIs per library (nucleus), was calculated as the average relative expression of each gene per library. *Malat1*, a lncRNA gene, was expressed in all libraries with a mean fraction of total UMI counts per library of about 7%. Polyadenylate captured RNA-seq libraries, irrespective of the used protocol, show a relatively high detection of *Malat1* reads. However, compared to other methods, the detection is even higher in snRNA-seq due to its nuclear localization and therefore *Malat1* expression is used to estimate the nuclear proportion of total mRNA (Bakken et al., 2018). *Malat1* was filtered out for downstream analysis. Genes with reads present in fewer than 10 nuclei (libraries) were removed. Libraries with detected genes below 500 and above 10,000 (doublets), and with percent_mito above 0.5 were also removed from the analysis.

All-nuclei dataset

After the filtering steps (above), 68914 nuclei remained (36,051 intact, 32,863 lesioned). The remaining nuclei after the quality control constituted the all-nuclei dataset and had UMIs per library in the range of 574 - 122753, genes per library in the range of 501 - 9991, with average UMIs per library of 17630, and average genes per library of 4762. The average UMI/gene was 3.58, with a range of 1.14 - 13.31 (Figure 1-figure supplement 2).

Dopaminergic nuclei dataset

The filtered dataset (all-nuclei, above) was then subject to selection based on the expression of two genes, *Th* and *Slc6a3 (DAT)*. Nuclei expressing either *Th* or *DAT* (Th^+/DAT^+) were selected. The remaining 24809 and 8243 nuclei from the respective intact and lesioned hemispheres were used to create the dopaminergic dataset.

Combined intact and control nuclei dataset

Count matrices for the 6 intact samples as well as the 6 control samples were loaded separately and merged to create a dataset of 12 samples (6 intact, 6 control). After the QC steps, described above, and selection of libraries based on either *Th* or *DAT* (Th^+ | DAT^+) expression, 24809 nuclei from the intact hemisphere and 6001 nuclei from the control remained which were used to make the final intact-control dataset. Interestingly, the ratio of the Th^+ | DAT^+ filtering for both groups was 0.7.

Normalization, variance stabilization and dimensionality reduction

SCTransform package (Hafemeister and Satija, 2019) was used to normalize, scale, and find the top 1000 variable genes, with the intension to replace the heuristic approaches of adding pseudo-counts and log transformation. This method of normalization and variance stabilization is meant to improve the disentanglement of biological heterogeneity from the confounding effects of the technical variables, using the Pearson residuals of the regularized negative binomial regression, with the sequencing depth as a covariate in a generalized linear model. The residuals of the SCTransform were used as input for dimensionality reduction. Principal component analysis (PCA) was applied, followed by the JackStraw procedure and ElbowPLot to visualize the standard deviations of the principal components. PCs 1:30 were used in the UMAP embedding.

SNN graph-based Clustering

Seurat FindNeighbors function was applied for SNN graph construction, followed by FindClusters function, using the Louvain algorithm, to identify the clusters. For the all-nuclei dataset, the resolution 0.1 was used since the goal was to identify the major cell classes.

BuildClusterTree function was used to make a phylogenetic tree, with assay = "SCT", slot = "data" parameters for the all-nuclei dataset. For the different datasets in this research, a range of resolutions were tested to estimate the optimal cluster numbers, as a heuristic approach based on known markers, neuroanatomical regions and hierarchical dendrograms. These estimated values were used in determining k (centers) parameters in k-means clustering when applicable (below).

K-means and hierarchical clustering

A matrix of PCA values based on PCs 1:30 was generated to be used as input for the kmeans clustering function, with “MacQueen” algorithm. The computed clusters (centroids) were added to the object metadata and then were subject to hierarchical clustering, with distance matrix computation method set as “Euclidean” and the agglomeration method as “ward.D2”.

Integration analysis mouse dopaminergic dataset

For the dopaminergic dataset, it was split based on condition (intact / lesioned), SCTransform approach was used to independently, normalize and identify the top 1000 variable genes per condition. Same number of genes (1000) were selected as integration features to be used as “anchors” in data integration, using SCT as the normalization method. PCs 1:30 were used in the UMAP embedding. K-means and hierarchical clustering were performed as described above.

Calculation of cell loss

The ratio of the FANS yield for the nuclei passing the QC, for condition (intact to lesioned) was calculated and the quotient (1.097) was set as a yield normalization coefficient (β = 1.097) for lesioned subclusters. The calculation of cell loss was done per subcluster for the

original kmeans clusters in the dopaminergic nuclei dataset. Based on the hierarchical dendrogram and class annotation, the non-dopaminergic clusters were excluded from the analysis. Louvain modularity optimization algorithm was applied with resolution of 0.5 to obtain the subclusters. For a couple of clusters that did not resolve with res=0.5, incrementally higher resolutions were applied. Most clusters were resolved into subclusters with no or only one identified singleton, while no clusters were resolved into subclusters with higher than 3 singletons. The cell composition of each subcluster per condition was determined, considering the FANS yield coefficient. For each cluster the average normalized cell loss was calculated, considering the subcluster size, according to the formula below, with normalized cell loss per cluster; A, determined by subcluster size by condition; s_l and s_i lesioned and intact respectively, N; number of subclusters per cluster, and β.

$$A = \left(\sum_{s=1}^{s=n} 1 - \frac{(\beta * sl)}{si}\right) / N$$

The A = 1 denotes total cell loss while A = 0, indicates no cell loss at all. The subcluster cell loss values were used to calculate the normalized cell loss per cluster, neighborhood, and territory. Subclusters with $\left(1 - \frac{(\beta * sl)}{si}\right) < 0$, were set to NA. The statistical tests used for cell loss analyses include Shapiro-Wilk normality test, either Levene's Test or Fligner-Killeen (for non-normally distributed data) tests of homogeneity of variances, Kruskal-Wallis rank sum test, and the Conover-Iman test, with Benjamini-Hochberg p adjustment method, for multiple pairwise comparisons across clusters, territories, and the neighborhoods within *Sox6* territory. For all other territories of only two neighborhoods, Levene's Test for Homogeneity of Variance and either Two Sample t-test or Wilcoxon rank sum test were applied. For the Conover-Iman test, the default (altp=FALSE) was set as p-value = $P(T \geq |t|)$, and the null hypothesis (H_0) was rejected at $p \leq \alpha/2$, which is 0.025.

Vulnerability and resilience modules

Vulnerable (normalized cell loss > 90%) and resilient (normalized cell loss < 50%) clusters were identified and gene sets each module were created based on being commonly upregulated (adjusted p value < 0.05 & average Log2FC > 0.5) in either vulnerable or resilient clusters. The Seurat function AddModuleScore was used to calculate the module scores. This function assigns a numerical value to every library for a given module. A positive value indicates an expression level above the cell population average, while a negative value indicates the opposite. Shapiro-Wilk normality test, or for cases with n>5000, the D'Agostino Skewness and Kurtosis Normality Tests from the "burrm/lolcat" R package were applied. Next, either Levene's Test or Fligner-Killeen tests of homogeneity of variances were used. Because data was neither normally distributed nor homoscedastic (homogeneity of variances), both the Kruskal-Wallis rank sum test and Welch's ANOVA (One-way analysis of means, not assuming equal variances) were applied. Both tests rejected the null hypothesis, indicating that there is significant difference among groups. Finally, the Conover-Iman test, with Benjamini-Hochberg p adjustment method, for multiple pairwise comparisons across dopaminergic clusters, neighborhoods and territories was applied.

Correlation between cell loss and vulnerability module score

Two linear regression models were fitted using normalized cell loss as the response variable and vulnerability module either with or without DAT (*Slc6a3*) as the predictor. The module scores were square root (sqrt) transformed to make their distribution normal, prior to fitting. The plots show the R^2, the estimated model equations and the 95% confidence interval limits.

Differential gene expression analysis

Differentially expressed genes between two or more groups were identified using Wilcoxon Rank Sum test in the Seurat FindMarkers or FindAllMarkers functions with the following set

parameters; logfc.threshold = 0.25, min.pct = 0.1, min.diff.pct = -Inf, only.pos = FALSE, max.cells.per.ident = Inf. The calculated p-values were adjusted for multiple comparisons using the Bonferroni correction.

GO enrichment analysis

The ML territory enriched genes (806 upregulated and 661 downregulated) were used for this analysis by removing the overlapping genes enriched in the DE between all lesioned and all intact (Figure 7-figure supplement 1). The "enrichR" package as the R interface to the Enrichr database, version 3.1 was used with GO_Biological_Process_2021 as the imported gene-set library (Kuleshov et al., 2016). The top 20 enriched terms, ranked by p value, were selected per group for visualization.

Generation of human dataset

Human dataset including count matrix, barcodes, features and metadata files were downloaded from the link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178265. Genes with reads present in fewer than 1 nucleus (libraries), and *Malat1* were removed. Libraries with detected UMIs below 650 and UMI/Gene below 1.2 and, with percent_mito above 10 were also removed. Reciprocal PCA (RPCA) approach was used to integrate individual SNc samples separately per condition (control vs diseased), by which each sample is projected into the others' PCA space, and the integration anchors are defined by the requirements of mutual neighborhood. Next the control and diseased dataset were integrated into one dataset, using RPCA approach. Louvain resolution 0.1 was used for clustering and the clusters of dopaminergic nuclei were identified by plotting typical markers. These dopaminergic clusters of 25003 nuclei (control=17039, LBD=4642, PD=3322 nuclei) were subset for downstream analysis.

Integration of human and mouse dataset

Orthology mapping package "Orthology.eg.db", version 3.13.0 was used to map NCBI gene IDs between species. Human genes were converted to their mouse orthologs and the ones without a mouse ortholog, were also kept. Data from human dopaminergic nuclei, subset in the previous step (above), containing 18 donors (8 control, 6 PD, 4 LBD) and data of the 12 mouse samples from 6 dissected midbrains (6 intact, 6 lesioned) from the dopaminergic dataset (Th^{+}/DAT^{+}), was first log-normalized, highly variable genes found (method=vst) separately per species, prior to being integrated using the Seurat canonical correlation analysis (CCA).

Mapping and annotating human control nuclei query dataset

Unimodal UMAP Projection approach from Seurat v4. was used to compute a UMAP reference from the mouse intact, mDA territories nuclei, using the first 30 PCs after log-normalization, finding highly variable genes (method=vst) and scaling of the mouse reference data. Then human control dopaminergic nuclei were used as query to be projected onto the reference UMAP structure, using the MapQuery() function. Next, the mouse intact only *Sox6* territory nuclei were used to make the reference UMAP, with the same parameters as above and the human nuclei with the transferred label “Sox6” from the previous step were used as a subset query to be mapped with a higher resolution to the *Sox6* neighborhoods (Figure 8-figure supplement 1).

Data availability

Sequencing Fastq files, raw and normalized count matrixes, and supplementary files are accessible via GEO accession number xxxxx. All-nuclei and mDA-nuclei datasets are

accessible for visualization and exploration via the links to our lab's local server, http://193.10.16.198:5005/ and http://193.10.16.198:5006/ respectively.

Acknowledgements

This work was supported by Swedish Research Council (VR 2020-00884), The Swedish Brain Foundation (Hjärnfonden), and Torsten Söderbergs Foundation to T.P., and a postdoctoral grant from Karolinska Institutet's Strategic research area neuroscience (StratNeuro, Karolinska Institutet) to B.Y.S. The computational analyses, data handling, and data curation were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) in the project NAISS 2023/23-181 and the Swedish National Infrastructure for Computing (SNIC) in the project SNIC 2022/23-171 at UPPMAX, Uppsala University , partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973. The authors wish to thank SciLifeLab Eukaryotic Single Cell Genomics as well as SciLifeLab In Situ Sequencing Infrastructure Units. We also would like to thank Petter Woll, at the Center for Hematology and Regenerative Medicine (HERM) in Neo, Karolinska Institutet, for advice.

Figure legends

Figure 1. Generation of a snRNAseq dataset from the mouse midbrain

A) Schematic design of the study, unilateral injection of low doses of 6-OHDA (0.7, 1.5 µg/µl) in the medial forebrain bundle (MFB), followed by dissection, nuclei isolation and enrichment via FANS, before library preparation, sequencing, and data analysis. B) Transgenic, Cre-dependent mCherryTRAP mouse model. Dat^{Cre} and $TrapC^{fl}$ lines were intercrossed to generate $Dat^{Cre/+}$; $TrapC^{fl/fl}$ mice. C) Immunohistochemical staining of the mouse ventral midbrain coronal section, showing the overlap of TH and mCherry, scalebar

100 μm. D) UMAP projection of the all-nuclei dataset, with cluster color-coding. E) Hierarchical dendrogram of the identified clusters and their descriptions with the same color-coding.

Figure 2 Midbrain DA dataset, cell types and composition

A) UMAP projection of the dopaminergic dataset, with numerical labels of the 71 clusters and color-coded for the major cell types. B) Typical dopaminergic markers plotted in UMAP separately, with the top 1 percentile of the normalized gene expression range clipped for better visualization. C) Hierarchical dendrogram of clusters from the dopaminergic nuclei dataset, with the cell-type enriched markers as a dotplot and cell-type color-coding. mODC = mature oligodendrocyte, HyDA = hypothalamic dopaminergic.

Figure 3 Dopaminergic territories and neighborhoods

A) UMAP projection of the dopaminergic dataset with territory color-coding and size. B) Co-expression of *Sox6* and *Calb1* plotted in UMAP, roughly delineating SNc and VTA respectively. C) Hierarchical dendrograms with labels and color-coded territories and their respective neighborhoods. Dotplot showing territory (TER) and neighborhood (NH) specific markers.

Figure 4 Mapping of mDA territories in the adult mouse midbrain

A) A color-coded UMAP projection of the seven mDA territories and of the mostly-lesion (ML) clusters, with individual enriched genes plotted on smaller UMAPs. B-D) Immunohistochemical staining with antibodies indicated in the ventral midbrain. E) Fluorescent RNA in situ hybridization with *Tacr3* probe combined with immunohistochemistry. Insets are shown as higher magnification on the right. F) Schematic

presentation of the localization of the seven mDA territories across the midbrain, with the most anterior section uppermost. SNc, Substantia Nigra pars compacta, SNcL, Substantia Nigra lateral, SNcV, Substantia Nigra ventral, SNcM, Substantia Nigra medial, VTAR, Ventral Tegmental Area rostral part, PBP, parabrachial pigmented nucleus, PN, paranigral nucleus, PIF, parainterfascicular nucleus, PAG, periaqueductal grey, RRF, retrorubral field, RLi, rostral linear nucleus, CLi, caudal linear nucleus of the raphe, IF, interfascicular nucleus, RN, red nucleus, IPN, interpeduncular nucleus, PN, pontine nucleus, fr, fasciculus retroflexus, ml, medial lemniscus. Scalebars are 200 μm.

Figure 5 Normalized cell loss across territories and neighborhoods in mDA dataset

A, B) UMAP projection of the dopaminergic dataset nuclei from the intact (un-lesioned) and the lesioned hemisphere, respectively. C) Calculated normalized cell loss for territories, visualized as percentages. D) Pairwise comparison of normalized cell loss across territories. E) Calculated normalized cell loss for neighborhoods, visualized as percentages. Comparisons were made between neighborhoods within each territory. See methods for the statistical tests used. F) Percentages of nuclei in each major cell type from either intact or lesioned samples. ns = not significant, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. For the Conover-Iman test; p-value = $P(T \geq |t|)$, and null hypothesis (H_0) was rejected at $p \leq \alpha/2$, which is 0.025. Error bars show standard deviation (SD).

Figure 6 Vulnerability and resilience modules in territories, neighborhoods, and the ML clusters

A-B) Violin plots of the vulnerability module across territories and neighborhoods, with percentage of normalized cell loss shown at the bottom per territory and neighborhood, respectively. C-D) Violin plots of the resilience module across territories and neighborhoods,

with percentage of normalized cell loss shown at the bottom per territory and neighborhood, respectively. Pairwise comparisons of vulnerability module scores across territories and neighborhoods, respectively. E) Pairwise comparisons of resilience module scores across neighborhoods. F) Vulnerability module gene list, ranked by average Log2FC. G) Resilience module gene list, ranked by average Log2FC. Summary statistics center lines are mean values. na = not applicable. ns = not significant. See Methods for the statistical tests used.

Figure 7 Transcriptional kinship and enriched markers of ML clusters

A) UMAP projection of the dopaminergic dataset nuclei from the intact (non-lesioned) hemisphere. B) UMAP projection of the dopaminergic dataset nuclei from the lesioned hemisphere. ML clusters are marked with dashed circles in A and B. C) UMAP projection of the integrated dopaminergic nuclei dataset, color-coded for the territory IDs of the merged dataset. D, E) *Sox6* and *Calb1* expression plotted in intact and lesioned nuclei respectively. F) Same UMAP projection as in C, with ML clusters nuclei highlighted. G-H) Bar graphs showing the numbers and percentages of ML clusters nuclei per territory of the integrated dataset, respectively. More than 80% of ML nuclei cluster in *Sox6* and *Pcsk6* territories of the integrated data. I) Mostly lesioned (ML) clusters enriched markers plotted in UMAP. J-M) ML-enriched markers in I, detected by fluorescent RNA in situ hybridization, combined with TH and mCherry immunohistochemistry in lesioned versus untreated tissue. Scalebar = 50 µm.

Figure 8 Integration of human and mouse datasets

A, B) UMAP projections of the integrated dopaminergic human (control + diseased) and mouse (intact + lesioned) nuclei, split per condition. This integrated dataset includes 25,003 human nuclei and 33,052 mouse nuclei. C) UMAP projections of the human and mouse nuclei

grouped by condition. Dataset was generated using the Seurat canonical correlation analysis (CCA). Ctrl = control neurotypical, PD = Parkinson's disease, LBD = Lewy body dementia, intact = non-lesioned.

References

Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, Barkan E, Bertagnolli D, Casper T, Dee N, Garren E, Goldy J, Graybuck LT, Kroll M, Lasken RS, Lathia K, Parry S, Rimorin C, Scheuermann RH, Schork NJ, Shehata SI, Tieu M, Phillips JW, Bernard A, Smith KA, Zeng H, Lein ES, Tasic B. 2018. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *Plos One* 13:e0209648. doi:10.1371/journal.pone.0209648

Björklund A, Dunnett SB. 2007. Dopamine neuron systems in the brain: an update. *Trends in neurosciences* 30:194–202. doi:10.1016/j.tins.2007.03.006

Blesa J, Przedborski S. 2014. Parkinson's disease: animal models and dopaminergic cell vulnerability. *Front Neuroanat* 8:155. doi:10.3389/fnana.2014.00155

Dougalis AG, Matthews GAC, Bishop MW, Brischoux F, Kobayashi K, Ungless MA. 2012. Functional properties of dopamine neurons and co-expression of vasoactive intestinal polypeptide in the dorsal raphe nucleus and ventro-lateral periaqueductal grey. *Eur J Neurosci* 36:3322–3332. doi:10.1111/j.1460-9568.2012.08255.x

Ekstrand MI, Terzioglu M, Galter D, Zhu S, Hofstetter C, Lindqvist E, Thams S, Bergstrand A, Hansson FS, Trifunovic A, Hoffer B, Cullheim S, Mohammed AH, Olson L, Larsson N-G. 2007. Progressive parkinsonism in mice with respiratory-chain-deficient dopamine neurons. *Proceedings of the National Academy of Sciences of the United States of America* 104:1325–1330. doi:10.1073/pnas.0605208103

FALCK B, HILLARP N-Å, THIEME G, TORP A. 1962. FLUORESCENCE OF CATECHOL AMINES AND RELATED COMPOUNDS CONDENSED WITH FORMALDEHYDE. *J Histochem Cytochem* 10:348–354. doi:10.1177/10.3.348

Fu Y, Yuan Y, Halliday G, Rusznák Z, Watson C, Paxinos G. 2011. A cytoarchitectonic and chemoarchitectonic analysis of the dopamine cell groups in the substantia nigra, ventral tegmental area, and retrorubral field in the mouse. *Brain structure & function*. doi:10.1007/s00429-011-0349-2

Garritsen O, Battum EY van, Grossouw LM, Pasterkamp RJ. 2023. Development, wiring and function of dopamine neuron subtypes. *Nat Rev Neurosci* 1–19. doi:10.1038/s41583-022-00669-3

Gyllborg D, Langseth CM, Qian X, Choi E, Salas SM, Hilscher MM, Lein ES, Nilsson M. 2020. Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res* 48:gkaa792-. doi:10.1093/nar/gkaa792

Hafemeister C, Satija R. 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 20:296. doi:10.1186/s13059-019-1874-1

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. 2021. Integrated analysis of multimodal single-cell data. *Cell* 184:3573-3587.e29. doi:10.1016/j.cell.2021.04.048

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. *The American Journal of Human Genetics* 102:427–446. doi:10.1016/j.ajhg.2018.02.001

Hupe M, Li MX, Gillner KG, Adams RH, Stenman JM. 2014. Evaluation of TRAP-sequencing technology with a versatile conditional mouse model. *Nucleic acids research* 42:e14. doi:10.1093/nar/gkt995

Jiang H-Y, Wek SA, McGrath BC, Lu D, Hai T, Harding HP, Wang X, Ron D, Cavener DR, Wek RC. 2004. Activating Transcription Factor 3 Is Integral to the Eukaryotic Initiation Factor 2 Kinase Stress Response. *Mol Cell Biol* 24:1365–1377. doi:10.1128/mcb.24.3.1365-1377.2004

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* 25:588–595. doi:10.1038/s41593-022-01061-1

Kanaan NM, Collier TJ, Cole-Strauss A, Grabinski T, Mattingly ZR, Winn ME, Steece-Collier K, Sortwell CE, Manfredsson FP, Lipton JW. 2015. The Longitudinal Transcriptomic Response of the Substantia Nigra to Intrastriatal 6-Hydroxydopamine Reveals Significant Upregulation of Regeneration-Associated Genes. *Plos One* 10:e0127768. doi:10.1371/journal.pone.0127768

Kouwenhoven WM, Fortin G, Penttinen A-M, Florence C, Delignat-Lavaud B, Bourque M-J, Trimbuch T, Luppi MP, Salvail-Lacoste A, Legault P, Poulin J-F, Rosenmund C, Awatramani R, Trudeau L-É. 2020. VGluT2 Expression in Dopamine Neurons Contributes to Postlesional Striatal Reinnervation. *J Neurosci Official J Soc Neurosci* 40:8262–8275. doi:10.1523/jneurosci.0823-20.2020

Kozareva V, Martin C, Osorno T, Rudolph S, Guo C, Vanderburg C, Nadaf N, Regev A, Regehr WG, Macosko E. 2021. A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. *Nature* 598:214–219. doi:10.1038/s41586-021-03220-z

Krishnaswami SR, Grindberg RV, Novotny M, Venepally P, Lacar B, Bhutani K, Linker SB, Pham S, Erwin JA, Miller JA, Hodge R, McCarthy JK, Kelder M, McCorrison J, Aevermann BD, Fuertes FD, Scheuermann RH, Lee J, Lein ES, Schork N, McConnell MJ, Gage FH, Lasken RS. 2016. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nature protocols* 11:499–524. doi:10.1038/nprot.2016.015

Krzywkowski T, Hauling T, Nilsson M. 2017. In Situ Single-Molecule RNA Genotyping Using Padlock Probes and Rolling Circle Amplification. *Methods in molecular biology (Clifton, NJ)* 1492:59–76. doi:10.1007/978-1-4939-6442-0_4

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44:W90–W97. doi:10.1093/nar/gkw377

Liao W, Luo H, Ruan Y, Mai Y, Liu C, Chen J, Yang S, Xuan A, Liu J. 2022. Identification of candidate genes associated with clinical onset of Alzheimer's disease. *Front Neurosci-switz* 16:1060111. doi:10.3389/fnins.2022.1060111

Liss B, Haeckel O, Wildmann J, Miki T, Seino S, Roeper J. 2005. K-ATP channels promote the differential degeneration of dopaminergic midbrain neurons. *Nat Neurosci* 8:1742–1751. doi:10.1038/nn1570

Luppi MP, Azcorra M, Caronia-Brown G, Poulin J-F, 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 Reports* 37:109975. doi:10.1016/j.celrep.2021.109975

Manno GL, Gyllborg D, Codeluppi S, Nishimura K, Saltó 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* 167:566-580.e19. doi:10.1016/j.cell.2016.09.027

Mogi M, Togari A, Kondo T, Mizuno Y, Komure O, Kuno S, Ichinose H, Nagatsu T. 1999. Brain-derived growth factor and nerve growth factor concentrations are decreased in the substantia nigra in Parkinson's disease. *Neurosci Lett* 270:45–48. doi:10.1016/s0304-3940(99)00463-2

Motoike T, Long JM, Tanaka H, Sinton CM, Skach A, Williams SC, Hammer RE, Sakurai T, Yanagisawa M. 2016. Mesolimbic neuropeptide W coordinates stress responses under novel environments. *Proc National Acad Sci* 113:6023–6028. doi:10.1073/pnas.1518658113

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 reports* 8:1018–1025. doi:10.1016/j.celrep.2014.07.016

Poulin J-F, Gaertner Z, Moreno-Ramos OA, Awatramani R. 2020. Classification of Midbrain Dopamine Neurons Using Single-Cell Gene Expression Profiling Approaches. *Trends in neurosciences* 43:155–169. doi:10.1016/j.tins.2020.01.004

Poulin J-F, Zou J, Drouin-Ouellet J, Kim K-YA, Cicchetti F, Awatramani RB. 2014. Defining Midbrain Dopaminergic Neuron Diversity by Single-Cell Gene Expression Profiling. *Cell reports* 1–26. doi:10.1016/j.celrep.2014.10.008

Preibisch S, Saalfeld S, Tomancak P. 2009. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* 25:1463–1465. doi:10.1093/bioinformatics/btp184

Qian X, Harris KD, Hauling T, Nicoloutsopoulos D, Muñoz-Manchado AB, Skene N, Hjerling-Leffler J, Nilsson M. 2020. Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat Methods* 17:101–106. doi:10.1038/s41592-019-0631-4

Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, Rivera H de, 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* 174:1015-1030.e16. doi:10.1016/j.cell.2018.07.028

Seijffers R, Mills CD, Woolf CJ. 2007. ATF3 increases the intrinsic growth state of DRG neurons to enhance peripheral nerve regeneration. *J Neurosci Official J Soc Neurosci* 27:7911–20. doi:10.1523/jneurosci.5313-06.2007

Steinkellner T, Zell V, Farino ZJ, Sonders MS, Villeneuve M, Freyberg RJ, Przedborski S, Lu W, Freyberg Z, Hnasko TS. 2018. Role for VGLUT2 in selective vulnerability of midbrain dopamine neurons. *J Clin Invest* 128:774–788. doi:10.1172/jci95795

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* 177:1888-1902.e21. doi:10.1016/j.cell.2019.05.031

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller J, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirschstein D, Pendergraft J, Sullivan HA, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H. 2018. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563:72–78. doi:10.1038/s41586-018-0654-5

Tiklova K, Björklund AK, Lahti L, Fiorenzano A, al B et, 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. *Nature communications* 10:581. doi:10.1038/s41467-019-08453-1

Toskas K, Yaghmaeian-Salmani B, Skiteva O, Paslawski W, Gillberg L, Skara V, Antoniou I, Södersten E, Svenningsson P, Chergui K, Ringnér M, Perlmann T, Holmberg J. 2022. PRC2-mediated repression is essential to maintain identity and function of differentiated dopaminergic and serotonergic neurons. *Sci Adv* 8:eabo1543. doi:10.1126/sciadv.abo1543

Yao Z, Velthoven CTJ van, Nguyen TN, Goldy J, Sedeno-Cortes AE, Baftizadeh F, Bertagnolli D, Casper T, Chiang M, Crichton K, Ding S-L, Fong O, Garren E, Glandon A, Gouwens NW, Gray J, Graybuck LT, Hawrylycz MJ, Hirschstein D, Kroll M, Lathia K, Lee C, Levi B, McMillen D, Mok S, Pham T, Ren Q, Rimorin C, Shapovalova N, Sulc J, Sunkin SM, Tieu M, Torkelson A, Tung H, Ward K, Dee N, Smith KA, Tasic B, Zeng H. 2021. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*. doi:10.1016/j.cell.2021.04.021

Zaborszky L, Vadasz C. 2001. The Midbrain Dopaminergic System: Anatomy and Genetic Variation in Dopamine Neuron Number of Inbred Mouse Strains. *Behav Genet* 31:47–59. doi:10.1023/a:1010257808945

Zetterberg H. 2017. Applying fluid biomarkers to Alzheimer's disease. *Am J Physiol-cell Ph* 313:C3–C10. doi:10.1152/ajpcell.00007.2017

Zhao S, Maxwell S, Jimenez-Beristain A, Vives J, Kuehner E, Zhao J, O'Brien C, Felipe C de, Semina E, Li M. 2004. Generation of embryonic stem cells and transgenic mice expressing green fluorescence protein in midbrain dopaminergic neurons. *The European journal of neuroscience* 19:1133–1140. doi:10.1111/j.1460-9568.2004.03206.x

Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8:14049. doi:10.1038/ncomms14049

Figure 1

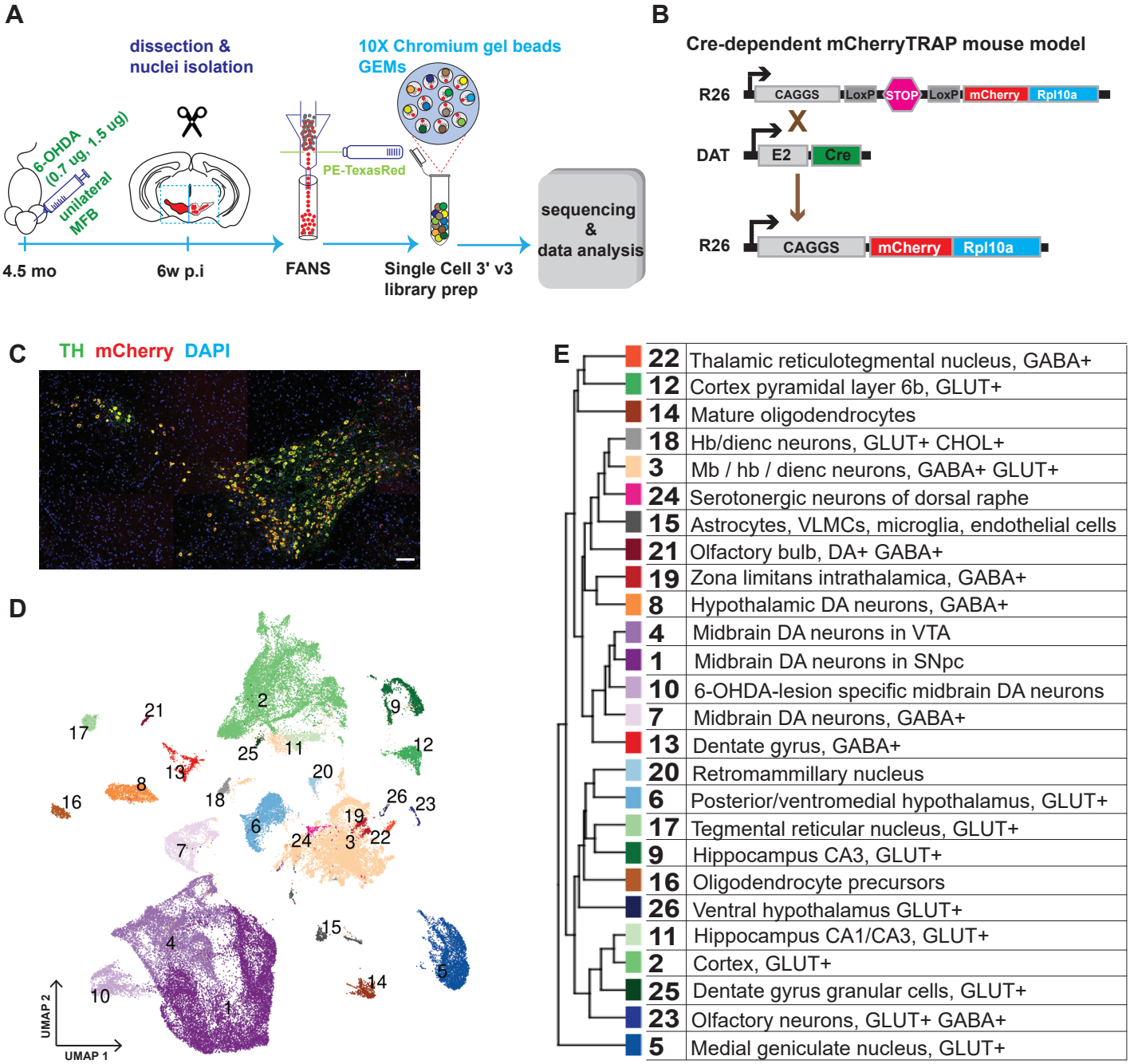

A
dissection &
nuclei isolation
10X Chromium gel beads
GEMs
6-OHDA
(0.7 ug, 1.5 ug)
unilateral
MFB
PE-TexasRed
sequencing
&
data analysis
4.5 mo
6w p.i
FANS
Single Cell 3' v3
library prep
B
Cre-dependent mCherryTRAP mouse model
R26
CAGGS
LoxP
STOP
LoxP
mCherry
Rpl10a
X
DAT
E2
Cre
R26
CAGGS
mCherry
Rpl10a
C
TH mCherry DAPI
D
UMAP 2
UMAP 1
E
22 Thalamic reticulotegmental nucleus, GABA+
12 Cortex pyramidal layer 6b, GLUT+
14 Mature oligodendrocytes
18 Hb/dienc neurons, GLUT+ CHOL+
3 Mb / hb / dienc neurons, GABA+ GLUT+
24 Serotonergic neurons of dorsal raphe
15 Astrocytes, VLMCs, microglia, endothelial cells
21 Olfactory bulb, DA+ GABA+
19 Zona limitans intrathalamica, GABA+
8 Hypothalamic DA neurons, GABA+
4 Midbrain DA neurons in VTA
1 Midbrain DA neurons in SNpc
10 6-OHDA-lesion specific midbrain DA neurons
7 Midbrain DA neurons, GABA+
13 Dentate gyrus, GABA+
20 Retromammillary nucleus
6 Posterior/ventromedial hypothalamus, GLUT+
17 Tegmental reticular nucleus, GLUT+
9 Hippocampus CA3, GLUT+
16 Oligodendrocyte precursors
26 Ventral hypothalamus GLUT+
11 Hippocampus CA1/CA3, GLUT+
2 Cortex, GLUT+
25 Dentate gyrus granular cells, GLUT+
23 Olfactory neurons, GLUT+ GABA+
5 Medial geniculate nucleus, GLUT+

Figure 2

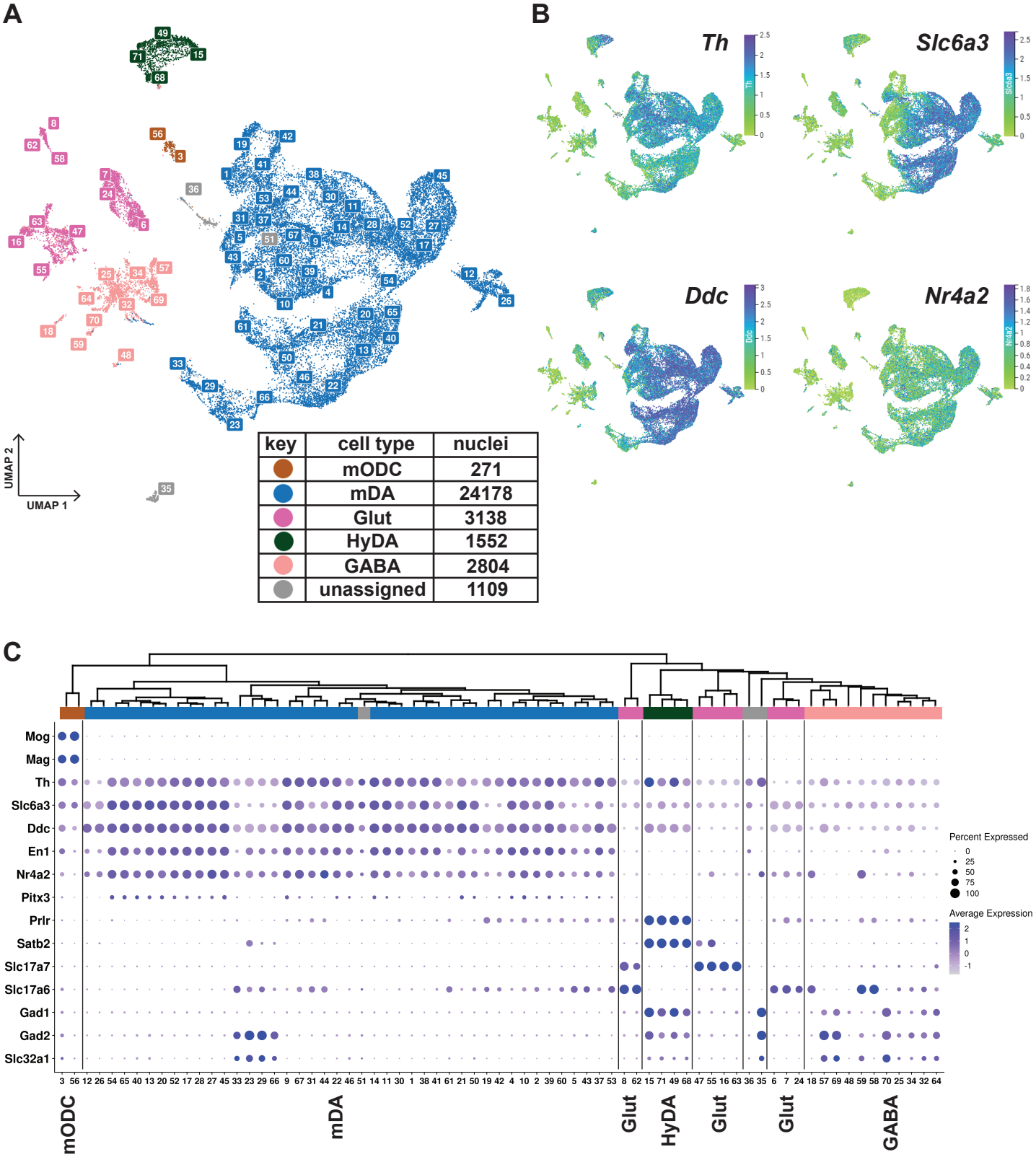
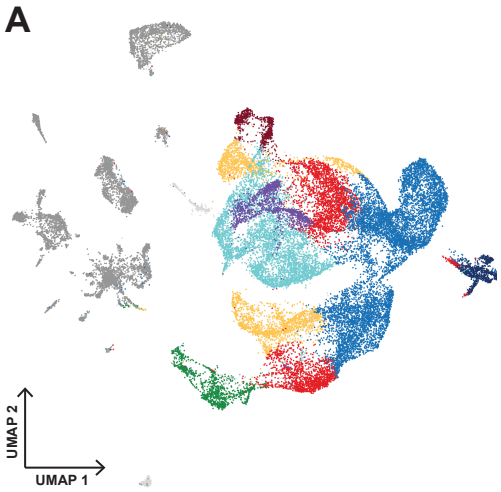
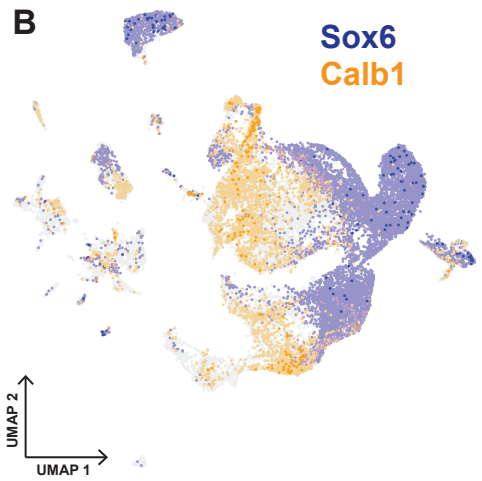

A
B
C
UMAP 2
UMAP 1
Th
Slc6a3
Ddc
Nr4a2
key | cell type | nuclei
mODC | 271
mDA | 24178
Glut | 3138
HyDA | 1552
GABA | 2804
unassigned | 1109
Mog
Mag
Th
Slc6a3
Ddc
En1
Nr4a2
Pitx3
Prlr
Satb2
Slc17a7
Slc17a6
Gad1
Gad2
Slc32a1
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
-1
mODC
mDA
Glut
HyDA
Glut
Glut
GABA

Figure 3



key	territory	nuclei
	Sox6	8514
	Gad2	1291
	Fbn2	1300
	Pcsk6	4008
	Pdia5	2845
	Ebf1	791
	Otx2	4563
	non-mDA	7765
	unassigned	1109

	ML clusters	866



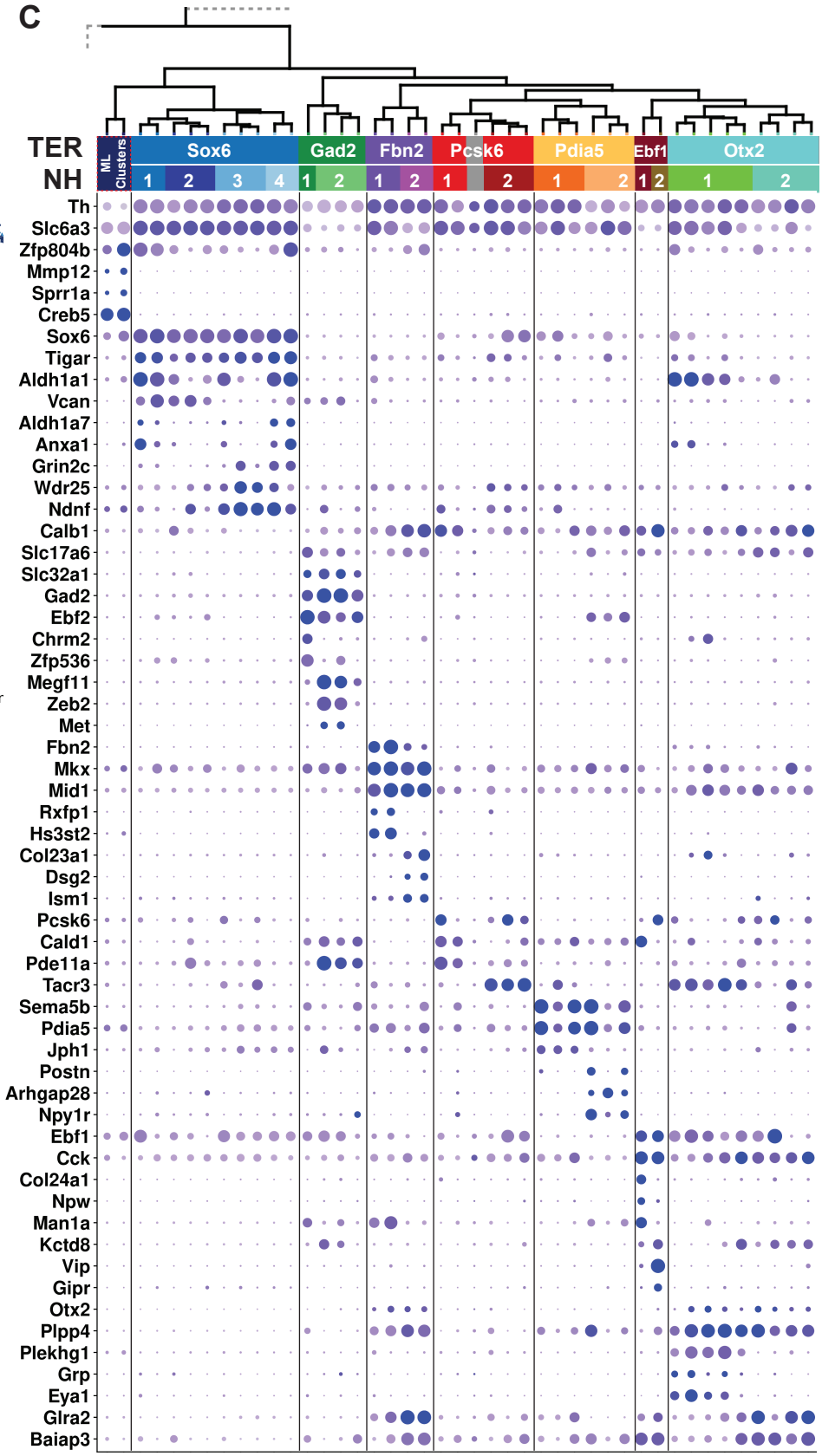


Figure 4

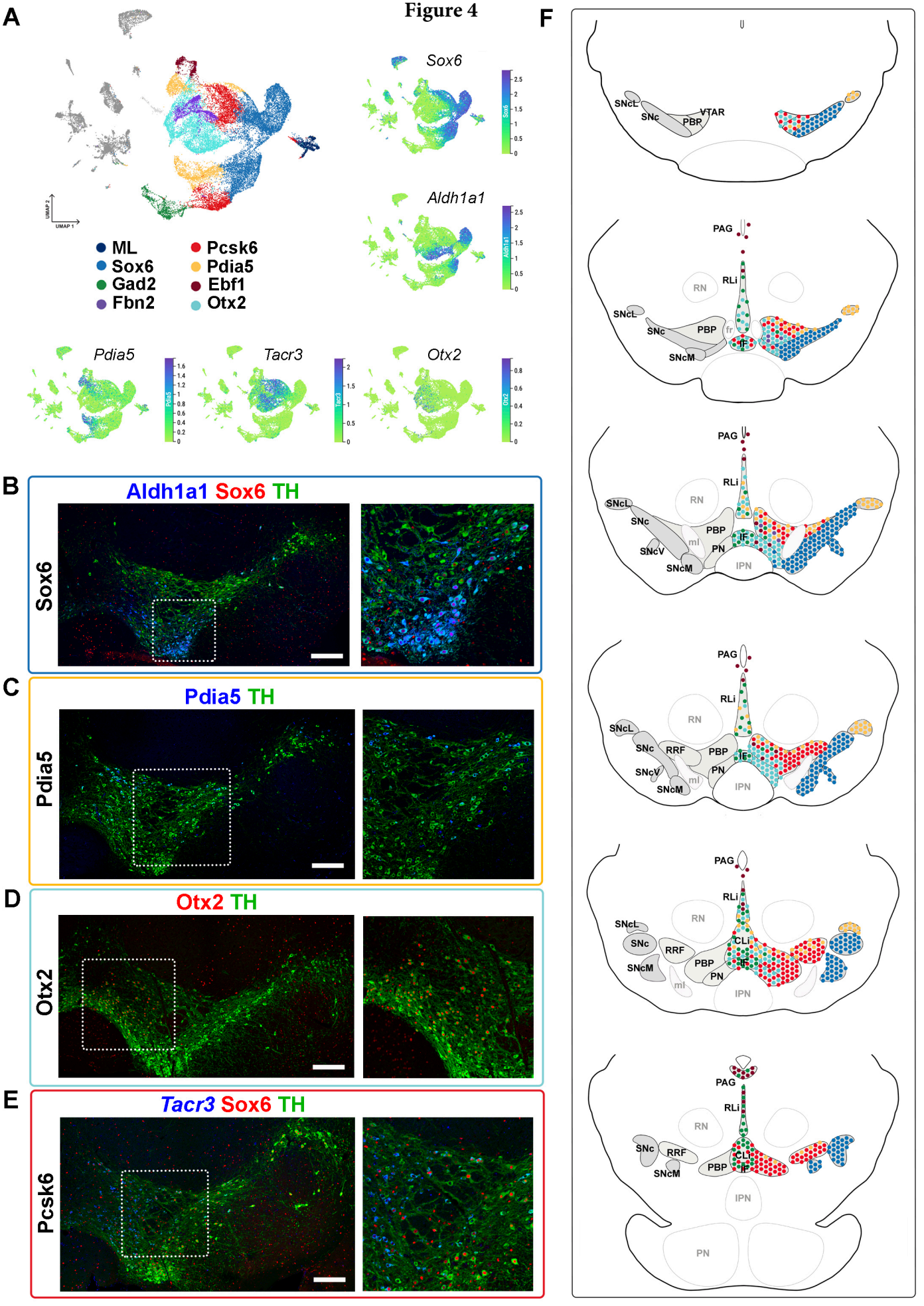
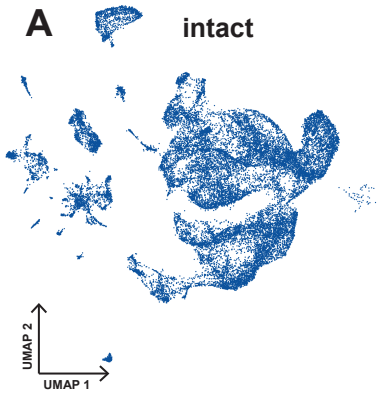
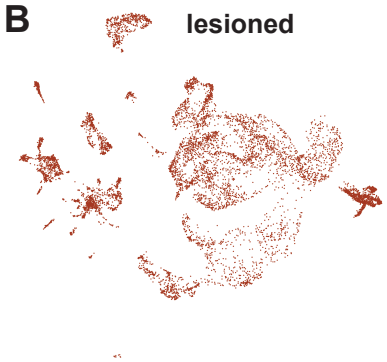
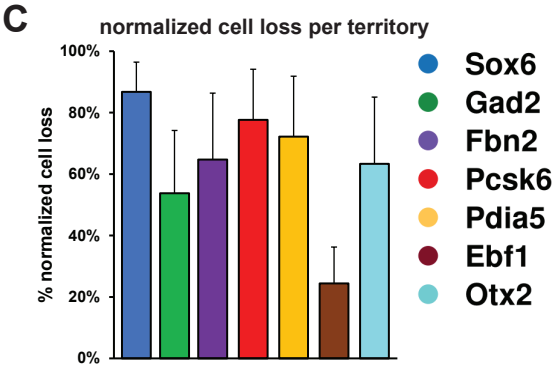

A
UMAP 2
UMAP 1
ML
Sox6
Gad2
Fbn2
Pcsk6
Pdia5
Ebf1
Otx2
Sox6
Aldh1a1
Pdia5
Tacr3
Otx2
B
Aldh1a1 Sox6 TH
Sox6
C
Pdia5 TH
Pdia5
D
Otx2 TH
Otx2
E
Tacr3 Sox6 TH
Pcsk6
F
SNcL
SNc
PBP
VTAR
PAG
RLi
RN
fr
IF
SNcM
SNcV
ml
PN
IPN
RRF
CLi
PN

Figure 5

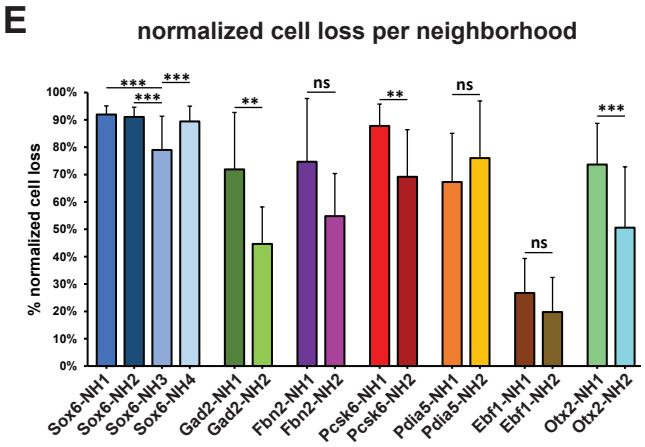






D

Key	Comparisons		P.adjusted
	Ebf1	Pdia5	****
	Ebf1	Otx2	**
	Pdia5	Otx2	ns
	Ebf1	Fbn2	**
	Pdia5	Fbn2	ns
	Otx2	Fbn2	ns
	Ebf1	Sox6	****
	Pdia5	Sox6	****
	Otx2	Sox6	****
	Fbn2	Sox6	****
	Ebf1	Pcsk6	****
	Pdia5	Pcsk6	ns
	Otx2	Pcsk6	**
	Fbn2	Pcsk6	*
	Sox6	Pcsk6	**
	Ebf1	Gad2	ns
	Pdia5	Gad2	**
	Otx2	Gad2	ns
	Fbn2	Gad2	ns
	Sox6	Gad2	****
	Pcsk6	Gad2	***



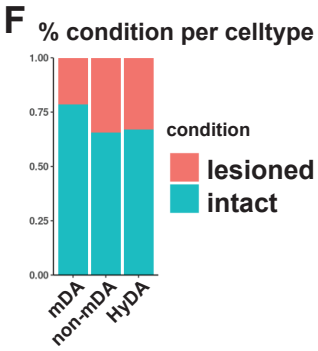
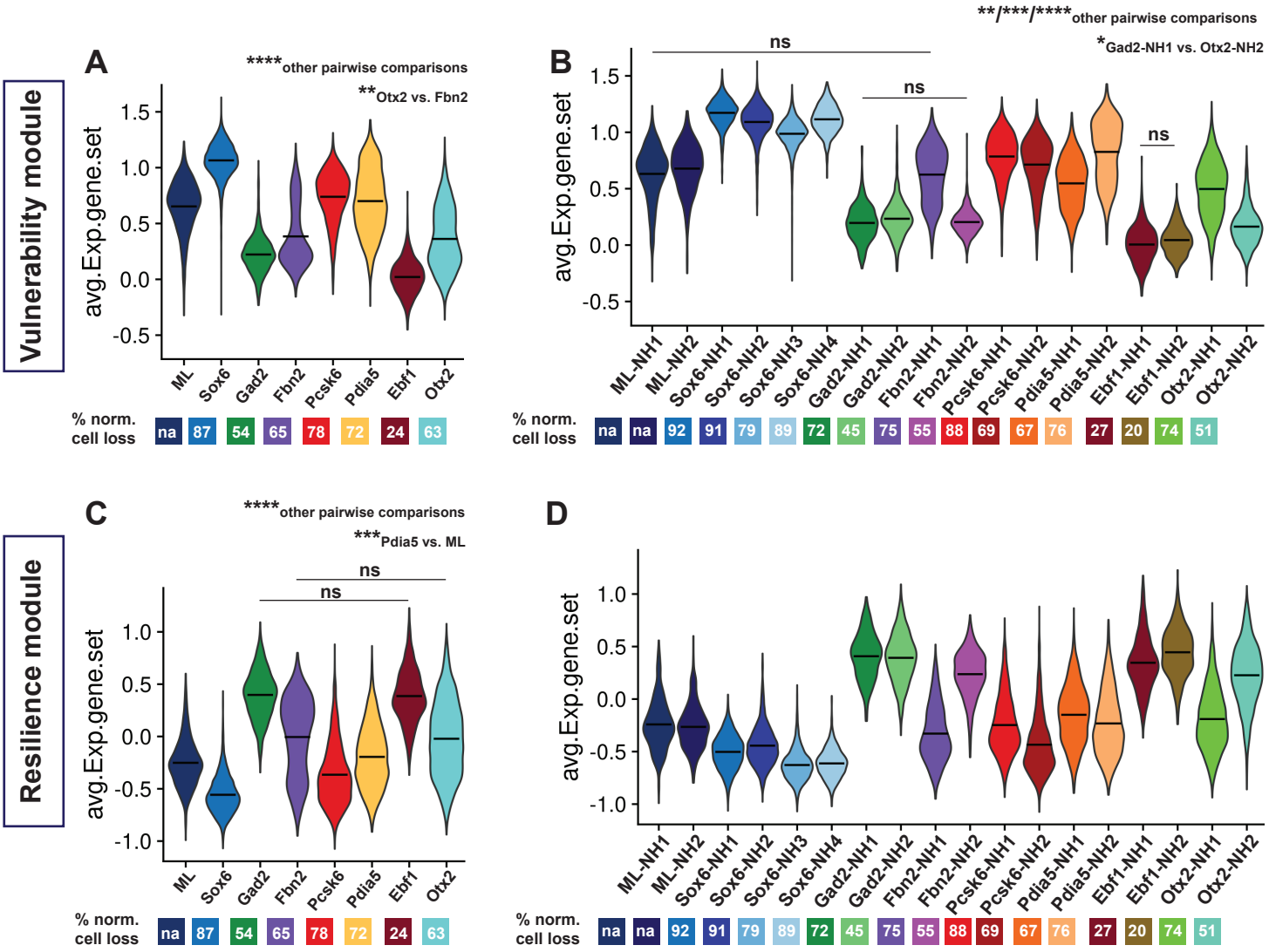
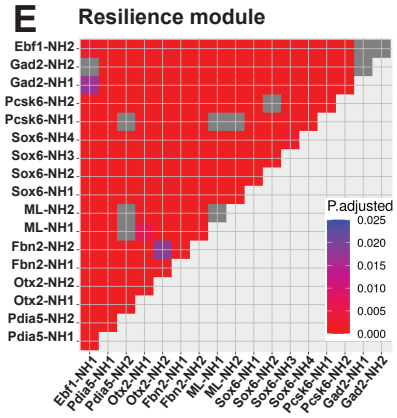


Figure 6





F

Module Genes	AvgLog2FC
Kcnj6	1.10
Rgs6	0.97
Lrp1b	0.91
9530059O14Rik	0.84
Lrrc3b	0.80
Nos1ap	0.79
Slc6a3	0.77
Cyyr1	0.68
Colgalt2	0.61
Ddc	0.59
Ripor2	0.58
Plcb4	0.58
Fam135b	0.57
Bnc2	0.57
Pou3f2	0.57
Pbx1	0.56
Klhl1	0.56
Gpr158	0.55
1700042O10Rik	0.51
Vsnl1	0.51

G

Module Genes	AvgLog2FC
Kctd8	0.87
Frmd5	0.87
Atp8a1	0.70
Scg2	0.64
C130073E24Rik	0.63
Dtna	0.63
Lmx1a	0.61
Nav3	0.59

Figure 7

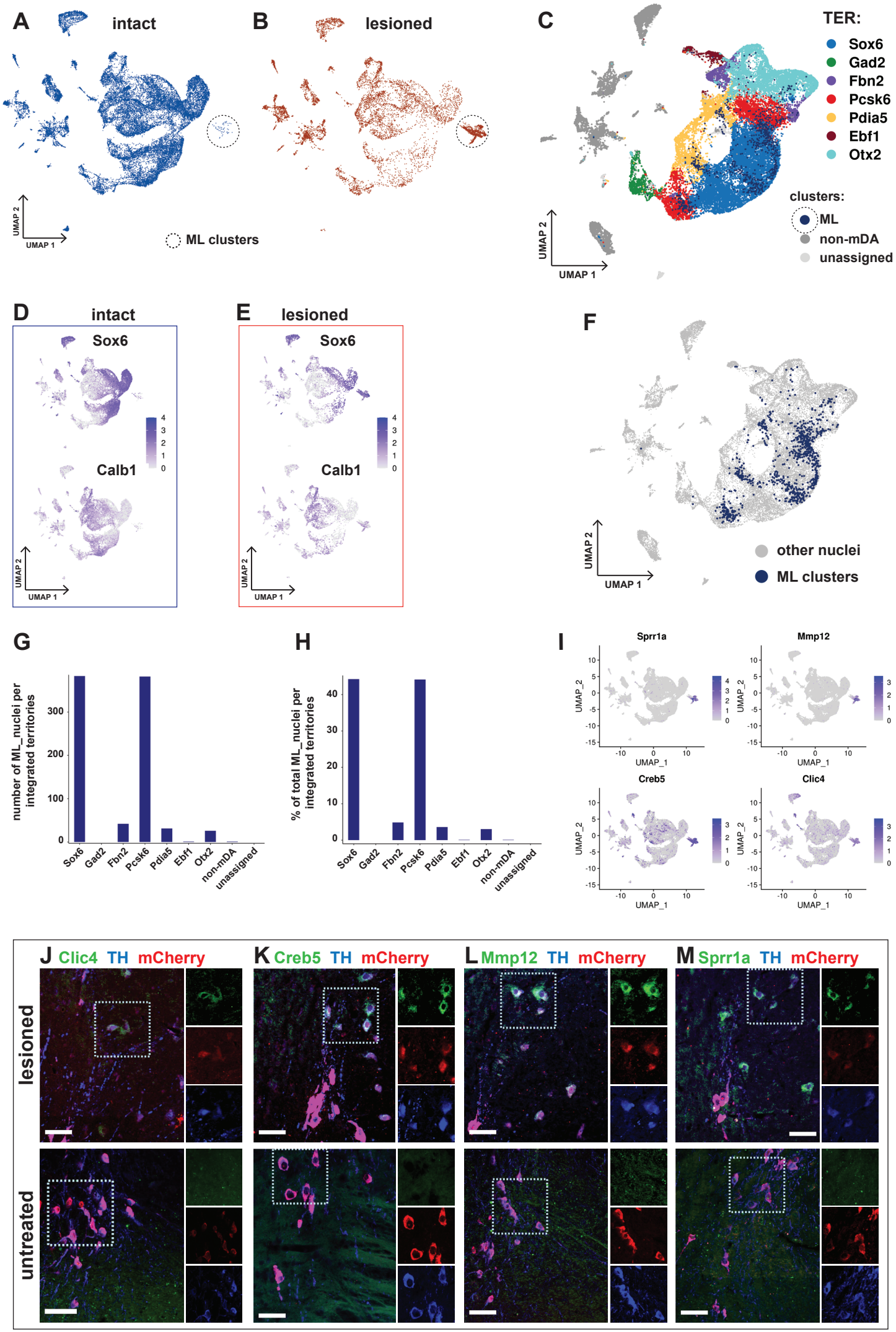
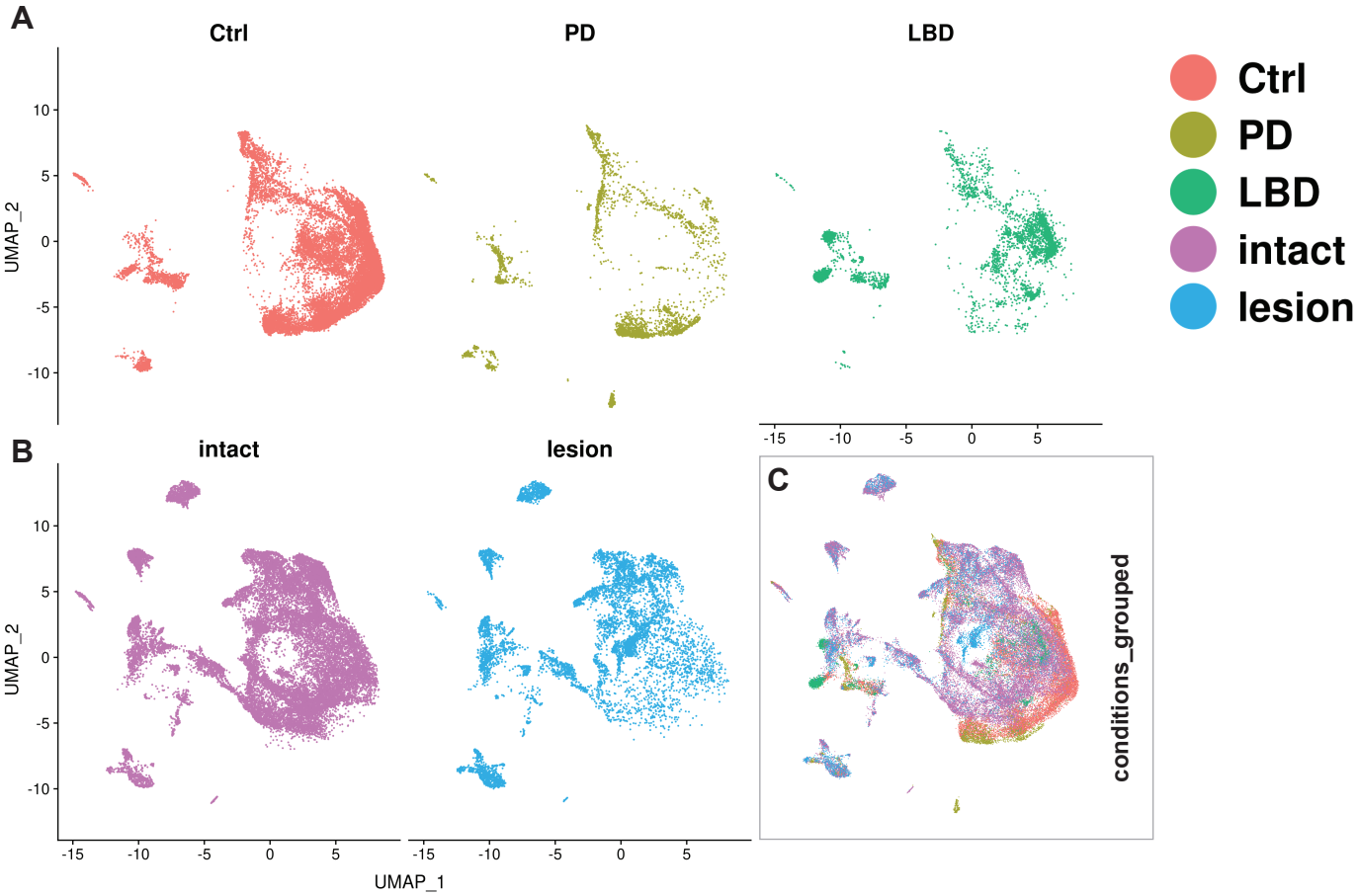

A
intact
B
lesioned
C
TER:
Sox6
Gad2
Fbn2
Pcsk6
Pdia5
Ebf1
Otx2
clusters:
ML
non-mDA
unassigned
UMAP 2
UMAP 1
ML clusters
D
intact
Sox6
Calb1
E
lesioned
Sox6
Calb1
F
other nuclei
ML clusters
G
number of ML_nuclei per integrated territories
Sox6
Gad2
Fbn2
Pcsk6
Pdia5
Ebf1
Otx2
non-mDA
unassigned
H
% of total ML_nuclei per integrated territories
I
Sprr1a
Mmp12
Creb5
Clic4
UMAP_1
UMAP_2
J Clic4 TH mCherry
K Creb5 TH mCherry
L Mmp12 TH mCherry
M Sprr1a TH mCherry
lesioned
untreated

Figure 8


A
Ctrl
PD
LBD
B
intact
lesion
C
UMAP_2
UMAP_1
conditions_grouped
Ctrl
PD
LBD
intact
lesion