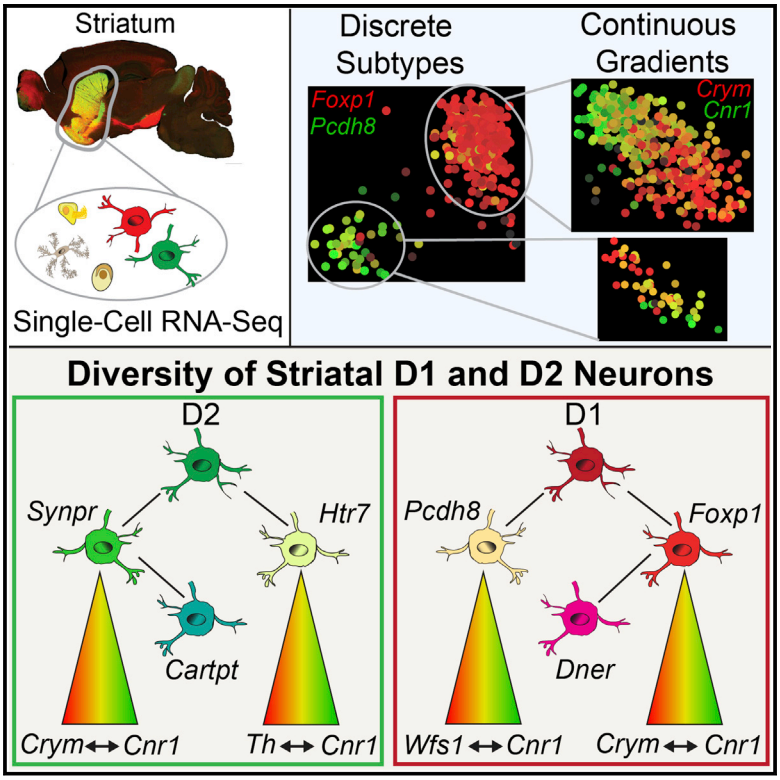
Resource

Cell Reports

Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq

Graphical Abstract



Authors

Ozgun Gokce, Geoffrey M. Stanley, Barbara Treutlein, ..., Marc V. Fuccillo, Thomas C. Südhof, Stephen R. Quake

Correspondence

tcs1@stanford.edu (T.C.S.), quake@stanford.edu (S.R.Q.)

In Brief

The striatum, the gateway to basal ganglia circuitry, is critical for motor functions. However, its cell types are incompletely characterized. Gokce et al. reveal the diversity of striatal cells using scRNA-seq. They also describe continuous expression gradients within all MSN subtypes and astrocytes that may be fundamental to transcriptional diversity.

Highlights

- Transcriptomes of 1,208 single striatal cells
- Description of previously unknown medium spiny neuron subtypes
- Discrete cell types that exist in a continuous spectrum of transcriptional states
- Neurons that have the largest transcriptome and more complex splicing patterns

Accession Numbers

GSE82187

Gokce et al., 2016, Cell Reports *16*, 1126–1137
July 26, 2016
http://dx.doi.org/10.1016/j.celrep.2016.06.059

OPEN ACCESS
CellPress

Cell Reports
Resource

Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq

Ozgun Gokce,[1,7,8] Geoffrey M. Stanley,[2,3,8] Barbara Treutlein,[3,6,8] Norma F. Neff,[3] J. Gray Camp,[6] Robert C. Malenka,[4] Patrick E. Rothwell,[1,4] Marc V. Fuccillo,[1,4] Thomas C. Südhof,[1,5,*] and Stephen R. Quake[3,5,*]

[1]Department of Molecular and Cellular Physiology
[2]Biophysics Program
[3]Departments of Bioengineering and Applied Physics
[4]Nancy Pritzker Laboratory, Department of Psychiatry and Behavioral Sciences
[5]Howard Hughes Medical Institute
Stanford University, Stanford, CA 94305, USA
[6]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[7]Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität LMU, 81377 Munich, Germany
[8]Co-first author
*Correspondence: tcs1@stanford.edu (T.C.S.), quake@stanford.edu (S.R.Q.)
http://dx.doi.org/10.1016/j.celrep.2016.06.059

SUMMARY

The striatum contributes to many cognitive processes and disorders, but its cell types are incompletely characterized. We show that microfluidic and FACS-based single-cell RNA sequencing of mouse striatum provides a well-resolved classification of striatal cell type diversity. Transcriptome analysis revealed ten differentiated, distinct cell types, including neurons, astrocytes, oligodendrocytes, ependymal, immune, and vascular cells, and enabled the discovery of numerous marker genes. Furthermore, we identified two discrete subtypes of medium spiny neurons (MSNs) that have specific markers and that overexpress genes linked to cognitive disorders and addiction. We also describe continuous cellular identities, which increase heterogeneity within discrete cell types. Finally, we identified cell type-specific transcription and splicing factors that shape cellular identities by regulating splicing and expression patterns. Our findings suggest that functional diversity within a complex tissue arises from a small number of discrete cell types, which can exist in a continuous spectrum of functional states.

INTRODUCTION

The striatum, the gateway to basal ganglia circuitry, is involved in translating cortical activity into adaptive motor actions. Conversely, striatal dysfunction in neuronal and non-neuronal cells contributes to many neuropsychiatric disorders, including Parkinson's and Huntington's disease, schizophrenia, obsessive-compulsive disorder, addiction, and autism (Kreitzer and Malenka, 2008; Maia and Frank, 2011; Robison and Nestler, 2011).

The principal projection neurons in the striatum are the medium spiny neurons (MSNs), which constitute 90%–95% of all neurons in the striatum. The classical model of basal ganglia circuits proposes that MSNs are composed of two subtypes with opposing circuit functions. D1 MSNs preferentially express D1-dopamine receptors and promote movement, while D2 MSNs primarily express D2-dopamine receptors and inhibit movement (DeLong and Wichmann, 2009). Anatomical and functional evidence suggests that this model, while heuristically useful, may need to be modified by incorporating a detailed characterization of the phenotypic diversity of striatal MSNs (Calabresi et al., 2014; Cui et al., 2013; Kupchik et al., 2015; Nelson and Kreitzer, 2014). Previous efforts to characterize striatal diversity have been either low-dimensional, measuring a small number of transcripts in single cells, or reliant on pooling large numbers of striatal cells for bulk RNA sequencing (RNA-seq) and obscuring heterogeneity within the pooled populations (Fuccillo et al., 2015; Heiman et al., 2008; Lobo et al., 2006).

Technological advances in single-cell mRNA sequencing (scRNA-seq) have enabled description of the cellular diversity of tissues and allowed identification of distinct cell subtypes in the developing mouse lung (Treutlein et al., 2014b), the murine spleen (Jaitin et al., 2014), the mouse and human cortex and hippocampus (Darmanis et al., 2015; Zeisel et al., 2015), other neuronal tissues (Pollen et al., 2014; Usoskin et al., 2015), and the intestine (Grün et al., 2015). Here, we use scRNA-seq of 1,208 striatal cells combined with unbiased computational analysis to reconstruct the phenotypic heterogeneity of the striatum.

RESULTS

Identification of Major Striatal Cell Types by Transcriptome Clustering

We measured the transcriptomes of 1,208 single striatal cells using two complementary approaches: microfluidic single-cell RNA sequencing (Mic-scRNA-seq) and single-cell isolation by

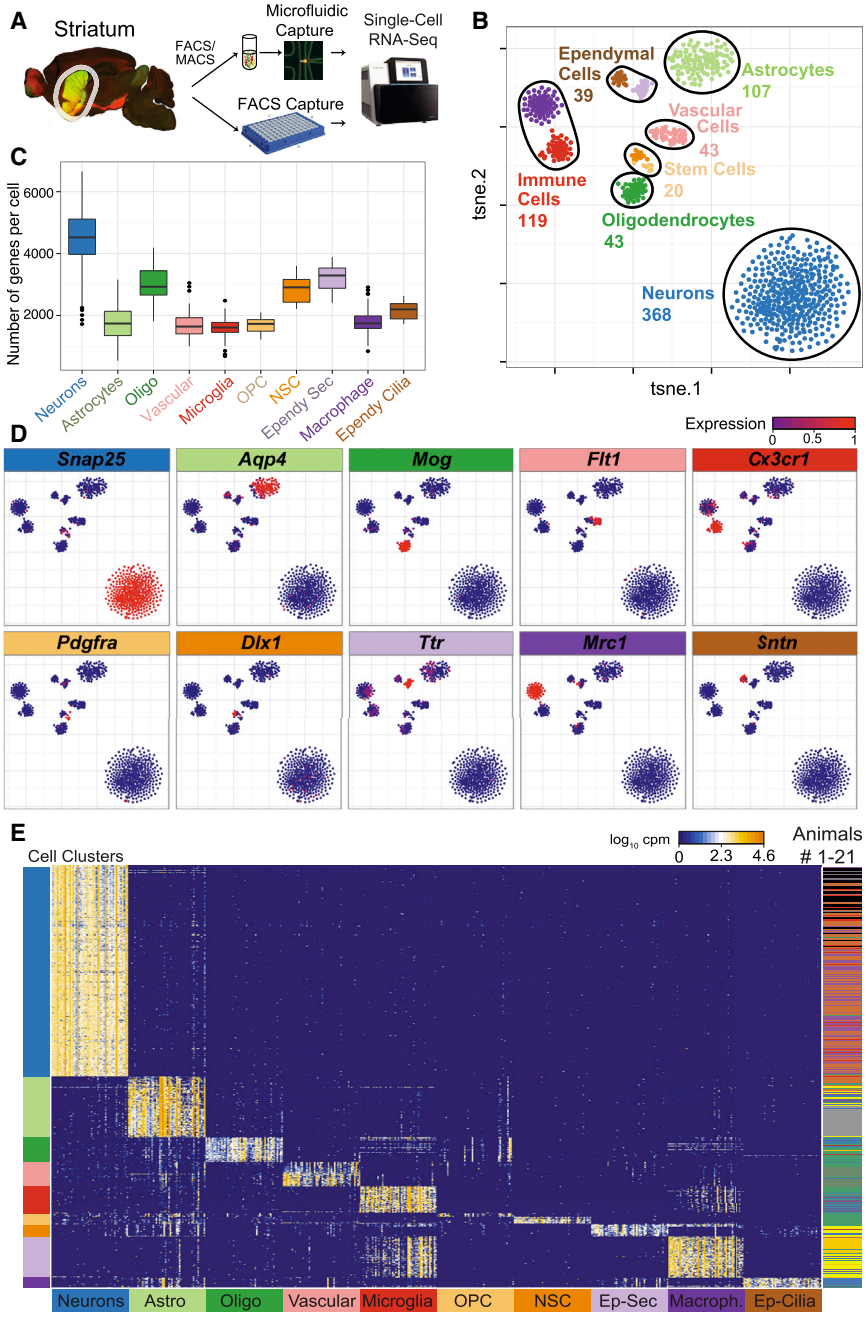


Figure 1. Reverse Engineering of Mouse Striatum by scRNA-Seq

(A) Workflow for obtaining and sequencing cDNA from single cells. Striatal slices from D1-tdTom/D2-GFP and *Aldh1l1*-GFP mice were dissociated and cells collected by FACS or MACS. Cells were then captured, imaged, and cDNA amplified in microfluidic chips.

(B) Unbiased clustering of ten major classes of cells using tSNE, which distributes cells according to their whole-transcriptome correlation distance. Each cell is represented as a dot and colored by a clustering algorithm (DBSCAN).

(C) Box-and-whisker plots showing total number of genes detected per cell for major cell types.

(D) Expression of putative marker genes for each of ten major cell types. Scaled expression of marker genes is shown by the color of the cell points. Each tSNE cluster is enriched for one marker, and we were able to assign cells to one of ten major cell types.

(E) Heatmap of the top 50 genes most highly correlated to each cell type. Each row is a single cell, and each column is a single gene. The bar on the right shows the experimental origin of cells. The bar on the left shows DBSCAN clustering of cells, and the bottom bar shows the cell type assignment for each set of 50 genes. Within each 50-gene set, the genes are ordered by increasing the p value of the correlation to that cell type from left to right.

fluorescence-activated cell sorting (FACS-scRNA-seq) (Table S1). We sampled cells either captured randomly or enriched specifically for MSNs or astrocytes using FACS from D1-tdTomato (tdTom)/D2-GFP or *Aldh1l1*-GFP mice, respectively (Figure 1A) (Heintz, 2004; Shuen et al., 2008). We assessed technical noise, cell quality, and dynamic range using RNA control spike-in standards (Figures S1A–S1D). Saturation analysis confirmed that our sequencing depth of 1×10^6–5×10^6 reads per cell is sufficient to detect most genes expressed (Figure S1E) and that the number of genes detected per cell is independent of the sequencing depth (Figures S1F–S1H).

To identify distinct cell populations in Mic-scRNA-seq cells, we visualized cells using 2D t-distributed stochastic neighbor embedding (tSNE) on whole-transcriptome data weighted by a Bayesian noise model (Figure 1B) (Kharchenko et al., 2014; van der Maaten and Hinton, 2008). By examining the expression of known cell type markers, we assigned cell type names to the major clusters (Figures 1B–1D) (Doetsch et al., 1999; Zhang et al., 2014). Our analysis revealed 368 neuronal cells, 119 immune cells (microglia and macrophages), 107 astrocytes, 43 oligodendrocytes, 43 vascular cells, 39 ependymal cells (ciliated and secretory), and 20 stem cells (Figure S3A). The stem cells were composed of neuronal stem cells (NSCs), likely captured from the rostral migratory stream (Aguirre et al., 2010), and oligodendrocyte precursor cells (OPCs).

The number of expressed genes per cell significantly differed among cell types independent of sequencing depth. Single striatal neurons expressed more genes than all other striatal cell types, suggesting that neurons possess a higher level of functional complexity (Figures 1C and S1E–S1H).

Cells captured in FACS enrichment of *Aldh1l1*-GFP mice were mostly confined to the astrocyte cluster, and cells from FACS enrichment of *Drd1a*-tdTom/Drd2-GFP cells were confined

largely to the neuronal cluster. However, neither tSNE nor whole-transcriptome principal-component analysis (PCA) (Figures 1B and S2), separated D1 MSNs from D2 MSNs within the neuronal cluster, suggesting that their overall gene expression pattern is similar. All identified cell types were sampled in multiple experiments involving a total of 26 mice (Figures 1E and S2; Table S1).

Unbiased Identification of Cell Type-Specific Genes

We identified marker genes for these cell types by the Spearman correlation of genes to cell cluster (Figure 1E; Table S2). NSCs expressed many astrocyte- and neuron-specific genes, further evidence for their proposed origin from astrocytes and eventual fate as neurons.

Striatal neurons show the highest correlation with *Trank1* (also called *Lba1*), whose genetic variants have been associated with bipolar disorder (Mühleisen et al., 2014). The second-highest correlating genes for striatal neurons is *Atp1a3*, which encodes the α3 subunit of Na^+/K+-ATPase and is associated with rapid-onset dystonia parkinsonism and alternating hemiplegia of childhood (Heinzen et al., 2014). The highest correlating gene for striatal astrocytes is *Gjb6*, which is required for intercellular trafficking of glucose and its metabolites to neurons (Pannasch et al., 2014) and is associated with non-syndromic autosomal dominant deafness (Rouach et al., 2008). For perivascular macrophages, the highest correlating gene is *Pf4* (also known as *Cxcl4*), which plays a role in coagulation and inflammation (Amiral et al., 1996). Microglia are defined by specific expression of *Tmem119*, a gene that has been proposed to clearly discriminate microglia from perivascular macrophages (Satoh et al., 2016). The highest correlating gene with OPC is *A930009A15Rik*, a gene that has not yet been functionally characterized (Table S2). Finally, ependymal cells were divided into two subtypes: ciliated ependymal cells are correlated with *Spag17*, *Armc4*, *Hydin*, and *Dnali1*, genes that are involved in cilium movement. The second subtype of secretory ependymal cells is correlated with *Npr3*, *Prlr*, and *Slc4a5*, genes that are involved in cellular secretion (Table S2).

Gene Ontology (GO) analysis based on cell type-specific correlating genes found terms associated with the given cell type function, such as synaptic transmission for neurons, vascular biology for vascular cells, cilium movement for ciliated ependymal cells, and cell division or cell-cycle exit for NSCs. GO analysis returned only two terms for OPC and none for secretary ependymal cells, highlighting the limited knowledge of their biology (Tables S3 and S4).

Discrete Subtypes of Striatal MSNs

The neuronal composition of the striatum is defined by a large population of D1- and D2 MSNs and a small population of interneurons (Kreitzer and Malenka, 2008). Using a set of known interneuron and MSN markers, we identified five interneurons, too few to explore their diversity (Figure S3B; Table S5). Using Mic-scRNA-seq, we analyzed a total of 334 MSNs either captured randomly or enriched using FACS (Figures S3C–S3E). The fluorescence signal emitted by each cell was monitored on the microfluidic chip (Figure S3E). To identify MSN subtype-specific genes with an unbiased approach, we used pairwise correlation analysis to identify genes that showed high correlation of expression within a group of genes and strong anticorrelation to another group of correlated genes. This revealed two sets of genes that are specific to D1- or D2 MSNs (Figure 2A). *Tac1*, *Drd1a*, and *Isl1* are known to be required for D1 MSN differentiation and function (Ehrman et al., 2013; Heiman et al., 2008). The D2-specific gene cluster included the known D2 MSN markers *Drd2*, *Adora2a*, *Penk*, *Gpr6*, and *Gpr52* (Heiman et al., 2008; Lobo et al., 2006), as well as the transcription factor (TF) *Sp9* (Figure 2A). We performed robust principal-component analysis (rPCA) (Todorov and Filzmoser, 2009) using those genes and observed a clearly bimodal distribution of PC1 scores, indicating that the D1/D2 division represented discrete subtypes and was independent of cell isolation methods (Figure S3C).

We confirmed our findings using an independent approach in which we sequenced an additional 570 single MSNs from five mice using FACS-scRNA-seq. We scored MSNs based on their expression of the D1 and D2 marker genes (D1-D2 scores) (Supplemental Experimental Procedures) and observed a clearly bimodal distribution of cells based on these scores (Figures 2B and 2C), indicating the existence of two discrete subtypes: D1- and D2 MSNs (Figure 2D). Some MSNs coexpressed several D1 and D2 markers. We confirmed their existence in the tissue by in situ imaging of the striatum of D2-GFP/D1-TdTom double-reporter mice. We could identify tdTom-GFP double-positive MSNs in both nucleus accumbens and dorsal striatum using fluorescence microscopy (Figures 2E, 2F, and S3F). The rPCA of D1 MSNs revealed a set of genes that separated D1 MSNs into two subpopulations: a majority subpopulation expressing high levels of *Foxp1* and *Camk4* and a minor subpopulation defined by expression of genes such as *Pcdh8*, *Tacr1*, and *Adarb2* (*Pcdh8*-MSN) (Figures 2G–2J; Table S6). The Pcdh8-MSN subpopulation coexpressed the D2 neuropeptide *Pcdh8* along with *Tac1* (Figure 2H). We scored D1 MSNs based on expression of rPCA-identified genes (*Pcdh8-Foxp1* scores), which revealed a clearly bimodal population distribution defining two discrete subpopulations (Figure 2I). Many genes enriched in *Pcdh8*-MSNs are related to neurological disorders (*Nrxn2*, *Sema3e*, *Sema4a*, *Sema5a*, *Sema5b*, *Sema6d*, *Pcdh7*, *Pcdh8*, *Ptprg*, *Ptprm*, *Ptpro*, and *Ptpru*) (Redies et al., 2012; De Rubeis et al., 2014; Südhof, 2008), tachykinin signaling (*Tac2* and *Tacr1*) (Steinhoff et al., 2014), and RNA binding (*Elavl4*, *Adarb2*, *Khdrbs3*, *Rbm20*, *Aff2*, *Lrpprc*, and *Celf4*) (Li et al., 2007; Noureddine et al., 2005). *Pcdh8*-MSNs are also depleted of important neuronal genes like *Nlgn1* and *Calb1* (Table S6).

Similar to D1 MSNs, rPCA of D2 MSNs revealed two discrete subpopulations: a small subpopulation marked by unique expression of many genes, including *Htr7* and *Agtr1a*, and the coexpression of D1 and D2 neuropeptides *Penk* and *Tac1* (*Htr7*-MSN) (Figures 2K–2N; Table S6). Genes enriched in the *Htr7*-MSN subpopulation included receptors *Htr7*, *Ptprt*, *Ngfr*, *Grik3*, *Cacng5*, and *Tmeff2* and TFs *Sox9*, *Sp8*, *Runx1*, *Mafb*, and *Litaf*. *Htr7*-MSNs are significantly depleted of important neuronal genes like *Cacna2d3* and *Synpr* compared to the major D2 population (Table S6) and showed coexpression of the *Gfp* and *tdTom* transgenes.

Continuous Transcriptional Gradients within MSN Subtypes

We continued our analysis of heterogeneity within subpopulations of MSNs. The rPCA on *Foxp1*-high D1 MSNs revealed a

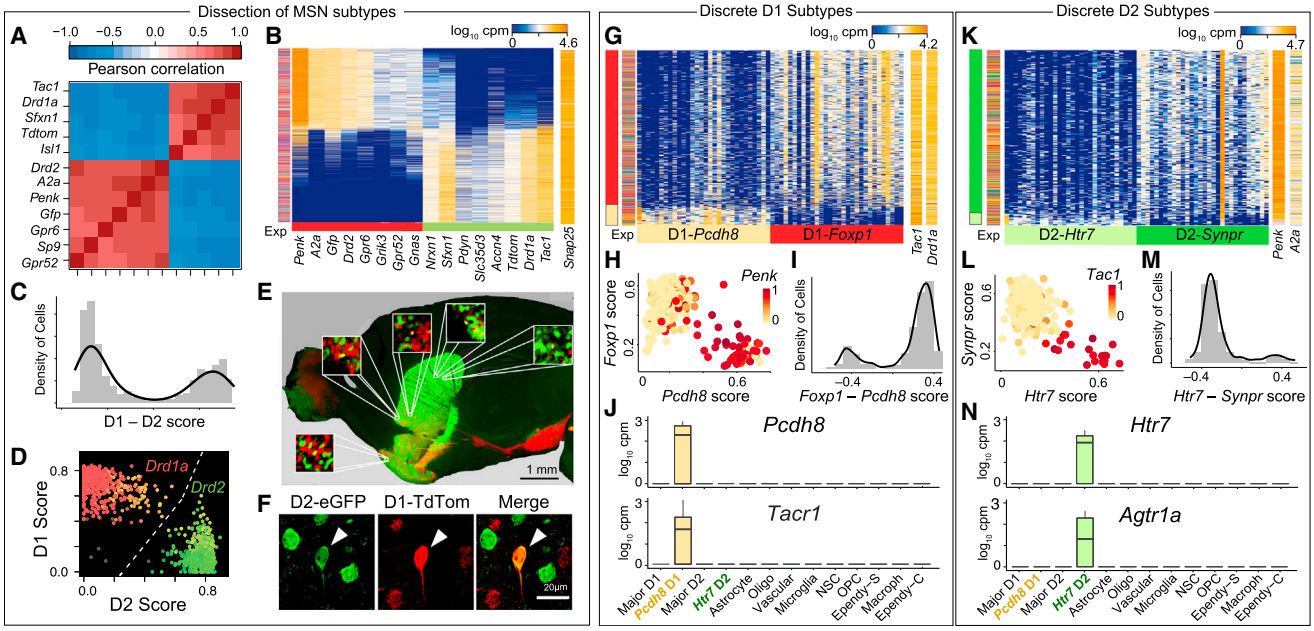


Figure 2. Characterization of Discrete MSN Subtypes

(A) Hierarchical clustering of the highest pairwise gene correlations in MSNs shows two strongly anticorrelated clusters of genes that include known MSN subtype markers for D1- and D2 MSNs.

(B) rPCA of MSNs. Cells (rows) are ordered by their projection onto PC1, and genes (columns) are ordered by their positive (left) or negative (right) contribution to PC1. This identifies three molecularly distinct populations, assigned as D1 MSN (red), D2 MSN (green), and D1/2 hybrid MSN (yellow). The bar on the left shows the experimental origin of cells.

(C) Distribution of single MSNs projected onto D1-D2 scores. The D1-D2 score is calculated by summing the scaled expression values of the genes shown in (B). D1- and D2 MSNs form distinct peaks, and the novel MSNs are distributed between the two peaks.

(D) Biplot of D1- and D2 MSN cells by their expression of D1 genes (y axis) and of D2 genes (x axis). Scaled expression of *Drd1a* is shown by the color of the cell points.

(E) Sagittal brain section of a D2-GFP/D1-TdTom double-reporter mouse showing tdTom-GFP double-positive novel MSNs in both nucleus accumbens and dorsal striatum.

(F) Confocal imaging of striatal slices of D2-GFP/D1-TdTom double-reporter mice demonstrating the existence of D1/2 hybrid MSNs in the striatum.

(G) rPCA of MSNs in the D1 part of (D) using all expressed genes. Cells (rows) are ordered by their projection onto PC1, and genes (columns) are ordered by their positive (left) or negative (right) contribution to PC1. This identifies two molecularly distinct populations assigned as major (*Foxp1*) D1 MSN (red) and *Pcdh8*-MSN (yellow).

(H) Biplot of *Foxp1*-D1 and *Pcdh8*-MSN cells by their expression of D1 genes (y axis) and *Pcdh8* genes (x axis). Scaled expression of *Penk* is shown by the color of the cell points.

(I) Distribution of MSNs projected onto *Pcdh8-Foxp1* scores. Major-D1 and *Pcdh8*-MSNs form distinct peaks.

(J) Boxplots showing specific *Pcdh8* and *Tacr1* expression in *Pcdh8*-MSNs compared to other cell types in MSN.

(K) rPCA of MSNs in the D2 part of (D). Cells (rows) are ordered by their projection onto PC1, and genes (columns) are ordered by their positive (left) or negative (right) contribution to PC1. This identifies two molecularly distinct populations assigned as D2 MSN (green) and *Htr7*-MSN (light green).

(L) Biplot of D2- and *Htr7*-MSN cells by their expression of *Htr7* genes (y axis) and *Synpr* genes (x axis). Scaled expression of *Tac1* is shown by the color of the cell points.

(M) Distribution of MSNs projected onto *Htr7-Synpr* scores. Major D2- and *Htr7*-MSNs form distinct peaks.

(N) Boxplots showing specific *Htr7* and *Agtr1a* expression in *Htr7*-MSNs compared to other cell types in MSN.

subpopulation expressing high levels of *Dner*, *Cxcl14*, and *Tnnt1* and lacking *Meis2* (D1-*Dner*) (Figures 3A–3C and S4A; Table S6). Similarly, rPCA on *Synpr*-high D2 MSNs revealed a small population expressing *Cartpt* and *Kcnip1* but lacking *Calb1* (Figures 3G–3H and S4C). The rPCA on the *Meis2*+/*Dner*− subpopulation of D1 MSNs revealed a gradient of transcriptome states (Figure 3D), indicated by the unimodal distribution of cells' *Crym-Cnr1* score (Figures 3E and S4B). Some genes defining the opposing gradients are known to be differentially expressed between striatal matrix (e.g., *Cnr1*) and striosome (e.g., *Nnat* and *Gfra1*) microzones (Figures 3D, 3E, and S4B) (Crittenden and Graybiel, 2011). The rPCA revealed similar unimodal gene expression gradients for the other main MSN subtypes (Figures 3J–3R and S4D). The transcriptional gradients within all main MSN subtypes shared several genes (e.g., *Cnr1*, *Crym*, and *Wfs1*).

Subtypes of Non-neuronal Striatal Cell Types

We next characterized heterogeneity within vascular cells, immune cells, oligodendrocytes, and astrocytes. Within vascular cells, pairwise correlation analysis revealed two large anti-correlated gene groups (Figure 4A). The rPCA using these

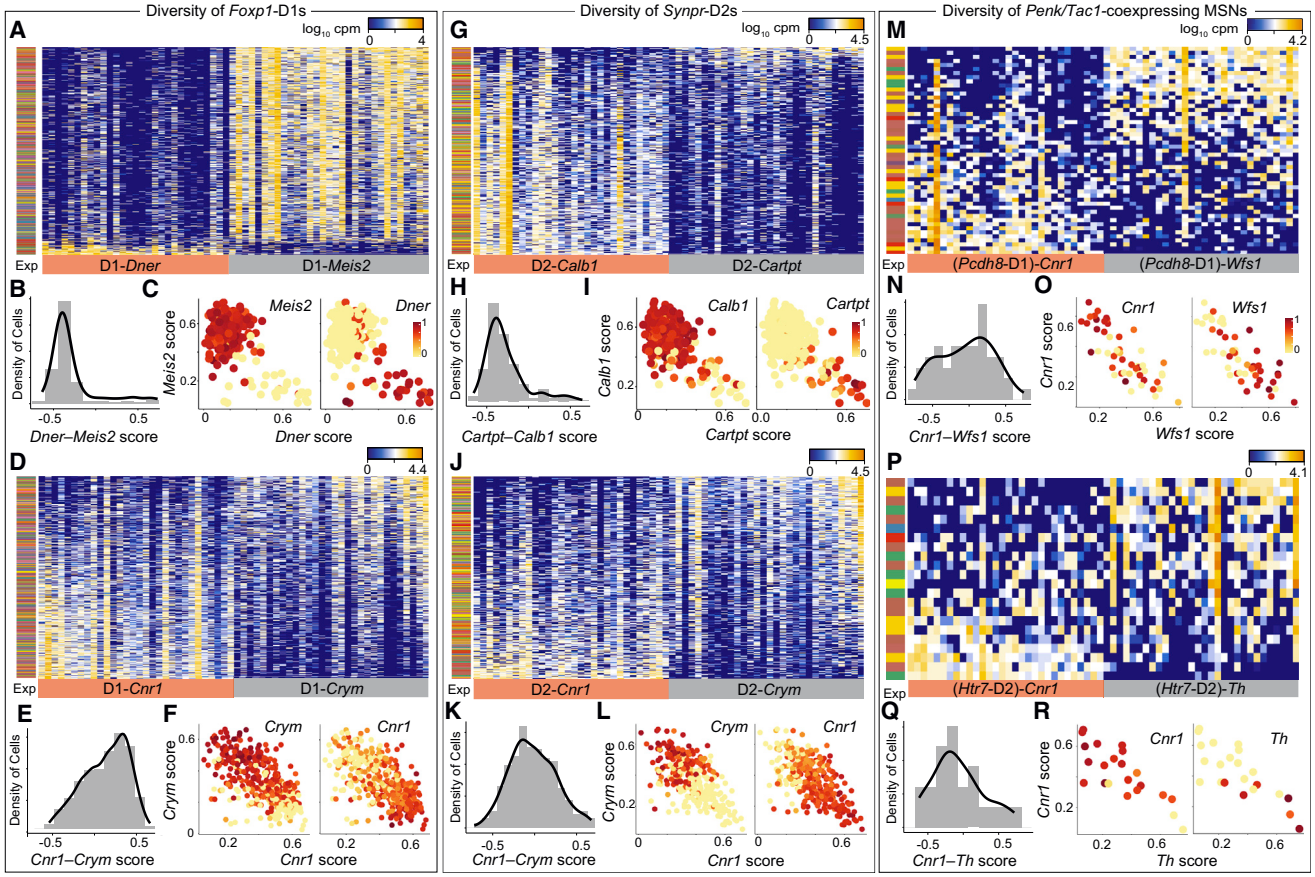


Figure 3. Identification of Heterogeneity within MSN Subtypes

(A) rPCA of major D1 MSNs using all expressed genes, revealing a molecularly distinct subpopulation of D1-*Dner* MSNs. The bar on the left shows the experimental origin of cells.

(B) Distribution of D1 MSNs projected onto *Meis2*-*Dner* gene group scores.

(C) Biplot of D1- and D1-*Dner* MSNs by their expression of D1-*Meis2* genes (y axis) and D1-*Tnnt1* genes (x axis). Scaled expression of *Meis2* and *Tnnt1* is shown by the color of the cell points.

(D) rPCA of D1 MSNs using all expressed genes, revealing a continuous transcriptional gradient marked by opposing expression gradients of *Cnr1* and *Crym*. The bar on the left shows the experimental origin of cells.

(E) Distribution of D1 MSNs projected onto *Cnr1*-*Crym* gradient scores.

(F) Biplot of D1 MSNs by their expression of *Crym* gradient genes (y axis) and *Cnr1* gradient genes (x axis). Scaled expression of *Cnr1* and *Crym* is shown by the color of the cell points.

(G) rPCA of D2 MSNs using all expressed genes, revealing a subpopulation of D2-*Cartpt* MSNs. The bar on the left shows the experimental origin of cells.

(H) Distribution of D2 MSNs projected onto *Calb1*-*Cartpt* gene group scores.

(I) Biplot of D2-*Calb1* and D2-*Cartpt* MSNs by their expression of *Cartpt* group genes (y axis) and *Calb1* group genes (x axis). Scaled expression of *Calb1* and *Cartpt* is shown by the color of the cell points.

(J) rPCA of D2 MSNs using all expressed genes, revealing a continuous transcriptional gradient marked by opposing expression gradients of *Cnr1* and *Crym*. The bar on the left shows the experimental origin of cells.

(K) Distribution of D1 MSNs projected onto *Cnr1*-*Crym* gradient scores.

(L) Biplot of D1 MSNs by their expression of *Crym* gradient genes (y axis) and *Cnr1* gradient genes (x axis). Scaled expression of *Cnr1* and *Crym* is shown by the color of the cell points.

(M) rPCA of *Pcdh8*-MSNs using all expressed genes, revealing a continuous transcriptional gradient marked by opposing expression gradients of *Cnr1* and *Wfs1* similar to 2D. The bar on the left shows the experimental origin of cells.

(N) Distribution of *Pcdh8*-MSNs projected onto *Cnr1*-*Wfs1* gradient scores.

(O) Biplot of *Pcdh8*-MSNs by their expression of *Wfs1* gradient genes (y axis) and *Cnr1* gradient genes (x axis). Scaled expression of *Cnr1* and *Wfs1* is shown by the color of the cell points.

(P) rPCA of *Htr7*-MSNs, revealing a continuous transcriptional gradient marked by opposing expression gradients of *Cnr1* and *Th*. The bar on the left shows the experimental origin of cells.

(Q) Distribution of *Htr7*-MSNs projected onto *Cnr1*-*Th* gradient scores.

(R) Biplot of *Htr7*-MSNs by their expression of *Th* gradient genes (y axis) and *Cnr1* gradient genes (x axis). Scaled expression of *Cnr1* and *Th* is shown by the color of the cell points.

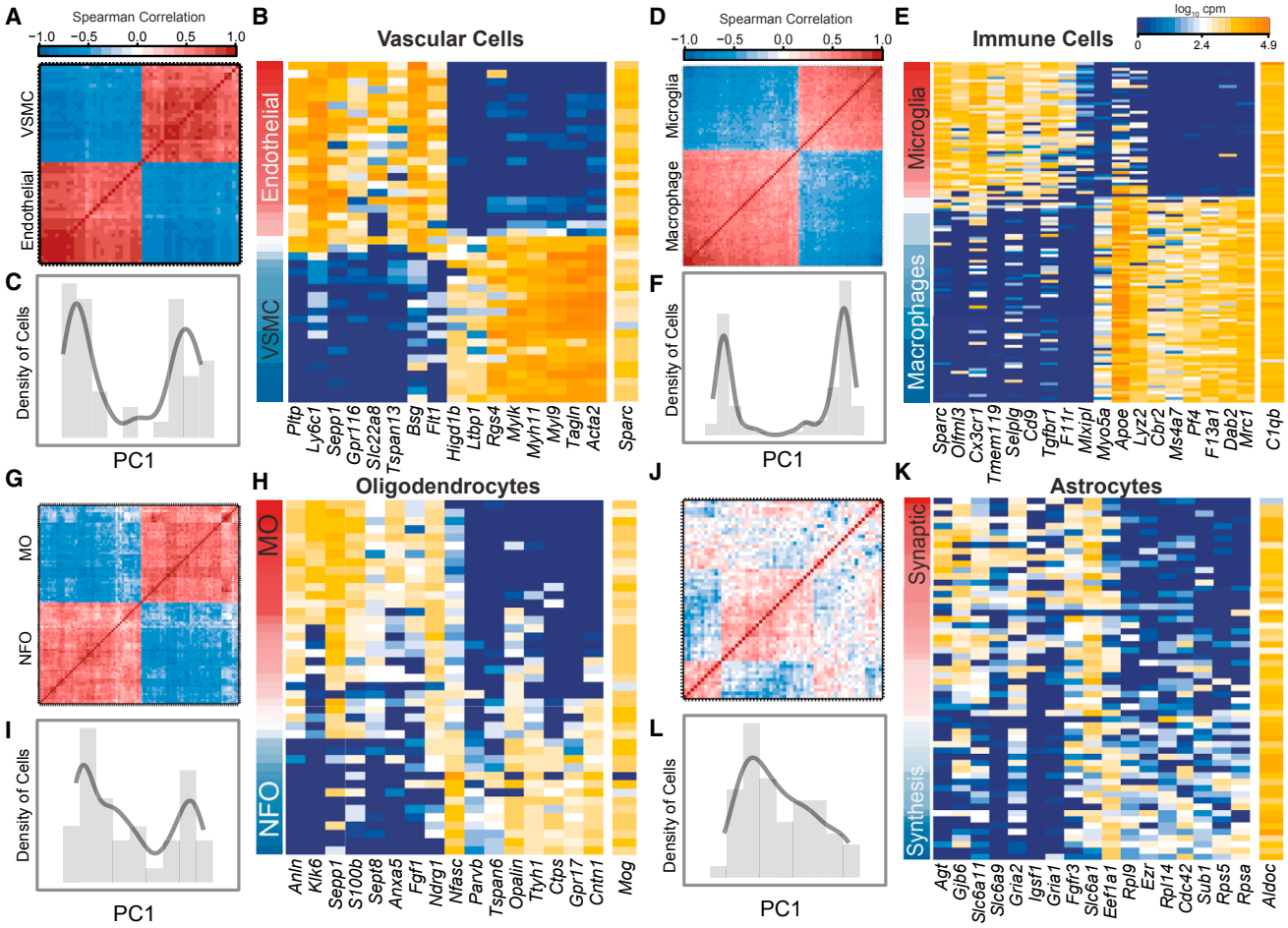


Figure 4. Characterization of Vascular Cells, Astrocytes, and Oligodendrocytes

(A) Hierarchical clustering of the highest pairwise gene correlations within vascular cells reveals two large clusters of subtype-specific transcripts.

(B) rPCA of 43 striatal vascular cells identifies two molecularly distinct populations, assigned to VSMCs that express *Myl9* and *Tagln* and endothelial cells that express *Ly6e* and *Pltp*.

(C) Histogram of vascular cells' PC1 scores shows a clearly bimodal distribution and confirms the existence of two distinct subtypes.

(D) Hierarchical clustering of the highest pairwise gene correlations within immune cells reveals two large clusters of interconnected transcripts.

(E) rPCA of 119 immune cells (microglia and macrophages) identifies two molecularly distinct populations, assigned to microglia that express *Sparc* and macrophages that express *Mrc1*.

(F) Distribution of single immune cells along the first PC1 of rPCA. Cells form two distinct peaks.

(G) Hierarchical clustering of the highest pairwise gene correlations within oligodendrocytes reveals two large clusters of interconnected transcripts.

(H) rPCA of 43 striatal oligodendrocytes identifies two distinct oligodendrocyte populations: MOs, which express *Klk6* and *Sec11c*, and NFOs, which express *Nfasc* and *Ckb*. There are also a significant number of oligodendrocytes with intermediate expression of both sets of genes, which are likely transitioning between NFOs and MOs.

(I) Distribution of single oligodendrocytes along the first PC1 of rPCA. MO and NFO cells form two distinct peaks, with transitioning oligodendrocytes bridging the two peaks.

(J) Hierarchical clustering of the highest pairwise gene correlations within astrocytes did not reveal interconnected transcripts, instead showing low correlation values between genes.

(K) rPCA of 107 single striatal astrocytes reveals transcripts that increase or decrease continuously without defining distinct subpopulations. Cells (rows) are ordered by their PC1 scores, and genes (columns) are ordered by their PC1 loading. The continuum of astrocyte transcriptional states is marked on one side by higher expression of transcripts related to synaptic communication (*Slc6a11*, *Slc6a1*, *Slc6a9*, *Gria1*, and *Gria2*) and on the other side by higher expression of transcripts related to translation (*Rpl9*, *Rpl14*, *Rps5*, and *Rpsa*) and cell polarity regulators (*Cdc42*).

(L) The distribution of single astrocytes along PC1 is unimodal, indicating striatal astrocytes exhibit continuous transcriptional variation within one discrete subtype.

genes revealed two discrete subpopulations: a population of vascular smooth muscle cells (VSMCs) (expression of *Myl9* and *Tagln*), and a population of endothelial cells (expression of *Pltp* and *Ly6c1*) (Figures 4B and 4C) (Zeisel et al., 2015). Likewise, within immune cells, pairwise correlation analysis and rPCA revealed two discrete subpopulations: microglia cells

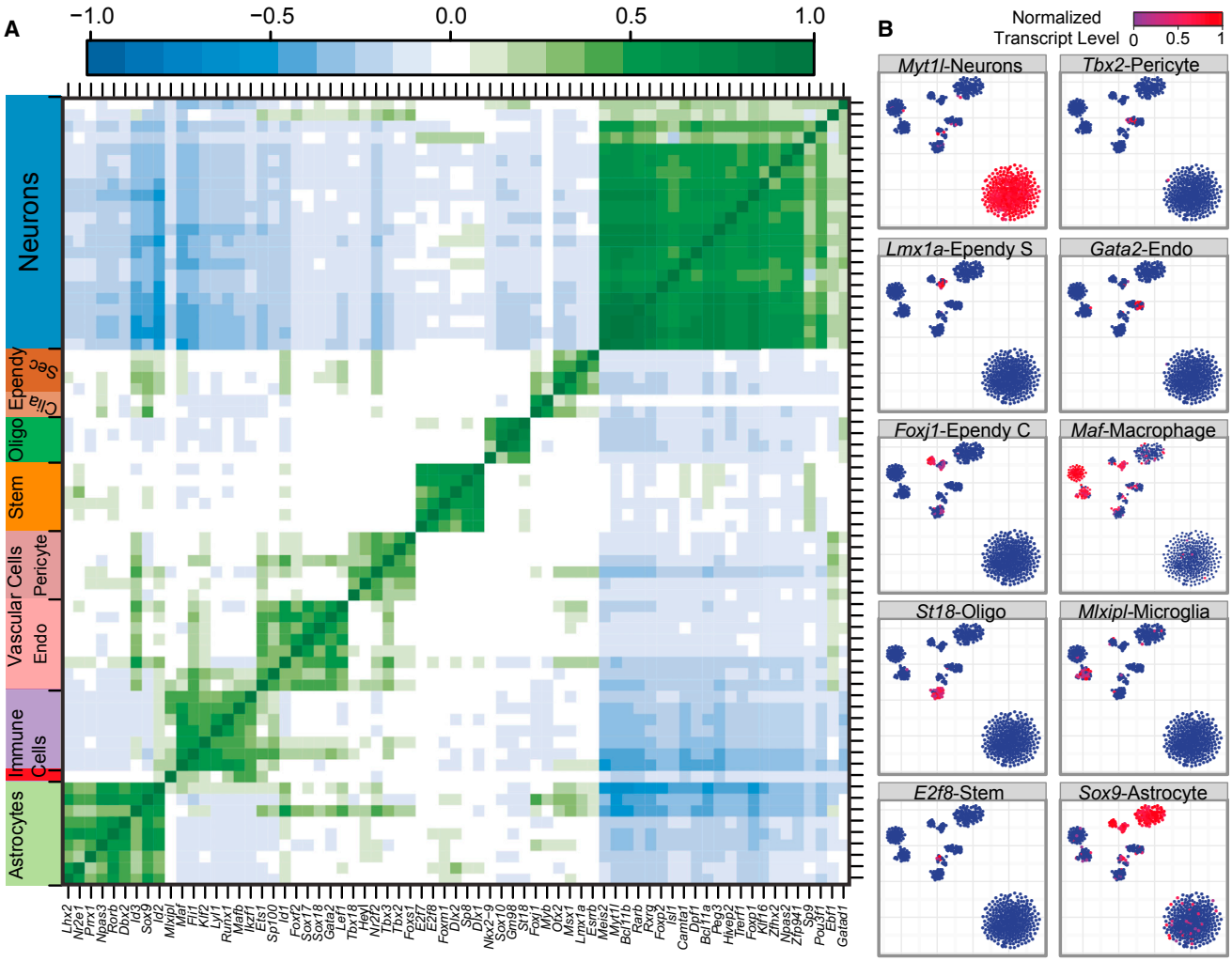


Figure 5. Cell Type-Specific TFs Identified by Single-Cell Transcriptome Analysis

(A) Correlogram visualizing correlation of single-cell gene expression between cell type-specific TFs.

(B) Scaled expression of the TFs most specific to discrete cell types. Scaled expression of marker genes is shown by the color of the cell points.

expressing *Cx3cr1*, *Olfml3*, and *Mlxipl* and macrophages expressing *Mrc1* and *Dab2* (Figures 4D–4F).

Within oligodendrocytes, we found two large anticorrelated gene groups (Figure 4G). Projecting the cells onto PC1 revealed two discrete oligodendrocyte subtypes: a population of newly formed oligodendrocytes (NFOs) expressing *Nfasc* and a population of mature oligodendrocytes (MOs) expressing *Klk6* (Zhang et al., 2014). These two discrete populations were connected by cells that coexpressed both NFO and MO genes and may represent a transitioning population (Figures 4H and 4I).

The pairwise correlation analysis did not identify two clearly anticorrelated gene groups for astrocytes (Figure 4J); instead, whole-transcriptome rPCA revealed a continuous transcriptional gradient defined by a unimodal population distribution (Figures 4K and 4L). The genes defining this gradient were related on one side to synaptic communication, such as neurotransmitter transporters (*Slc6a11*, *Slc6a1*, and *Slc6a9*) and glutamate receptors (*Gria1* and *Gria2*), and on the other side to translation, like ribosomal proteins (*Rpl9*, *Rpl14*, *Rps5*, and *Rpsa*) and cell polarity regulators (*Cdc42*) (Etienne-Manneville and Hall, 2003).

Distinct Groups of TFs Maintain Cell Type Identity

We used correlation analysis to identify TFs expressed specifically (Spearman correlation coefficient > 0.5) in single cells of a given type and visualized the pairwise correlation of the resulting 69 TFs in a correlogram (Figure 5A; Table S7). We found that striatal neurons correlated with the largest set of TFs (22) (Figure 1C), independent of the number of cells per subtype (Figure S5). The most specific neuronal TF, *Myt1l*, is able to induce neuronal cell identity in vitro when expressed in fibroblasts with two additional TFs (Vierbuchen et al., 2010). Generally, many of the cell type-specific TFs identified by our analysis have been used in TF-based direct conversion protocols. Among them are *Sox9* for astrocytes (Caiazzo et al., 2015), *Sox10* for oligodendrocytes (Yang et al., 2013), *Gata2* for endothelial cells (Elcheva et al., 2014), and *Maf* for macrophages (Hegde et al.,

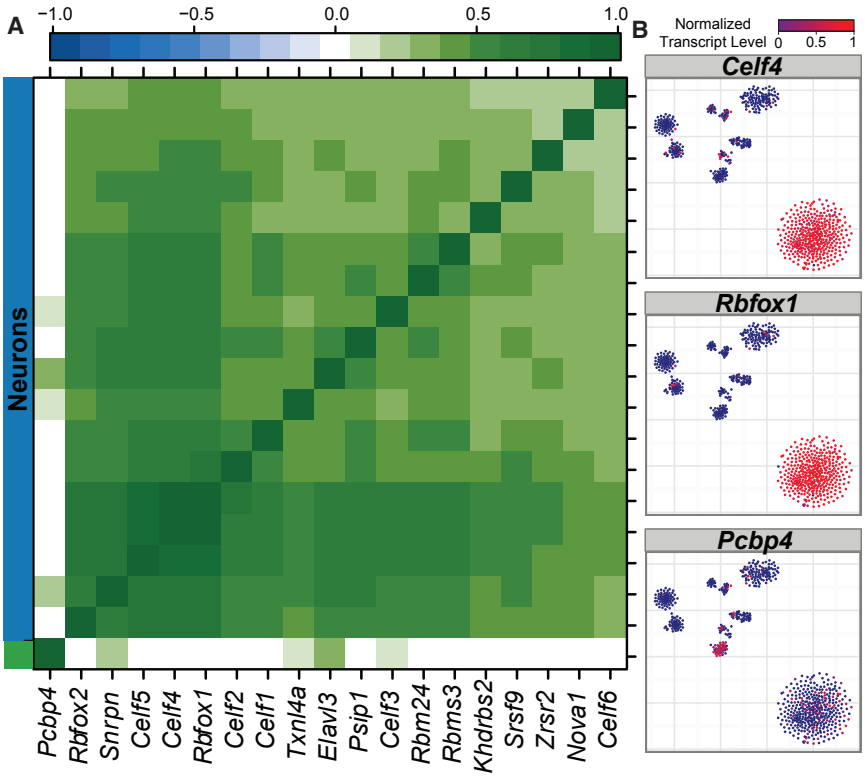


Figure 6. Single-Cell Transcriptome Analysis Reveals Splicing Factors Specific to Neurons

(A) Hierarchical clustering of pairwise gene correlations for cell type-specific splicing factors.

(B) tSNE plot of single cells colored by the scaled expression selected splicing factors: the neuronal-enriched splicing factors *Celf4* and *Rbfox1* and the oligodendrocyte-enriched splicing factor *Pcbp4*.

1999). Thus, we also provide candidate TFs for stem cell transformation into cell types for which there are no existing protocols (*Lmx1a* for secretory ependymal cells, *Foxj1* for ciliated ependymal cells, *Tbx2* for VSMCs, *E2f8* for NSCs, *Mlxipl* for microglia, *Isl1* for D1 MSNs, and *Sp9* for D2 MSNs (Figure 5A; Table S7).

While *Isl1* is known to be required for normal functioning of D1 MSNs (Ehrman et al., 2013), the role of *Sp9* in MSNs had not been described previously. Therefore, we analyzed the effect of *Sp9* overexpression on expression of MSN markers in primary striatal cultures. The overexpression of *Sp9* significantly lowered the expression of *Drd1a* without significantly altering the expression of D1 MSN markers *Tac1*, *Pdyn* or D2 MSN marker *Drd2*. These results suggest that *Sp9* overexpression may disrupt the normal development of D1 MSNs (Figure S7).

Cell Type-Specific Regulation of RNA Splicing

Alternative splicing is a key mechanism for generating transcriptomic diversity. We used correlation analysis (Supplemental Experimental Procedures) to explore cell type-specific expression of RNA splicing factors. We found 18 splicing factors that were specific to neurons and only 1 splicing factor (*Pcbp4*) that correlated with oligodendrocytes (Figures 6A and 6B). No splicing factors were found with high correlation to any other cell types. The factors specific to striatal neurons include all six members of the CUG-BP, Elav-like family (*Celf*), which are associated with autism, epilepsy, and Alzheimer's disease (Ladd, 2013), as well as all three members of the RNA-binding Fox family of splicing factors: *Rbfox1* (also known as Fox-1 or A2BP1), *Rbfox2* (also known as Fox-2 or RBM9), and *Rbfox3* (also known as Fox-3, HRNBP3, or NeuN). The RNA-binding Fox family of splicing factors has been shown to maintain mature neuronal physiology and is linked to autism (Gehman et al., 2012). *Khdrbs2* (also called SLM1), which may contribute to alternative splicing of neurexins, was also enriched in neurons (Iijima et al., 2014). Many other neuron-enriched splicing factors have not yet been well characterized for their neuronal roles (*Snrpn*, *Txnl4a*, *Psip1*, *Rbm24*, and *Rbm3*). In summary, analysis of cell type-specific splicing factors showed that the regulatory diversity of splicing is higher in neurons than in other cell types in the striatum.

To complement this analysis, we identified 45,843 sites of alternative splicing by the reads mapping to exon junctions at splice sites. Using Fisher's exact test, we defined splice junctions that are differentially spliced ($p < 10^{-5}$) in one or more cell types (Figures 7A and S7A; Table S8). These splice sites included genes such as *Csde1*, whose splicing is dysregulated in schizophrenia and for which a de novo loss-of-function mutation causes autism (Sanders et al., 2012); *Hnrnpk* (heterogeneous nuclear ribonucleoprotein K), which is associated with intellectual disability (Au et al., 2015); and *Hspa5*, which is linked to bipolar disorders and schizophrenia (Kakiuchi et al., 2003). The visualization of cell type-specific alternative splicing revealed many single cells that expressed both the major and the minor variants, a phenomenon we termed compound splicing (Figures 7A and S7A–S7D; Table S8).

It had been previously reported that single immune cells exclusively express either the major or the minor splice variant (Shalek et al., 2013). However, we had observed that individual neurons were capable of expressing both splice versions of neurexins using single-cell qPCR (Fuccillo et al., 2015). Here, we performed a global analysis to quantify and compare genes with compound splice sites by cell type (Figure 7B). We detected few compound splice sites in immune cells of the striatum but higher numbers in single neurons, as well as in NSCs, oligodendrocytes, and secretory ependymal cells. There was no relationship between sequencing depth and number of compound splice sites detected per cell, ruling out systematic differences in sequencing depth as an explanation (Figure S7B). Moreover, 3′–5′ RNA coverage bias was similar across all cell types and between bulk controls and single cells (Figures S1H and S1I).

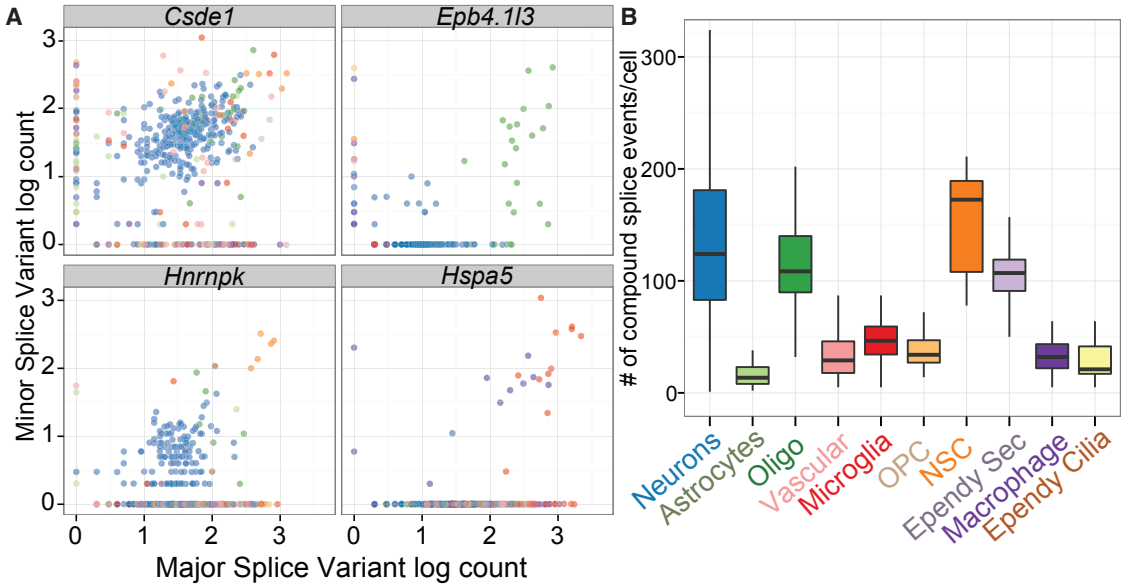


Figure 7. Differential Splicing Analysis Reveals Several Modes of Single-Cell Splicing Regulation

(A) Expression of selected splice sites that have significant cell type-specific regulation. The number of reads per cell (log_{10}) aligning to the major and minor variants is plotted on the x axis and y axis, respectively. Many single cells are located on the diagonal of the plots, indicating they express both the major and the minor variants of that splice site (compound splice sites).

(B) Boxplots showing the total number of compound splice sites detected per cell across cell types.

DISCUSSION

Previous studies on the heterogeneity of striatal MSNs provided various classification schemes, including the classical distinction between D1- and D2 MSN subtypes based on expression of the D1-dopamine receptor versus the D2-dopamine receptor (Kreitzer and Malenka, 2008), the anatomical location of MSNs in dorsal striatum or nucleus accumbens (O'Doherty et al., 2004), and the position of a MSN within a striosome or matrix microzone (Crittenden and Graybiel, 2011).

We analyzed MSNs using scRNA-seq and found that D1 MSNs could be subdivided into two discrete populations: *Pchd8*-D1s and *Foxp1*-D1s. The *Foxp1*-D1 population could be further subdivided into *Foxo1*-high or *Dner*-high populations, although it was not clear how discrete the *Foxo1-Dner* division is. Similarly, we could subdivide D2 MSNs into two discrete subpopulations: *Htr7*-D2s and *Synpr*-D2s. The *Synpr*-D2 neurons could be further subdivided into *Calb1*-high or *Cartpt*-high subpopulations, although as with *Dner*-D1s, it was not clear how discrete this division is. Two marker genes for *Pchd8*-D1 and *Htr7*-D2 subtypes, *Htr7* and *Tacr1*, are linked to alcohol dependence (Blaine et al., 2013; Zlojutro et al., 2011), and repeated cocaine exposure regulates epigenetic control of *Pcdh8*-MSN marker *4932411E22Rik* (Feng et al., 2015). The *Pchd8*-MSN subtype coexpresses *Drd1a* and *Drd2* and therefore may possess atypical signaling and distinct properties. Previous studies also suggested the possible existence of a third MSN population that coexpresses *Drd1a* and *Drd2*. However, none of these studies provided specific markers for this population or resolved them into distinct subtypes (Ade et al., 2011; Bertran-Gonzalez et al., 2010; Frederick et al., 2015; Surmeier et al., 1998). Important results of our work are that we have conclusively proved the existence of a MSN subtype that coexpress several D1- and D2-specific genes and shown that there are two distinct populations, each with specific markers. Given the involvement of MSNs in cognitive disorders, novel and discrete MSN subtypes expressing neuronal plasticity and addiction-related genes are likely to have important functions, which can be characterized in future studies using genetic tools based on the markers identified here.

Within the discrete MSN subtypes, we identified a large degree of heterogeneity that did not further separate the cells into clear subtypes. We termed this a continuous transcriptional gradient, because the neurons had a range of expression of the opposing gene sets and a roughly unimodal expression distribution. The continua within all main subtypes shared several genes (particularly *Cnr1*, *Crym*, and *Wfs1*), suggesting a common origin of these gradients. Previous scRNA-seq analyses have largely assumed that the identity of post-mitotic adult neurons is discrete: every neuron can be assigned a single subtype (Darmanis et al., 2015; Tasic et al., 2016; Usoskin et al., 2015; Zeisel et al., 2015). Here, we find that there are at least two distinct aspects to neuronal identity: discrete subtypes, in which two or more groups of neurons are clearly separated with few intermediates, and continuous gradients, in which neurons within a subtype lie on a spectrum of gene expression, with most having a range of intermediate expression values of two gene sets.

We found continuous transcriptional gradients within glial cells as well. Astrocytes have been reported to be a heterogeneous cell type (Hochstim et al., 2008; Matthias et al., 2003). We discovered a continuum (Figures 4J–4L) wherein genes for neurotransmitter transporters and receptors are anticorrelated to genes

coding for ribosomal proteins and the cell polarity regulator *Cdc42*. This may therefore represent a continuum of astrocyte states, ranging from more actively involved in synaptic communication (Matthias et al., 2003) to more inactive or more developmentally immature (Etienne-Manneville and Hall, 2003). We applied our analysis approach to other striatal cell types identified by tSNE. Pairwise correlation analysis of the vascular cells, the immune cells, and the oligodendrocytes identified subtypes and large set of subtype-specific markers (Figures 4A–4L).

We also identified cell type-specific TFs (Figures 5A and 5B; Table S7), which provides a mechanistic explanation for the maintenance of discrete cell type identities. The most cell type-specific TFs we found were often previously observed to fuel direct conversion of cells into that cell type (Caiazzo et al., 2015; Elcheva et al., 2014; Hegde et al., 1999; Pang et al., 2011; Vierbuchen et al., 2010; Yang et al., 2013). We found that the D2 MSN-specific TF identified here first, *Sp9*, disrupted D1 MSN-specific gene expression, confirming the functional relevance of these cell type-specific TFs (Figure S6).

Finally, we observed that neurons contribute to the brain's uniquely high-level alternative splicing more than any other cell type (Grosso et al., 2008; Yeo et al., 2005). We found that neurons express more specific splicing factors and that the overall complexity of alternative splicing is higher in neurons than in other cell types. For many sites of alternative splicing, we detected both variants in each neuron (compound splice sites), revealing an alternative splicing machinery that can increase the diversity of the transcriptome by altering alternative splicing ratios.

Our results show that the phenotypic diversity of striatal neurons arises from a small number of discrete subtypes, within which neurons lie on a continuum of gene expression states. Our analysis methods distinguish between discrete subtypes (with transitioning intermediates when the cells are actively differentiating) and continuous heterogeneity. These distinctions may prove fundamental to understanding the phenotypic diversity of the cells that form complex tissues.

EXPERIMENTAL PROCEDURES

Animals

All procedures conformed to the NIH Guidelines for the Care and Use of Laboratory Animals and were approved by the Stanford University Administrative Panel on Laboratory Animal Care.

Single-Cell Transcriptional Profiling

Acute brain slices were cut from the 5- to 7-week-old male mice and, after papain treatment, dissociated mechanically (Supplemental Experimental Procedures). Live cells were purified by either magnetic bead-activated cell sorting (MACS; Miltenyi) or FACS. For cell type-specific isolation, genetically labeled MSN subtypes D1 MSN and D2 MSN and astrocytes were purified by FACS. Single cells were captured on a microfluidic chip on the C1 system (Fluidigm), and whole-transcriptome-amplified cDNA was prepared on chip using the SMARTer Ultra Low RNA Kit for Illumina (Clontech Laboratories). For the smart-seq2 protocol, three MSN populations (D1 MSN tdTom+, D1 MSN GFP+, and tdTom+/GFP+ MSNs) were sorted individually into 96-well plates with lysis buffer, spun down, frozen at −80°C, and amplified using the protocol described previously (Picelli et al., 2013). Single-cell libraries were constructed as described previously (Supplemental Experimental Procedures) (Treutlein et al., 2014a).

Immunohistochemistry

Vibratome sections from perfusion-fixed mouse brains were stained with antibody for GFP (Rockland) and immunofluorescence imaging done on an Olympus fluorescent microscope or Nikon confocal microscope.

ACCESSION NUMBERS

The accession number for the single-cell sequencing raw data reported in this paper is GEO: GSE82187.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.06.059.

AUTHOR CONTRIBUTIONS

O.G., G.M.S., B.T., T.C.S., and S.R.Q. designed the study. O.G., G.M.S., B.T., and N.F.N. performed the experiments. O.G., G.M.S., and B.T. analyzed datasets. T.C.S. and S.R.Q. provided intellectual guidance in the interpretation of the data. P.E.R., R.C.M., and M.V.F. provided reagents. O.G., G.M.S., B.T., G.J.C., T.C.S., and S.R.Q. wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank all members of the S.R.Q. and T.C.S. laboratories for helpful discussions and Ben Barres for providing *Aldh1l1*-GFP mice. This work was supported by grants from the NIDA (K99DA038112 to O.G.), the NIH (R37MH52804 to T.C.S.), and the Brain and Behavior Research Foundation (to O.G.). Authors disclose the following: Stephen R. Quake is a founder, consultant, and shareholder of Fluidigm Corporation.

Received: November 6, 2015
Revised: May 13, 2016
Accepted: June 11, 2016
Published: July 14, 2016

REFERENCES

Ade, K.K., Wan, Y., Chen, M., Gloss, B., and Calakos, N. (2011). An improved BAC transgenic fluorescent reporter line for sensitive and specific identification of striatonigral medium spiny neurons. Front. Syst. Neurosci. *5*, 32.

Aguirre, A., Rubio, M.E., and Gallo, V. (2010). Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal. Nature *467*, 323–327.

Amiral, J., Wolf, M., Fischer, A., Boyer-Neumann, C., Vissac, A., and Meyer, D. (1996). Pathogenicity of IgA and/or IgM antibodies to heparin-PF4 complexes in patients with heparin-induced thrombocytopenia. Br. J. Haematol. *92*, 954–959.

Au, P.Y.B., You, J., Caluseriu, O., Schwartzentruber, J., Majewski, J., Bernier, F.P., Ferguson, M., Valle, D., Parboosingh, J.S., Sobreira, N., et al. (2015). GeneMatcher aids in the identification of a new malformation syndrome with intellectual disability, unique facial dysmorphisms, and skeletal and connective tissue abnormalities caused by de novo variants in HNRNPK. Hum. Mutat. *36*, 1009–1014.

Bertran-Gonzalez, J., Hervé, D., Girault, J.A., and Valjent, E. (2010). What is the degree of segregation between striatonigral and striatopallidal projections? Front. Neuroanat. *4*, 1–9.

Blaine, S., Claus, E., Harlaar, N., and Hutchison, K. (2013). TACR1 genotypes predict fMRI response to alcohol cues and level of alcohol dependence. Alcohol. Clin. Exp. Res. *37* (*Suppl 1*), E125–E130.

Caiazzo, M., Giannelli, S., Valente, P., Lignani, G., Carissimo, A., Sessa, A., Colasante, G., Bartolomeo, R., Massimino, L., Ferroni, S., et al. (2015). Direct

conversion of fibroblasts into functional astrocytes by defined transcription factors. Stem Cell Reports *4*, 25–36.

Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V., and Di Filippo, M. (2014). Direct and indirect pathways of basal ganglia: a critical reappraisal. Nat. Neurosci. *17*, 1022–1030.

Crittenden, J.R., and Graybiel, A.M. (2011). Basal ganglia disorders associated with imbalances in the striatal striosome and matrix compartments. Front. Neuroanat. *5*, 59.

Cui, G., Jun, S.B., Jin, X., Pham, M.D., Vogel, S.S., Lovinger, D.M., and Costa, R.M. (2013). Concurrent activation of striatal direct and indirect pathways during action initiation. Nature *494*, 238–242.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. USA *112*, 7285–7290.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al.; Consortium (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

DeLong, M., and Wichmann, T. (2009). Update on models of basal ganglia function and dysfunction. Parkinsonism Relat. Disord. *15* (*Suppl 3*), S237–S240.

Doetsch, F., Caillé, I., Lim, D.A., García-Verdugo, J.M., and Alvarez-Buylla, A. (1999). Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. Cell *97*, 703–716.

Ehrman, L.A., Mu, X., Waclaw, R.R., Yoshida, Y., Vorhees, C.V., Klein, W.H., and Campbell, K. (2013). The LIM homeobox gene Isl1 is required for the correct development of the striatonigral pathway in the mouse. Proc. Natl. Acad. Sci. USA *110*, E4026–E4035.

Elcheva, I., Brok-Volchanskaya, V., Kumar, A., Liu, P., Lee, J.-H., Tong, L., Vodyanik, M., Swanson, S., Stewart, R., Kyba, M., et al. (2014). Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators. Nat. Commun. *5*, 4372.

Etienne-Manneville, S., and Hall, A. (2003). Cdc42 regulates GSK-3beta and adenomatous polyposis coli to control cell polarity. Nature *421*, 753–756.

Feng, J., Shao, N., Szulwach, K.E., Vialou, V., Huynh, J., Zhong, C., Le, T., Ferguson, D., Cahill, M.E., Li, Y., et al. (2015). Role of Tet1 and 5-hydroxymethylcytosine in cocaine action. Nat. Neurosci. *18*, 536–544.

Frederick, A.L., Yano, H., Trifilieff, P., Vishwasrao, H.D., Biezonski, D., Mészáros, J., Urizar, E., Sibley, D.R., Kellendonk, C., Sonntag, K.C., et al. (2015). Evidence against dopamine D1/D2 receptor heteromers. Mol. Psychiatry *20*, 1373–1385.

Fuccillo, M.V., Földy, C., Gökce, Ö., Rothwell, P.E., Sun, G.L., Malenka, R.C., and Südhof, T.C. (2015). Single-cell mRNA profiling reveals cell-type-specific expression of neurexin isoforms. Neuron *87*, 326–340.

Gehman, L.T., Meera, P., Stoilov, P., Shiue, L., O'Brien, J.E., Meisler, M.H., Ares, M., Jr., Otis, T.S., and Black, D.L. (2012). The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. Genes Dev. *26*, 445–460.

Grosso, A.R., Gomes, A.Q., Barbosa-Morais, N.L., Caldeira, S., Thorne, N.P., Grech, G., von Lindern, M., and Carmo-Fonseca, M. (2008). Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res. *36*, 4823–4832.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Hegde, S.P., Zhao, J., Ashmun, R.A., and Shapiro, L.H. (1999). c-Maf induces monocytic differentiation and apoptosis in bipotent myeloid progenitors. Blood *94*, 1578–1589.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suárez-Fariñas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*, 738–748.

Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). Nat. Neurosci. *7*, 483.

Heinzen, E.L., Arzimanoglou, A., Brashear, A., Clapcote, S.J., Gurrieri, F., Goldstein, D.B., Jóhannesson, S.H., Mikati, M.A., Neville, B., Nicole, S., et al.; ATP1A3 Working Group (2014). Distinct neurological disorders with ATP1A3 mutations. Lancet Neurol. *13*, 503–514.

Hochstim, C., Deneen, B., Lukaszewicz, A., Zhou, Q., and Anderson, D.J. (2008). Identification of positionally distinct astrocyte subtypes whose identities are specified by a homeodomain code. Cell *133*, 510–522.

Iijima, T., Iijima, Y., Witte, H., and Scheiffele, P. (2014). Neuronal cell type-specific alternative splicing is regulated by the KH domain protein SLM1. J. Cell Biol. *204*, 331–342.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Kakiuchi, C., Iwamoto, K., Ishiwata, M., Bundo, M., Kasahara, T., Kusumi, I., Tsujita, T., Okazaki, Y., Nanko, S., Kunugi, H., et al. (2003). Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder. Nat. Genet. *35*, 171–175.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods *11*, 740–742.

Kreitzer, A.C., and Malenka, R.C. (2008). Striatal plasticity and basal ganglia circuit function. Neuron *60*, 543–554.

Kupchik, Y.M., Brown, R.M., Heinsbroek, J.A., Lobo, M.K., Schwartz, D.J., and Kalivas, P.W. (2015). Coding the direct/indirect pathways by D1 and D2 receptors is not valid for accumbens projections. Nat. Neurosci. *18*, 1230–1232.

Ladd, A.N. (2013). CUG-BP, Elav-like family (CELF)-mediated alternative splicing regulation in the brain during health and disease. Mol. Cell. Neurosci. *56*, 456–464.

Li, Q., Lee, J.-A., and Black, D.L. (2007). Neuronal regulation of alternative pre-mRNA splicing. Nat. Rev. Neurosci. *8*, 819–831.

Lobo, M.K., Karsten, S.L., Gray, M., Geschwind, D.H., and Yang, X.W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. Nat. Neurosci. *9*, 443–452.

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Maia, T.V., and Frank, M.J. (2011). From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. *14*, 154–162.

Matthias, K., Kirchhoff, F., Seifert, G., Hüttmann, K., Matyash, M., Kettenmann, H., and Steinhäuser, C. (2003). Segregated expression of AMPA-type glutamate receptors and glutamate transporters defines distinct astrocyte populations in the mouse hippocampus. J. Neurosci. *23*, 1750–1758.

Mühleisen, T.W., Leber, M., Schulze, T.G., Strohmaier, J., Degenhardt, F., Treutlein, J., Mattheisen, M., Forstner, A.J., Schumacher, J., Breuer, R., et al. (2014). Genome-wide association study reveals two new risk loci for bipolar disorder. Nat. Commun. *5*, 3339.

Nelson, A.B., and Kreitzer, A.C. (2014). Reassessing models of basal ganglia function and dysfunction. Annu. Rev. Neurosci. *37*, 117–135.

Noureddine, M.A., Qin, X.J., Oliveira, S.A., Skelly, T.J., van der Walt, J., Hauser, M.A., Pericak-Vance, M.A., Vance, J.M., and Li, Y.J. (2005). Association between the neuron-specific RNA-binding protein ELAVL4 and Parkinson disease. Hum. Genet. *117*, 27–33.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452–454.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Südhof, T.C., and Wernig, M. (2011). Induction of human neuronal cells by defined transcription factors. Nature *476*, 220–223.

Pannasch, U., Freche, D., Dallérac, G., Ghézali, G., Escartin, C., Ezan, P., Cohen-Salmon, M., Benchenane, K., Abudara, V., Dufour, A., et al. (2014). Connexin 30 sets synaptic strength by controlling astroglial synapse invasion. Nat. Neurosci. *17*, 549–558.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat. Biotechnol. *32*, 1053–1058.

Redies, C., Hertel, N., and Hübner, C.A. (2012). Cadherins and neuropsychiatric disorders. Brain Res. *1470*, 130–144.

Robison, A.J., and Nestler, E.J. (2011). Transcriptional and epigenetic mechanisms of addiction. Nat. Rev. Neurosci. *12*, 623–637.

Rouach, N., Koulakoff, A., Abudara, V., Willecke, K., and Giaume, C. (2008). Astroglial metabolic networks sustain hippocampal synaptic transmission. Science *322*, 1551–1555.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

Satoh, J., Kino, Y., Asahina, N., Takitani, M., Miyoshi, J., Ishida, T., and Saito, Y. (2016). TMEM119 marks a subset of microglia in the human brain. Neuropathology *36*, 39–49.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Shuen, J.A., Chen, M., Gloss, B., and Calakos, N. (2008). Drd1a-tdTomato BAC transgenic mice for simultaneous visualization of medium spiny neurons in the direct and indirect pathways of the basal ganglia. J. Neurosci. *28*, 2681–2685.

Steinhoff, M.S., von Mentzer, B., Geppetti, P., Pothoulakis, C., and Bunnett, N.W. (2014). Tachykinins and their receptors: contributions to physiological control and the mechanisms of disease. Physiol. Rev. *94*, 265–301.

Südhof, T.C. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. Nature *455*, 903–911.

Surmeier, D.J., Yan, Z., and Song, W.J. (1998). Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. Adv. Pharmacol. *42*, 1020–1023.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346.

Todorov, V., and Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. J. Stat. Softw. *32*, 1–47.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014a). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Treutlein, B., Gokce, O., Quake, S.R., and Südhof, T.C. (2014b). Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. Proc. Natl. Acad. Sci. USA *111*, E1291–E1299.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P.V., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat. Neurosci. *18*, 145–153.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035–1041.

Yang, N., Zuchero, J.B., Ahlenius, H., Marro, S., Ng, Y.H., Vierbuchen, T., Hawkins, J.S., Geissler, R., Barres, B.A., and Wernig, M. (2013). Generation of oligodendroglial cells by direct lineage conversion. Nat. Biotechnol. *31*, 434–439.

Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T., and Burge, C.B. (2005). Identification and analysis of alternative splicing events conserved in human and mouse. Proc. Natl. Acad. Sci. USA *102*, 2850–2855.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure: cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947.

Zlojutro, M., Manz, N., Rangaswamy, M., Xuei, X., Flury-Wetherill, L., Koller, D., Bierut, L.J., Goate, A., Hesselbrock, V., Kuperman, S., et al. (2011). Genome-wide association study of theta band event-related oscillations identifies serotonin receptor gene HTR7 influencing risk of alcohol dependence. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. *156B*, 44–58.

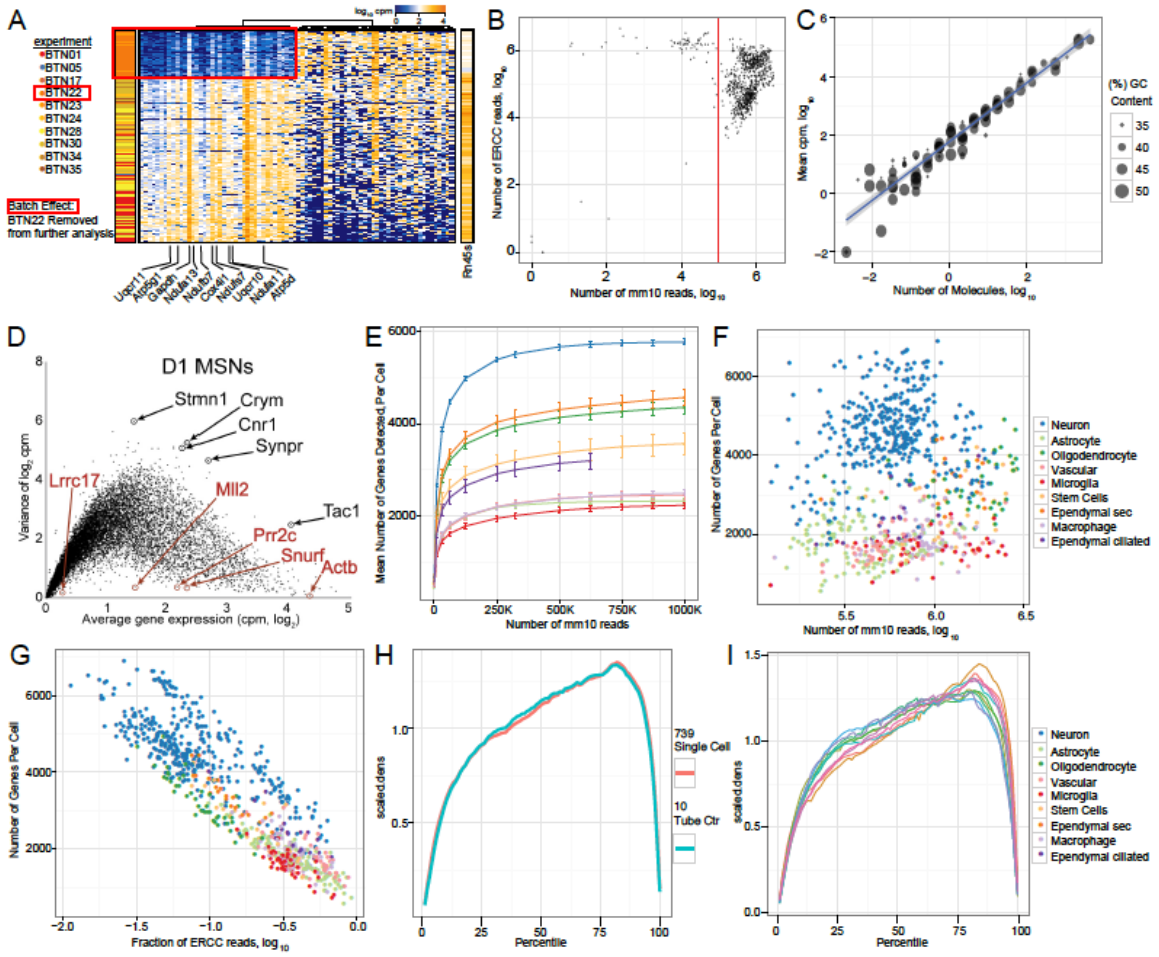
Cell Reports, Volume *16*

Supplemental Information

Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq

Ozgun Gokce, Geoffrey M. Stanley, Barbara Treutlein, Norma F. Neff, J. Gray Camp, Robert C. Malenka, Patrick E. Rothwell, Marc V. Fuccillo, Thomas C. Südhof, and Stephen R. Quake

Supp. Fig. 1 Quality Control and Characterization of Single-Cell RNA Sequencing, Related to Figure 1



A) Controlling for batch effect. We removed experiments from downstream rPCA only if most cells from a particular experiment were found as outliers by rPCA and the cells had dramatically reduced expression of housekeeping genes.

B) The number of mm10-aligned and ERCC-aligned reads per cell. Cells with less than 10^5 mm10-aligned reads were removed before any further analysis.

C) Mean detected expression levels ($\log_{10}$ mean cpm) for 92 ERCC RNA spike-ins as a function of the expected number of molecules per lysis reaction for all experiments and cell types. A linear fit is shown. The GC content of each

transcript is encoded as the size of the point, and no amplification bias towards low-GC transcripts is seen. Amplification is linear over 6 orders of magnitude, which is consistent with previous scRNAseq studies.

D) Noise profile of RNA-sequencing data for the D1 neurons. Variance of expression is plotted on the y-axis and mean expression on the x-axis. Housekeeping genes, labeled in red, are detected with very low cell-cell noise over a range of mean expression values, indicating the reliability of scRNAseq. Many highly variable genes shown in black are detected by rPCA as correlating to a continuous gradient within D1 neurons (Fig. 3 and Supp Fig. 5).

E) Saturation curves generated by subsampling the mm10 reads of cells to the depth on the x-axis and counting the number of genes detected with at least 10 reads. 500,000 mm10-aligning reads are sufficient to detect nearly all of the genes expressed by single cells, similar to previous reports (Treutlein et al., 2014). This establishes neurons as expressing by far the most genes of the sampled brain cell types. Each point represents the average and SEM of the number of genes per cell, where each cell is subsampled by randomly selecting reads from the pool of reads that aligned to the mm10 exome.

F) The number of genes per cell by the sequencing depth of each cell shows no relationship of genes detected to read depth for cells with between 3 x 10^5 and 3 x 10^6 aligning to the mm10.

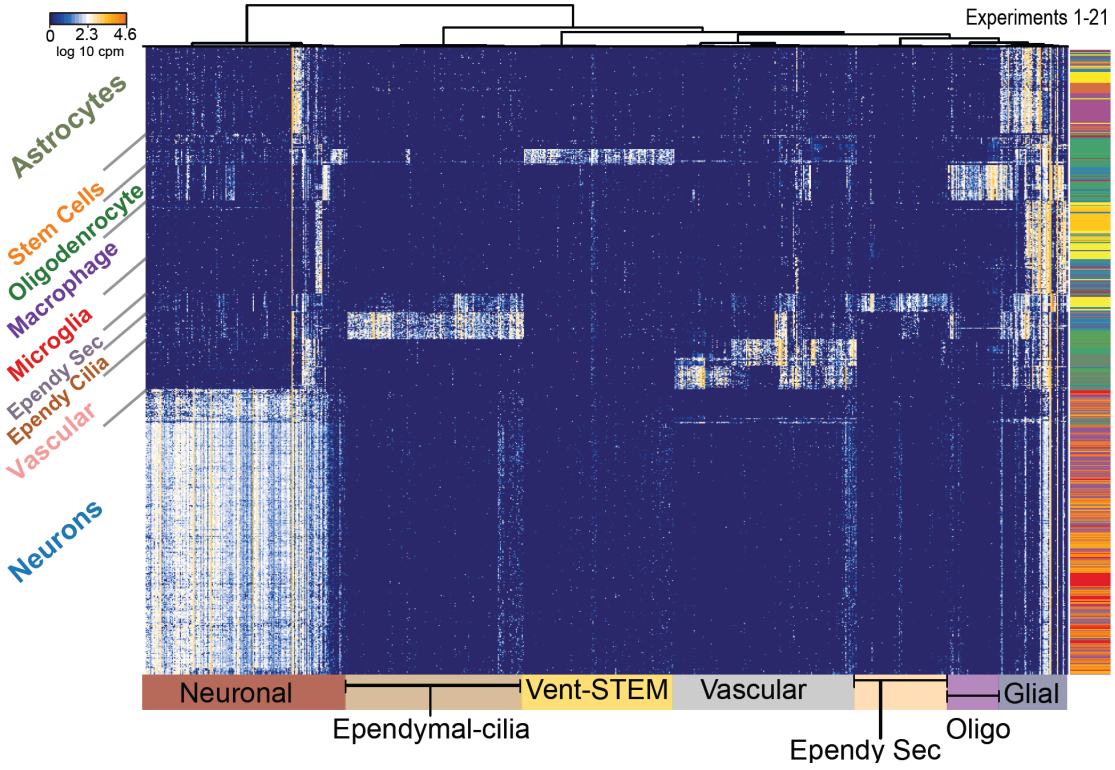
G) Number of genes per cell by the log fraction of reads aligning to the control ERCC sequences = $\log_{10} \frac{\#\text{ERCC reads}}{\#\text{ ERCC reads}+\#\text{mm10 reads}}$. The log fraction ERCC represents the ratio of ERCC RNA to cellular mRNA in each cell. Because the

amount of spiked-in ERCC RNA was constant for every experiment, this corresponds to the amount of RNA in the cell. A linear fit returns an r^2 value of 0.80 for this relationship, indicating that 80% of the variance over all brain cell types of total genes expressed can be predicted by the quantity of mRNA in the cell.

H) Plot of normalized RNA-Seq coverage at each normalized transcript position from 5' (left) to 3' (right) for single cells and tube controls. Both the single cell and the population controls have similar level of 3' bias.

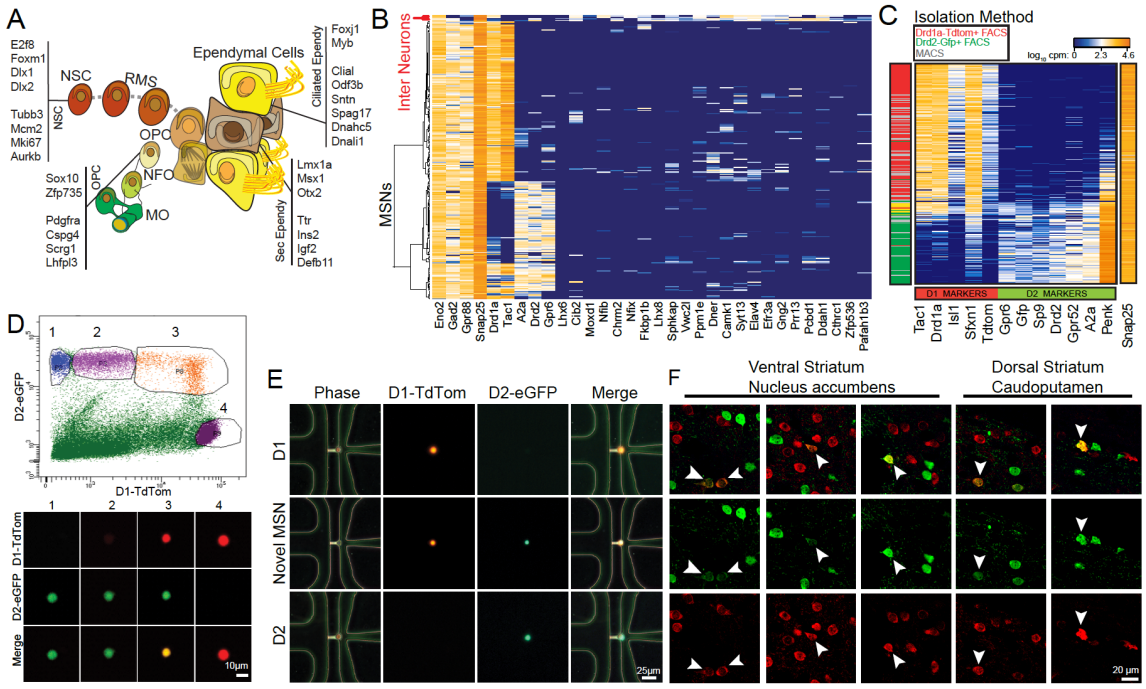
I) Plot of normalized RNA-Seq coverage at each normalized transcript position from 5' (left) to 3' (right) for single cells from different cell types. Regardless of cell type, single cells have similar level of 3' bias.

Supp. Fig. 2 PCA and Hierarchical Clustering of All Cells, Related to Figure 1



Cells were hierarchically clustered (HC) (Spearman correlation distance, Ward clustering) using the top 400 genes correlated to PC dimensions 1,2,3, and 4. Hierarchical clustering returned a similar number of cell types as did tSNE, and there was high degree of agreement in assigning cells to clusters between the two techniques. As with tSNE, D1 and D2 neurons were not separated.

Supp. Fig. 3 Ependymal and Stem Cell Types of the Ventricle and Rostral Migratory Stream, Hierarchical Clustering of Neurons with Known MSN and Interneuron Markers and Characterization of MSNs Subtypes, Related to Figure 2



A) The cells that reside at the ventricular wall of the striatum, a thin layer of cells separating the striatum from the ventricles, consists of ependymal cells and neuronal or oligodendrocyte progenitor cells. Transcription factors (TF, red) and non-TF marker genes (black) that exhibit restricted expression between cell types are shown.

B) Neurons are clustered by a panel of markers genes for neuron specific, MSNs specific and interneuron specific marker genes. We identify 5 neurons that likely belong to the striatal interneuron population (INs).

C) rPCA of RNAseq data from 330 single striatal neurons using the genes identified by pairwise correlation in Figure 2A. Cells (rows) are ordered by their projection onto PC1 and genes (columns) by their positive (left) or negative (right) contribution to PC1.

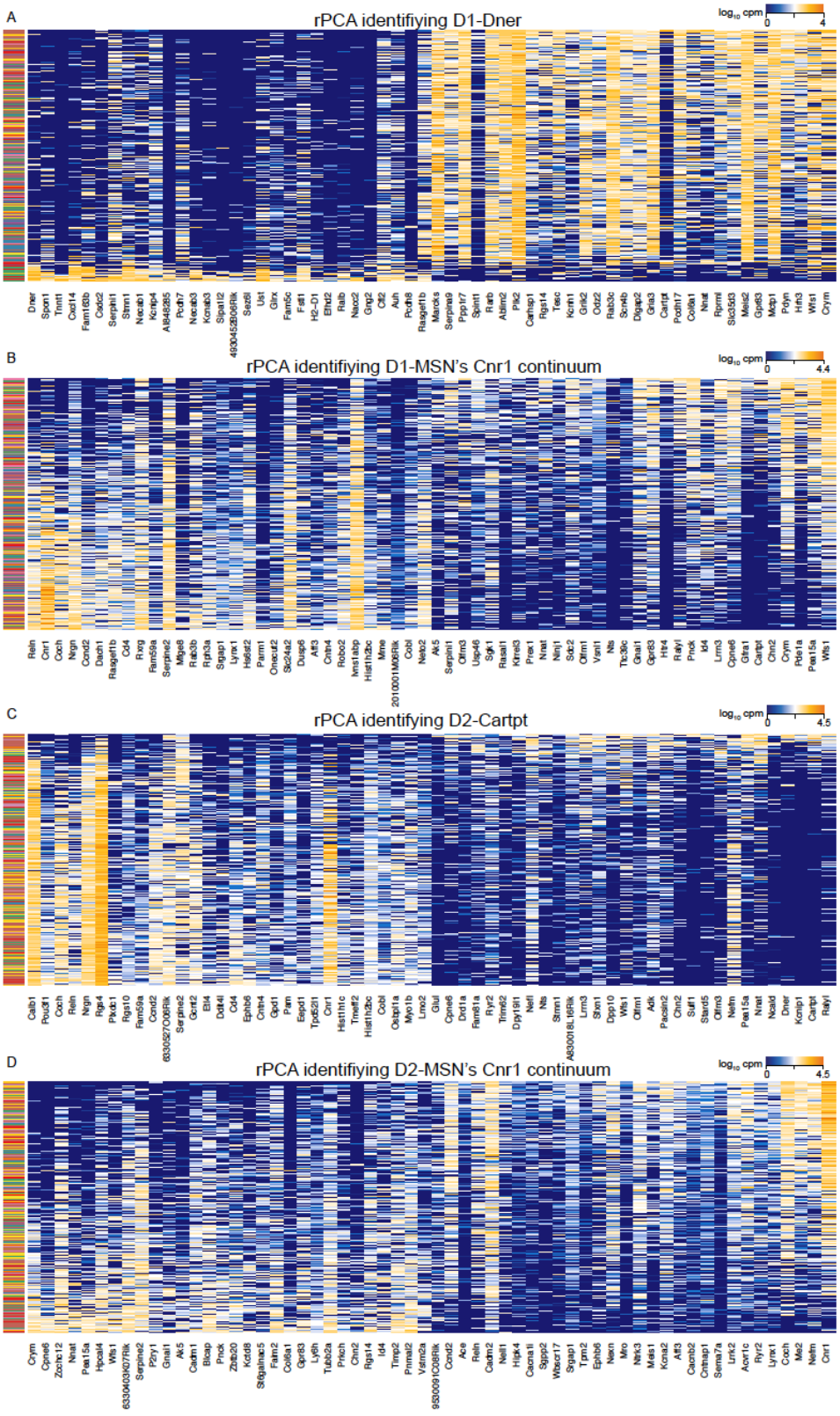
D) A single-cell suspension from Drd1a-Tdtom/Drd2-Gfp animals was prepared as described in the Supplementary Methods. Single cells were gated using forward and side scatter and live cells gated from the DAPI-low population.

A typical fluorescence scatter plot showing distinct clusters for D1-Tdtom, D2-Gfp and coexpressing cells. Each dot represents one event detected by the laser and the labeled populations were collected separately as indicated by the numbers. FACS purified populations imaged by confocal microscopy.

E) Single MSNs from D2-GFP/D1-TdTom double reporter mice captured in microfluidic chips with capture sites designed to trap single cells. Coexpressing MSNs were fluorescent for both GFP and TdTom.

F) Confocal imaging of striatal slices of D2-GFP/D1-TdTom double reporter mice demonstrating the existence of coexpressing MSNs in ventral striatum and dorsal striatum.

Supp. Fig. 4 rPCA clustering of MSN subtypes, Related to Figure 3



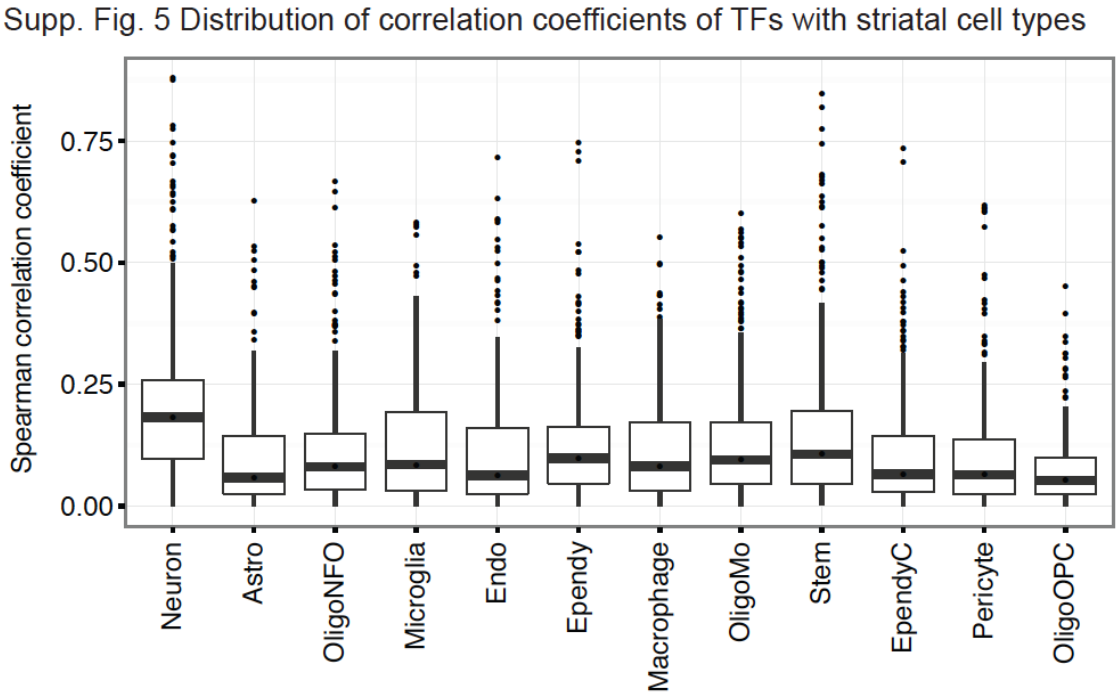
A) rPCA of D1 neurons as in main text Figure 3A with gene names.

B) rPCA of D1 neurons as in main text Figure 3D with gene names.

C) rPCA of D2 neurons as in main text Figure 3G with gene names.

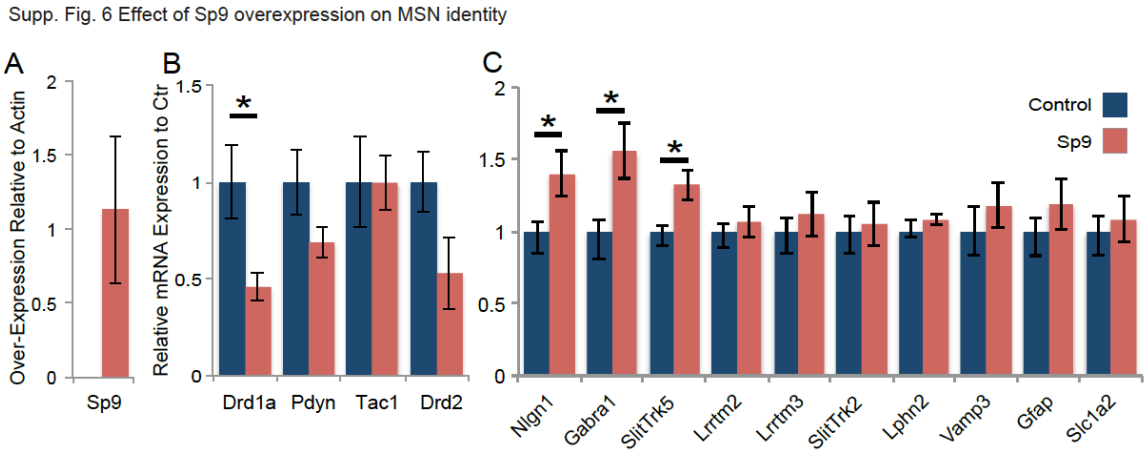
D) rPCA of D2 neurons as in main text Figure 3J with gene names.

Supp. Fig. 5 Distribution of correlation coefficients of TFs to major striatal cell types, Related to Figure 5



Cells of a given cell type were randomly subsampled 10 times down to 20 cells and Spearman's correlation coefficients of all TFs with cells grouped by cell type were calculated each time. Mean correlation coefficients were calculated for each TF per cell type and the distributions are shown as box plot with the vertical thick line denoting the mean.

Supp. Fig. 6 Effects of Sp9 overexpression on MSN identity, Related to Figure 5

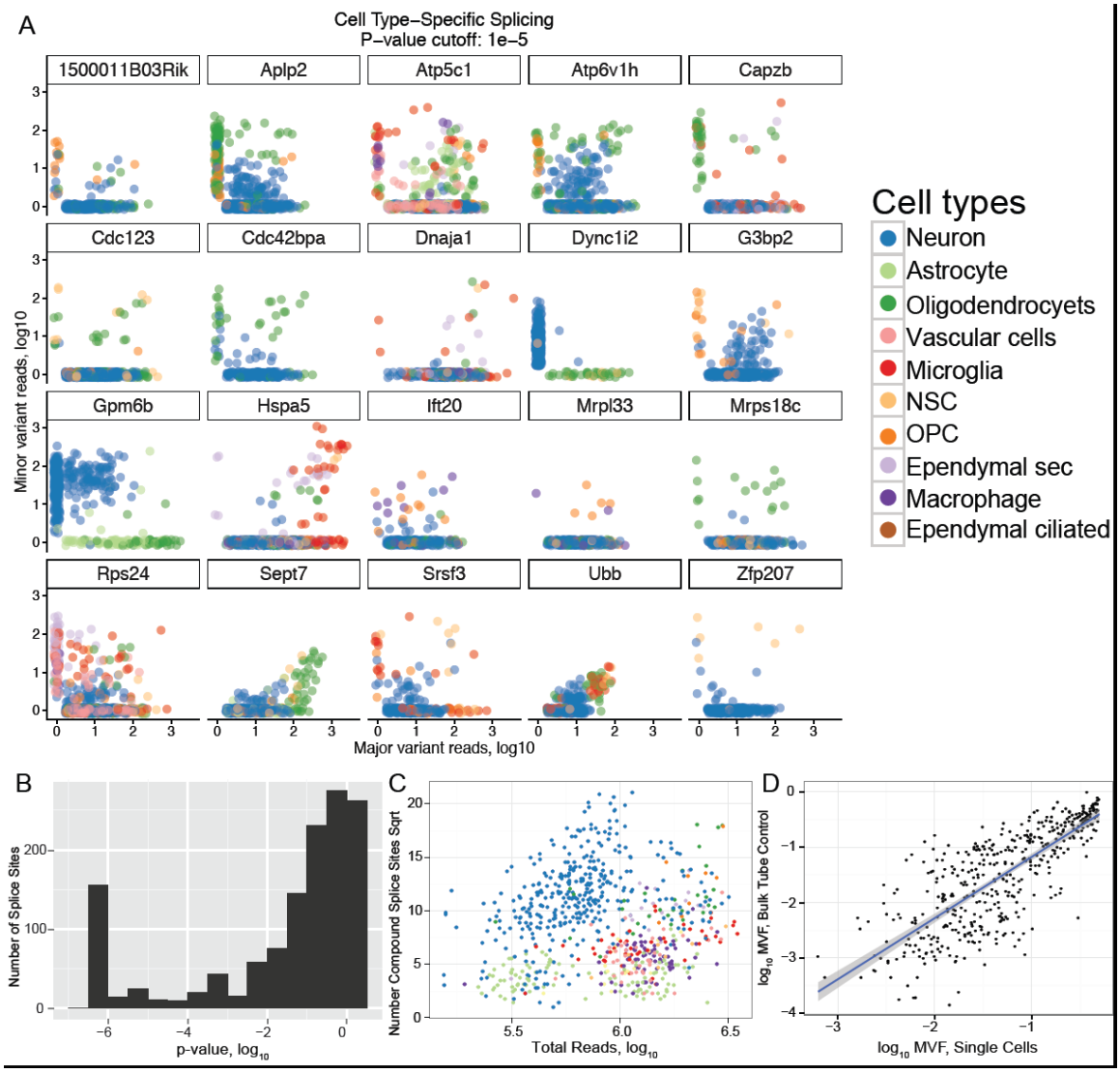


A) Q-PCR detection of the lentiviral *Sp9* overexpression in mouse striatal primary neurons shown relative to β-actin.

B) Lentiviral *Sp9* overexpression effect on D1-MSN and D2-MSN markers. Drd1a expression was significantly reduced in the *Sp9* expressing cultures.

C) Lentiviral *Sp9* overexpression effect on mRNA expression of neuronal and glial markers. Among tested targets, *Nlgn1*, *Gabra1* and *SlitTrk5* showed significant increase in the *Sp9* expressing cultures. mRNA levels were determined by real-time RT–PCR as described in the Experimental section. Values are means ±S.D. ($n \geq 4$). Statistical significance was determined using a Student's *t* test for unpaired data. *$P<0.05$.

Supp. Fig. 7 Cell-type Specific Splicing at Single Cell Resolution and Statistics and Coverage of Compound Splicing, Related to Figure 7



A) Visualization of all splice sites that were significantly alternatively spliced with a Fisher Exact Test p-value $<10^{-5}$, with log10 counts of the major and minor variants on the x- and y-axes, respectively. Each cell is represented as a point colored by cell type as described in Supplementary Methods. All spliced sites included here have examples of single cells with compound splicing.

B) The distribution of p-values from the Fisher Exact Test (the bump near 10^{-6} is a result of the p-values being "simulated" to reduce computational expense).

C) The number of compound splice sites versus the total number of mm10-aligned reads per cell. There no relationship between coverage depth and number of compound splice sites detected for between 3 x 10^5 and 3 x 10^6 reads per cell.

D) Log_{10} minor variant frequency (MVF) of each splice site used in the Fisher Exact test. The value of the x-axis was generated by summing over all the junction counts for every single cell sample, and the y-axis by summing over the junction counts of every bulk RNA tube control, and $\log_{10}$ MVF calculated as $\log_{10} \frac{1+c_{miv}}{1+c_{miv}+c_{mav}}$, where c_{miv} and c_{mav} are the number of counts of the minor and major variant, respectively. The MVF of the splice sites we analyzed in single cells correlates highly with bulk RNAseq results.

Supplementary Table 1. List of all 27 experiments, Related to Figure 1

The experimental method and number cells of each cell type are shown for 27 experiments.

Supplementary Table 2. Top 50 Cell-Type Specific Genes, Related to Figure 1

Cell type-specific genes were generated as describe in the Supplementary Methods. The p-value is the unadjusted p-value of Spearman correlation to the corresponding cell type.

Supplementary Table 3. Gene Ontology (GO) Enrichment of Striatal Cell Type-Specific Genes, Related to Figure 1

Each set of 50 cell type-specific genes was analyzed with default settings for biological process, cellular component, and molecular function at geneontology.org [ref]. Bonferroni-adjusted p-values of enrichment are shown in the last column. The default mouse gene background was used.

Supplementary Table 4. Summary of Gene Expression by Cell Type, Related to Figure 1

Mean, variance and fano factor (variance normalized to the mean) was calculated for each gene across the cells of a given type. Further, we calculated for each gene the fraction of cells of a given cell type that expressed the gene higher than $\log_2(\text{TPM}) = 1$ ("fraction expressed"). This table can be used to examine the specificity of genes to be expressed in a given cell type in the striatum. Most cell-type specific genes have a fraction expressed of almost 1 for one cell type and of nearly 0 for all other cell types.

Supplementary Table 5. Genes correlating specifically with MSN subtypes, Related to Figure 2 and 3

MSN type-specific genes were generated as describe in the Supplementary Methods. The p-value is the unadjusted p-value of Spearman correlation to the corresponding cell type.

Supplementary Table 6. TFs Expressed Specifically in Single Cells of a Given Type, Related to Figure 5

List of cell type specific TFs identified by correlation analysis (Spearman correlation coefficient >0.5) (see Supplemental Experimental Procedures).

Supplementary Table 7. Differentially expressed Splice sites, Related to Figure 7

The list of alternative spliced sites that are differentially spliced ($p<10^{-5}$) in one or more of the cell types and their start site. For each splice site, the number of single cells expressing only major or only minor or both versions of the splice variants shown.

Supplementary Methods

Animals

Generation of D1-tdTom/D2-GFP, D1-tomato, and Aldhl1-GFP BAC reporter mice was previously described (Heintz, 2004; Shuen et al., 2008). All lines were maintained by backcrossing to C57Bl/6J. All other experiments were conducted on 5-7 week old male C57/Bl6 mice from Jackson Labs. Mice were weaned at ~21 days of age and housed in groups of 2-5 on a 12-hour light/dark cycle (lights on 0700-1900h), with free access to food. All procedures conformed to the National Institutes of Health *Guidelines for the Care and Use of Laboratory Animals* and were approved by the Stanford University Administrative Panel on Laboratory Animal Care.

Tissue Dissociation

Parasagittal slices (350 μm) containing striatum were prepared using standard procedures. Briefly, mice were anesthetized with isoflurane and decapitated, brains were quickly removed and placed in ice-cold low-sodium/high-sucrose

dissecting solution. Slices were cut by adhering the lateral surface of the brain to the stage of a Leica vibroslicer, and placed in ice-cold hibernate-A solution (Gibco, Themo Fisher). The striatum was dissected from each slice, and the tissue was dissociated using papain-based dissociation. Briefly, striatal tissue was incubated with frequent agitation at 30°C for 20 min 2mg/ml papain solution in Hibernate-A, followed by washing with hibernate-A including 1mg/ml Ovomucoid protease inhibitor (Sigma). The striatal slices were triturated briefly using fire polished glass Pasteur pipettes until a single-cell suspension was obtained. The suspension was cleaned from small debris by centrifuging for 3 min at 450g through a three-step density gradient of Percoll (Sigma), top layer containing small debris was discarded and remaining cells were pelleted for 5 min at 550g. The pellet was resuspended in 0.1 ml of hibernate A with DAPI (4',6-diamidino-2-phenylindole, 1:4000 dilution; Sigma) to label dead cells (Guez-barber et al., 2012). Striatal tissue was kept in ice-cold solution throughout the processing until loading onto the microfluidic chip, except during the 20 min dissociation with papain at 30°C.

Purification of Striatal Cells by MACS.

Live striatal cells were purified by MACS from cell debris and dead cells using the Dead Cell Removal Kit (Miltenyi Biotec) in accordance with the protocol provided by the vendor. Before loading, the single-cell suspension was passed through a 35-µm cell strainer (BD Biosciences).

Purification of Striatal Cells by FACS.

The dissociated cell suspension from adult D1-tdTom/D2-GFP or Aldhl1-GFP mouse was passed through a 70-µm cell strainer (BD Biosciences). Viable cells (DAPI negative) were sorted by FACS (Aria II; BD Biosciences). For specific enrichment of genetically labeled cell types, wild type and fluorescent striatal cells were analyzed to determine optimal parameters and gating for FACS. Finally, test sorted cells were imaged on confocal microscope (Figure S3D). Cells were sorted into ice-cold hibernate-A and pelleted by centrifugation for 5 min at

300g. For the Smart-seq2 protocol (FACS-scRNAseq) (Picelli et al., 2013), Single cells were sorted into 96 well plates filled with 4 µl lysis buffer composed of 0.05% Triton X-100 (Sigma) and, ERCC spike-ins (1:24000000 dilution), 2.5 µM oligo-dT, 2.5 mM dNTP and 2U/µl of RNase inhibitor (Clontech) then spun down and frozen at -80°C. Plates were thawed and libraries prepared as described below.

Capturing of Single Cells and Preparation of cDNA.

Single striatal cells were captured on medium sized (10–17µm cell diameter) mRNA-seq or STA microfluidic chips (Fluidigm) using the Fluidigm C1 system. Cells were loaded onto the chip at a concentration of 500–700 cells/µl and imaged by phase-contrast and fluorescence (DAPI and fluorescent reporters for D1 and D2, or DAPI and fluorescent reporter for astrocyte) microscopy to assess the number and viability of cells per capture site. Only single, live cells were included in the analysis. Single cell cDNA was amplified on chip using the SMARTer Ultra Low RNA kit for Illumina (Clontech). ERCC (External RNA Controls Consortium) RNA spike-in Mix (Ambion, Life Technologies) was added to the lysis reaction and processed in parallel to cellular messenger RNA. Bulk RNAseq tube controls were prepared with 500-700 striatal cells per tube and lysed, reverse transcribed, and preamplified with the same reagents and thermal protocol as for the microfluidic chip (Fluidigm). Tube controls from MACS experiments had a variety of striatal cell types; tube controls from FACS experiments had primarily neurons or astrocytes.

For the Smart-seq2 protocol (FACS-scRNAseq), plates were incubated for 3min at 72°C and immediately placed on ice. To perform reverse transcription (RT) we added each well a mix of 0.59µl H_2O, 0.5µl SMARTScribe™ Reverse Transcriptase (Clontech), 2µl 5x First Strand buffer, 0.25µl Recombinant RNase Inhibitor (Clontech), 2µl Betaine (5M Sigma), 0.5µl DTT (100mM) 0.06µl MgCL2 (1M Sigma), 0.1µl Template-switching oligos (TSO) (100µM AAGCAGTGGTATCAACGCAGAGTACrGrG+G). Next reaction RT reaction mixes were incubated at 42°C for 90 min followed by 70°C for 5 min. Pre-

amplification of cDNA done by adding 12.5µl KAPA HiFi Hotstart 2x (KAPA Biosystems), 2.138µl H_2O, 0.25µl ISPCR primers (10 µM, 5′-AAGCAGTGGTATCAACGCAGAGT-3), 0.1125µl Lambda Exonuclease under the following conditions: 37°C for 30 min, 95°C for 3 min, 21 cycles of (98°C for 20 sec, 67°C for 15 sec, 72°C for 4 min), final extension at 72°C for 5 min. Fallowed by Ampure XP PCR Purification (Beckman Coulter) using the protocol supplied by Fluidigm.

RNA-seq Library Construction and DNA Sequencing.

The quality of single-cell cDNA was assessed on a capillary electrophoresis-based fragment analyser (Advanced Analytical) by analyzing size distribution and concentration. Illumina libraries were constructed in 96 or 384-well plates using the Illumina Nextera XT DNA Sample Preparation kit as described previously (Treutlein et al., 2014) using the protocol supplied by Fluidigm. Libraries were quantified by Agilent Bio-analyzer, using the High Sensitivity DNA analysis kit, and also fluorometrically using Qubit's DNA HS assay kits and a Qubit 2.0 Fluorometer (Invitrogen, LifeTechnologies). Single-cell Nextera XT (Illumina) libraries of each of the 27 independent experiments were pooled and sequenced 2x75 reads base pairs (bp) paired-end on Illumina HiSeq 2000 or NextSeq to a depth of $1x10^6$–$5x10^6$ reads/sample as previously described (Treutlein et al., 2014).

Processing, Analysis and Graphic Display of Single-cell RNA-seq Data.

Raw reads were aligned using rnaSTAR (Dobin and Gingeras, 2002) to the mm10 genome with Tdtomato and Gfp transgenic mRNA sequences and ERCC synthetic RNA added. Gene expression was quantified as the number of paired-end reads (counts) per gene divided by the total number of mm10 or ERCC counts, respectively, times 10^6 to get expression in counts per million (cpm). Values of $\log_{10}(\mathrm{cpm}+1)$ were used for analysis. See Supplementary Data for information about the number of mm10-aligned and ERCC-aligned reads per cell (Figure S1B). From a total of 820 cells captured by mic-scRNA-Seq, 704 single

cells had more than 10^5 reads aligning to the mouse genome and passed quality controls and therefore were included in the downstream analysis. 502 cells captured by FACS-scRNA-Seq passed similar quality control and were included in the analysis. Limit and linearity of detection of microfluidic scRNA-seq was determined by detected expression levels for ERCC RNA spike-ins as a function of the expected number of molecules (Figure S1C), which indicates similar performance to pervious report (Treutlein et al., 2014).

In order to confirm that our sequencing depth was sufficient to detect most genes expressed by single cells, saturation analysis was done by subsampling the mm10-aligned reads of all cells and counting at each depth the number of genes detected with at least 10 reads. This analysis showed that for all striatal cell types $5x10^5$ reads are sufficient to detect nearly all of the genes expressed (Figure S1C). Moreover, there was no correlation between the number of genes per cell and the sequencing depth of each cell (Figure S1E). Technical noise and biological variation in scRNA-seq were visualized in D1 cells by plotting the relationship between mean expression level and variance of expression of each gene (Figure S1D). Housekeeping genes are detected with very low cell-cell noise over a range of mean expression values, indicating the reliability of scRNAseq and low technical noise.

Identification of Major Cell Types by tSNE and Density-Based Clustering

Custom R scripts were used to identify the major cell types using tSNE. First, cell-cell distances were calculated using expression values weighted by the scde noise-reduction algorithm (scde::mode.fail.dist) (Kharchenko et al). We applied 2-dimension tSNE (van der Maaten L, 2008) using the Kharchenko cell-cell distances to map cells onto 2 dimensions for visualization and clustering. We used density-based clustering (DBSCAN) to cluster cells based on their proximity to a cluster of cells in tSNE space (cells are colored by cluster in Figure 1B). Each cell cluster was assigned an identity by its unique expression of one of 10 marker genes. DBSCAN also identified outliers that were not close to any cluster (21 cells), and those were removed from further analysis. Cell type-specific

genes were identified by calculating the Spearman correlation and p-value of all genes' expression values to each cluster. The top 50 correlated genes per cell cluster were selected by the lowest p-values and displayed on a heatmap. Cells were ordered by cluster (Figure 1E). GO analysis of cell type-specific genes was performed with default settings at geneontology.org with a cutoff of adjusted p-value < 0.05.

Pairwise Correlation Analysis to Identify Discrete Subtypes

Custom R functions were developed to identify subtypes of the major cell type. First, the pairwise Spearman correlations of all genes across a cell type were computed. Genes were selected if they had a correlation of greater than 0.6 to at least one other gene and less than -0.6 to at least one gene. Pairwise correlations of these genes were visualized using a heatmap with rows and columns ordered by hierarchical clustering (Euclidean distance, Ward method). We defined an "interconnected" set of genes as two clusters of genes where each cluster has at least 4 genes, each having a correlation of at least 0.6 to another gene in that cluster and an anticorrelation of at least -0.6 to a gene in the other cluster. If we identified an interconnected set of genes within that cell types, rPCA (rrcov::PcaHubert) reference was performed using those genes. Gene expression was visualized by heatmap where the cells (rows) were ordered by their projection onto the first principal component (PC1 score). The genes (columns) were ordered by their projection onto the first principal component (PC1 loading), such that the genes contributing the most to the principal component were on the far left (negative loadings) and far right (positive loadings). The distribution of gene expression across cells was visualized by plotting a histogram of PC1 scores. For discrete subtypes, the histograms were clearly bimodal. The two resulting subtypes were then analyzed separately using further rounds of pairwise correlation analysis and rPCA.

Robust PCA (rPCA) to Identify Subtypes and Continuous Transcriptional States

We performed rPCA on the cells using the whole transcriptome. Heatmaps of gene expression were plotted using the top genes by rPC loading. For figures 2 and 3, cells were scored by their expression of the top 30 genes by rPC loading as follows: each gene was scaled by dividing its expression by its maximum over all cells analyzed, returning values ranging from 0 to 1. For each cell, the scaled expression of the top 30 genes were summed (every gene weighted equally) and divided again by 30 to get a score in the range of 0 to 1. Heatmaps and density plots were based on the score of the positive-loading genes minus the score of the negative-loading genes.

Cell-type Specific Transcription and Splice Factors

To search for cell type specific transcriptional regulators, splice factors and genes of other gene categories, we defined a "perfect marker gene" for each unbiasedly identified cell type with a high transcript level ($\log_2$(TPM) = 10) in all cells of interest and no expression ($\log_2$(TPM) = 0) in all other cells. We then determined the pairwise Pearson correlation between the single cell expression profile of each perfect marker gene and every other transcribed gene of the transcriptional regulator or splice factor lists. The list of murine transcription factors that was screened was obtained from the online animal transcription factor database http://www.bioguo.org/AnimalTFDB/13 and the list for splicing factors generated from splice factor database http://geneontology.org and literature search.

Summary of Transcriptome Data by Cell Type

To summarize the expression of each gene per cell type, we calculated for all cells of a given cell type the mean $\log_2$(TPM), the variance of $\log_2$(TPM) and the Fano factor (variance divided by the mean). Further, we calculated for each gene the fraction of cells of a given cell type that expressed the gene higher than $\log_2$(TPM) = 1.

Primary striatal cultures, Virus Generation and Q-PCR analysis

Dissociated primary striatal cultures were prepared from ganglionic eminences of E16 mice embryos as described in (Gokce et al., 2009). Cultures were infected with *Sp9* expressing or empty lentivirus at DIV1. RNA was collected from striatal cultures at DIV 13 and analyzed for expression of D1- and D2-MSN markers by real-time PCR.

To generate the lentiviral vector, *Sp9* codon optimized for in-vitro synthesis then cloned into the FSW lentiviral vector by Gibson assembly (NewEngland Biolabs). Lentiviral particles were produced by calcium phosphate transfection of HEK293T cells (ATCC) as described previously (Gokce and Südhof). The effect of *Sp9* expression on mRNA levels analyzed using the Fluidigm Biomark dynamic array according to the manufacturer's protocol. Briefly, RNA was collected from cultures infected with control or *Sp9* at DIV13 using the RNAqueous kit (Ambion). One nanogram of the isolated RNAs were subjected to target-specific reverse transcription and 18 cycles of PCR pre-amplification with a mix of primers specific to the target genes. Pre-amplified cDNAs were then processed for real-time PCR analysis on BioMark 96.96 Dynamic Array integrated fluidic circuits (Fluidigm). To quantify total Nrx mRNA levels, probes were selected to detect all target protein variants. All assays for quantitative RT-PCR analyses were purchased from IDT.

Target	Forward Primer	Probe	Reverse Primer
Sp9-opti	GTGTGCAACAAGCGTTTCAT	AGCGACCACCTGAGCAAACACATTA	CACTGTCGCTGCCCTTT
ActinB	ATGCCGGAGCCGTTGTC	CCGCCACCAGTTCGCCATG	GCGAGCACAGCTTCTTTG
DRD1a	CTTCCTGGTCAATCTCAGTCAC	TGCCACATCTCTCCAAATGCCCG	GAGCGTGGTCTCCCAGA
Drd2	TGACAGCATCTCCATTTCCAG	AGCATCCATTCTCCGCCTGTTCA	TCTGCACCGTTATCATGAAGT
PDYN	CATGTCTCCCACTCCTCTGA	TCAACCCCCTGATTTGCTCCCTG	GTGCAGTGAGGATTCAGGATG
Tac1	GCCCATTAATCCAAAGAACTGC	AGAGAATCGCCCGAAGACCCAAG	GACAGTGACCAGATCAAGGAG
NLGN1	GGTTGGGTTTGGTATGGATGA	TGAGGAACTGGTTGATTTGGGTCACC	GATGTTGAGTGCAGTAGTAATGAC
Gabbr1	CTCCGATCCACTTGTCTGTT	TCTTGTAGCTGCCGCCCTGTAG	ATGGCATGGACGCTTATCG
SlitTrks5	GGTCCTGTTCCAAAGTTACTGG	ACGTGCATTTTACCTCTGTTCCCATCT	TGGGTCGGAGAAAGTTGC
LRRTM2	GGCCACTTGAAATGTAAGCC	TGCAGCCTCCAATGTGCTCAGAA	CACTGCGTTGAGTCTGACAA
LRRTM3	CATATGCCAGAAAGGTTGACAC	AGGCTCCAGGGAATGTGAGATACCT	GAGATGCTGCTGAACGGAA

SlitTrks2	TCTTTAAGCACCTTCGATGTCA	CCGTCTTCTCCTGATGTGTTTGCCT	GCTAAACTACTGGTGGACTGAA
Lphn2	CTCGTGGTGGATATTGTGGTT	TGACCCTGCCCAAGTGCCTAC	TTACGGTATTCCCTGGAGTTTG
VAMP3	GGCTTCAGTTTCATCGGTTCT	TGATCCCTATCGCCCACATCTTGC	GCTGCCAAGTTGAAGAGAAAG
gfap	TTCACTTCAGTTACACGCTCTC	CCTGGTCTCAGACGATTGTCATAACCG	CAGCCCTCCCAAGATGAAC
Slc1a2	ATCGCCCACCACATTGAC	ATTCTATCCAGCAGCCAGTCCACC	CCATGCTCCTCATTCTCACAG

Supplementary Table 1. RT-PCR assays used to detect *Sp9* overexpression effect on primary striatal cultures.

REFERENCES

Dobin, A., and Gingeras, T.R. (2002). Mapping RNA-seq Reads with STAR. In Current Protocols in Bioinformatics, (John Wiley & Sons, Inc.),.

Gokce, O., and Südhof, T.C. (2013) Membrane-Tethered Monomeric Neurexin LNS-Domain Is Sufficient for Triggering Synapse Formation. The Journal of neuroscience 33-36, 14617-28

Gokce, O., Runne, H., Kuhn, A., and Luthi-Carter, R. (2009). Short-term striatal gene expression responses to brain-derived neurotrophic factor are dependent on MEK and ERK activation. PLoS One *4*, e5292.

Guez-barber, D., Fanous, S., Harvey, B.K., Zhang, Y., Lehrmann, E., Becker, K.G., Picciotto, M.R., and Hope, B.T. (2012). FACS purification of immunolabeled cell types from adult rat brain. J. Neurosci. Methods *203*, 10–18.

Heintz, N. (2004). Gene Expression Nervous System Atlas (GENSAT). Nat Neurosci *7*, 483.

van der Maaten L, H.G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Shuen, J. a, Chen, M., Gloss, B., and Calakos, N. (2008). Drd1a-tdTomato BAC transgenic mice for simultaneous visualization of medium spiny neurons in the direct and indirect pathways of the basal ganglia. J. Neurosci. *28*, 2681–2685.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M. a., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature.