Cell Reports

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

Graphical Abstract

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

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.

Accession Numbers
GSE82187
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. Sudhof,1,5,8 and Stephen R. Quake3,5,*

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

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 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 (Jaatin et al., 2014), the mouse and human cortex and hippocampus (Damalis 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
We sampled cells either captured randomly or enriched specifically for MSNs or astrocytes using FACS from D1-tdTomato/D2-GFP and 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 $1 \times 10^6 \sim 5 \times 10^6$ 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).

Figure 1. Reverse Engineering of Mouse Striatum by scRNA-Seq
(A) Workflow for obtaining and sequencing cDNA from single cells. Striatal slices from D1-tdTomato/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.

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 Pdh (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 A830009A15Rik, 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 Dnai1, genes that are involved in cilium movement. The second subtype of secretory ependymal cells is correlated with Npr3, Pthr, 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 secretory 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, Tac1, and Adarb2 (Pcdh8-MSN) (Figures 2G–2J; Table S6). The Pcdh8-MSN subpopulation coexpressed the D2 neuropeptide Penk, 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 2J). Many genes enriched in Pcdh8-MSNs are related to neurological disorders (Nrxn2, Sema3e, Sema4a, Sema5a, Sema5b, Pcdh7, Pcdh8, Ptp4g, Ptpmr, Ptpor, and Ptnru) (Redies et al., 2014; De Rubeis et al., 2014; Sudhof, 2006), tachykinin signaling (Tac2 and Tac1) (Steinhoff et al., 2014), and RNA binding (Elavl4, Adarb2, Khdrbs3, Rbm20, Altf2, Lrp2pc, and Celf4) (Li et al., 2007; Nourse et al., 2005). Pcdh8-MSNs are also depleted of important neuronal genes like Ngn1 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, Ptptr, Ngfr, Grik3, Cacng5, and Tmeff2 and TFs Sox9, Sp8, Runx1, Mafb, and Lita1. 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
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 anticorrelated gene groups (Figure 4A).
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-Thn1 genes (x axis). Scaled expression of Meis2 and Thn1 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 Cartpt group genes (y axis) and Calb1 group genes (x axis). Scaled expression of Cartpt and Calb1 is shown by the color of the cell points.

(I) 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.

(J) Distribution of D2 MSNs projected onto Cnr1-Crym gradient scores.

(K) 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.

(L) rPCA of Pcdh8-MSNs using all expressed genes, revealing a continuous transcriptional gradient similar to 2D. The bar on the left shows the experimental origin of cells.

(M) Distribution of Pcdh8-MSNs projected onto Cnr1 gradient scores.

(N) 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.

(O) 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.

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

(Q) 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.
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 Ptp 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...
expressing \( \text{Cxcr1} \), \( \text{Olfm3} \), and \( \text{Mxipl} \) and macrophages expressing \( \text{Mrc1} \) and \( \text{Dab2} \) (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 \( \text{Nfasc} \) and a population of mature oligodendrocytes (MOs) expressing \( \text{Klk6} \) (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 (\( \text{Slc6a11, Slc6a1, and Slc6a9} \)) and glutamate receptors (\( \text{Gria1} \) and \( \text{Gria2} \)), and on the other side to translation, like ribosomal proteins (\( \text{Rpl9, Rpl14, Rps5, and Rpsa} \)) and cell polarity regulators (\( \text{Cdc42} \)) (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 corerlogram (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, \( \text{Myt1l} \), 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 \( \text{Sox9} \) for astrocytes (Caiazzo et al., 2015), \( \text{Sox10} \) for oligodendrocytes (Yang et al., 2013), \( \text{Gata2} \) for endothelial cells (Elcheva et al., 2014), and \( \text{Maf} \) for macrophages (Hegde et al., 2015).
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, Psp1, Rbm24, and Rbm31). 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); Hnmpk (heterogeneous nuclear ribonucleoprotein K), which is associated with intellectual disability (Au et al., 2015); and Hspa5, which is linked to bipolar disorders and schizophrenia (Kakuci 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 S1l).
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 sub-populations: Htr7-D2s and Synpr-D2s. The Synpr-D2 neurons could be further subdivided into Calb1-high or Cartpt-high sub-populations, 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; Zijlstra et al., 2011), and repeated cocaine exposure regulates epigenetic control of Pchd8-MSN marker 493241E22Rik (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).

The most significant 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**


**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 (R57MH52804 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**


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


