

## RESEARCH ARTICLE SUMMARY

## BICCN

## Transcriptomic diversity of cell types across the adult human brain

Kimberly Siletti\*, Rebecca Hodge†, Alejandro Mossi Albiach†, Ka Wai Lee, Song-Lin Ding, Lijuan Hu, Peter Lönnerberg, Trygve Bakken, Tamara Casper, Michael Clark, Nick Dee, Jessica Gloe, Daniel Hirschstein, Nadiya V. Shapovalova, C. Dirk Keene, Julie Nyhus, Herman Tung, Anna Marie Yanny, Ernest Arenas, Ed S. Lein, Sten Linnarsson\*

**INTRODUCTION:** The mammalian brain comprises billions of neurons and glia capable of executing highly complex behaviors. These cells are organized into several major functional regions with distinct developmental origins. Although the cerebral cortex is the most well studied because of its role in cognition, the other regions are no less essential. In the past several years, single-cell genomic methods have revolutionized our understanding of the brain's cellular diversity, revealing hundreds of transcriptomic cell types across the mouse brain. Prior work has shown that transcriptional cell types can be aligned with other modalities—e.g., electrophysiology, morphology, and connectivity—as well as across large evolutionary distances. However, the human brain has not been comprehensively surveyed, and few regions outside the cerebral cortex have been profiled. Thus, the overall number, distribution, and region-specificity of human neurons and glia remain unknown.

**RATIONALE:** As a first step toward a brain-wide census of cell types, we used single-nucleus

RNA sequencing to profile cells sampled from throughout the entire human brain. We isolated postmortem tissue from three donors and enriched for neurons from approximately 100 locations across the forebrain (the cerebral cortex, hippocampus, cerebral nuclei, hypothalamus, and thalamus), midbrain, and hindbrain (the pons, medulla, and cerebellum). The final dataset comprised more than three million cells, including more than two million neurons, which we clustered iteratively into 31 superclusters, 461 clusters, and 3313 subclusters. This top-down approach enabled us to examine and compare heterogeneity within and across cell classes and regions.

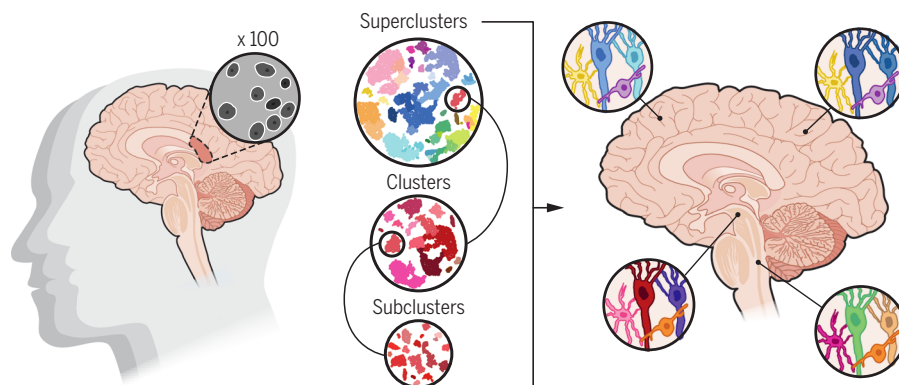
**RESULTS:** Neurons varied extensively across brain regions. Many neuronal superclusters comprised cells mainly localized to specific brain regions. Cell states moreover broadly mirrored their developmental history. For example, several superclusters distributed across the telencephalon—the developmental compartment that produces the cortex, hippocampus, and cerebral nuclei. Cortical clusters

comprised layer-specific excitatory neurons as well as distinct inhibitory interneurons with distinct developmental origins. Other superclusters reflected cellular migration during development, including midbrain-derived inhibitory neurons located in the thalamus, which transcriptionally aligned with midbrain neurons.

Within regions, neuronal subtypes were not distributed according to simple rules. Dissections differed from one another according to both specific cell types and cell type proportions. Notably, neurons were particularly diverse outside the cortex. The hypothalamus, midbrain, and hindbrain contained markedly high neuronal heterogeneity, consistent with their diverse functions. These neurons were also organized less hierarchically compared with cortical neurons: Many belonged to a single supercluster that uniquely contained both inhibitory and excitatory neurons along with serotonergic and dopaminergic neurons. These neurons combinatorially expressed many neurotransmitters and neuropeptides.

Glia also varied across brain regions, and their diversity similarly reflected development. In particular, both astrocytes and oligodendrocyte precursors formed two major groups enriched within and outside the telencephalon. By contrast, mature oligodendrocytes exhibited two major types found across the entire brain. However, these oligodendrocyte types existed in different proportions inside and outside the telencephalon, which suggests that even the relatively homogeneous oligodendrocyte lineage exhibits regional variation.

**CONCLUSION:** Our findings suggest that each brain area contains a specific complement of cell types and states, which implies that a complete characterization of cell types will require deep tissue sampling, particularly outside the cortex. The telencephalon appears unique with respect to other brain regions, across both neurons and glial cells, whereas the brainstem comprises an extremely diverse set of neurons that may support innate behaviors. These observations have implications for a range of human diseases that exhibit regional variation, including cancer and neurodegenerative disease. Our work therefore provides a basis for exploring the role of neuroepithelial diversity in human health and disease. ■



**Cellular diversity across the entire human brain.** Approximately 100 anatomical locations were dissected from the human brain in three donors followed by single-nucleus RNA sequencing (left). Three levels of clustering revealed the cell type composition of the human brain (center). Whereas cortical neurons varied more gradually, brainstem neurons were unexpectedly diverse with a large number of small clusters indicating distinct cell types organized by combinatorial gene expression (right). Although less heterogeneous than neurons, glial cells, such as astrocytes and oligodendrocyte precursors, also differed between the cortex and the brainstem (right).

The list of author affiliations is available in the full article online.  
\*Corresponding author. Email: k.a.siletti@umcutrecht.nl (K.S.); sten.linnarsson@ki.se (S.L.)  
†These authors contributed equally to this work.  
Cite this article as K. Siletti et al., *Science* **382**, eadd7046 (2023). DOI: 10.1126/science.add7046

**S READ THE FULL ARTICLE AT**  
<https://doi.org/10.1126/science.add7046>

## RESEARCH ARTICLE

## BICCN

## Transcriptomic diversity of cell types across the adult human brain

Kimberly Siletti<sup>1\*†</sup>, Rebecca Hodge<sup>2‡</sup>, Alejandro Mossi Albiach<sup>1‡</sup>, Ka Wai Lee<sup>1</sup>, Song-Lin Ding<sup>2</sup>, Lijuan Hu<sup>1</sup>, Peter Lönnerberg<sup>1</sup>, Trygve Bakken<sup>2</sup>, Tamara Casper<sup>2</sup>, Michael Clark<sup>2</sup>, Nick Dee<sup>2</sup>, Jessica Gloe<sup>2</sup>, Daniel Hirschstein<sup>2</sup>, Nadiya V. Shapovalova<sup>2</sup>, C. Dirk Keene<sup>3</sup>, Julie Nyhus<sup>2</sup>, Herman Tung<sup>2</sup>, Anna Marie Yanny<sup>2</sup>, Ernest Arenas<sup>1</sup>, Ed S. Lein<sup>2</sup>, Sten Linnarsson<sup>1\*</sup>

The human brain directs complex behaviors, ranging from fine motor skills to abstract intelligence, but the diversity of cell types that support these skills has not been fully described. In this work, we used single-nucleus RNA sequencing to systematically survey cells across the entire adult human brain. We sampled more than three million nuclei from approximately 100 dissections across the forebrain, midbrain, and hindbrain in three postmortem donors. Our analysis identified 461 clusters and 3313 subclusters organized largely according to developmental origins and revealing high diversity in midbrain and hindbrain neurons. Astrocytes and oligodendrocyte-lineage cells also exhibited regional diversity at multiple scales. The transcriptomic census of the entire human brain presented in this work provides a resource for understanding the molecular diversity of the human brain in health and disease.

The mammalian brain controls sensory-motor function, maintains physiological homeostasis, and supports cognition and memory. Single-cell sequencing has revealed extensive cell type diversity in the entire mouse brain, as well as regions of the human brain (1–5). We performed single-nucleus RNA sequencing on tissue from across three entire human brains from male donors (Fig. 1A and Materials and methods). We aimed to sample 100 anatomically distinct locations (dissections) from the three brains in technical duplicate (table S1). We also included one motor cortex dissection from a female donor published in a previous work (M1C; two technical replicates, 55,591 cells) (3). We ultimately included 606 high-quality samples covering 105 dissections across 10 brain regions obtained from four brains (fig. S1 and table S2).

We used graph-based clustering to hierarchically divide cells first into 31 superclusters, then 461 clusters, and finally 3313 subclusters (Fig. 1, B to D, and Materials and methods). We named superclusters based on the literature and their regional composition (table S3) with one exception. One supercluster contained neurons from most brain regions, and we named these neurons *splatter neurons* based solely on their appearance on the two-

dimensional embedding (Fig. 1B). A dendrogram of all clusters mirrored previous mouse work: Neurons split from nonneuronal cells and into two main clades (Fig. 1E) (1). One clade contained telencephalic excitatory neurons expressing *SLC17A7*; the other contained telencephalic inhibitory neurons (*SLC32A1+*) as well as all excitatory and inhibitory diencephalic, midbrain, and hindbrain neurons (Fig. 1E).

### Supercluster distributions reflect brain development

The distributions of superclusters across dissections likely reflected their developmental history (Fig. 2A). Most superclusters derived from the telencephalon, which indicates that we sampled this developmental compartment deeply relative to its complexity. Several superclusters corresponded to the six layers of cortical excitatory neurons: intratelencephalic (IT)–projecting [*upper-layer* and *deep-layer IT* (italics indicating supercluster names)], near-projecting (NP) (*deep-layer NP*), and corticothalamic-projecting (CT/6b) (*deep-layer CT/6b*). Both *upper-layer IT* and *deep-layer IT* superclusters contained putative layer-four clusters. Putative layer-five extratelencephalic (ET)–projecting neurons belonged to *miscellaneous*, along with lymphocytes and other rare cell types (~25,000 cells). Two superclusters corresponded to the cerebral nuclei's dopamine receptor–expressing *medium spiny neurons (MSNs)* and *CASZ1+ eccentric MSNs* (Fig. 2A; fig. S2, A to F; and Materials and methods) (5). Most *eccentric MSNs* expressed only *DRD1*, but many in the basal ganglia co-expressed *DRD1* and *DRD2*, suggesting regional specialization (fig. S2, G and H). Most other

superclusters were region-specific excitatory neurons.

The presence of three superclusters in multiple regions likely reflected migration during development. First, cortical interneurons were distributed across the telencephalon, reflecting their migration from the medial and caudal ganglionic eminences (*MGE* and *CGE interneurons*, respectively) (6). Our data suggested that *CGE* interneurons migrate more widely, including to the hypothalamus, thalamus, and midbrain. This observation might reflect transcriptomic convergence during development—*LHX6* generates local inhibitory neurons in the hypothalamus (7)—but migration from the ganglionic eminences into the thalamus occurs specifically in humans (8). Second, a supercluster found mainly in the thalamus and midbrain likely corresponded to *SOX14+* inhibitory neurons that migrate from the embryonic midbrain to the thalamus (9). Our data suggested that some of these cells also migrate to the hypothalamus and pons; we confirmed the latter with in situ hybridization (fig. S3). Third, two superclusters appeared to derive from the rhombic lip, which produces not only the cerebellum but also neurons that migrate to specific nuclei in the pons and medulla (10).

### Dissections reveal supercluster-specific anatomical diversity within brain regions

Because superclusters were enriched in specific developmental compartments and regions, we wondered whether anatomical relationships also explained heterogeneity within superclusters. As expected, telencephalic excitatory neurons varied considerably more than interneurons (fig. S4, A and B, and Materials and methods) (11). Nevertheless, even some broadly distributed clusters were distributed unevenly across dissections. To determine whether clusters were distributed according to systematic rules, we examined how dissections within the same regions related to one another. We calculated neighborhood graphs separately for the cerebral cortex, cerebral nuclei, and thalamus, connecting each dissection to the two dissections with the most similar cluster proportions. Dissections were often linked to anatomical neighbors, which demonstrates that cell type composition broadly reflects spatial position within developmental compartments (Fig. 2B and fig. S4, C and D). Although imprecise dissection might partly drive this result in the cerebral nuclei and thalamus (table S4), the more unambiguous cortical dissections were similarly arranged (Fig. 2B).

However, superclusters might distribute differently across brain regions. We therefore compared dissections within each supercluster separately, focusing on the cortex because of its relatively high number of dissections and superclusters. Neighboring dissections were still

<sup>1</sup>Karolinska Institute, 171 77 Stockholm, Sweden. <sup>2</sup>Allen Institute for Brain Science, Seattle, WA 98109, USA.

<sup>3</sup>Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA 98195, USA.

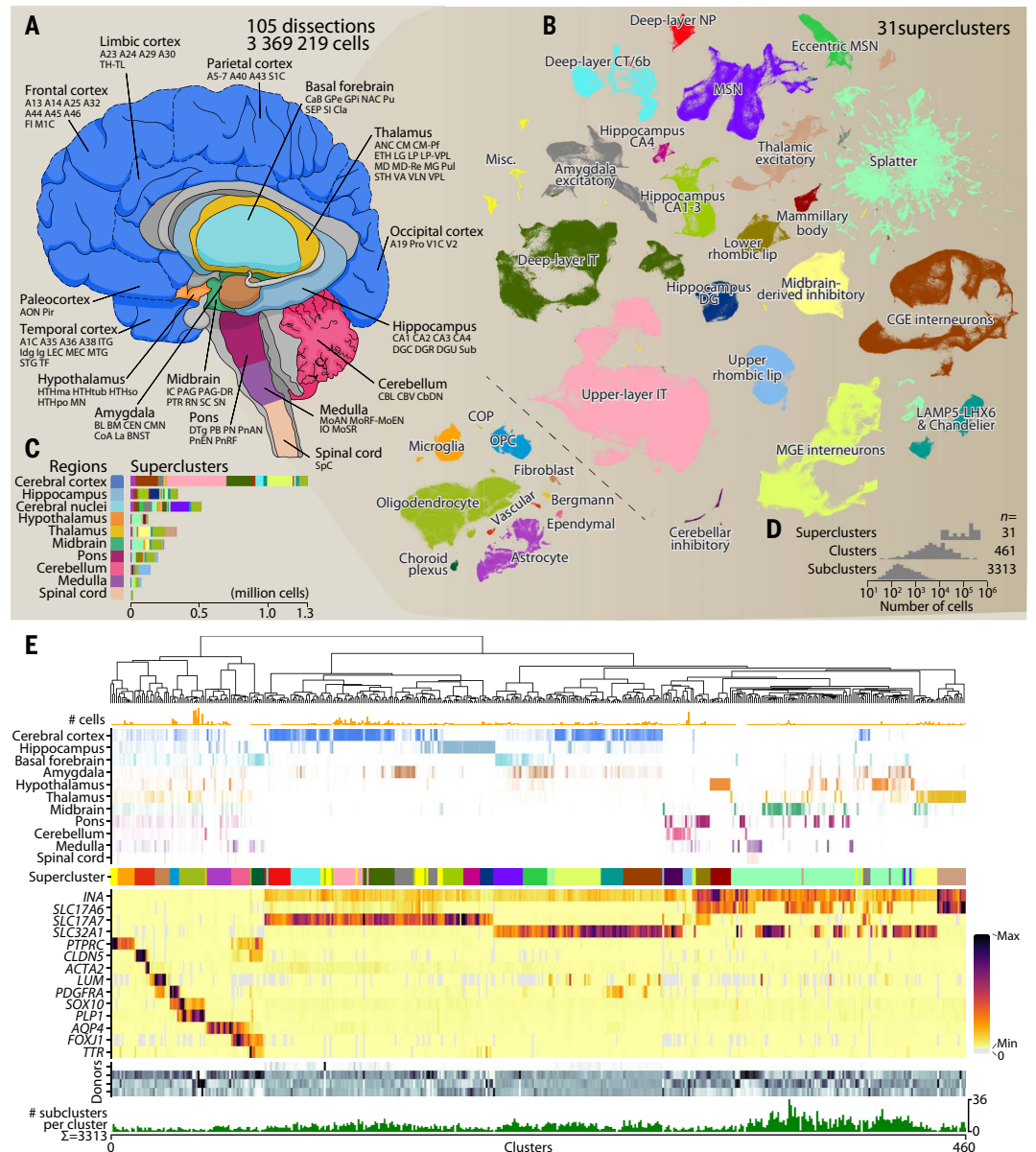
\*Corresponding author. Email: k.a.siletti@umcutrecht.nl (K.S.); sten.linnarsson@ki.se (S.L.)

†Present address: University Medical Center Utrecht, 3584 CG Utrecht, Netherlands.

‡These authors contributed equally to this work.

### Fig. 1. Single-nucleus RNA sequencing of the adult human brain.

(A) Schematic showing dissections. Abbreviations correspond to the Allen Brain Reference Atlas or cortical Brodmann areas. (B) t-SNEs showing neurons (right) and nonneuronal cells (inset, left) colored by supercluster. CA, cornu ammonis area; CGE, caudal ganglionic eminence; CT, corticothalamic; DG, dentate gyrus; IT, intratelencephalic; NP, near-projecting; MGE, medial ganglionic eminence; MSN, medium spiny neuron; OPC, oligodendrocyte precursor cell. (C) Stacked bars colored as in (B) showing the number of cells in each supercluster. (D) Histograms showing the distributions of the number of cells in each supercluster, cluster, or subcluster as indicated. (E) Dendrogram of 461 clusters, with attributes from top to bottom: number of cells; color saturation representing the proportion of cells per cluster from each brain region; supercluster colored as in (B); heatmap showing gene expression that marks neurons (*INA*), glutamatergic (*SLC17A6*, *SLC17A7*) and GABAergic (*SLC32A1*) neurons, immune cells (*PTPRC*), vascular cells (*CLDN5*, *ACTA2*), fibroblasts (*LUM*), OPCs (*PDGFRA*, *SOX10*), oligodendrocytes (*SOX10*, *PLP1*), astrocytes (*AQP4*), ependymal cells (*FOXJ1*), and choroid plexus (*FOXJ1*, *TTR*); grayscale representing the proportion of cells per cluster from each donor; bars representing the number of subclusters per cluster.



broadly more correlated, but superclusters exhibited different correlation patterns and were not always smoothly distributed across the brain (Fig. 2C and fig. S4E). Excitatory neurons produced diverse patterns, which implies different degrees of anatomical specialization. *Upper-layer IT* neurons appeared most specialized and were distributed distinctively in many dissection groups like Brodmann areas 44 to 46 (A44–A45, A46), frontal insula and anterior cingulate cortex (FI, A24), motor cortex (MIC) and somatosensory cortex (SIC), and visual cortex (VIC). *Deep-layer NP* neurons were distributed in two major patterns that split within the parietal cortex. Some dissections consistently grouped together, including FI and A24, as well as A44–A45 and A46. The entorhinal cortex and paleocortex [medial en-

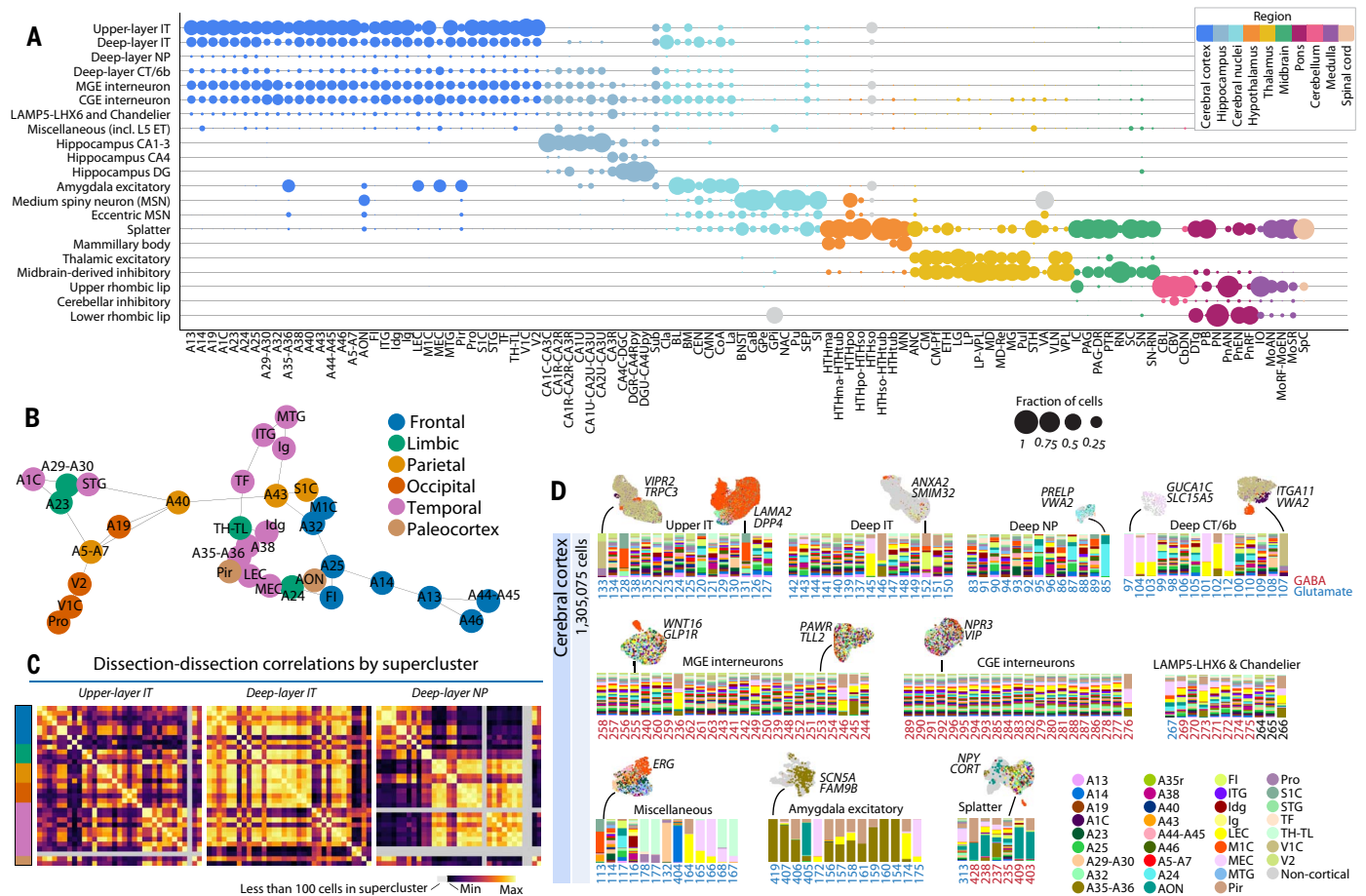
torhinal cortex (MEC), lateral entorhinal cortex (LEC), piriform cortex (Pir), anterior olfactory nucleus (AON)], considered the more evolutionarily conserved part of the cortex that contains only three cellular layers, were particularly distinct.

We next explored whether particular clusters drove these correlation patterns. Whereas most interneuron clusters contributed to all cortical dissections (Fig. 2D), some excitatory clusters were enriched in specific dissections that explained previous observations. MIC and SIC shared two *upper-layer IT* clusters (no. 128 and no. 131) that likely drove their high correlation in this supercluster and their tight relationship on the neighborhood graph (Fig. 2, B and D, and fig. S4E). VIC exhibited distinctive *deep-layer CT/6b* and *upper-layer*

*IT* clusters (no. 107 and no. 133), including VIC's distinctive *TRPC3+* layer-four neurons (12). The entorhinal and piriform cortex contained distinctive *deep-layer IT* and *deep-layer CT/6b* clusters as well as *amygdala excitatory* and interneuron clusters (Fig. 2, B and D). However, no clusters were uniquely enriched in FI and A24 or A44–A45 and A46, which indicates that these highly correlated dissections instead contained similar proportions of more broadly distributed clusters.

Outside the cortex, superclusters and clusters also distinguished dissections according to different patterns (fig. S5). Some superclusters distributed relatively broadly, including migratory *MGE* and *CGE interneurons* and *midbrain-derived inhibitory* neurons. By contrast, many region-specific superclusters were





**Fig. 2. Superclusters and clusters are differentially distributed across brain regions.** (A) Supercluster distributions across dissections, normalized by column and colored by region (gray, potential index-hopping and dissection imprecision). (B) Cortical dissections colored by lobe and connected to the two dissections with the most similar cluster proportions. (C) Heatmaps showing correlations of cluster proportions between dissections, in the same order on both axes. (Left) Color bar

enriched in specific nuclei. Many amygdala nuclei contained a relatively specific *amygdala excitatory* cluster. The striatum [caudate (CaB), nucleus accumbens (NAc), and putamen (Pu)] contained relatively specific *MSN* clusters. *Thalamic excitatory* clusters were highly enriched in the lateral geniculate nucleus (LG) and anterior nuclear complex (ANC). Even within these superclusters, however, many clusters were found across multiple dissections (figs. S4A and S5). Because dissections often overlapped in these regions, future work will likely localize an even greater number of cell types to specific nuclei. Nevertheless, our observations suggest that individual brain nuclei are transcriptionally distinct as a result of both varying cell type proportions and specialized cell types.

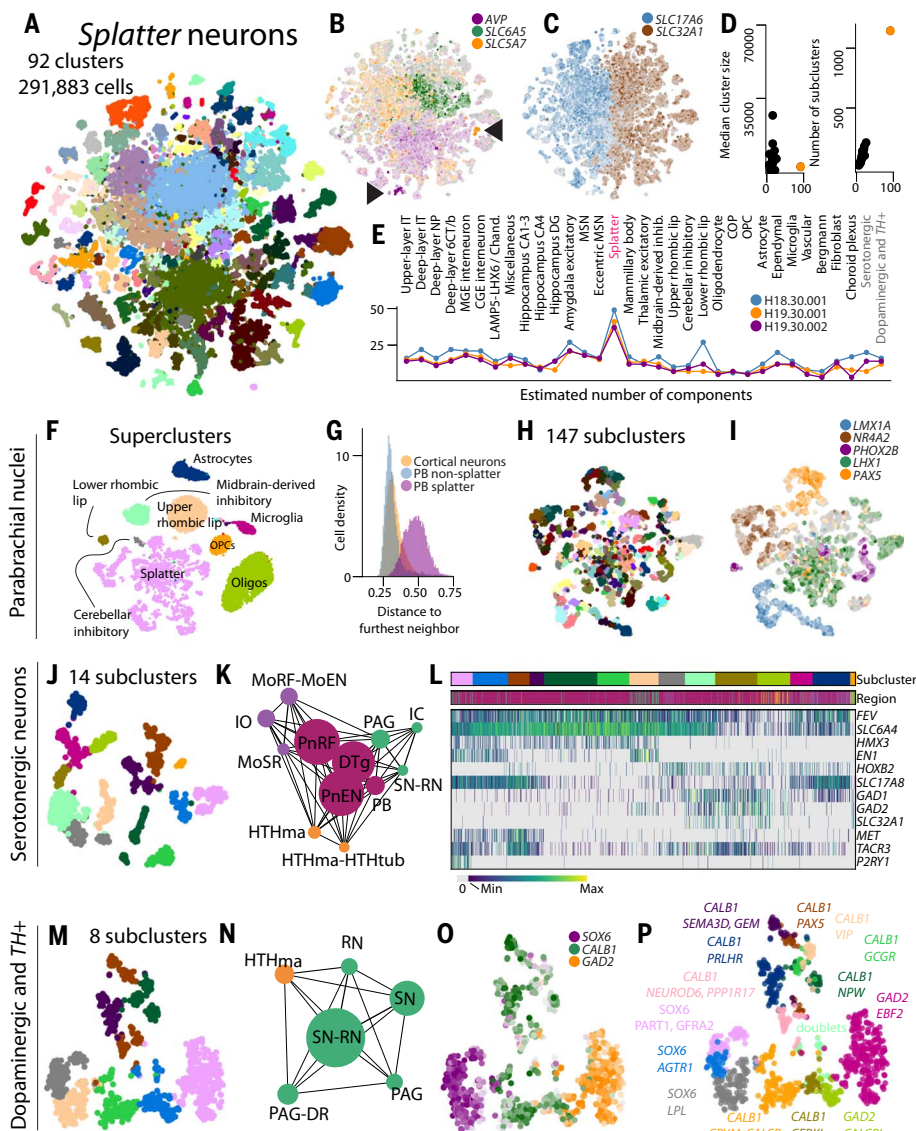
### Most neuronal diversity is found outside the telencephalon

*Splatter* neurons comprised a heterogeneous group of cells from most brain regions, especial-

ly midbrain, hindbrain, and more than half of all hypothalamic neurons (Fig. 3A and fig. S6A). Only one *splatter* cluster contained cells from across the cortex and likely represented long-range projecting  $\gamma$ -aminobutyric acid-releasing (GABAergic) neurons (Fig. 2D; no. 235 expressing *SST* and *CHODL*) (11). The supercluster expressed diverse neurotransmitters, containing both inhibitory and excitatory cells—22% expressed the GABA transporter *SLC32A1*, 39% expressed the glutamatergic transporter *SLC17A6*, and 1.4% coexpressed these genes (Fig. 3, B and C, and fig. S6B)—as well as cholinergic, glycinergic, serotonergic, and dopaminergic neurons (fig. S6C). Illustrating the complexity of these cells relative to other superclusters, 92 *splatter* clusters yielded 1145 subclusters, >900 more than any other supercluster (Fig. 3 and fig. S6D). Moreover, nearly twice as many principal components were required to describe *splatter* neurons (Fig. 3E, fig. S6E, and Materials and methods).

indicating cortical lobe colored as in (B). (D) Stacked bar plots per cluster (excluding clusters with less than 100 cells or less than 1% cells from cortex) showing the relative contributions of neurons from each cortical dissection, colored according to the legend. A UMAP embedding is shown for select clusters (gray, noncortical cells). Clusters are labeled with two of their most enriched genes. Red labels, GABAergic (*SLC32A1*); blue labels, excitatory glutamatergic (*SLC17A6* or *SLC17A7*).

Furthermore, principal components did not reveal a clear hierarchical structure to the supercluster and instead primarily distinguished a specific telencephalic subtype (fig. S6, F and G). Even within dissections, the first components in *splatter*-enriched regions were less correlated to neurotransmitter identity compared with cortical dissections, even when *splatter* neurons were analyzed separately (fig. S6, H and I). We explored whether this atypical molecular architecture was an artifact of our methods and observed that *splatter* neurons were consistent across donors and did not stand out in terms of quality metrics (fig. S6, J and K). *Splatter* neurons consistently grouped together after down-sampling, without donor-batch correction and with different gene sets (fig. S6, L and M). Moreover, a Gaussian mixture model (GMM) clustered *splatter* neurons together independently of the cell-neighborhood graph, which implies that these cells occupy a distinct transcriptional subspace (fig. S7A).



**Fig. 3. Splatter neurons are highly complex.** (A) Splatter neurons colored on the t-SNE by cluster. (B and C) Cells colored by their expression of the gene they express most highly among those in the legend (gray, no expression). Arrows indicate clusters with high AVP and SLC5A7 expression. (D) For each supercluster or subset, median cluster size (left) or total subclusters (right) is plotted against total number of clusters. Splatter is orange. (E) The optimal number of principal components to represent each supercluster, calculated separately for each donor as indicated by the key. Serotonergic and dopaminergic indicate splatter subsets. (F) Cells from the parabrachial nuclei colored and labeled by supercluster on the t-SNE. (G) Distributions of Euclidean distances between cells and their 25th nearest neighbors, normalized by each dataset's maximum distance. (H) Splatter cells from the parabrachial nuclei colored by subcluster on the t-SNE. (I) Same as (B). (J) Serotonergic neurons colored by cluster on the t-SNE. (K) Dissections that constitute 95% of the serotonergic neurons shown in (J). The node size represents the number of cells from each dissection, and the edge width represents the percentage of cells that were neighbors. (L) Expression across serotonergic cells for select genes. (Top bar) Subcluster colored as in (J). (Second bar) Region colored as in (K). (M) TH<sup>+</sup> and dopaminergic neurons colored by subcluster. (N) Same as (K), but for TH<sup>+</sup> neurons. (O) Same as (B). (P) Finer subtypes, indicated by color and marker genes.

We therefore used an alternative factorization method to identify transcriptional modules across these cells. Topic modeling [latent Dirichlet allocation (LDA) (13)] uncovered a *splatter*-enriched module that included neuronal activity-related genes, including sero-

nergic receptor *HTR2C*, voltage-gated channels *SCN7A* and *SCN9A*, and the actin-organizing protein *KLHL1* (fig. S7, B and C). However, the genes were not highly specific, and LDA models fit with fewer topics grouped *splatter* neurons with other superclusters (fig. S7D). We

therefore hypothesized that some other superclusters were similar to *splatter* neurons, but were more abundant, and thus were more easily separated by clustering analysis. To test this idea, we reanalyzed a down-sampled dataset but retained only five *mammillary body* neurons, which LDA had grouped with *splatter* neurons. The five cells clustered with *splatter* neurons (fig. S7E). Accordingly, most nontelencephalic neurons were second-most similar to *splatter* neurons (fig. S7F). These observations suggest that *splatter* neurons comprise diverse nontelencephalic cell types that were not sampled deeply enough to be modeled well in the current dataset. Some might form their own superclusters if sampled more deeply; when we down-sampled our whole dataset by cluster, *splatter* neurons were relatively more common and exhibited more structure on the embedding (fig. S6L).

Therefore, this supercluster might be better understood by considering subsets. We first examined whether these neurons appeared less complex in specific dissections and analyzed the parabrachial nuclei, where we sampled a high proportion of *splatter* neurons. We found *midbrain-derived inhibitory*, *upper rhombic lip*, *lower rhombic lip*, *cerebellar inhibitory*, and *splatter* neurons (Fig. 3F). *Splatter* neurons were most heterogeneous, exhibiting greater distances from their 25th nearest neighbors than did neurons from other superclusters, whose distances were more like those between cortical neurons (Fig. 3G). We found 147 *splatter* subclusters with more than five cells in the parabrachial nuclei (Fig. 3H). The subclusters spanned populations known to neighbor (*PHOX2B*<sup>+</sup>) and spatially segregate within (*FOXP2*<sup>+</sup>, *LMX1B*<sup>+</sup>, and *PAX5*<sup>+</sup>) the murine parabrachial nuclei (14, 15). *NR4A2* and *LMX1A* more specifically labeled *FOXP2*<sup>+</sup> and *LMX1B*<sup>+</sup> neurons in our data (Fig. 3I). Parabrachial cell types are therefore extremely diverse but broadly conserved across mice and humans.

We also explored *splatter* clusters defined by neurotransmitter type, such as serotonin. Serotonergic markers *FEV* and *SLC6A4* were highly expressed in cluster 397, which comprised 14 subclusters derived primarily from the pons, midbrain, medulla, and hypothalamus (Fig. 3, J and K, and fig. S8, A to I). All cells expressed *FEV* except a small cluster of 20 *PAX6*<sup>+</sup> cells and some *PITX2*<sup>+</sup> hypothalamus cells (Fig. 3L and fig. S8I). As previously described in mice, subclusters could be grouped into rostral- (*HMX3*<sup>+</sup>) and caudal- (*HOX*<sup>+</sup>) derived populations, and rostral neurons could be further split into *EN1*<sup>+</sup> and *EN1*<sup>-</sup> populations (Fig. 3L) (16). Similar to a recent single-cell study of mouse serotonergic neurons (17), our analysis revealed combinatorial expression of glutamatergic and GABAergic genes in addition to neuropeptides like *PENK* and *TRH* (Fig. 3L and fig. S8, G to I). One subcluster in



our analysis appeared to correspond to the distinct *P2RY1+MET+TACR3+* cells located near the cerebral aqueduct in mice (17). This sub-cluster derived mostly from the periaqueductal gray, which suggests that these neurons and their specific electrophysiological and projection properties are conserved in humans.

We next investigated *splatter* cluster 395 that expressed dopaminergic markers *TH*, *SLC6A3*, and *DDC* and contained mostly midbrain—substantia nigra, red nucleus, and periaqueductal gray—and very few hypothalamic neurons (Fig. 3, M and N, and fig. S8, J to N). This analysis yielded three neuronal classes (*TH+SOX6+*, *TH+CALB1+*, and *TH+GAD2+*) and 14 subtypes (Fig. 3, O and P; fig. S8O; and Materials and methods). We trained a classifier on a recent dataset that defines 10 subtypes and two classes (*TH+SOX6+* and *TH+CALB1+*) of dopaminergic neurons among *NR4A2+* nuclei from the substantia nigra (4). We found that seven of our 14 neuronal subtypes corresponded to eight previously defined dopaminergic neuron subtypes (fig. S8, P to S, and Materials and methods). Integration analysis of both datasets confirmed these results and the presence of seven additional neuronal subtypes (fig. S8, T to V, and Materials and methods), including one in the substantia nigra (*TH+SOX6+LPL+*) and four *TH+CALB1+* subtypes (*PAX5+*, *NPW+*, *GCCR+*, or *VIP+*) in the periaqueductal gray and ventral tegmental area surrounding the red nucleus (Fig. 3P and fig. S8O). Similar to serotonergic neurons, *TH+CALB1+* subtypes in the periaqueductal gray expressed different combinations of neuropeptides, such as *VIP*, *NPPC*, *NPW*, and *PENK* (fig. S8W). In addition, we found two *TH+GAD2+* clusters (*CALCR+* or *EBF2+*) that defined the molecular signature of a third class of neurons that coexpress some dopaminergic and GABAergic markers (18) (fig. S8, O and X). Our results are thus consistent with midbrain *TH+* neurons exhibiting transcriptional profiles ranging from a complete dopaminergic phenotype (12 subtypes of the *TH+SOX6+* and *TH+CALB1+* classes) to *TH+* neurons with a partial dopaminergic and GABAergic phenotype (two subtypes of the *TH+GAD2+* class).

Thus, even *splatter* subtypes exhibit extensive regional and molecular heterogeneity. In fact, serotonergic and dopaminergic neurons exhibited similar complexity to some superclusters, despite comprising only about 1000 cells (Fig. 3E). Similarly, enhanced electric fluorescence in situ hybridization (EEL-FISH) (Materials and methods) revealed GABAergic and glutamatergic neurons in the midbrain and pons that expressed a number of neuropeptides (fig. S9, B and C: B1-13, C1; and fig. S10). We also observed some very rare cell populations: In the substantia nigra *pars reticulata*, we found cells coexpressing *GHR* and *SST* (fig. S9A: A4), whereas in the retrorubral field, we

found cells expressing peptides such as *ADCYAPI*, *CARTPT*, *NPX*, or *NTS* (fig. S9B: B10-12). Cells across the brain stem are therefore defined by combinatorial neuropeptide and neurotransmitter expression.

Our observations suggest that *splatter* neurons cluster together because neurons outside the telencephalon are (i) less similar to telencephalic neurons than to one other and (ii) not hierarchically organized by neurotransmitter identity. This similarity is weak and does not imply a biological category. Our previous adolescent mouse brain atlas did not reveal any *splatter*-like clusters (1). However, we observed *splatter*-like neurons outside the cortex and hippocampus when we included cells that had initially been excluded from that dataset (fig. S11, A to C). The population included adjacent glutamatergic, GABAergic, and glycinergic subpopulations that expressed genes that our LDA model ranked highly for *splatter* neurons (fig. S11D). These neurons were possibly excluded previously as a result of their lack of distinct markers because most were similar in quality to other mouse neurons (fig. S11E). Accordingly, regions like the preoptic hypothalamus are highly heterogeneous in the mouse (19). These and our findings underline that the mammalian brain contains highly diverse populations that are challenging to cluster and organize hierarchically without deeper sampling.

### Oligodendroglia and astrocytes are specific to brain regions

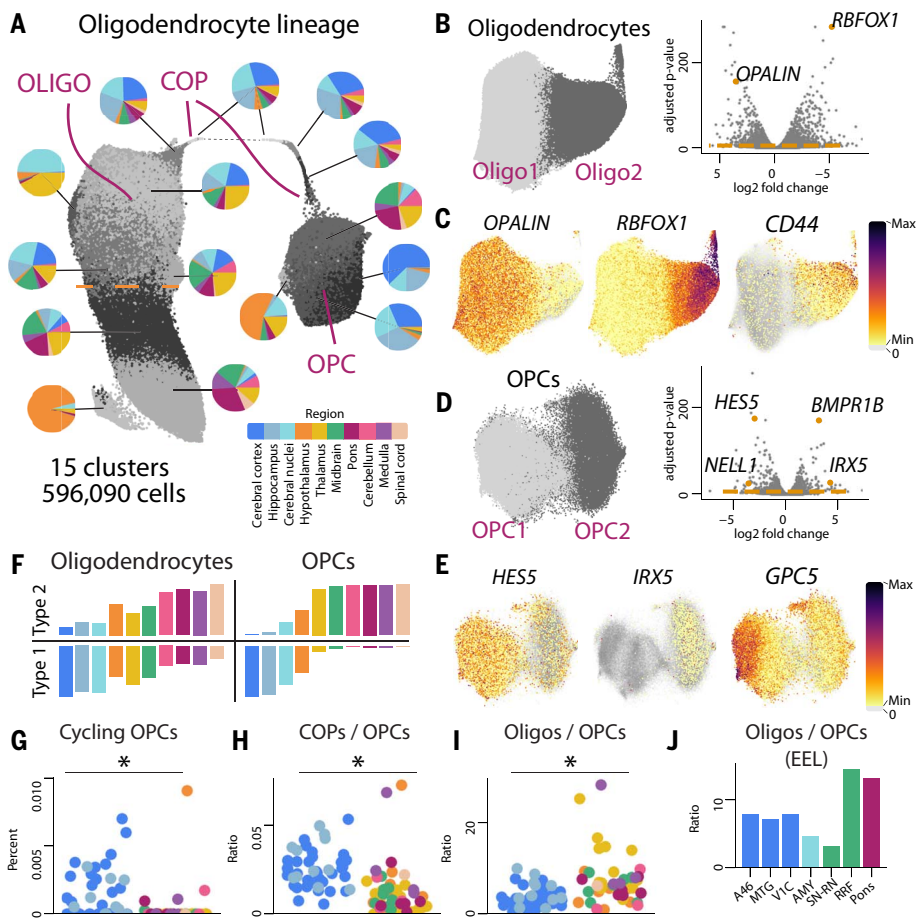
Oligodendrocyte-lineage superclusters included *oligodendrocyte precursor cells* (OPCs), *committed oligodendrocyte precursors* (COPs), and *oligodendrocytes* and projected onto a single uniform manifold approximation and projection (UMAP) embedding that resembled lineage differentiation (Fig. 4A). A dendrogram grouped eight *oligodendrocyte* clusters into two types that expressed *OPALIN* or *RBFOX1*, as previously described (Oligo1 and Oligo2, respectively) (Fig. 4, B and C; fig. S12A; and table S5) (20). One distinct *RBFOX1+* cluster largely (88%) derived from a single dissection in one donor and was likely low quality (fig. S12, B to F). *RBFOX1+* oligodendrocytes were previously postulated to be a mature end state, consistent with their UMAP position in our analysis.

Four OPC clusters were characterized by regional specificity rather than maturation state (Fig. 4A and figs. S12G and S13). Two major types, OPC1 and OPC2 (Fig. 4, D and E), were enriched within and outside the telencephalon, respectively, and likely corresponded to the *NELL1+* cortical and *PAX3+* spinal cord populations recently described (21). One OPC2 cluster was 54% mammillary nucleus, reminiscent of an oligodendrocyte cluster (fig. S13, C to E). OPC1 and OPC2 differentially expressed 128 genes with more than a fourfold change in either population, including axon-

guidance molecules, region-specific transcription factors, and Notch signals (fig. S13A and table S6). Although we previously found only one OPC cluster in the adolescent mouse, reanalysis revealed that mouse OPCs expressed type-specific markers *Hes5* and *Irx5* mutually exclusively (fig. S13B). Thus, OPCs exhibit relatively little transcriptomic variation—few principal components explained most of their variance (Fig. 3E)—but are not homogeneous across the brain. Accordingly, clustering within OPC1 appeared driven by two expression gradients. Telencephalon-specific transcription factor *FOXG1* was evenly expressed across OPC1, whereas the secreted protein *NELL1* and cell-adhesion molecule *CNTN5* increased smoothly (fig. S13, F to I). The second gradient was defined by the proteoglycan *GPC5*, a marker for gray matter astrocytes (Fig. 4E) (22). Telencephalic OPC heterogeneity might therefore reflect interactions with the local environment.

All regions contained substantial numbers of both oligodendrocytes but not both OPC types, which argues against a simple model where OPC1 generates Oligo1 and OPC2 generates Oligo2. For example, OPC1 was largely absent from the midbrain, but both Oligo1 and Oligo2 were present. Nevertheless, oligodendrocytes exhibited regional differences: Oligo1 predominated in the telencephalon; Oligo2 cells progressively increased toward the brain's more posterior regions (Fig. 4F). Because Oligo1 was also putatively less mature than Oligo2, we investigated whether the differential distributions of OPCs and oligodendrocytes along the antero-posterior axis might partly reflect different rates of oligodendrocyte turnover inside and outside the telencephalon. Telencephalic OPCs expressed more cell cycle-related genes than nontelencephalic OPCs, which suggests proliferation (Fig. 4G).

We therefore investigated the ratios of OPC, COP, and *oligodendrocyte* populations across dissections, reasoning that these ratios would reflect relative turnover rates. We found an elevated ratio of COPs to OPCs in the telencephalon, which indicates more frequent initiation of differentiation (Fig. 4H). The ratio of oligodendrocytes to OPCs was instead higher in the brainstem (Fig. 4I; confirmed by EEL-FISH in Fig. 4J), which indicates lower turnover of mature cells relative to the progenitor population. Thus, in our dataset, telencephalic OPCs appear more proliferative, differentiate more frequently, and form a larger population relative to oligodendrocytes. Elevated OPC proliferation might therefore more frequently replenish the oligodendrocyte population in the telencephalon, lowering the proportion of more mature Oligo2 cells. However, Oligo2 is not experimentally confirmed as a mature end state, and OPCs play a variety of roles in the brain. These ratios might reflect other regional differences.



**Fig. 4. Oligodendrocyte lineage.** (A) Oligodendrocytes (*RBFOX1*+ below dotted orange line), COPs, and OPCs on a UMAP embedding, grayscale-colored by cluster. Pie charts show regional compositions. (B) Type 1 and type 2 oligodendrocytes separately re-embedded by t-SNE (left). Differential gene expression (right) (orange line,  $q < 10^{-5}$ ). The y axis shows  $-\log_{10}$  of the adjusted  $P$  values. (C) Oligodendrocytes colored by gene expression. (D) Same as (B), but for OPCs. (E) OPC gene expression. (F) Proportions of type 1 and type 2 oligodendrocytes and OPCs found in each region colored as in (A). (G to I) Dissections colored by region as in (A). (Left) Telencephalon. (Right) Rest of the brain. Vertical axis represents a dissection's percentage of cycling OPCs (G), ratio of COPs to OPCs (H), or ratio of oligodendrocytes to OPCs (I). Asterisks indicate significant differences (Mann-Whitney,  $P < 10^{-4}$ ). The y axis in (G) was trimmed to 0.01; one telencephalic dissection contained 0.04% cycling cells. (J) Ratios of oligodendrocytes to OPCs in each tissue section profiled with EEL; SN-RN originated from an anterior position.

As expected, astrocytes exhibited more sharply delineated regional heterogeneity than oligodendrocytes. The mouse has two major astrocyte types—telencephalic and nontelencephalic—that contain gray matter (*Gfap*-negative) and white matter (*Gfap*-positive) subtypes (I). We found this molecular architecture conserved in humans: 13 *astrocyte* clusters were organized into telencephalic and nontelencephalic types (type 1 and 2 respectively), each containing *GFAP*-low and -high populations (Fig. 5, A and B, and fig. S14). Type 1 included known human cortical populations: *WIF1*+ gray matter, *TNC*+ white matter, and *LMO2*+ interlaminar astrocytes (2). Interlaminar astrocytes appeared especially distinct.

However, our data revealed additional regional heterogeneity, with astrocyte subtypes

organized according to developmental and anatomical compartments (Fig. 5, D and E, and fig. S14, D to F). Mirroring cortical neurons, gray matter cortical astrocytes included two clusters enriched in the neocortex and in the entorhinal and piriform cortex, respectively (no. 52 and no. 54). Whereas the former was mostly specific to the cortex, the latter was also found in the hippocampus and amygdala. Striatal astrocytes formed their own distinct cluster (no. 55). Globus pallidus astrocytes resembled midbrain and thalamic astrocytes and therefore represented the only major telencephalic type 2 cells (fig. S14F); other type 2 cells derived from the diencephalon, midbrain, and hindbrain. LDA revealed additional heterogeneity orthogonal to clusters, including a topic enriched in the globus pallidus and well re-

presented by *FOS* and *JUN* expression (Fig. 5, F to H, and fig. S15). These genes are downstream of many cellular processes including ischemia, which suggests that pathological conditions drove the classification of many globus pallidus astrocytes. Two other topics (11 and 8) were well represented by potential markers for reactive astrocytes (23) (Fig. 5H and fig. S15D). Together, these results emphasize both the regional heterogeneity of astrocytes and the diverse molecular programs with which they react to their environments.

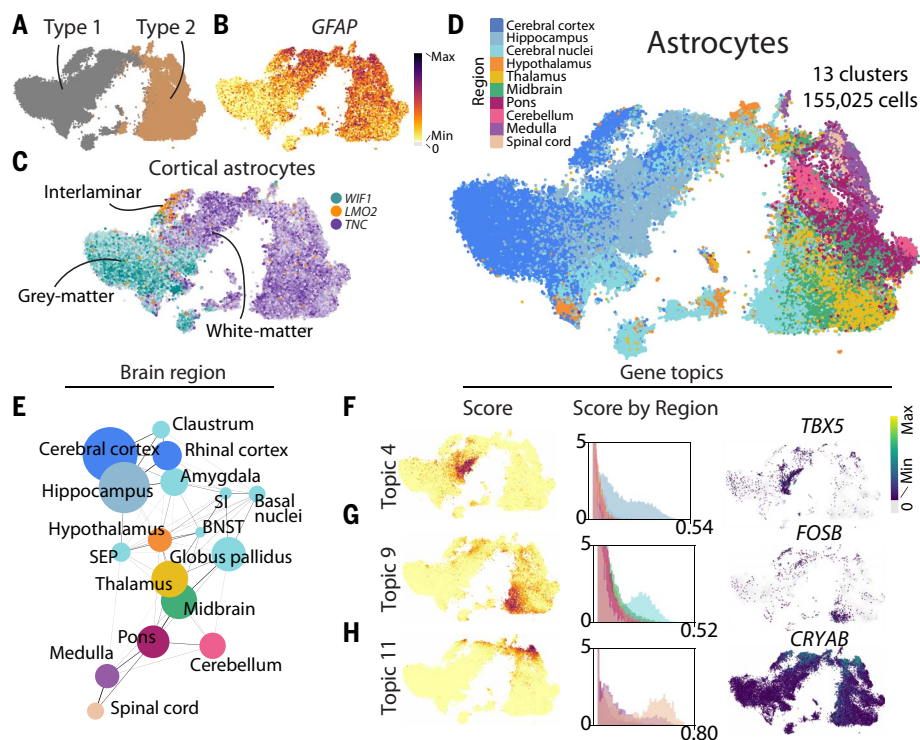
## Discussion

In this work, we provide an overview of transcriptomic diversity across the brain. Future work must disentangle whether these distinct transcriptomic states exhibit different epigenomic and functional properties. Our study highlights some of the challenges associated with profiling human tissue. Dissections were difficult to distinguish and replicate across donors, preventing us from definitively mapping cell types to their precise anatomical locations (see the “Technical notes on the dataset” section in the Materials and methods; table S4). We also likely observed pathological states, such as ischemia (in astrocytes, for example) (Fig. 5), but cannot confirm this observation without more extensive donor sampling. Our dataset does not permit conclusions about donor-specific diversity; how cell types vary by age, sex, and behavioral and disease states will be exciting avenues for future research.

Nevertheless, our study design enabled us to make important observations about regional variation across the brain. We identified a notable gap in our current understanding of neuronal diversity outside the telencephalon, reflecting these regions' anatomical complexity. *Splatter* neurons expressed neuropeptides, neurotransmitters, and other neuronal genes in unexpectedly complex patterns. For instance, *TH*+, dopaminergic, and serotonergic neurons expressed diverse neuropeptides and a graded range of GABAergic markers, generating extraordinary molecular diversity. This result indicates that future work must profile individual brain nuclei to fully ascertain their cellular diversity.

Our results underscore how strongly regional and developmental origins shape adult transcriptomic types. Both astrocytes and OPCs exhibited telencephalic and nontelencephalic types that transitioned near the hypothalamus (figs. S12G and S14D). Astrocytes showed additional region-specific types, suggesting specialized functions in support of local neurons. Some superclusters reflected migratory cell types that appeared in multiple dissections. A thorough understanding of region-specific cellular diversity and its development will therefore be key for treating disease. For example, the distinctive composition of the telencephalon's





**Fig. 5. Astrocyte diversity.** (A to D) t-SNE showing two major astrocyte types (A); GFAP expression (B); expression of the gene expressed most highly among *WIF1*, *LMO2*, and *TNC* (C); and regions of origin (D). (E) Brain region similarity by astrocyte diversity. Node size represents number of cells, and edge width represents percentage of cells in the regions or dissections that were neighbors. BNST, bed nucleus of the stria terminalis; SEP, septal nuclei; SI, substantia innominata and nearby nuclei. (F to H) (Left) Cells colored as in (B) by their scores for LDA topics. (Center) Score distributions for each region colored as in (D). The y axis is trimmed at 5. (Right) Expression of topic-representative genes.

oligodendrocyte and the midbrain's dopaminergic lineages might be relevant for diseases such as multiple sclerosis and Parkinson's. Our work therefore provides a critical foundation for exploring the brain's diverse neural circuitry and its implications for human health.

## Materials and methods

### Human postmortem tissue specimens

Deidentified postmortem adult human brain tissue was obtained after receiving permission from the deceased's next-of-kin. Tissue collection was performed in accordance with the provisions of the United States Uniform Anatomical Gift Act of 2006 described in the California Health and Safety Code section 7150 (effective 1/1/2008) and other applicable state and federal laws and regulations. The Western Institutional Review Board reviewed tissue collection procedures and determined that they did not constitute human subjects research requiring institutional review board (IRB) review. The collection and processing of human tissue was approved under Swedish law by the Swedish Ethical Review Authority (2019-03054).

Male and female donors 18 to 68 years of age with no known history of neuropsychiatric or neurological conditions were considered for

inclusion in the study. Routine serological screening for infectious disease (HIV, hepatitis B, and hepatitis C) was conducted using donor blood samples and donors testing positive for infectious disease were excluded from the study. Specimens were screened for RNA quality and samples with average RNA integrity (RIN) values  $\geq 7.0$  were considered for inclusion in the study. We sought to minimize the post-mortem interval, with further exclusionary criteria based on evidence of head trauma, intubation, neuropathology, and homicide. Brain specimens meeting these criteria, and for which whole brain hemispheres could be obtained for the current study, were quite rare and with a heavy male bias. Between 2018 and 2022, a total of 16 donors met these criteria, with only three female donors that ultimately failed quality control (QC) criteria. The three donors passing all exclusionary and QC criteria were male.

Postmortem brain specimens were processed as previously described (24). Briefly, coronal brain slabs were cut at 1-cm intervals, frozen in dry ice-cooled isopentane, and transferred to vacuum-sealed bags for storage at  $-80^{\circ}\text{C}$  until the time of further use. To isolate regions of interest, tissue slabs were briefly transferred to  $-20^{\circ}\text{C}$ , and the region of interest was removed

and subdivided into smaller blocks on a custom temperature controlled cold table. Tissue blocks were stored at  $-80^{\circ}\text{C}$  in vacuum-sealed bags until later use.

### Tissue sampling and single-nucleus RNA sequencing

We sampled all but 12 dissections from three postmortem donors, although dissections were sometimes combined differently across donors to achieve this goal, e.g., PAG became PAG-DR in one donor. For most dissections, we enriched for neurons with fluorescence-activated nuclei sorting. Nucleus isolation was conducted as described (25). Gating on 4',6-diamidino-2-phenylindole (DAPI) and NeuN fluorescence intensity was also described previously (2). NeuN+ and NeuN- nuclei were sorted into separate tubes and pooled at a defined ratio after sorting. Sorted samples were centrifuged, frozen in a solution of 1X phosphate-buffered saline (PBS), 1% bovine serum albumin (BSA), 10% dimethyl sulfoxide (DMSO), and 0.5% RNasin Plus RNase inhibitor (Promega, N2611), and stored at  $-80^{\circ}\text{C}$  until the time of shipment on dry ice from the Allen Institute to the Karolinska Institute for 10X chip loading. One shipment was accidentally switched for airplane parts by the courier and sent to a company in Ireland (we received the airplane parts); once the error was discovered and rectified, the samples arrived thawed but nevertheless yielded data that passed QC and are included in the dataset (table S2, notes in column M).

Immediately before loading on the 10x Chromium instrument, frozen nuclei were thawed in a  $37^{\circ}\text{C}$  water bath, spun down briefly, and pipetted several times to mix (26). Nuclei were then quantified and loaded according to the 10X Genomics protocol, targeting 5000 cells. We aimed to sequence two replicates per sample. Nearly all samples were processed with the 10x Genomics V3 kit; 12 samples were sequenced with v3.1 (table S2). Samples were first sequenced to a shallow depth ( $\sim 1000$  reads per cell) on the Illumina NextSeq platform to validate sample concentrations. Samples were then sequenced to  $\sim 100,000$  reads per cell on the Illumina NovaSeq platform. Sequencing saturation was examined for each sample using the preseq package (<https://github.com/smithlabcode/preseq>). Any samples that were not saturated to 60% were sequenced more deeply using preseq predictions.

### Data preprocessing

Sequencing runs were demultiplexed with cellranger mkfastq version 4.0.0 (10x Genomics) and filtered through the index-hopping-filter tool version 1.1.0 (10x Genomics). Unique molecular identifier (UMI) counts were determined using STARsolo version 2.7.10a with the following parameters: `-soloFeatures Gene` Velocyto; `-soloBarcodeReadLength 0`; `-solo`



Type CB\_UMI\_Simple; -soloCellFilter Empty Drops\_CR %s 0.99 10 45000 90000 500 0.01 20000 0.01 10000; -soloCBmatchWlType IMM\_multi\_Nbase\_pseudocounts; -soloUMI filtering MultiGeneUMI\_CR; -soloUMIdedup IMM\_CR; -clipAdapterType CellRanger4; -outFilterScoreMin 30.

Barcode whitelists were downloaded from the 10x Genomics website. Exonic, intronic, and ambiguous counts were summed for cluster-ing analysis.

The reference genome and transcript annotations were based on the human GRCh38. p13 gencode V35 primary sequence assembly. However, we filtered the reference. Because our pipeline only counted reads that uniquely aligned to one gene, reads that aligned to more than one gene were lost. Related genes therefore exhibited few or zero counts. To minimize this loss, we discarded overlapping genes or transcripts that overlapped or mapped to other genes or noncoding RNAs' 3' untranslated regions (3'UTRs), leaving only one of these transcripts in the genomic reference. We used BLAST to align the last 400 nucleotides (3'UTR) of all protein-coding and non-coding transcripts to all other genes (maximum 4 mismatches, minimum alignment length 300 nucleotides). We resolved all the matches by the following procedure:

First, fusion genes were filtered based on their names: genes with names that contained both fusion genes ["gene1-gene2"] were discarded. Second, noncoding transcripts that matched a coding transcript were discarded. Third, for overlapping transcripts where both transcripts were either coding or noncoding:

(i) If one of the gene names matched the pattern "XX#####." its transcript was discarded.  
(ii) If one of the transcripts belonged to a gene with one or more transcripts that were already discarded during the procedure, it was discarded as well.

(iii) We discarded transcripts of the gene with fewest splice variants.

(iv) Otherwise, the transcript with a 3'UTR that overlapped the other transcript was kept in the genomic reference.

(v) For paralogs, we mapped all related genes (that aligned to one another) and selected the paralog with most splice variants. All other highly similar paralogs were discarded. In special cases, we manually chose the gene.

Altogether we filtered 387 fusion genes, 1140 overlapping transcripts, 414 noncoding transcripts, 1127 coding paralogs, and 350 non-coding paralogs. The reference .gtf file and a list of the filtered genes and transcripts can be found at <https://github.com/linnarsson-lab/adult-human-brain>.

#### Quality control

10x Chromium samples that captured more than 15,000 cells were not analyzed further

(table S2). All remaining samples were pooled with their replicates and analyzed with "cytograph qc" (<https://github.com/linnarsson-lab/adult-human-brain>, commit 5f07e59768844e3c84fceb8fd7d1993059e13a7e), which uses a modified version of DoubletFinder to calculate a doublet score for each cell (27). We captured a median 6224 cells from each sample and then filtered cells based on their total number of mRNA molecules—as counted by unique molecular identifiers (UMIs)—and percentages of unspliced RNA, as well as doublet scores. Cells with fewer than 800 UMIs, fewer than 40% unspliced molecules, or a doublet score below 0.3 were removed from further analysis. These thresholds were determined by examining distributions across the dataset (fig. S1).

Total genes captured per cell varied across dissections partly due to biological variation in cell size. For example, nonneuronal cells are small and yield fewer molecules. Although we aimed to collect 90% neurons and 10% non-neuronal cells from each sample, we were not able to achieve this proportion in many non-cortical regions due to an abundance of oligodendrocytes (fig. S1B). We also noted variability across donors that likely reflected tissue quality (fig. SIC).

#### Clustering

Cells were clustered using an updated version of the Cytograph package (<https://github.com/linnarsson-lab/adult-human-brain>, commit 22a83ebcf79a9fdcdff34beb0300aa1695d7039b) (7). The command "cytograph process" was run with configuration *min\_umis*: 0, *doublets\_action*: None, *features*: variance, *remove\_low\_quality*: False, *remove\_doublets*: False, *batch\_keys*: ["Donor"], *clusterer*: sknetwork, *steps*: nn, embeddings, clustering. Default values were used for other parameters. In brief, principal components analysis (PCA) was performed on the most variable genes and used to construct a k-nearest neighbors graph, which provided input for Louvain clustering (scikit-network). Harmony with default parameters integrated principal components across donors before constructing the graph (28).

This clustering pipeline was used at four levels of analysis: (i) All cells that passed QC were pooled into a single dataset and clustered. (ii) This top-level clustering was split by the dendrogram into two groups that corresponded to neuronal and nonneuronal cells, with one exception: lymphocytes clustered with neurons. Each of these two groups was separately analyzed. (iii) Within each group, Paris clustering was used on the neighborhood graph to find "superclusters" for further analysis (Fig. 1) (29). The Paris dendrogram was arbitrarily cut so that the superclusters mirrored large islands on the t-distributed stochastic neighbor embedding (t-SNE). Each superclus-

ter was separately analyzed to produce "clusters." (iv) Each of the 461 clusters was separately analyzed to produce "subclusters."

At level 3, additional cluster-level quality control was performed. A cluster was removed from further analysis if: (i) its mean percentage of mitochondrial UMIs was greater than the 99th percentile mean (0.056); or (ii) if the Cytograph pipeline tagged the cluster with two or more of the following auto-annotations: NEUR, ASTRO, OLIGO, OPC, MGL, ENDO, BERG, or VLNC. Auto-annotations and further information about the auto-annotation procedure are available at <https://github.com/linnarsson-lab/auto-annotation-ah>. After removing these clusters, the remaining cells in each supercluster were reclustered. This procedure was repeated until no clusters met these criteria.

Finally, clusters and subclusters at levels 3 and 4, respectively, underwent a merging procedure ("cytograph merge"). A cluster that did not express at least one statistically enriched gene was merged with the nearest cluster on the t-SNE. In short, false discovery rate (FDR)-corrected *P* values were calculated by comparing Cytograph's enrichment scores with a null distribution obtained by permuting cluster labels across all cells in the supercluster.

A t-SNE and dendrogram were calculated for the final 461 clusters (Fig. 1). The dendrogram was calculated based on Euclidean distances between cluster means for the 2000 most variable genes. Cluster means were normalized to median cluster means across the dataset and then standardized (*scikit-learn* StandardScaler).

The script for splitting data into superclusters and creating Cytograph punchcards is available as *split.py* at <https://github.com/linnarsson-lab/adult-human-brain>.

#### Annotation

Superclusters were manually annotated based on the literature and on their regional composition. The *splatter* supercluster was named for its splatter-like appearance on the t-SNE. The *miscellaneous* supercluster contained lymphocytes and very rare neuronal cell types (in total only 25,000 cells); the sparsity of these cells on the neighborhood graph likely drove this grouping.

Clusters were not named with a single annotation but instead tagged with auto-annotations (<https://github.com/linnarsson-lab/auto-annotation-ah>) from four categories: neurotransmitters (auto-annotation-ha/Human\_adult/Neurotransmission), neuropeptides (auto-annotation-ha/Neuropeptides), class (auto-annotation-ha/Human\_adult/Class), and subtype (auto-annotation-ha/Human\_adult/Subtype). Neuropeptide-related tags were manually compiled based primarily on the NeuroPep database (30). Subtype auto-annotations

included manually selected cell types and states previously described in the literature. To identify cortical cell types, we also transferred labels from a single-cell analysis of the middle temporal gyrus (MTG) (31). We performed PCA (50 components) on our full dataset, trained a random forest classifier (scikit-learn, `class_weight='balanced'`, `max_depth=50`) on the MTG labels, and then predicted labels for all cells. We labeled each cluster with the mode of its constituent cells if two conditions were met: more than 0.8 of predicted labels matched the mode, and the mean probability of these predictions was greater than 0.8. Clusters that did not meet these criteria were labeled N/A.

#### Technical notes on the dataset

Our annotation effort revealed two experimental caveats of our work. First, distinguishing anatomical borders between dissections was challenging. Dissections were therefore not always identical across donors and might contain adjacent tissue. Two diencephalic dissections contained telencephalon, for example, and therefore contributed to telencephalic superclusters (HTHso and VA; Fig. 2A). We accordingly further annotated each dissection to indicate such ambiguities (table S4).

Second, we sequenced multiple dissections together at once, and sequencing reads were occasionally misassigned across samples due to index hopping (32). Because index hopping usually yields cells with very few UMIs, we flagged additional index-hopped cells by examining each supercluster for samples that only contributed cells with unusually few UMIs. Note that such samples might alternatively represent low-quality cells, so this approach does not exclusively identify index-hopped cells. For each supercluster, we calculated the fifth percentile of the distribution of total UMIs per cell. Then we flagged samples if they contributed more than one cell to the supercluster and 90% of these cells fell below this UMI threshold. We identified 64 such samples, but most contributed only a few cells. Only two samples contributed more than 100 such cells to a supercluster: 10X249-1 (V2) to *deep-layer IT*; 10X385-3 (GPi) to *lower rhombic lip*. These results suggest that index hopping minimally affected the dataset.

However, these observations suggest that the data should not be used to definitively map cell types to dissections. Instead, the dissection labels facilitate a minimal estimate of how extensively cell types differ across neighboring anatomical locations.

#### Integrating human and mouse striatal data

We merged our striatal datasets (CaB, Pu, NAC) with striatal data from Saunders *et al.* (5) (striatum at [http://dropviz.org/?\\_state\\_id\\_=39c454601e4bd7c2](http://dropviz.org/?_state_id_=39c454601e4bd7c2)), retaining genes with homologs in <http://www.informatics.jax.org/downloads/>

[reports/HMD\\_HumanPhenotype.rpt](#). To integrate the data, we selected genes that were among the top 2000 highly variable genes in either dataset. We then used the Cytograph pipeline as described under “Clustering” but with `batch_keys`: [“Species”], which prompts Harmony to operate along the cell attribute “Species” that we added to the joint dataset.

The script is available as `prepare_integration.py` at <https://github.com/linnarsson-lab/adult-human-brain>.

#### RNAscope multiplex fluorescence in situ hybridization (mFISH)

Fresh-frozen human postmortem brain tissues were sectioned at 10  $\mu$ m onto Superfrost Plus glass slides (Fisher Scientific). Sections were dried for 10 min at  $-20^{\circ}\text{C}$  and then vacuum sealed in plastic slide boxes and stored at  $-80^{\circ}\text{C}$  until use. The RNAscope multiplex fluorescent v2 kit was used per the manufacturer’s instructions for fresh-frozen tissue sections (ACD Bio), except that fixation was performed for 60 min in 4% paraformaldehyde in 1X PBS at  $4^{\circ}\text{C}$  and protease treatment was shortened to 15 min. Sections adjacent to those used for RNAscope were fixed for 15 min in 4% paraformaldehyde in 1X PBS, briefly washed in PBS, and stained with NeuroTrace 500/525 Green Fluorescent Nissl Stain and DAPI to aid in region localization. Nissl-stained sections were imaged using a 10X objective on a Nikon TiE fluorescence microscope equipped with NIS-Elements Advanced Research imaging software (v4.20, RRID:SCR\_014329). RNAscope sections were imaged using a 40X oil immersion lens on the same microscope. For mFISH experiments, positive cells were called by manually counting RNA spots for each gene. Cells were called positive for a gene if they contained  $\geq 5$  RNA spots for that gene. Lipofuscin autofluorescence was distinguished from RNA spot signal based on the larger size of lipofuscin granules and broad fluorescence spectrum of lipofuscin. Staining for the probe combination was repeated with similar results on at least three separate sections from one human donor. Images were assessed with FIJI distribution of ImageJ v1.52p and Nikon NIS-Elements imaging software. Probes used were Hs-OTX2-C3 (no. 484581-C3), Hs-SOX14-C2 (no. 1055351-C2), and Hs-CASR (no. 411931) from ACD Bio.

#### Assessing anatomical heterogeneity within superclusters

For each dissection that contributed more than 100 cells to a neuronal supercluster, we calculated the fraction of cells that derived from each cluster in the supercluster. We calculated this “cluster-proportion matrix” separately for each supercluster, excluding *miscellaneous* immune-cell clusters. Figure S4B shows these matrices as heatmaps. We used the matrices to

plot fig. S4A, counting for each cluster the number of dissections that comprised at least 5% cells from that cluster. Each cluster is thus its own dot on the plot.

These cluster-proportion matrices were also used to calculate neighborhood graphs (Fig. 2B and fig. S4, C and D). Each dissection was connected to the two dissections with the most similar cluster proportions. Nonneuronal cells were excluded, because similar proportions of neurons and nonneuronal cells were not sorted for all samples. Similarity was calculated using the *sklearn* NearestNeighbors function, `metric='correlation'`.

Correlations between dissections were visualized (fig. S4E) by generating correlation matrices from the cluster-proportion matrices (the *numpy* `corrcoef` function). To generate the final heatmap in the figure panel, which indicates dissections that were always more correlated to one another than to any other dissections, mean and standard deviation were calculated for each row of the correlation matrix. Then dissections in each row were flagged if their correlation values were more than 1.5 standard deviations away from the mean in all seven cortical superclusters. We moreover enforced symmetry: dissections must have been flagged for one another to be indicated on the heatmap.

#### Assessing supercluster complexity with PCA

We used a permutation test to determine the number of principal components necessary to describe each supercluster, reasoning that this value would be proportional to the supercluster’s complexity. We performed PCA twice on cells sampled from each supercluster (*scikit-learn* PCA, `n_comps=50`): once on the original expression matrix, and once on a permuted expression matrix, where the expression values of each gene were randomly redistributed across cells. We compared the percentage of variance that each principal component explained (attribute “`explained_variance_ratio_`”) in the original and unpermuted matrix. We considered the first component where `explained_variance_ratio_` was similar between the two analyses (less than 10% higher in the unpermuted analysis than in the permuted) to represent the optimal number of components to describe the supercluster (Fig. 3E).

This procedure was performed separately for each donor. 10,000 cells were randomly sampled from each donor for the analysis. Some superclusters or subsets did not contain 10,000 cells from a donor. From H18.30.002: Deep-layer NP, Hippocampus CA4, Mammillary body, Cerebellar inhibitory, COP, Ependymal, Vascular, Bergmann, Fibroblast, Choroid plexus, Serotonergic, Dopaminergic. From H19.30.001: Deep-layer NP, Miscellaneous, Hippocampus CA4, Mammillary body, Cerebellar inhibitory, COP, Ependymal, Vascular,



Bergmann, Fibroblast, Choroid plexus, Serotonergic, Dopaminergic. From H19.30.002: Deep-layer NP, Miscellaneous, Hippocampus CA4, Mammillary body, Cerebellar inhibitory, Lower rhombic lip, COP, Ependymal, Vascular, Bergmann, Fibroblast, Choroid plexus, Serotonergic, Dopaminergic.

The script is available as `permuted_pca.py` at <https://github.com/linnarsson-lab/adult-human-brain>.

### Correlating principal components with GABAergic identity

Neurons from all dissections were analyzed separately using the Cytograph parameters described under “Clustering.” For each dissection, we calculated the absolute value of the correlation between the first principal component and the log of *GAD2* expression. For dissections with more than 5000 *splatter* neurons, we additionally analyzed the *splatter* neurons separately and calculated the correlation value with the same approach.

### Assessing robustness of the splatter supercluster

We down-sampled our whole-brain dataset by randomly selecting 10, 100, or 1000 cells per cluster. We then analyzed each down-sampled dataset using the Cytograph parameters described under “Clustering.”

We performed additional analyses with the dataset generated by sampling 100 cells per cluster. We analyzed it using the Cytograph parameters described under “Clustering” but: (i) without Harmony (*batch\_keys*: []); (ii) with the 2000 most variable genes in the *splatter* supercluster; (iii) with 2000 randomly selected genes.

The script to prepare these variations of the dataset for Cytograph analysis is available as `downsample_data.py` at <https://github.com/linnarsson-lab/adult-human-brain>.

Lastly, we performed additional analyses with the dataset generated by sampling 10 cells per cluster. We (i) reclustered the data by fitting the Cytograph-generated principal components with a Gaussian Mixture Model [*sklearn* *GaussianMixture*(*n\_components*=10, *n\_init*=50)]; (ii) performed LDA as detailed in the section below, changing *k* until a *splatter*-enriched topic appeared at *k* = 11; (iii) removed all but five *mammillary body* neurons, selected the top 2000 most variable genes previously determined by Cytograph, normalized the data to the median of total UMIs per cell, calculated 50 principal components with *sklearn*, and recalculated the t-SNE using Cytograph’s `art_of_tsne` function; (iv) used the *sklearn* *KNeighborsClassifier* to predict the supercluster for each cell, restricting the function “fit” to use five of the cell’s neighbors outside its own supercluster to make the prediction. We then calculated for each supercluster the fraction of predictions that fell in each supercluster.

### Topic modeling with LDA

We used *tomotopy* version 0.12.2 to perform topic modeling with LDA. In brief, we trained a model with a specified number of topics “*k*” until perplexity stabilized (function “*LDAModel*” with parameters *alpha*=50/*k*, *eta*=0.1). We trained with no more than 200,000 cells and all protein-coding genes in the genome that were expressed in at least 100 cells and fewer than 60% of all cells. Some topics were donor-specific or quality-specific, e.g., mitochondrial or ribosomal, and some did not have an apparent biological meaning. We therefore identified topics of interest manually. We also identified representative genes for each topic by using the topic probabilities reported for each gene (function “*get\_topic\_word\_dist*”). We filtered genes by specificity (topic probabilities for each gene normalized by that gene’s probability across all topics) and then sorted the remaining genes by their unnormalized probabilities.

A script is available as `optimize_lda.py` at <https://github.com/linnarsson-lab/adult-human-brain>.

### Extended analysis of dopaminergic and TH+ neurons

To identify additional dopaminergic subtypes, subclusters 1870, 1871, 1873, and 1876 were reanalyzed separately as described under “Clustering.”

To compare our data with a recently published census of midbrain dopaminergic neurons from Kamath *et al.* (4), we trained a random forest classifier on nuclei from their dataset annotated as dopaminergic neurons from control samples with nonneurological disorders (donor IDs: 3298, 3322, 3345, 3346, 3482, 4956, 5610, 6173). We trained the classifier with features that Cytograph selects when run with the option to select cluster-enriched rather than variable genes (default parameters except *n\_factors*: 20 and *features*: enrichment). The *scikit-learn* *RandomForestClassifier* was regularized with the parameters *max\_features* and *min\_samples\_leaf*. Five-fold cross-validation was performed on data from Kamath *et al.* with *scikit-learn* *StratifiedShuffledSplit* for *max\_features* 0.3, 0.4, 0.5 and *min\_samples\_leaf* between 3 and 50. We selected *max\_features*=0.3 and *min\_samples\_leaf*=4 to transfer labels to our data; *oob\_score* was set true, and other parameters were left as default.

We also used Harmony to integrate our TH+ neurons and 1000 randomly sampled neurons from Kamath *et al.* We selected features for integration by using Cytograph’s Feature SelectionByVariance with default parameters on each dataset separately, and then taking the intersection of the two gene sets. Clusters were obtained using Cytograph’s *UnpolishedLouvain* with default parameters on the harmonized principal components. We

then transferred labels from Kamath *et al.* to our TH+ neurons by using *NNDescent* on the harmonized principal components. A nearest-neighbor graph was constructed using *NNDescent* with default parameters. Cell types defined by Kamath *et al.* were considered for transfer if they contributed more than 30% to one integrated cluster. Cells were assigned the most common label among their five nearest neighbors.

### EEL-FISH

EEL-FISH was performed as recently described (12). Donor H18.30.002 was used for the V1C sample; H20.30.002 for the AMY, MTG, and A46 samples; H19.30.001 for the pons sample; and H18.30.001 for the SN-RN and RRF samples. For all figure panels, genes were only plotted for an individual cell if the cell contained more than two molecules of that gene. Expression below this threshold was considered experimental noise.

We manually selected all 445 genes: 168 previously described markers for forebrain neurons and nonneuronal cells and 277 genes selected from our data. The latter were expressed in subtypes of *splatter* cells (200 genes), OPCs (40 genes), and astrocytes (37 genes). Genes were selected only if their mean expression was more than 5 UMIs in the target cluster.

Nuclei were segmented using *CellPose* based on the DAPI signal, and the segmentation masks were expanded by 8  $\mu$ m without overlapping (33). The RNA dots were only assigned to a segmentation mask if they were found inside.

### Reanalysis of adolescent mouse brain data

Adolescent mouse data were downloaded from *mousebrain.org*, including all L1 files from the amygdala, cerebellum, cortex, hippocampus, hypothalamus, medulla, midbrain, pons, striatum, and thalamus. Neuronal cells were reanalyzed with Cytograph as described under “Clustering,” but with *batch\_keys*: []. Zeisel *et al.* (1) removed a large number of cells, for several reasons. Most cells (>200,000 out of 492,949 cells) were oligodendrocytes, removed because they otherwise would skew the gene selection especially in hindbrain. Another ~100,000 cells were removed as suspected doublets, and as clusters without unique markers. The procedure was complex, proceeded in several stages, and is described in great detail in the original publication (1). We now realize that many cells removed in that analysis probably correspond to *splatter* neurons in the present work. Chromium v1 was 10 times less sensitive than current v3 chemistry (i.e., ~2000 UMI counts per cell for v1 versus ~20,000 for v3, even though we used whole cells then and only nuclei now). Clustering was therefore extremely difficult, and clusters without distinct

markers were removed in an effort to be conservative.

### Differential expression testing

5000 cells were randomly sampled from each of two types in each of three donors. Counts were summed within each group to generate six pseudobulk samples. DESeq2 was used to test for differential expression across the samples.

Code is available in a notebook at <https://github.com/linnarsson-lab/adult-human-brain>.

### Gene ontology enrichment analysis

The *gget* package (<https://github.com/pachterlab/gget>) was used to perform all gene ontology enrichment analysis (function “enrichr” with database=“ontology”) (34).

### REFERENCES AND NOTES

1. A. Zeisel *et al.*, Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018). doi: [10.1016/j.cell.2018.06.021](https://doi.org/10.1016/j.cell.2018.06.021); PMID: 30096314
2. R. D. Hodge *et al.*, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019). doi: [10.1038/s41586-019-1506-7](https://doi.org/10.1038/s41586-019-1506-7); PMID: 31435019
3. T. E. Bakken *et al.*, Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021). doi: [10.1038/s41586-021-03465-8](https://doi.org/10.1038/s41586-021-03465-8); PMID: 34616062
4. T. Kamath *et al.*, Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. *Nat. Neurosci.* **25**, 588–595 (2022). doi: [10.1038/s41593-022-01061-1](https://doi.org/10.1038/s41593-022-01061-1); PMID: 35513515
5. A. Saunders *et al.*, Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015–1030.e16 (2018). doi: [10.1016/j.cell.2018.07.028](https://doi.org/10.1016/j.cell.2018.07.028); PMID: 30096299
6. H. Wichterle, J. M. Garcia-Verdugo, D. G. Herrera, A. Alvarez-Buylla, Young neurons from medial ganglionic eminence disperse in adult and embryonic brain. *Nat. Neurosci.* **2**, 461–466 (1999). doi: [10.1038/8131](https://doi.org/10.1038/8131); PMID: 10321251
7. D. W. Kim *et al.*, Gene regulatory networks controlling differentiation, survival, and diversification of hypothalamic Lhx6-expressing GABAergic neurons. *Commun. Biol.* **4**, 95 (2021). doi: [10.1038/s42003-020-01616-7](https://doi.org/10.1038/s42003-020-01616-7); PMID: 33479483
8. K. Letinic, P. Rakic, Telencephalic origin of human thalamic GABAergic neurons. *Nat. Neurosci.* **4**, 931–936 (2001). doi: [10.1038/nm0901-931](https://doi.org/10.1038/nm0901-931); PMID: 11528425
9. P. Jager *et al.*, Tectal-derived interneurons contribute to phasic and tonic inhibition in the visual thalamus. *Nat. Commun.* **7**, 13579 (2016). doi: [10.1038/ncomms13579](https://doi.org/10.1038/ncomms13579); PMID: 27929058
10. C. F. Kratochwil, U. Maheshwari, F. M. Rijli, The Long Journey of Pontine Nuclei Neurons: From Rhombic Lip to Cortico-Ponto-Cerebellar Circuitry. *Front. Neural Circuits* **11**, 33 (2017). doi: [10.3389/fncir.2017.00033](https://doi.org/10.3389/fncir.2017.00033); PMID: 28567005
11. B. Tasic *et al.*, Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018). doi: [10.1038/s41586-018-0654-5](https://doi.org/10.1038/s41586-018-0654-5); PMID: 30382198
12. L. E. Borm *et al.*, Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nat. Biotechnol.* **41**, 222–231 (2023). doi: [10.1038/s41587-022-01455-3](https://doi.org/10.1038/s41587-022-01455-3); PMID: 36138169
13. D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
14. S. Karthik *et al.*, Molecular ontology of the parabrachial nucleus. *J. Comp. Neurol.* **530**, 1658–1699 (2022). doi: [10.1002/cne.25307](https://doi.org/10.1002/cne.25307); PMID: 35134251
15. J. L. Pauli *et al.*, Molecular and anatomical characterization of parabrachial neurons and their axonal projections. *eLife* **11**, e81868 (2022). doi: [10.7554/eLife.81868](https://doi.org/10.7554/eLife.81868); PMID: 36317965
16. C. J. Wylie *et al.*, Distinct transcriptomes define rostral and caudal serotonin neurons. *J. Neurosci.* **30**, 670–684 (2010). doi: [10.1523/JNEUROSCI.4656-09.2010](https://doi.org/10.1523/JNEUROSCI.4656-09.2010); PMID: 20071532
17. B. W. Okaty *et al.*, A single-cell transcriptomic and anatomic atlas of mouse dorsal raphe *Pet1* neurons. *eLife* **9**, e55523 (2020). doi: [10.7554/eLife.55523](https://doi.org/10.7554/eLife.55523); PMID: 32568072
18. N. X. Tritsch, A. J. Granger, B. L. Sabatini, Mechanisms and functions of GABA co-release. *Nat. Rev. Neurosci.* **17**, 139–145 (2016). doi: [10.1038/nrn.2015.21](https://doi.org/10.1038/nrn.2015.21); PMID: 26865019
19. J. R. Moffitt *et al.*, Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018). doi: [10.1126/science.aau5324](https://doi.org/10.1126/science.aau5324); PMID: 30385464
20. S. Jäkel *et al.*, Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019). doi: [10.1038/s41586-019-0903-2](https://doi.org/10.1038/s41586-019-0903-2); PMID: 30747918
21. L. A. Seeker *et al.*, Brain matters: Unveiling the distinct contributions of region, age, and sex to glia diversity and CNS function. *Acta Neuropathol. Commun.* **11**, 84 (2023). doi: [10.1186/s40478-023-01568-z](https://doi.org/10.1186/s40478-023-01568-z); PMID: 37217978
22. L. Schirmer *et al.*, Neuronal vulnerability and multilineage diversity in multiple sclerosis. *Nature* **573**, 75–82 (2019). doi: [10.1038/s41586-019-1404-z](https://doi.org/10.1038/s41586-019-1404-z); PMID: 31316211
23. C. Escartin *et al.*, Reactive astrocyte nomenclature, definitions, and future directions. *Nat. Neurosci.* **24**, 312–325 (2021). doi: [10.1038/s41593-020-00783-4](https://doi.org/10.1038/s41593-020-00783-4); PMID: 33589835
24. Allen Institute for Brain Science, Human Post Mortem Brain Processing (protocols.io, 2020); doi: [10.17504/protocols.io/bf4ajqse](https://doi.org/10.17504/protocols.io/bf4ajqse)
25. LeinU01 BRAIN grant, Isolation of Nuclei from Adult Human Brain Tissue for 10x Genomics Platform (protocols.io, 2019); doi: [10.17504/protocols.io/y6rfzd6](https://doi.org/10.17504/protocols.io/y6rfzd6)
26. K. Siletti, S. Linnarsson, Preparation of frozen nuclei for single-nucleus RNA seq (protocols.io, 2022); doi: [10.17504/protocols.io.eq2lyrdpgx9k/v2](https://doi.org/10.17504/protocols.io.eq2lyrdpgx9k/v2)
27. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337.e4 (2019). doi: [10.1016/j.cels.2019.03.003](https://doi.org/10.1016/j.cels.2019.03.003); PMID: 30954475
28. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). doi: [10.1038/s41592-019-0619-0](https://doi.org/10.1038/s41592-019-0619-0); PMID: 31740819
29. T. Donald, B. Charpentier, A. Galland, A. Hollocou, Hierarchical Graph Clustering using Node Pair Sampling. *arXiv:1806.01664 [cs.LG]* (2018).
30. Y. Wang *et al.*, NeuroPep: A comprehensive resource of neuropeptides. *Database* **2015**, bav038 (2015). doi: [10.1093/database/bav038](https://doi.org/10.1093/database/bav038); PMID: 25931458
31. N. L. Jorstad *et al.*, Comparative transcriptomics reveals human-specific cortical features. *Science* **382**, eade9516 (2023). doi: [10.1126/science.ade9516](https://doi.org/10.1126/science.ade9516)
32. R. Farouni, H. Djambazian, L. E. Ferri, J. Ragoussis, H. S. Najafabadi, Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing. *Nat. Commun.* **11**, 2704–2708 (2020). doi: [10.1038/s41467-020-16522-z](https://doi.org/10.1038/s41467-020-16522-z); PMID: 32483174
33. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: A generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021). doi: [10.1038/s41592-020-01018-x](https://doi.org/10.1038/s41592-020-01018-x); PMID: 33318659
34. L. Luebbert, L. Pachter, Efficient querying of genomic reference databases with *gget*. *Bioinformatics* **39**, btac836 (2023). doi: [10.1093/bioinformatics/btac836](https://doi.org/10.1093/bioinformatics/btac836); PMID: 36610989
35. S. Linnarsson *et al.*, linnarsson-lab/adult-human-brain: Revision 1, version rev20230123, Zenodo (2023); doi: [10.5281/zenodo.7561827](https://doi.org/10.5281/zenodo.7561827)

### ACKNOWLEDGMENTS

We thank the Linnarsson group for feedback and discussion, especially A. Johnsson for laboratory management, E. Kvedaraitė for immune cell annotation, E. Vinsland for fibroblast and vasculature annotation, and I. Kapustová for artwork. We acknowledge support from the National Genomics Infrastructure in Stockholm. We thank the tissue procurement, tissue processing, and facilities teams at the Allen Institute for Brain Science for assistance with the transport and processing of human brain specimens and the technology team at the Allen Institute for assistance with data management. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This publication was supported by and coordinated through the NIH's Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative - Cell Census Network (BICCN). This publication is part of the Human Cell Atlas ([www.humancellatlas.org/publications/](http://www.humancellatlas.org/publications/)). **Funding:** This study was supported by European Molecular Biology Organization fellowship ALTF 1015-2018 (K.S.); US NIH grant U01 MH114812-02 (S.L. and E.S.L.); the Chan Zuckerberg Initiative and the Silicon Valley (NDON 2018-191929) (E.A. and S.L.); the Knut and Alice Wallenberg Foundation (2015.0041, 2018.0172, and 2018.0220 to S.L. and 2018.0232 to E.A.); the Swedish Foundation for Strategic Research (SB16-0065) (S.L. and E.A.); the Torsten Söderberg Foundation (S.L.); European Union grants ERC-ADG 884608 (E.A.), H2020 874758 (NSC-Reconstruct) (E.A.); and Vetenskapsrådet 2020-01426, Karolinska Institutet (StratRegen SFO 2018) and Hjärtfonden F02019-0068 (E.A.). **Author contributions:** Study design: S.L., E.S.L., T.B., S.-L.D., R.H., and K.S. Tissue dissections and nuclei isolation: R.H., S.-L.D., M.C., T.C., N.D., J.G., C.D.K., J.N., H.T., A.M.Y., D.H., and N.V.S. RNA sequencing: L.H. and K.S. RNA data preprocessing: P.L. RNA data analysis: K.S. and S.L. RNA data annotation: K.S. and A.M.A. EEL experiments and analysis: A.M.A. Dopaminergic analysis: K.W.L. and E.A. RNAScope: R.H. Writing: K.S., A.M.A., E.A., K.W.L., and R.H. Editing: K.S., S.L., and E.A. **Competing interests:** S.L. is a paid scientific advisor to Molecular, Combigen, and the Oslo University Center of Excellence in Immunotherapy. S.L. and A.M.A. are shareholders in EEL Transcriptomics AB, which owns the intellectual property to the EEL method (“Spatial RNA localization method,” provisional US patent application 63/139,701). The authors declare no other competing interests. **Data and materials availability:** Raw data were generated for the BRAIN Initiative Cell Census Network (RRID:SCR\_015820) and are available for download from the Neuroscience Multi-omics Archive (RRID:SCR\_016152) at [https://data.nemoarchive.org/biccn/grant/u01\\_lein/linnarsson/transcriptome/sncell/10x\\_v3/human/raw/](https://data.nemoarchive.org/biccn/grant/u01_lein/linnarsson/transcriptome/sncell/10x_v3/human/raw/). The analyzed data can be browsed and downloaded at CELLxGENE (<https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443>). Code used for analysis and figure generation is available at <https://github.com/linnarsson-lab/adult-human-brain> [archived at Zenodo (35)]. Auto-annotations are available at <https://github.com/linnarsson-lab/auto-annotation-ah>. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

### SUPPLEMENTARY MATERIALS

[science.org/doi/10.1126/science.add7046](https://science.org/doi/10.1126/science.add7046)

Figs. S1 to S15

Tables S1 to S6

MDAR Reproducibility Checklist

Submitted 29 September 2022; accepted 7 September 2023  
10.1126/science.add7046