RESEARCH ARTICLE SUMMARY

BICCN

Transcriptomic cytoarchitecture reveals principles of human neocortex organization

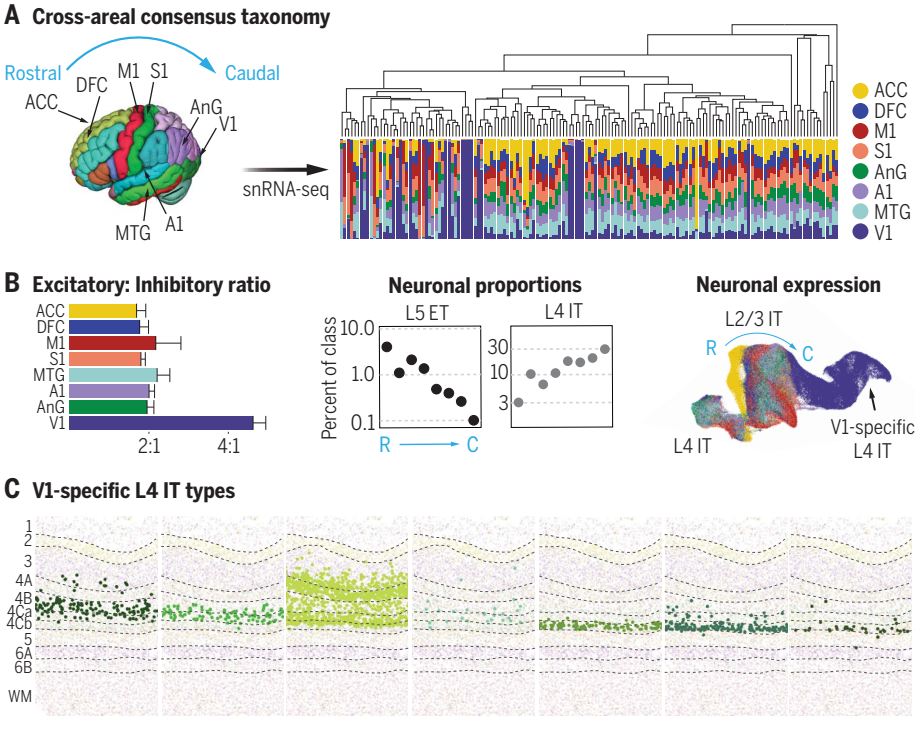
Nikolas L. Jorstad, Jennie Close, Nelson Johansen, Anna Marie Yanny, Eliza R. Barkan, Kyle J. Travaglini, Darren Bertagnolli, Jazmin Campos, Tamara Casper, Kirsten Crichton, Nick Dee, Song-Lin Ding, Emily Gelfand, Jeff Goldy, Daniel Hirschstein, Katelyn Kiick, Matthew Kroll, Michael Kunst, Kanan Lathia, Brian Long, Naomi Martin, Delissa McMillen, Trangthanh Pham, Christine Rimorin, Augustin Ruiz, Nadiya Shapovalova, Soraya Shehata, Kimberly Siletti, Saroja Somasundaram, Josef Sulc, Michael Tieu, Amy Torkelson, Herman Tung, Edward M. Callaway, Patrick R. Hof, C. Dirk Keene, Boaz P. Levi, Sten Linnarsson, Partha P. Mitra, Kimberly Smith, Rebecca D. Hodge*, Trygve E. Bakken*, Ed S. Lein*

INTRODUCTION: The human neocortex is generally organized into six layers of neurons but the size and cellular composition of these layers varies across the cortex, and this variation is thought to underlie differences in connectivity that impart specific functional specialization to distinct cortical areas. However, the degree to which cortical areas have a canonical versus noncanonical organization has proved difficult to reliably quantify. Single-nucleus and spatial transcriptomic methods enable high-resolution characterization of the cellular structure of the human neocortex providing a means to quantitatively compare the molecular and cellular structure and specialization of distinct cortical areas.

RATIONALE: Eight cortical areas that are representative of major variation in cellular architecture and include primary sensory and association cortices were sampled using single-nucleus transcriptomics to generate a dataset comprised of more than 1.1 million nuclei.

RESULTS: Nuclei were grouped based on gene expression similarity into one of 24 cellular subclasses, which were found in all cortical areas. Layer 4 intratelencephalic excitatory neurons were present even in agranular areas that lacked a histologically distinct layer 4, suggesting a common subclass-level cellular blueprint across the cortex. However, gene expression and subclass proportions varied substantially between cortical areas, with more differences in excitatory projection neurons than inhibitory neurons. All non-neuronal subclasses were shared across cortical areas but their laminar distributions varied between areas, and astrocytes also expressed regional marker genes. Variation as a function of rostrocaudal location in the cortex was a clear organizational feature where neighboring cortical areas were most similar, in line with previous observations of gene expression similarity by topographic proximity in the cortex. At a finer cell-type level of analysis, area-enriched and area-specific cell types were apparent in multiple cortical areas, but most notably in the primary visual cortex (V1) that had many distinct excitatory neuron types and several distinct inhibitory neuron types that reflect the specialized cellular architecture of this area in humans and other primates. V1 specialized inhibitory cell types were mostly Somatostatin-expressing neurons likely originating from the medial ganglionic eminence during development. Layer 4 in V1, which is visibly enlarged and has multiple sublayers, was notably different from other areas with discrete sublaminar distributions of specialized excitatory and inhibitory neurons revealed by spatial transcriptomics.

CONCLUSION: A common set of cell types are found across human cortical areas that have diverse functions. Excitatory projection neurons exhibit large spatial gradients and regional differences in proportions, laminar distributions, and gene expression that are less pronounced in inhibitory neurons or non-neuronal cells. V1 is molecularly distinct from other cortical areas and several excitatory and inhibitory neuronal types are found only in V1. The ratio of excitatory to inhibitory neurons in V1 is also more than double that of other cortical areas, reflecting specialization of the human cortex for processing visual information. ■



Regional specializations of human cortical cell types. (**A**) Single nucleus RNA-sequencing data from eight areas of the human neocortex were used to generate a cross-areal taxonomy with shared and area-specific types. (**B**) The relative number of excitatory and inhibitory neurons is similar across all areas except V1. Neuronal cell-type proportions and gene expression varied systematically from the rostral (R) to caudal (C) cortex with additional regional signatures. (**C**) Some excitatory neuron types that are exclusive to V1 are in the visual input layer 4, which is expanded compared to other cortical areas.

The list of author affiliations is available in the full article.
*Corresponding author. Email: edl@alleninstitute.org (E.S.L.); trygveb@alleninstitute.org (T.E.B.); rebeccah@alleninstitute.org (R.D.H.)
Cite this article as N. L. Jorstad *et al.*, *Science* **382**, eadf6812 (2023). DOI: 10.1126/science.adf6812

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adf6812

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

RESEARCH ARTICLE

BICCN

Transcriptomic cytoarchitecture reveals principles of human neocortex organization

Nikolas L. Jorstad[1], Jennie Close[1], Nelson Johansen[1], Anna Marie Yanny[1], Eliza R. Barkan[1], Kyle J. Travaglini[1], Darren Bertagnolli[1], Jazmin Campos[1], Tamara Casper[1], Kirsten Crichton[1], Nick Dee[1], Song-Lin Ding[1], Emily Gelfand[1], Jeff Goldy[1], Daniel Hirschstein[1], Katelyn Kiick[1], Matthew Kroll[1], Michael Kunst[1], Kanan Lathia[1], Brian Long[1], Naomi Martin[1], Delissa McMillen[1], Trangthanh Pham[1], Christine Rimorin[1], Augustin Ruiz[1], Nadiya Shapovalova[1], Soraya Shehata[1], Kimberly Siletti[2], Saroja Somasundaram[1], Josef Sulc[1], Michael Tieu[1], Amy Torkelson[1], Herman Tung[1], Edward M. Callaway[3], Patrick R. Hof[4], C. Dirk Keene[5], Boaz P. Levi[1], Sten Linnarsson[2], Partha P. Mitra[6], Kimberly Smith[1], Rebecca D. Hodge[1]*, Trygve E. Bakken[1]*, Ed S. Lein[1]*

Variation in cytoarchitecture is the basis for the histological definition of cortical areas. We used single cell transcriptomics and performed cellular characterization of the human cortex to better understand cortical areal specialization. Single-nucleus RNA-sequencing of 8 areas spanning cortical structural variation showed a highly consistent cellular makeup for 24 cell subclasses. However, proportions of excitatory neuron subclasses varied substantially, likely reflecting differences in connectivity across primary sensorimotor and association cortices. Laminar organization of astrocytes and oligodendrocytes also differed across areas. Primary visual cortex showed characteristic organization with major changes in the excitatory to inhibitory neuron ratio, expansion of layer 4 excitatory neurons, and specialized inhibitory neurons. These results lay the groundwork for a refined cellular and molecular characterization of human cortical cytoarchitecture and areal specialization.

Areal parcellation of the neocortex is premised on the idea that structural variations in cellular architecture (*1*–*3*) and myeloarchitecture (*4*) underlie functional divisions [reviewed in (*5*)]. Neocortex has a 6-layered organization common across species and areas, apart from agranular areas such as the primary motor cortex (M1) that lack layer 4. Cortical layers contain projection neurons with generally stereotyped input and output properties hypothesized to represent a "canonical" circuitry (*6*, *7*). However, cortical areas differ in positional topography, shape and size, laminar and columnar organization, and neuron proportions (*8*–*10*).

Advances in single cell transcriptomics have revealed a complex, hierarchical cortical cell-type architecture based on gene expression signatures that is conserved across species except at the finest cell-type distinctions (*11*–*15*). Prior work in M1 established a cellular hierarchy consisting of 24 neuronal and non-neuronal subclasses with distinct laminar patterning and correlated phenotypic properties (table S1) and revealed deeper cellular complexity in any given cortical area than previously appreciated (*12*, *13*, *15*–*19*). The current study aims to quantitatively define the cellular architecture of eight human neocortical areas representative of topographic, functional, and structural variation, using single nucleus RNA-seq (snRNA-seq) and spatial transcriptomics methods.

[1]Allen Institute for Brain Science, Seattle, WA 98109, USA. [2]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 171 77 Stockholm, Sweden. [3]Systems Neurobiology Laboratories, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [4]Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [5]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA 98195, USA. [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY 11724, USA. *Corresponding author. Email: edl@alleninstitute.org (E.S.L.); trygveb@alleninstitute.org (T.E.B.); rebeccah@alleninstitute.org (R.D.H.)

Within-area cell taxonomies demonstrate common subclass architecture

To sample major axes of cortical variation, we analyzed eight neocortical areas that included M1 and primary somatosensory (S1), auditory (A1), visual (V1) and association areas [dorsolateral prefrontal cortex (DFC), anterior cingulate cortex (ACC), middle temporal gyrus (MTG) angular gyrus (AnG)], which spanned the rostral to caudal (anterior to posterior in many mammals) extent of the cortical sheet, and represented major variations in cortical cytoarchitecture (Fig. 1A) (*20*). Cortical areas were identified across tissue donors using a combination of surface anatomical landmarks and histological verification of cytoarchitecture (Methods). Human postmortem brain samples were collected from 5 individuals (3 males, 2 females, table S2). Tissue photographs taken at the time of autopsy and tissue dissection were used to manually align tissue samples to three-dimensional (3D) reference atlases [Allen Human Reference Atlas 3D https://github.com/BICCN/cell-locator; Julich-Brain v2.9 parcellation, DOI:10.25493/VSMK-H94 (*21*)] and a 2D plate-based reference (Allen Human Reference Atlas http://atlas.brain-map.org/). The best matching structure in each reference atlas is reported (table S2, Methods) and secondary structures are reported when more than one cortical area is predicted according to the mapping results (table S2). Most tissue samples map to a single area in the Allen Human Reference Atlas but MTG samples included both the intermediate and caudal subdivisions of A21. Mapping to the probabilistic Julich Brain Atlas suggests that several areas may have been sampled for ACC (area 33 and area p24ab) and A1 (area TE 1.0 and area TE 1.1), and variation in the precise location of sampling might result in increased variability in the cellular compositions of these areas.

Three snRNA-seq datasets were generated: a 10x Chromium v3 (Cv3) dataset with >924,000 nuclei sampled from all cortical layers, a Cv3 dataset of >231,000 nuclei captured by microdissection of layer 5 to enrich for rare layer 5 extratelencephalic projecting (L5 ET) neurons (for all areas except AnG and M1), and a SMART-seqv4 (SSv4) dataset of over 60,000 nuclei sampled from individual cortical layers to provide laminar selectivity for all clusters. For AnG, only a Cv3 dataset of all cortical layers was generated (Fig. 1B).

Nuclei were assigned to one of 24 cell subclasses based on transcriptomic similarity to a reference taxonomy for human M1 (*12*, *13*), and subclasses were grouped into five neighborhoods (Fig. 1, C and D). For each area and neighborhood, nuclei profiled with Cv3 and SSv4 were integrated based on shared coexpression and clustered to identify transcriptomically distinct cell types. Neighborhood clusters were aggregated and organized into within-area taxonomies ranging between 120 and 142 cell types (Fig. 1C and figs. S1 to S8) with distinct marker expression (table S3). Cellular variation within subclasses was quantified as the average entropy of variably expressed genes. Entropy was higher for all neuronal than non-neuronal subclasses and did not differ between excitatory and inhibitory subclasses or across areas based on a two-way analysis of variance (ANOVA) followed by post-hoc Tukey HSD tests (Fig. 1E). The number of distinct cell types within a subclass was similar across areas for inhibitory and non-neuronal subclasses but varied for excitatory neuron subclasses (Fig. 1F). This within-subclass variation was not driven by differential sampling of nuclei across areas (Fig. 1, G and H). In V1, there were more layer (L)4 intratelencephalic projecting (IT) types, consistent with expansion and specialization of the thalamorecipient L4 in V1, and more L5 IT types. L6 IT Car3 neurons were more diverse in MTG, A1, and AnG compared with other areas. L5 ET neurons were least diverse in the rostral area ACC and in the

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

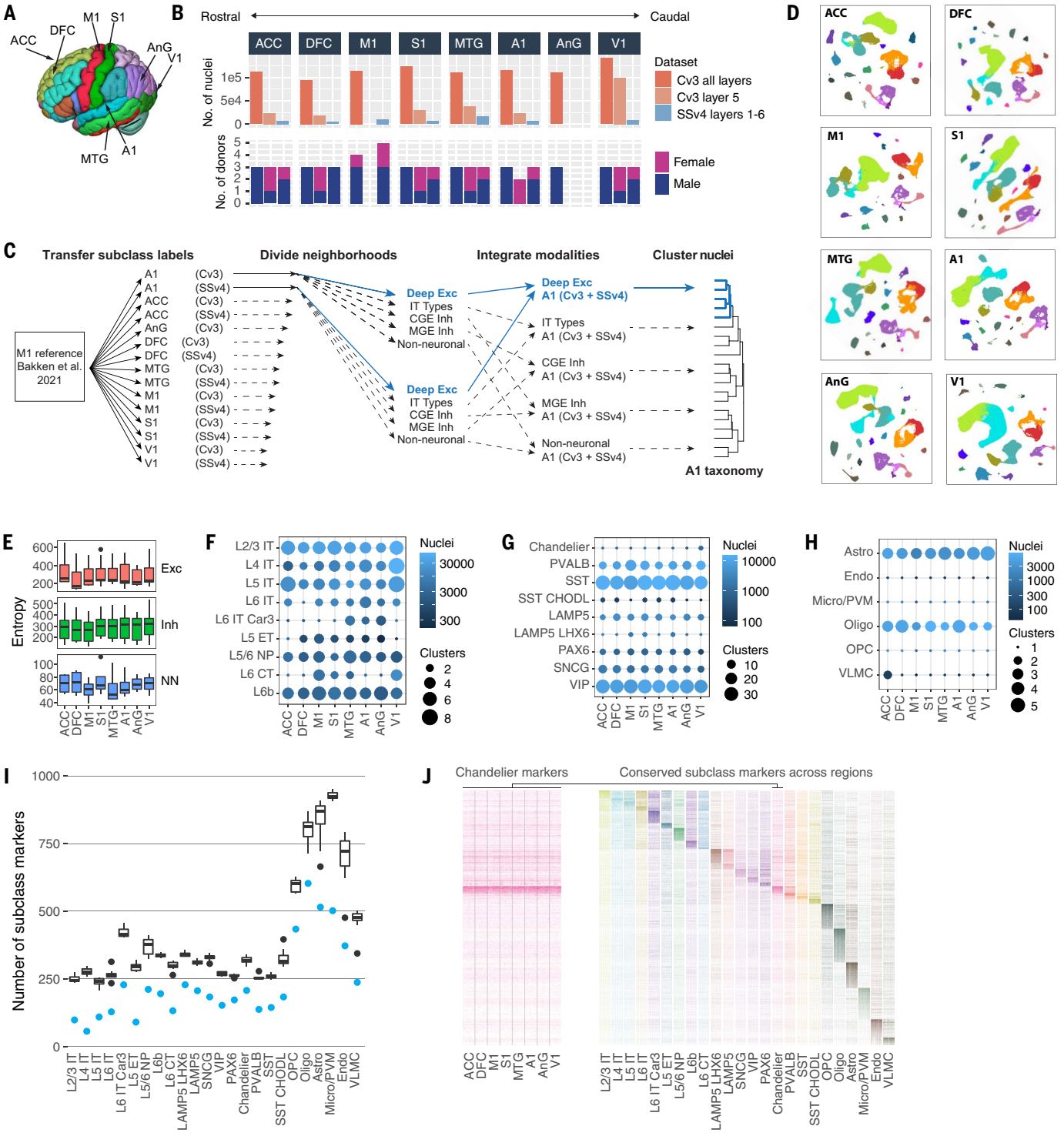


Fig. 1. Transcriptomic cell type diversity across human cortical areas. (**A**) Eight areas of the neocortex were sampled from four lobes of the adult human brain. (**B**) snRNA-seq sampling across areas grouped by RNA-seq platform, layer dissection strategy, and number of male and female donors. (**C**) Schematic of snRNA-seq clustering to generate cell-type taxonomies for each area. (**D**) UMAPs of single nuclei from each area based on variable gene expression and colored by cell subclass as in (**J**). (**E**) Distributions of subclass transcriptomic entropy differ between neuronal (Exc and Inh) and non-neuronal (NN) classes and not between areas. (**F**, **G**, and **H**) Summary of within-area taxonomies showing the number of nuclei sampled from each subclass and the number of distinct clusters (cell types) identified for excitatory (F) and inhibitory (G) neurons and non-neuronal cells (H). (**I**) Number of subclass markers in each area (box plots) and shared across areas (blue points). Box plots show median, interquartile range (IQR), up to 1.5*IQR (whiskers), and outliers (points). (J) Heatmaps of conserved marker expression for 50 random nuclei sampled from each area for chandelier interneurons and horizontally compressed for all subclasses.

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

most caudal area V1, whereas L6 corticothalamic-projecting (L6 CT) neurons were most diverse in M1, S1, MTG, and V1. Individual subclasses had hundreds of distinct markers in each area (table S4), and 20 to 70% of markers were conserved across areas (Fig. 1I). For example, Fig. 1J plots expression of a set of chandelier cell markers that were common across areas (left), and a set of common markers for all subclasses (right). Excitatory subclasses had the smallest fraction of conserved markers, pointing to more variable expression of excitatory neuron gene expression across the cortex as reported in mice (*15*).

Cross-areal abundance changes reveal areal specification

The areas analyzed have distinct cytoarchitecture based on Nissl staining that shows variation in cell size, shape, and laminar organization (Fig. 2A), and spanned the rostrocaudal axis of the cortical sheet (Fig. 2B). Relative proportions of transcriptionally defined neuronal subclasses varied across areas (Fig. 2C and table S5). Excitatory neuron subclasses had the greatest differences in proportions across areas and often reflected known differences in cellular architecture. Agranular M1 and ACC (Fig. 2A) had L4 IT neurons but at lower proportions than other areas (*12*), with the lowest proportion observed in ACC. By contrast, in V1 where L4 is visibly enlarged, L4 IT neuron proportion was increased. As described previously in mouse cortex (*15*) and between human M1 and MTG (*12*), inhibitory neuron subclasses were similar across areas except for a marked increase of medial ganglionic eminence (MGE)-derived PVALB neurons and fewer caudal ganglionic eminence (CGE)-derived interneurons (LAMP5 LHX6, LAMP5, SNCG, VIP, PAX6) in V1. These proportion differences in excitatory and inhibitory neurons were validated in situ by labeling of neuronal subclasses in MTG and V1 using MERFISH spatial transcriptomics (Fig. 2C, right panels, and table S5), demonstrating that they were not an artifact of nuclear isolation or snRNA-seq processing.

Subclass proportions were highly consistent across donors despite variation in the precise location sampled for areas such as MTG (Fig. 2D). Examined from a subclass perspective, the most obvious proportion differences were seen in L4 IT (range 10-fold, from 3 to 30% of excitatory neurons), and in the much sparser L5 ET neurons (range 50-fold, from 0.1 to 5%). Many of these proportion differences varied in a graded fashion generally along the rostrocaudal (R-C) axis. Pairwise correlations in excitatory neuron proportions revealed correlated R-C decreases in L5 ET, L6B, L6 IT, and L5/6 near-projecting neurons (L5/6 NP), with an anticorrelated R-C increase in L4 IT (Fig. 2E). Among inhibitory subclasses, the rarest types

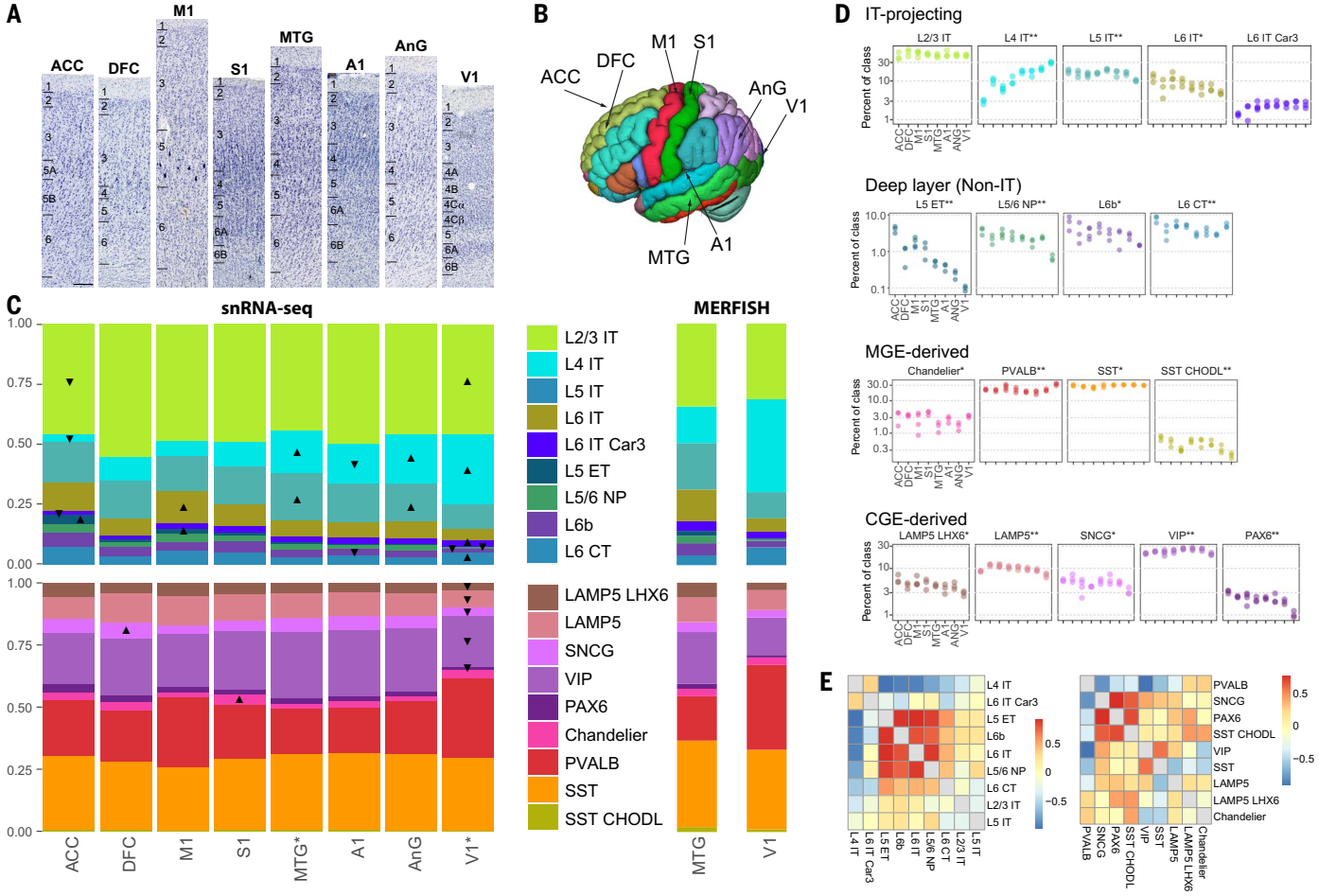


Fig. 2. Cell subclass composition reflects cytoarchitecture and varies systematically along the R-C axis. (**A**) Images of Nissl-stained sections of cortical areas are labeled with approximate layer boundaries and show distinct cytoarchitecture. Areas are ordered by position along the R-C axis of the cortex. (**B**) Representative cortical gyral locations of sampled tissue. (**C**) Relative proportions of neuronal subclasses as a fraction of all excitatory or inhibitory neurons in each area and estimated based on snRNA-seq profiling or in situ labeling using MERFISH. Arrowhead directions indicate subclasses that significantly increase (pointing up) or decrease (pointing down) across areas based on scCODA analysis (**D**) For each donor, subclass proportions were calculated as a fraction of all neurons in the same class (excitatory or inhibitory) and grouped by neighborhood (*nominal $P < 0.05$; **Bonferroni-corrected $P < 0.05$). (**E**) Spearman correlations of excitatory and inhibitory subclass proportions across areas. Scale bar on (A) is 200 μm.

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

(SNCG, PAX6, and SST CHODL) had the most correlated changes in proportions with a decreasing R-C gradient.

Smaller-scale areal specializations in proportions were overlaid on these broad trends of conservation or R-C gradients and many subclasses showed a particularly large difference in V1 (L5/6 NP, L6B, PAX6). Chandelier inhibitory neuron proportions were lowest in MTG and AnG and highest in S1 (Fig. 2, C and D). There were more L4 IT neurons and fewer L5 ET neurons in DFC, more L6 IT neurons in M1, and fewer L6 IT Car3 neurons in ACC than expected based on the broad trends. In summary, cell subclass proportions define a quantitative cytoarchitecture that is canonical in having all 24 subclasses in all areas, with varying proportions and gradient properties that likely reflect developmental gradients and specializations driven by the circuit requirements of functionally distinct cortical areas.

Excitatory to inhibitory neuron ratio varies across cortical areas and layers

In addition to areal specializations in neuronal subclass proportions, we found differences in relative proportions of excitatory and inhibitory neurons (E:I ratio) (table S5). As previously reported for M1 (*12*), the E:I ratio was 2:1 for most cortical areas, in contrast to the reported E:I ratio of 5:1 in mice. However, the E:I ratio in V1 (4.5:1) was much higher (Fig. 3A) and comparable to that of rodents. MERFISH analysis in MTG and V1 confirmed these values and areal differences (table S5).

Layer-specific dissections of nuclei from 7 regions (excluding AnG) allowed a deeper exploration of E:I ratio variation. E:I ratios varied by area and layer and were consistent across

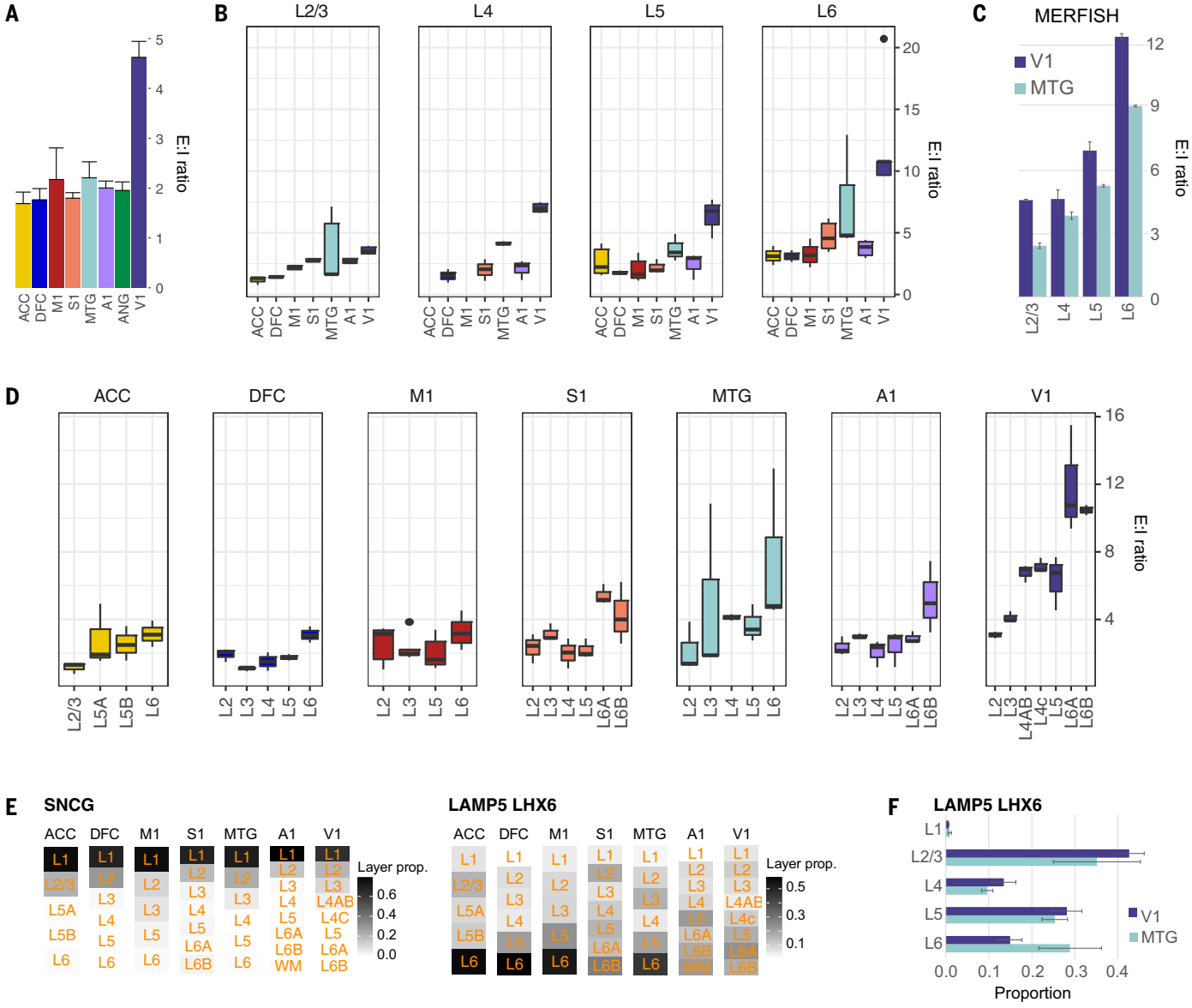


Fig. 3. E:I ratio variation across cortical areas and layers. (**A**) Relative number of excitatory neurons to inhibitory neurons (E:I ratio) in each area. Bar plots indicate average and standard deviation across donors. (**B**) E:I ratios estimated for a common set of layers dissected from each area. Box plots show median, interquartile range (IQR), up to 1.5*IQR (whiskers), and outliers (points) across multiple donors. (**C**) Validation of increased E:I ratios in all cortical layers in V1 compared with MTG based on MERFISH experiments. Bar plots and whiskers indicate average and standard deviation of E:I ratios across donors, respectively. (**D**) E:I ratios estimated for all layers dissected from each area. (**E**) Laminar distributions of interneurons were conserved (SNCG) or divergent (LAMP5 LHX6) across areas based on counts of layer-dissected nuclei. Note that primary sensory areas (S1, A1, and V1) have a distinct distribution of LAMP5 LHX6 neurons. (**F**) MERFISH in situ labeling of LAMP5 LHX6 cells shows a decreased proportion of cells in layer 6 of V1 compared with MTG.

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

donors (Fig. 3B). However, increased variability was seen in MTG perhaps due to sampling of both the intermediate and caudal subdivisions of A21. V1 had the highest E:I ratio across all layers, not just in L4 (7:1) but also in L5 and L6, where the highest E:I ratio of 10:1 was seen. Moreover, there was a monotonic increase in the E:I ratio of other areas along a R-C gradient, which was most apparent in L2/3. E:I ratios were more variable in L4 and L5, masking the trend in overall E:I ratios (Fig. 3A). In situ cell counts in MTG and V1 using MERFISH confirmed a higher E:I ratio in all layers of V1 (Fig. 3C). From a within-area perspective, E:I ratios increased with cortical depth, with the highest ratios in L6 for all areas (Fig. 3D). Furthermore, the E:I ratio in L4 was distinctly elevated in V1 relative to L2, L3, and L5, highlighting specialization of visual processing compared with other sensory modalities. Finally, laminar distributions of excitatory and inhibitory neurons were relatively consistent across cortical areas (fig. S9), such as SNCG in L1 (Fig. 3E), but some areal and laminar variation was apparent, such as LAMP5 LHX6 proportions in L6 (Fig. 3, E and F). Taken together, E:I ratios vary extensively both by layer and area, with markedly different ratios in V1 and areal variation that is masked by averaging across cortical layers.

Transcriptomic cellular topography

To characterize the transcriptomic landscape of neuronal subclass cortical areas, neuronal nuclei were integrated by donor for each of four neighborhoods [IT-projecting excitatory, deep layer (non-IT) excitatory, MGE-derived GABAergic and CGE-derived GABAergic] and visualized as UMAPs colored by subclass (Fig. 4A) and area (Fig. 4B). Three organizational principles were apparent. First, excitatory neurons had strong areal signatures, visualized as clear banding by area, whereas inhibitory neurons were mostly intermixed across areas similar to reports in the mouse cortex (*15*). Second, there were visible V1 specializations including substantial expansion of L4 IT neurons and distinct islands in the UMAPs for most IT-projecting subclasses and L6 CT neurons. Distinct V1 islands were also seen for parts of the PVALB and SST subclasses (arrows in MGE-derived UMAPs). Third, the areal similarity of excitatory neurons appeared to vary in a R-C topographic order for many subclasses, similar to prior reports of gene expression similarity across the human cortex (*22*). Neighboring areas were similar and intermixed despite known functional distinctiveness; for example, nuclei from M1 and S1 intermingled despite their specificity for motor and somatosensory functions, respectively.

Areal variation in gene expression mirrored the UMAP trends. The number of differentially expressed genes (DEGs, table S6) across areas was largest for excitatory neurons (Fig. 4C), and highest for L4 IT and L5 ET subclasses (over 1000 DEGs). DEGs for inhibitory neuron subclasses varied widely, from over 100 DEGs for SST and PVALB interneurons to fewer than 10 DEGs for SNCG and SST CHODL and a single DEG (*ADAMTS9-AS2*) for PAX6. Non-neuronal cell subclasses similarly displayed few areal DEGs. We next used a previously defined tau score (*23*) to identify area-specific markers (table S7), which were much less common. Excitatory neurons expressed the most areal markers and ACC and V1 were the most distinct areas (Fig. 4D and fig. S10, A and B). IT-projecting neurons were specialized in both ACC and V1 whereas Non-IT L6 CT and L5 ET neurons were specialized mostly in V1.

The topographic ordering of the excitatory neuron subclasses above suggested graded changes as a function of distance, similar to bulk tissue profiling studies reporting gradual changes in gene expression across the cortical sheet (*22*). We therefore calculated transcriptomic similarities of excitatory subclasses as a function of the approximate physical distance between pairs of areas on an unfolded cortical sheet (Fig. 4E and table S2). Because V1 was so distinct (Fig. 4A), we fit two linear models of subclass similarity versus areal distance, one that included pairwise comparisons to V1 and one that did not (Fig. 4E). All excitatory neuron subclasses showed the same monotonic decrease of similarity with distance but had different amounts of transcriptomic specialization in V1 (intercepts, but not slopes, are different in Fig. 4E). Interneuron similarity also decreased with distance at the same rate for all subclasses, albeit at about 40% the rate of excitatory neurons, and with much less specialization in V1 (fig. S10C). By contrast, non-neuronal expression did not change systematically with inter-areal distance and was not more specialized in V1 (fig. S10D).

To determine how gene expression varied across the cortical sheet, we performed a variance partitioning analysis for each subclass (fig. S10E and table S8). Expression of hundreds of genes was explained by area identity and spatial gradients for excitatory neurons compared with a few genes for inhibitory neurons and non-neuronal cells. These genes had a similar proportion of expression variation explained by area or gradients (median 5 to 10%) in all subclasses (fig. S10F). Therefore, the observed differences in transcriptomic topography (Fig. 4E) were mainly due to the number of genes with areal variation and not to the strength of that variation. Among IT-projecting neurons, some genes showed distinct patterning in a single subclass whereas other genes were topographically patterned in all IT subclasses (fig. S10G). We calculated the expression variance explained by gradients along three axes: rostrocaudal (R-C), midline-surface (M-S, anatomical left to right), and dorsoventral (D-V). The relative position of each cortical area along these axes was calculated using the voxel coordinates corresponding to the approximate center of each area in the Allen Human 3D Reference Atlas (table S2). For genes with at least 5% of expression variance explained by any gradient, we quantified the relative strength of gradients based on the relative proportion of expression variance that was explained (fig. S10H). For most subclasses, R-C gradients were dominant but non-IT subclasses also expressed many genes with M-S and D-V gradients (Fig. 4F and fig. S10H).

A set of R-C genes was defined for each subclass by requiring a Spearman correlation >0.7 between expression and areal position along the R-C axis and a correlation >0.5 after excluding V1 and ACC as these are transcriptionally distinct regions that might bias the identification of gradients. For the most varying L4 IT and L5 ET neurons, roughly equal numbers of genes increased and decreased expression rostrocaudally (Fig. 4G). For other subclasses, many more genes increased rather than decreased expression along the R-C axis. The correlations of R-C genes were greater than correlations to a randomly shuffled ordering of areas for most neuronal subclasses (fig. S10I). Genes with a R-C gradient in one subclass frequently had a gradient in the same direction in other subclasses that expressed the gene (fig. S10J), such as *CBLN2* in L2/3 IT and L4 IT neurons (Fig. 4H), which is expressed in a similar gradient in maturing cortical neurons during human prenatal development (*24*). However, some genes such as *DCC* had opposing gradients in different subclasses (L5/6 NP and VIP), and some functionally related genes had opposing gradients in the same subclass, such as the cell adhesion molecules Contactin 5 and 6 (*CNTN5* and *CNTN6*) in L5 IT neurons (Fig. 4H). Based on gene ontology (GO) analysis, genes with strong areal enrichment or R-C gradients included voltage-gated potassium and calcium channels (table S7). Notably, only R-C genes were associated with axon guidance pathways including SLIT/ROBO, ephrin, and semaphorin signaling molecules (table S7), likely reflecting developmental patterning of connectivity.

Cross-areal consensus taxonomy

Next, we sought to understand finer cell-type areal variation by clustering integrated cell neighborhoods (Fig. 4) to identify a set of cell types either common to or varying across cortical areas (Fig. 5A). We defined and organized 153 cell types by transcriptomic similarity into a consensus taxonomy (Fig. 5B). Consensus cell types had consistent markers across areas (table S9), were represented in all donors (Fig. 5C), and ranged from 0.01 to 20% of excitatory and inhibitory neurons and from 0.1 to 30% of non-neuronal cells (Fig. 5D). Most types were

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

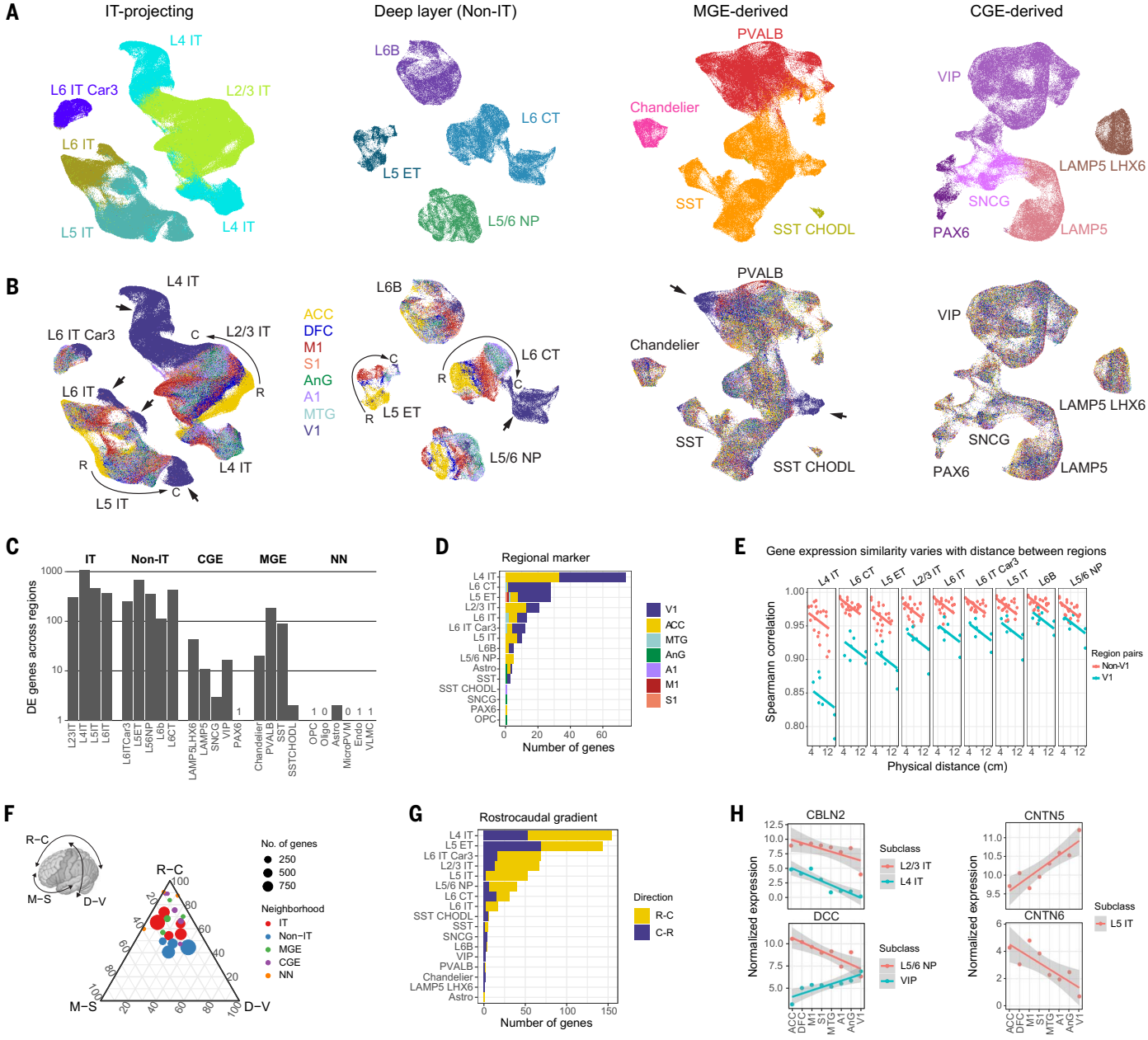


Fig. 4. Transcriptional topography across cortical areas. (**A** and **B**) UMAPs showing transcriptomic similarities of single nuclei dissected from eight cortical areas and colored by neuronal subclass (A) and area (B) for excitatory and inhibitory neuron neighborhoods. Arrows indicate V1-specialized neurons. Curved arrows illustrate R-C ordering of areas on the cortical sheet. (**C**) The number of genes that are significantly differentially expressed across areas for each subclass grouped by neighborhood. Subclasses with 0 or 1 DEG are labeled. See table S6 for all DEGs. (**D**) The number of genes that have highly enriched expression in a single area for each subclass. (**E**) Spearman correlations of expression similarity between pairs of areas as a function of the approximate physical distance along an unfolded neocortical sheet. Pairwise comparisons that include V1 (blue points) or do not include V1 (red) are grouped separately because V1 is so transcriptomically distinct. (**F**) Ternary plot summarizing the relative proportion of variance explained by expression gradients across areas along R-C, M-S (anatomical left to right), and D-V axes for each subclass. Point size indicates the number of genes with >5% of variance explained by at least one gradient, and point location shows the weighted mean proportion across all genes (shown in fig. S10H). Points are colored by cell neighborhood. (**G**) For each subclass, the number of genes with expression that increases (R-C) or decreases (C-R) in areas ordered by their position along the R-C axis. (**H**) Examples of genes with R-C gradient expression that have been previously described in development (*CBLN2*) (*32*), have opposing gradients in different subclasses for the same gene (*DCC*), or for two related genes (*CNTN5* and *CNTN6*) involved in neuronal connectivity for the same subclass.

found in all areas, with generally uniform representation across areas for inhibitory neuron types and non-neuronal cells (Fig. 5E). However, we also identified area-enriched or area-specific cell types, particularly in V1 (dark blue). V1-enriched clusters were seen in most excitatory subclasses, particularly L4 IT, as well as SST and several PVALB and VIP types. There was also one ACC-selective VIP type. Similarity by proximity was evidenced by cross-areal excitatory cell types common to neighboring regions (M1 and S1, MTG and AnG).

To study changes in the relative abundances of cell types while accounting for the compositional nature of the snRNA-seq data, we applied

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

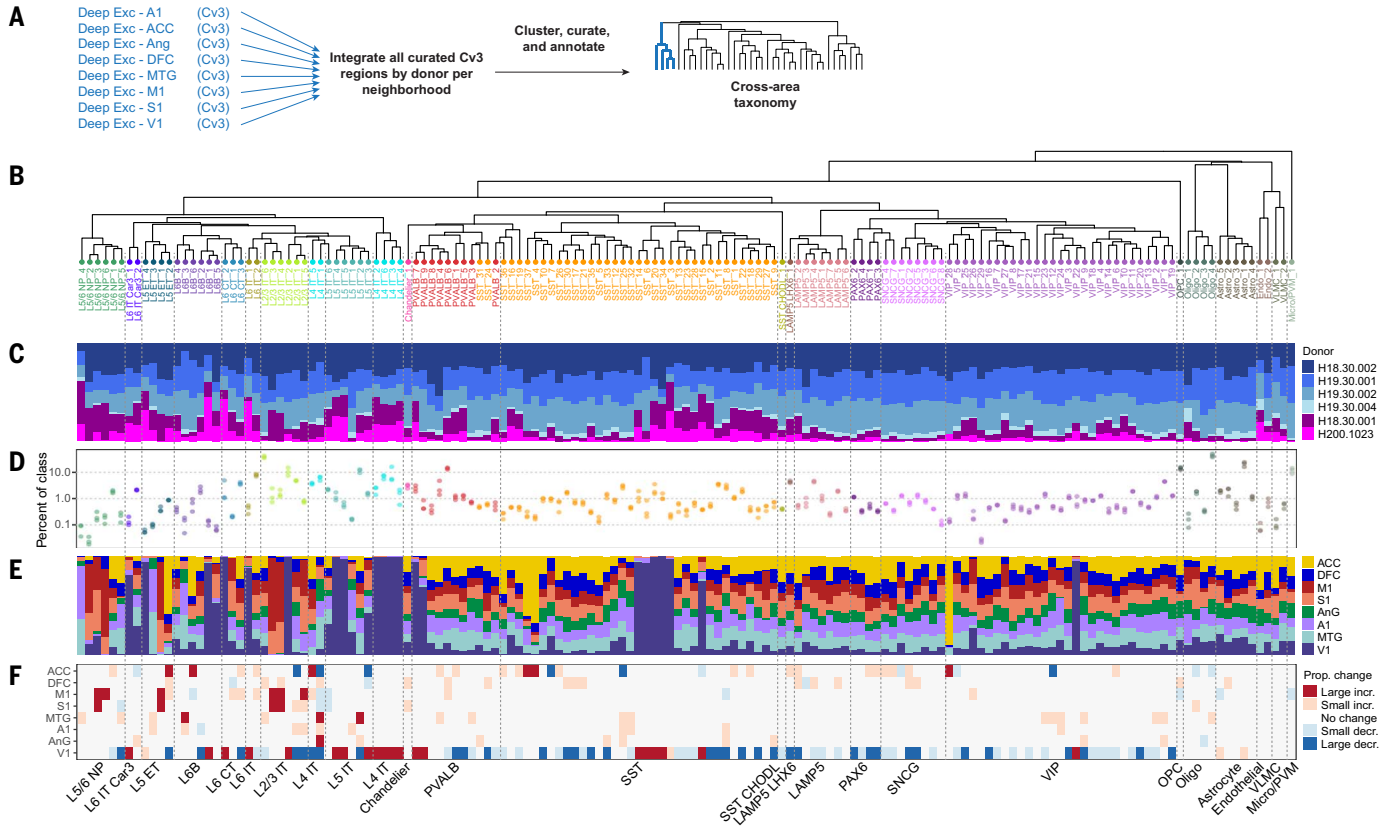


Fig. 5. Cross-areal consensus taxonomy. (**A**) Schematic of data integration across donors used for each neighborhood to generate the cross-area consensus taxonomy. (**B**) Consensus taxonomy of cell types across eight areas. (**C**) Proportion of nuclei in each consensus type dissected from each donor. (**D**) Consensus type proportion including nuclei from all areas as a fraction of cell class. Individual dots indicate proportions measured per donor. (**E**) The relative number of nuclei dissected from areas that contribute to each consensus cell type. (**F**) Changes in consensus type proportions across areas based on compositional analyses of neurons and non-neuronal cells using scCODA. Larger magnitude changes are indicated by darker colors. See table S10 for proportion effect sizes.

a Bayesian model (scCODA) (*25*). Nuclei were grouped by consensus types and iteratively tested for consistent differences using each type as the "unchanged" reference population. All subclasses included consensus types with both increased and decreased proportions (table S10), except for PAX6 inhibitory type abundances that were uniformly decreased in V1. V1 had the most consensus types with abundance changes (92 of 153, 60%), including two types with the largest changes (L4 IT_5 and L2/3 IT_2). Excitatory neurons were the most specialized in V1, but several SST, PVALB, and VIP consensus types were also specific to V1. Specialized types were also found in other areas, including L2/3 IT and L5/6 NP excitatory types (L2/3 IT_3, L2/3 IT_4, L5/6 NP_3 and L5/6 NP_6) in M1 and S1, SST types (SST_4 and SST_10) in ACC, and distinct L5 ET types across the R-C axis. These substantial changes, along with more subtle abundance changes (median = 17 consensus types affected in each area), are likely important determinants of the functional role of each area.

V1 specializations

The distinctiveness of V1 was reflected in the transcriptomic data for specific cell types. Considering cell types with >60% membership in V1 compared with other areas to be V1-specialized, we identified specialized cell types in every excitatory subclass except L5/6 NP, with the greatest number of V1-specialized types in the L2/3 IT and L4 IT subclasses (Fig. 6A and table S11). Unexpectedly, given prior reports of common GABAergic neurons across the mouse neocortex (*15*, *17*), V1 had a number of specialized CGE- and MGE-derived types.

MERFISH analysis of V1 demonstrated the spatial organization of all cell types (fig. S11, A and B). L2/3 IT types had distinct markers (table S12), sublaminar distributions, and relative proportions (Fig. 6B). L2/3 IT5 and L2/3 IT2 clearly delineated L2 and L3 from one another, respectively. Other L2/3 IT types were more sparsely distributed in L2 (L2/3 IT4), L3 (L2/3 IT3), or both (L2/3 IT1 and 6). L2/3 IT types were also found in layer 4A (L2/3 IT2) and the superficial part of layer 4B (L2/3 IT3), and these types were V1-specialized. Conversely, several L4 IT types were found in L4A and L4B and into the deep part of L3 (L4 IT1 and 3, Fig. 6D). Thus, the specialized L4A and L4B contain not only L4 IT-type neurons, but also L2/3 IT-type neurons. This finding may help resolve ongoing questions about primate V1 L4A and L4B, which contain both stellate (L4 IT-like) and pyramidal corticocortical projection neurons (L2/3 IT-like) (*26*).

L4 in V1 is highly distinctive even in unlabeled tissues as a result of the band of myelinated thalamocortical axons entering L4 that form the stria of Gennari. This distinctiveness was also seen at the level of L4 IT neuron types, all but one of which were V1-specialized (Fig. 6, C and D). L4 IT types had specific markers (table S12) and sublaminar distributions, from dense pan-L4 (L4 IT3) to sublayer-specific distributions. Layers 4Cα and 4Cβ receive selective inputs from magnocellular and parvocellular layers of the thalamic lateral geniculate nucleus, respectively (*27*), and MERFISH revealed localization of specific types to each sublayer. L4 IT5 was selectively localized in 4Cβ, whereas L4 IT2 was enriched in 4Cα but extended into L4B, consistent with the fuzzy boundary between 4Cα and 4B described in other human studies (*28*). Sparser L4 IT types were scattered across layers. Together, these results illustrate the cellular specialization of the distinctive input layer of V1 and reveal a complexity of putative thalamorecipient stellate neurons that offers many avenues for future exploration.

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

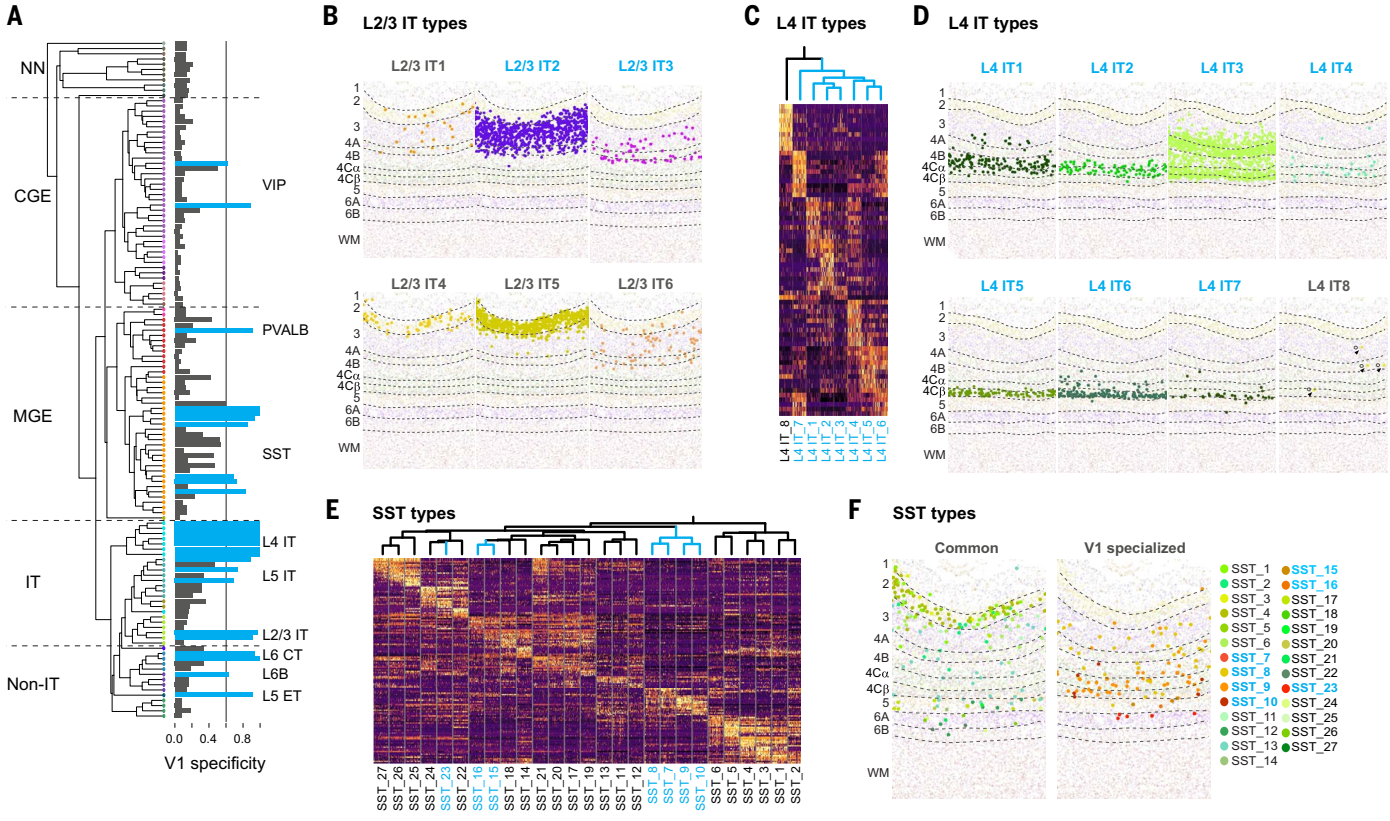


Fig. 6. V1 cell-type specialization. (**A**) Transcriptional distinctiveness of cell types in the V1 taxonomy. Cell types with specificity >0.6 are considered V1-specialized and are highlighted in blue (see table S11). (**B**) Laminar distributions of specialized (blue text) and common (gray) L2/3 IT types based on MERFISH in situ labeling experiments. (**C** and **E**) Scaled expression of marker genes of V1 specialized (blue labels) and common (black) L4 IT (C) and SST (E) types across areas. Dendrograms were pruned from the V1 taxonomy in (A). (**D** and **F**) Laminar distributions of specialized and common L4 IT (D) and SST (F) types based on MERFISH experiments.

L6 CT neurons that send reciprocal projections to the LGN were also highly specialized in V1 (Fig. 6A), with two distinct types that expressed many V1-enriched genes (fig. S11C). Gene set enrichment analysis showed enrichment for calcium signaling, axon guidance, and axonal and synaptic compartments, including axon guidance molecules *CDH7*, *EPHA6*, and *SEMA6A* (fig. S11D). Various ion channels (*KCNT2* and *SCN1B*) and synaptic genes (*SYT6*), as well as calcium and calmodulin signaling–associated genes (*PCP4*, *NPY2R*), were similarly enriched, and several of these have conserved V1 enrichment in monkeys (*29*). Myelin basic protein (MBP), normally described in oligodendrocytes but known to function in certain neurons as part of a Golli-MBP complex (*30*), was also enriched in V1 L6 CT neurons.

V1 also contained specialized GABAergic interneuron types. Most were SST types (Fig. 6, E and F, and table S12), as well as one PVALB and two VIP types (fig. S11A). The SST types common across V1 and other cortical areas were concentrated in L2 with sparser representation in other layers. By contrast, the V1-specialized SST types were concentrated in L4 near the V1-specialized L4 IT types, suggestive of a relationship between these specialized excitatory and inhibitory types.

L5 ET neuron diversity

Neocortical L5 ET neurons are sparse and capturing them required additional L5-specific sampling. L5 ET neurons were most abundant in ACC and their abundance generally decreased along the R-C axis (Fig. 2D). V1 had the lowest proportion of L5 ET neurons (~0.1% of excitatory neurons), consistent with data from macaque monkeys demonstrating projections to subcortical targets such as the superior colliculus from large, very sparse neurons (Meynert cells) localized to deep layers in V1 (fig. S12A) (*31*–*33*).

We identified 4 consensus L5 ET types (Fig. 5 and Fig. 7A), several of which were dominated by nuclei derived from cortical areas near each other. M1 and S1 predominantly contributed to L5 ET 1, whereas L5 ET 3 was largely composed of nuclei from MTG and A1 (and to a lesser extent AnG), again suggesting similarity based on topographic position. V1 specialization was also apparent in L5 ET types, with only a single type (L5 ET 4) consisting of nuclei predominantly from V1 (Fig. 7, A and B). L5 ET neurons could be divided into at least two transcriptomically distinct subtypes in most regions (Fig. 7B). M1 had 3 distinct subtypes and we showed previously (*12*) that at least two L5 ET M1 subtypes included Betz cells. Notably, despite having the highest proportion of L5 ET neurons of all areas, only one subtype was identified in ACC, implying that VENs in ACC likely do not represent a distinct transcriptomic cluster, consistent with our previous findings in frontoinsula (*34*).

L5 ET neurons had more genes with variable expression across areas than any other cell type (fig. S10F). Up to 32% of variation in gene expression across areas was explained by gradients along the M-S (anatomical left to right), D-V, and R-C axes (Fig. 7C). Top gradient genes included a glutamate receptor subunit (*GRID2*), a semaphorin (*SEMA3D*), and a neuropilin (*NRP1*) that are involved in trans-synaptic signaling and connectivity (Fig. 7C). Some gene expression varied across areas but not as a gradient, such as *DGKB*, which was selectively down-regulated in primary sensory areas (A1, S1, and V1). L5 ET neurons also expressed distinct areal markers (Fig. 7D and table S13), including the voltage-gated potassium channel *KCNG2* in ACC, glutamate receptor subunit *GRIK1* in MTG and AnG, and *ANK1* in V1, a gene that encodes for the scaffolding protein Ankyrin 1 (*35*). Applying GO enrichment analysis to L5 ET areal markers identified enriched pathways associated with synaptic signaling, connectivity, and intrinsic neuronal firing

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

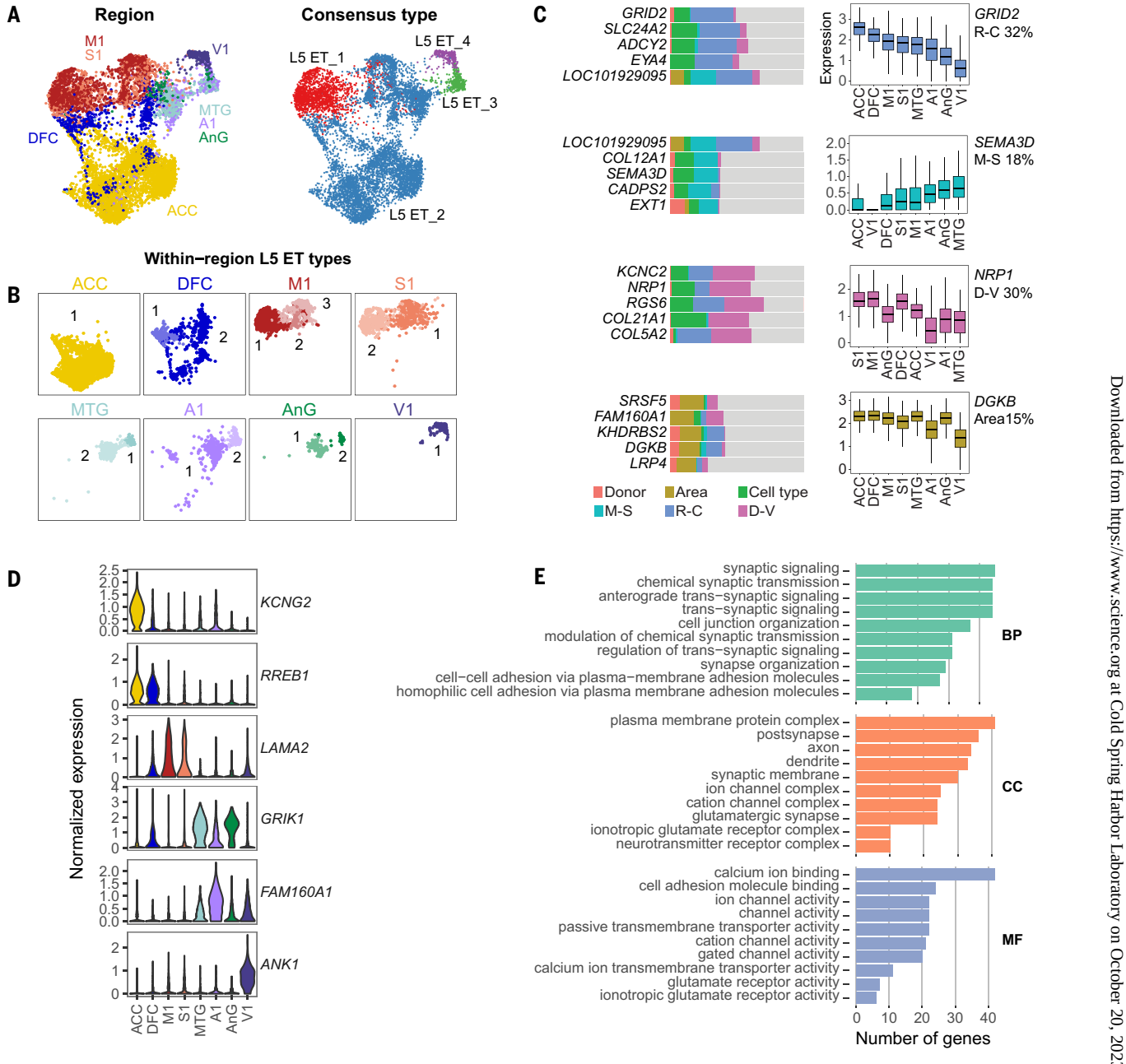


Fig. 7. L5 ET-projecting neuronal diversity. (**A**) UMAPs of L5 ET neurons labeled by area and cross-area consensus type. (**B**) Within-area L5 ET subtypes for each area shown in the same UMAP space as (A). (**C**) Bar plots summarizing the expression variance explained by human donor, L5 ET subtype, and four types of variation across areas: R-C, M-S (anatomical left to right), and D-V gradients and more complex patterns or in a single area (area). For the four types of areal variation, the distribution of expression across areas is shown for one of the top five genes. (**D**) Examples of genes with L5 ET neuron expression restricted to one or a few areas. (**E**) Number of genes in the top 10 significantly enriched terms from gene ontology (GO) analyses (biological process, BP; cellular component, CC; molecular function, MF) of L5 ET areal markers (table S13).

properties (Fig. 7E), consistent with known areal variation in firing properties.

Glial specialization

Non-neuronal cells comprised at least 40 to 65% of cortical cells across areas based on flow cytometric analysis of dissociated nuclei labeled with the neuronal marker NeuN and gated based on NeuN fluorescence intensity (fig. S12A). However, these proportions underestimate the total non-neuronal population because vascular cells, including endothelial cells and VLMCs, are difficult to dissociate (*36*) and are under sampled in the snRNA-seq dataset based on in situ labeling with MERFISH (fig. S12B) (*37*). M1 and S1 had a higher proportion of non-neuronal (NeuN-negative) cells than other areas, and snRNA-seq data showed that this was driven by an expansion of oligodendrocytes relative to OPCs, astrocytes, and microglia (fig. S12, B, C, and D), consistent with neuroimaging studies showing that these areas are the most heavily myelinated in the cortex (*38*, *39*). By contrast, areas described to be among the most lightly myelinated in the cortex (ACC and DFC) (*38*) had the lowest proportion of oligodendrocytes (fig. S12, B and C).

Non-neuronal cells were grouped into major subclasses based on conserved marker

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

expression (fig. S12F), and many subclasses could be further divided into distinct subtypes. Astrocytes were subdivided into previously described protoplasmic, interlaminar (ILM), and fibrous types, which also had robust markers across areas (fig. S12G). Consistent with previous reports of shared non-neuronal types across the cortex (*15*, *17*), there was little areal expression signature for most non-neuronal cell types (Fig. 8A and fig. S12E). However, areal variation in protoplasmic—but not ILM or fibrous astrocytes—was apparent, consistent with previous descriptions of brain-wide astrocyte heterogeneity (*40*, *41*) and variation in astrocytes across cortical and hippocampal areas in mice (*42*). Protoplasmic astrocytes from ACC showed clear banding on the UMAP (Fig. 8A) and distinct areal marker gene expression (Fig. 8B).

Laminar distributions varied across areas for all glial subclasses (Fig. 8C and fig. S12H). There was a notable depletion of astrocytes in L4A and L4B of V1 but not in L4 of other sensory or granular cortices (Fig. 8C). To validate this finding, we compared in situ expression of the astrocyte marker GFAP in V1 and DFC. GFAP protein and gene expression was reduced in L4B of V1 (Fig. 8D), and only protein expression was reduced in L4 of DFC (fig. S12I) based on immunofluorescence (IF) and in situ hybridization (ISH) labeling of adult human tissue. In V1, a band of dense GFAP labeling was apparent in L6A and L6B, which tapered off in the underlying white matter. GFAP IF in V1 revealed a population of astrocytes that extended long processes away from the white matter and into L5, similar to descriptions of varicose projection astrocytes (VPA) that are distinctive to humans and great apes and not found in the cortex of other anthropoid primates (*43*, *44*) (Fig. 8D and S12I). Deep-layer astrocytes in DFC did not extend long processes and had morphologies typical of protoplasmic and fibrous astrocytes (fig. S12I).

The spatial organization of astrocytes in V1 was further investigated using MERFISH (fig. S8). Based on laminar distributions (Fig. 8E) and marker gene expression (Fig. 8F), there were two subtypes of protoplasmic (Astro_1 and Astro_3) and ILM (Astro_2 and Astro_5) astrocytes and one fibrous subtype (Astro_4). In contrast to prior descriptions of protoplasmic astrocytes as relatively homogenous cells, protoplasmic subtypes in V1 displayed distinct laminar patterns with Astro_1 localized to the sublayers of L4 and Astro_3 spread across L2-6 but absent in L1, L6B, and white matter. Astro_1 markers were related to energy metabolism, including mitochondrial genes *COX1* (Fig. 8F), *COX2*, and *COX3*, and Astro_1 cells may represent highly active protoplasmic astrocytes (an Astro_3 cell state) rather than a developmentally distinct type. Astro_5 cells were mostly restricted to the L1-pial border, whereas Astro_2 cells were enriched in the deeper part of L1. These subtypes likely represent pial and subpial ILMs (*45*, *46*), respectively. The putative subpial ILM type (Astro_2) included a small number of cells localized to deep L6. Because ILMs and VPAs have previously been shown to express shared marker genes (*AQP4* and *CRYAB*, Fig. 8F) and have similar morphologies (*44*, *45*), these deep-layer Astro_2 cells may represent a type of VPA. However, further work will be needed to fully characterize the diversity of astrocyte morphologies across the cortex and their relationships to transcriptomic astrocyte types.

Discussion

The cellular complexity of the cortex has challenged generations of neuroscientists aiming to understand the structural basis of cognitive function. The BRAIN Initiative Cell Census Network established a paradigm for mapping cortical cellular diversity, developed methods to work across species, and established the

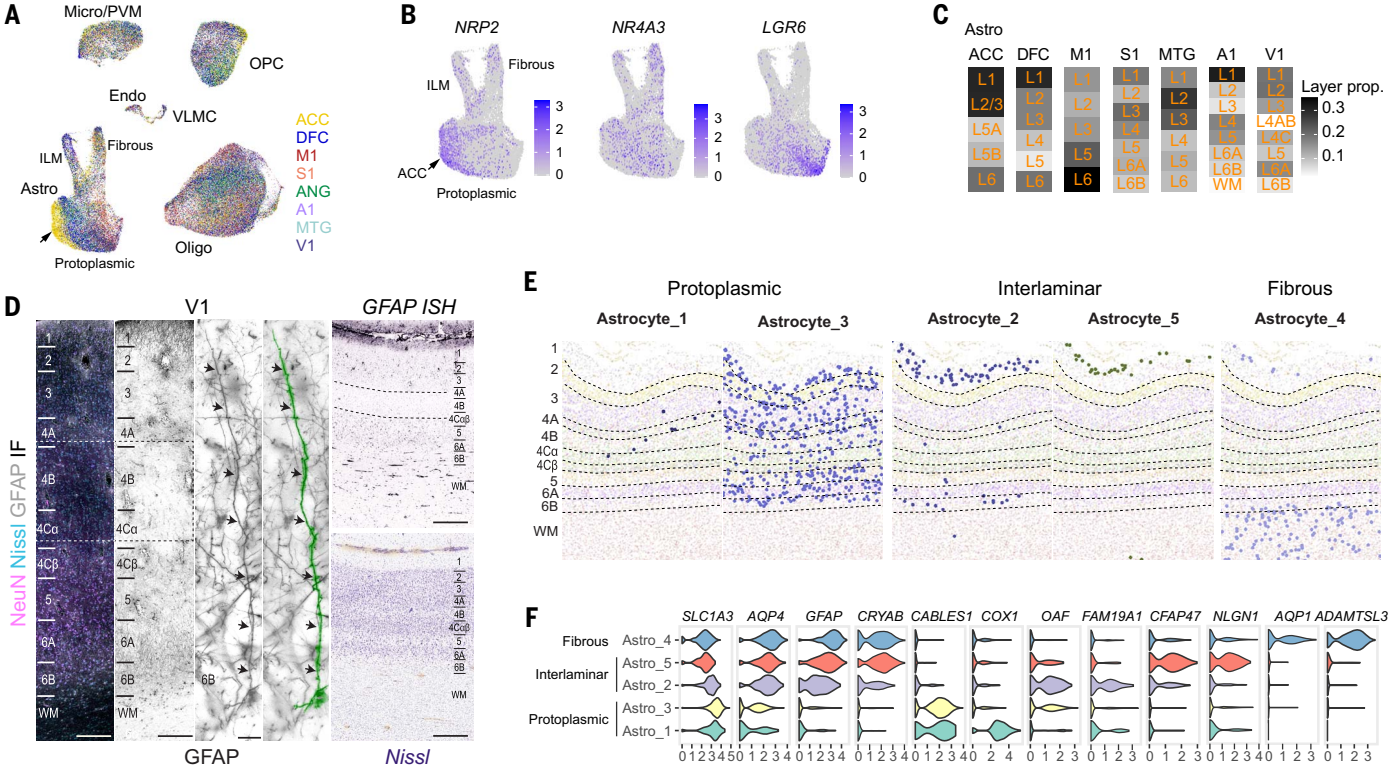


Fig. 8. Areal specialization of astrocytes. (**A**) UMAP of non-neuronal cells labeled by cortical area. (**B**) UMAPs of astrocyte expression for genes with areal enrichment. Arrows in (A and B) shows grouping of nuclei from ACC on the UMAP. (**C**) Laminar distributions of astrocytes vary across areas and are depleted in V1 L4A and L4B. (**D**) GFAP IF and ISH illustrate variable laminar distributions and morphologies of astrocytes in V1 and validates depletion in L4A and L4B. Single channel IF images were inverted to increase visibility of GFAP IF. Scale bars: IF columns (100 μm), GFAP tracing images (15 μm), ISH (200 μm). (**E**) Laminar distributions of astrocyte subtypes in V1 based on MERFISH in situ labeling experiments. (**F**) Pan-astrocyte and subtype marker expression.

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

concordance of a transcriptomic cellular classification with other cellular properties in a way that integrates prior literature while identifying greater cellular diversity than previously appreciated (*12*, *13*, *16*, *18*, *19*, *47*). We used these principles to analyze a series of human cortical areas, building on our highly annotated M1 taxonomy (*12*). Because the cortex has a common organization as well as graded changes and areal specializations, we applied two complementary strategies to define cell types. First, each area was analyzed independently, transferring labels from the M1 taxonomy to other areas, which provides the highest resolution clustering in each area and identifies a common cell subclass organization. Next, data from all areas was analyzed jointly, identifying a set of consensus clusters present in multiple areas while also capturing specialized cell types distinct to a single area. Similar joint analysis strategies have been used on mouse cortex with comparable results (*17*).

A key finding of this study is that all 24 subclasses first identified in M1 are found in all cortical areas analyzed here, substantiating the idea that there is a common cellular organization across the cortex. This was true even for L4 IT neurons, which were present in agranular ACC and M1 (*12*, *48*–*50*). Each cortical area analyzed could be defined as a distinct proportional makeup of cell subclasses. Proportional differences were mostly due to variation in excitatory neuron subclasses, which could be large (10- to 50-fold). Finer cell-type analysis demonstrated substantial areal variation where distant areas had distinct gene expression and some cell types clustered separately. Thus, both a canonical and a noncanonical architecture were apparent, depending on the granularity of cellular detail analyzed.

Topographic variation as a function of R-C position was a clear organizational feature. Prior microarray-based analysis of human (*22*) and macaque (*29*) cortex showed that molecular similarity varies as a function of distance on the cortical sheet that likely mirrors early developmental gradients (*51*, *52*). Here we see comparable variation by R-C position and similarity as a function of distance, predominantly in select cell types. As in the mouse cortex (*15*), areal variation was mostly in excitatory and not inhibitory neurons (except in V1). These results are consistent with the fact that most inhibitory neurons migrate from the ganglionic eminences and are relatively homogeneous across the cortex, whereas excitatory neurons are generated from progenitor cells with developmental gradients that are maintained in postmitotic neurons. R-C variation was seen not just in gene expression but also in excitatory neuron proportions. L4 IT neuron proportions increased from rostral to caudal, whereas L5 ET neurons proportions showed negative correlation, suggesting that developmental gradients likely sculpt the cortical cellular makeup. Similar R-C variation in cellular morphology in macaque monkeys supports this idea (*53*).

The molecular and cell-type distinctiveness of V1 in the present study mirrors the specialized cytoarchitecture of V1 in humans, primates, and other binocular mammals, unlike in mice in which V1 is similar to other cortical areas (*17*). V1 was more molecularly distinct than expected by topographic position, consistent with previous bulk microarray analysis (*22*) and expansion of thalamorecipient L4 was reflected in increased L4 IT proportions. There has been a long-standing debate about the cellular makeup of L4A and L4B in V1, which have alternatively been called 3BP and 3C, respectively (*26*, *54*). Our results show that L4A and L4B contain both L2/3 IT and L4 IT neuron types, providing a potential explanation for this confusion. Additional cellular diversity that does not strictly obey laminar boundaries complicates this organization, similar to previous work showing lack of strict laminar organization in human MTG (*11*).

The balance of excitation and inhibition is thought to be critical to proper balance of neuronal circuitry (*55*). E:I ratios of ~4:1 in the human frontal cortex (*56*) and 4:1 in monkey V1 (*57*, *58*) have been reported based on GABA immunohistochemistry. Transcriptomic quantification of cell proportions indicates a 5:1 E:I ratio in mouse cortex (*12*) versus a 2:1 ratio in human MTG and M1, which was confirmed by MERFISH here and in (*37*), and by electron microscopy analysis of mouse and human L2 (*59*). We find that the human E:I ratio of 2:1 is consistent across all areas except V1 in which the ratio is 4.5:1, likely as a result of increased L4 IT neurons in V1. However, the E:I ratio varies substantially by layer and is as high as 10:1 in L6 of V1. Whether this variation can be compensated by homeostatic processes remains to be studied, but these results indicate that the E:I ratio can vary substantially in the human cortex.

The current results illustrate the potential of single cell transcriptomics to provide a comprehensive cellular map of the cortex that can be thought of as a form of quantitative cytoarchitectonics based on the genes that give the cell types their properties. These analyses place a cellular lens on thinking about cortical areal variation as variation in the proportions and properties of the component cell types that define the input-output properties of those areas. Recent studies have shown that morphological and anatomical characteristics are correlated with transcriptomic identity (*13*, *47*, *60*), indicating that transcriptomic maps can also be highly predictive of cell phenotype variation. The present study sampled a small number of human tissue donors and further work will be required to understand variation of cortical gene expression and cell types across diverse individuals. Another future challenge will be creation of a multimodal map inclusive of the entire human neocortex where areal sampling is guided by detailed anatomical and functional parcellations that will reveal graded features versus discrete boundaries and enable direct linkage between the cellular and functional architecture of the cortex.

Materials and Methods

Postmortem tissue donors

Males and females 18 to 68 years of age with no known history of neuropsychiatric or neurological conditions, evidence of head trauma, intubation, or neuropathology were considered for inclusion in this study. De-identified postmortem human brain tissue was collected after obtaining permission from the decedent's legal next-of-kin. Tissue collection was performed in accordance with the provisions of the US 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 (WIRB) reviewed the use of de-identified postmortem brain tissue for research purposes and determined that, in accordance with federal regulation 45 CFR 46 and associated guidance, the use of de-identified specimens from deceased individuals did not constitute human subjects research requiring IRB review. Routine serological screening for infectious disease (HIV, Hepatitis B, and Hepatitis C) was conducted where possible using donor blood samples and donors negative for these infectious diseases were considered for inclusion in the study. Tissue RNA quality was assessed using samples of total RNA derived from the frontal and occipital poles of each donor brain which were processed on an Agilent 2100 Bioanalyzer using the RNA 6000 Nano kit to generate RNA Integrity Number (RIN) scores for each sample. Specimens with average RIN values ≥7.0 were considered for inclusion in the study. Tissue samples from five individuals (3 males, 2 females, mean postmortem interval 12.8 hours, mean age 47 years, table S2) were used for snRNA-seq data generation. Tissue samples from 3 individuals (3 males, table S2) were used for MERFISH data generation.

Processing of postmortem brain specimens

Postmortem brain specimens were transported to the Allen Institute or the University of Washington on ice and processed as previously described (https://dx.doi.org/10.17504/protocols.io.bf4ajqse). Briefly, brain specimens were bisected through the midline and individual hemispheres were embedded in Cavex Impressional Alginate for slabbing. Coronal brain slabs were cut at 1 cm intervals for all donors except H20.30.002 (table S2), which

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

was processed at a slab interval of 4mm. Tissue photographs were acquired for all slabs prior to freezing. Individual slabs were frozen in a slurry of dry ice and isopentane. Frozen slabs were vacuum sealed and stored at −80°C until the time of use.

Tissue mapping and dissection

Cortical areas of interest were identified on tissue slab photographs taken at the time of autopsy and at the time of dissection. Tissue samples used for Cv3 and SSv4 data generation were on average 3mm wide, encompassed the full height of the cortical depth from pia to white matter of the sampled area (~5mm), and were 1cm in thickness. Tissue photographs were used to map the tissue blocks sampled for Cv3 data generation across donors and areas to several reference atlases (table S2). First, samples were pinned to the Allen Human Reference Atlas 3D in MNI volume space, which includes labeling of 141 brain structures drawn on the ICBM 152 2009b Nonlinear Symmetric reference volume, using the publicly available BICCN Cell Locator tool (https://github.com/BICCN/cell-locator). Table S2 lists the coordinates and structure name corresponding to the approximate center of each cortical area pinned using the Cell Locator tool. As the Allen Human 3D Reference makes use of a gyral structural ontology the best matching structure in the Allen Human Reference plate-based 2D and associated Modified Brodmann structural ontology (see documentation at http://atlas.brain-map.org/) was also determined (table S2). Additionally, samples were mapped to the Julich Brain Maximum Probability Maps in MNI ICBM 152 2009c Nonlinear Asymmetric reference space (DOI: 10.25493/TAKY-64D) using Connectome Workbench (https://www.humanconnectome.org/software/connectome-workbench) for file viewing and annotation and mapped structures were cross-referenced to the publicly available Julich-Brain v2.9 parcellation (DOI: 10.25493/VSMK-H94) in the same 3D reference volume using the EBRAINS Siibra Explorer (https://atlases.ebrains.eu/viewer/#/). Table S2 lists the best matching (primary) brain structure from the Julich-Brain v2.9 parcellation and, for cases where the Julich Maximum Probability Maps predict more than one cortical area at the reference coordinates corresponding to a mapped sample, a secondary structure term is listed. In the Allen Institute Modified Brodmann ontology, dissections of DFC mapped to the superior frontal gyrus corresponding to the lateral subdivision of Brodmann Area (A) 9 (A9l). Dissections of ACC corresponded to A24 in the rostral (anterior) cingulate gyrus. A1 was localized in the transverse temporal gyrus (Heschl's gyrus) corresponding to A41. MTG dissections were mostly targeted to the caudal subdivision of A21 (A21c), but some dissections mapped to the intermediate subdivision of A21 (A21i). M1, S1, and V1 dissections mapped to the primary sensory regions M1C, S1C, and V1C, respectively. Localization of V1 was also confirmed by identification of the Stria of Gennari on tissue slab photographs. For SSv4 data generation, M1 and S1 dissections targeted the putative hand and trunk-lower limb sub-regions of each cortical area. Confirmation of the localization of tissue blocks in M1 and S1 was also carried out by processing one block from each donor for cryosectioning and fluorescent Nissl staining (Neurotrace 500/525, ThermoFisher Scientific). Nissl-stained sections were screened for histological hallmarks of each cortical area (such as the presence of Betz cells in L5 of M1) to verify that dissected regions were appropriately localized to either M1 or S1. AnG dissections targeted the caudal subdivision of A39 (A39c). All tissue dissections from parent tissue slabs were carried out using a custom cold table maintained at -20°C for the duration of dissection.

Nuclear isolation and capture

For SMART-seqv4 (SSv4) and Cv3 with layer 5 microdissection, tissue blocks were placed in ice-cold 1X PBS supplemented with 10mM DL-Dithiothreitol (DTT, Sigma Aldrich) and mounted on a vibratome (Leica) for sectioning at 500 µm in the coronal plane. Sections were placed in fluorescent Nissl staining solution (Neurotrace 500/525, ThermoFisher Scientific) prepared in 1X PBS with 10mM DTT and 0.5% RNasin Plus RNase inhibitor (Promega) and stained for 5 min on ice. After staining, sections were visualized on a fluorescence dissecting microscope (Leica) and cortical layers were individually microdissected using a needle blade micro-knife (Fine Science Tools) as previously described (https://dx.doi.org/10.17504/protocols.io.bq6ymzfw). Nuclear suspensions were prepared from microdissected tissue pieces as described (https://dx.doi.org/10.17504/protocols.io.ewov149p7vr2/v2). Dissected L5 tissue pieces for Cv3 processing were pooled across multiple sections per tissue block to ensure adequate sample for Cv3 chip loading. For Cv3 processing of tissue blocks encompassing all cortical layers, samples were placed directly into a Dounce homogenizer after removal from the −80°C freezer and processed as described (https://dx.doi.org/10.17504/protocols.io.bq64mzgw).

All samples were immunostained for fluorescence activated nuclear sorting (FANS) with mouse anti-NeuN conjugated to PE (EMD Millipore, FCMAB317PE) at a dilution of 1:500 with incubation for 30 min at 4°C. Control samples were incubated with mouse IgG1,k-PE Isotype control (BD Pharmingen). A subset of SSv4 samples was immunostained with rabbit anti-SATB2 conjugated to Alexa Fluor 647 (Abcam, ab196536) at a dilution of 1:500 to discriminate excitatory (SATB2+/NeuN+) and inhibitory (SATB2-/NeuN+) nuclei. After immunostaining, samples were centrifuged to concentrate nuclei and were resuspended in 1X PBS, 1% BSA, and 0.5% RNasin Plus for FACS. DAPI (4′, 6-diamidino-2-phenylindole, ThermoFisher Scientific) was applied to samples at a concentration of 0.1 µg/ml. Single nucleus sorting was carried out on either a BD FACSAria II SORP or BD FACSAria Fusion instrument (BD Biosciences) using a 130 µm nozzle. A standard gating strategy was applied to all samples as previously described (*11*). Briefly, nuclei were gated on their size and scatter properties and then on DAPI signal. Doublet discrimination gates were applied to exclude multiplets. Lastly, samples were gated on NeuN signal (PE) and SATB2 (Alexa Fluor 647) signal where applicable. For Cv3 experiments, NeuN+ and NeuN- nuclei were sorted into separate tubes and combined at a defined ratio of neurons and non-neurons (80% NeuN+, 20% NeuN-), except for L5 dissected samples where only neuronal (NeuN+) nuclei were captured. Samples were then centrifuged and resuspended in 1XPBS, 1% BSA, 0.5% RNasin Plus, and 5 to 10% DMSO and frozen at −80°C until the time of chip loading. Samples were processed according to the following protocol for chip loading (https://dx.doi.org/10.17504/protocols.io.774hrqw). For SSv4, single nuclei were sorted into 8-well strip tubes containing 11.5 µl of SMART-seq v4 collection buffer (Takara) supplemented with ERCC MIX1 spike-in synthetic RNAs at a final dilution of 1×10-8 (Ambion). Strip tubes containing sorted nuclei were briefly centrifuged and stored at −80°C until the time of further processing.

SMART-seqv4 RNA-sequencing

We used the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara #634894) per the manufacturer's instructions for reverse transcription of RNA and subsequent cDNA amplification as described (https://dx.doi.org/10.17504/protocols.io.8epv517xdl1b/v2). Standard controls were processed alongside each batch of experimental samples. Control strips included: 2 wells without cells, 2 wells without cells or ERCCs (i.e., no template controls), and either 4 wells of 10 pg of Human Universal Reference Total RNA (Takara 636538) or 2 wells of 10 pg of Human Universal Reference and 2 wells of 10 pg Control RNA provided in the Clontech kit. cDNA was amplified with 21 PCR cycles after the reverse transcription step. cDNA libraries were examined on either an Agilent Bioanalyzer 2100 using High Sensitivity DNA chips or an Advanced Analytics Fragment Analyzer (96) using the High Sensitivity NGS Fragment Analysis Kit (1bp to 6000bp). Purified cDNA was stored in 96-well plates at −20°C until library preparation.

The NexteraXT DNA Library Preparation (Illumina FC-131-1096) kit with NexteraXT

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

Index Kit V2 Sets A to D (FC-131-2001, 2002, 2003, or 2004) was used for sequencing library preparation as described (*11*). NexteraXT DNA Library prep was done at either 0.5x volume manually or 0.4x volume on the Mantis instrument (Formulatrix, https://dx.doi.org/10.17504/protocols.io.brdjm24n). Samples were quantitated using PicoGreen on a Molecular Bynamics M2 SpectraMax instrument. Sequencing libraries were assessed using either an Agilent Bioanalyzer 2100 with High Sensitivity DNA chips or an Advanced Analytics Fragment Analyzer with the High Sensitivity NGS Fragment Analysis Kit for sizing. Molarity was calculated for each sample using average size as reported by Bioanalyzer or Fragment Analyzer and pg/µl concentration as determined by PicoGreen. Samples were normalized to 2 to 10 nM with Nuclease-free Water (Ambion). Libraries were multiplexed at 96 samples/lane and sequenced on an Illumina HiSeq 2500 instrument using Illumina High Output V4 chemistry.

SMART-seqv4 RNA-seq gene expression quantification

Raw read (fastq) files were aligned to the GRCh38 human genome sequence (Genome Reference Consortium, 2011) with the RefSeq transcriptome version GRCh38.p2 (current as of 4/13/2015) and updated by removing duplicate Entrez gene entries from the gtf reference file for STAR processing. For alignment, Illumina sequencing adapters were clipped from the reads using the fastqMCF program (*61*). After clipping, the paired-end reads were mapped using Spliced Transcripts Alignment to a Reference (STAR) (*62*) using default settings. Reads that did not map to the genome were then aligned to synthetic constructs (External RNA Controls Consortium, ERCC) and the *E. coli* genome (version ASM584v2). The results files included quantification of the mapped reads (raw exon and intron counts for the transcriptome-mapped reads), and percentages of reads mapped to the RefSeq transcriptome, to ERCC spike-in controls, and to *E. coli*. Quantification was performed using summerizeOverlaps from the R package GenomicAlignments (*63*).

Expression was calculated as counts per million (CPM) of exonic plus intronic reads, and log2(CPM + 1) transformed values were used for a subset of analyses as described below. Gene detection was calculated as the number of genes expressed in each sample with CPM > 0. CPM values reflected absolute transcript number and gene length. Short and abundant transcripts may have the same apparent expression as long but rarer transcripts. Intron retention varied across genes, so no reliable estimates of effective gene lengths were available for expression normalization. Instead, absolute expression was estimated as fragments per kilobase per million (FPKM) using only exonic reads so that annotated transcript lengths could be used.

10x Chromium RNA-sequencing and expression quantification

Samples were processed using the 10x Chromium Single-Cell 3′ Reagent Kit v3 following the manufacturer's protocol as described (https://dx.doi.org/10.17504/protocols.io.bq7cmziw). Gene expression was quantified using the default 10x Cell Ranger v3 (Cell Ranger, RRID:SCR_017344) pipeline. The human reference genome used included the modified genome annotation described above for SMART-seq v4 quantification. Introns were annotated as "mRNA" and intronic reads were included in expression quantification.

RNA-sequencing processing and clustering

Cell-type label transfer

Human M1 reference taxonomy subclass labels (*12*) were transferred to nuclei in the current MTG dataset using Seurat's label transfer (3000 high variance genes using the 'vst' method then filtered through exclusion list). This was carried out for each RNA-seq modality dataset; for example, human-Cv3 and human-SSv4 were labeled independently. Each dataset was subdivided into 5 neighborhoods—IT and Non-IT excitatory neurons, CGE- and MGE-derived interneurons, and non-neuronal cells—based on marker genes and transferred subclass labels from published studies of human and mouse cortical cell types and cluster grouping relationships in a reduced dimensional gene expression space.

Filtering low-quality nuclei

SSv4 nuclei were included for analysis if they passed all QC criteria:

> 30% cDNA longer than 400 base pairs

> 500,000 reads aligned to exonic or intronic sequence

> 40% of total reads aligned

> 50% unique reads

> 0.7 TA nucleotide ratio

QC was then performed at the neighborhood level. Neighborhoods were integrated together across all areas and modality; for example, deep excitatory neurons from human-Cv3, human-Cv3-Layer5 and human-SSv4 datasets were integrated using Seurat integration functions with 2000 high variance genes. Integrated neighborhoods were Louvain clustered into over 100 meta cells, and Low-quality meta cells were removed from the dataset based on relatively low UMI or gene counts (included glia and neurons with greater than 500 and 1000 genes detected, respectively), predicted doublets (include nuclei with doublet scores under 0.3), and/or subclass label prediction metrics within the neighborhood (excitatory labeled nuclei that clustered with majority inhibitory or non-neuronal nuclei).

RNA-seq clustering

Nuclei were normalized using SCTransform (*64*), and neighborhoods were integrated together within an area and across individuals and modalities by identifying mutual nearest neighbor anchors and applying canonical correlation analysis as implemented in Seurat (*65*). For example, deep excitatory neurons from human-Cv3 were split by individuals and integrated with the human-SSv4 deep excitatory neurons. Integrated neighborhoods were Louvain clustered into over 100 meta cells. Meta cells were then merged with their nearest neighboring meta cell until merging criteria were sufficed, a split and merge approach that has been previously described (*12*). The remaining clusters underwent further QC to exclude Low-quality and outlier populations. These exclusion criteria were based on irregular groupings of metadata features that resided within a cluster.

Defining cross-area consensus cell types

For each neighborhood, Cv3 nuclei were integrated together across individuals. The integrated latent space was Louvain clustered into over 100 meta cells. Meta cells were then merged with their nearest neighboring meta cell until merging criteria were sufficed, a split and merge approach that has been previously described (*12*) and was also used to define the within-area cluster identities. The process was repeated for each neighborhood, with an example diagram of the workflow shown in Fig. 5A.

Cell-type taxonomy generation

For each area, a taxonomy was built using the final set of clusters and was annotated using subclass mapping scores, dendrogram relationships, marker gene expression, and inferred laminar distributions. Within-area taxonomy dendrograms were generated using build_dend function from scrattch_hicat R package. A matrix of cluster median log2(cpm + 1) expression across the 3000 High-variance genes for Cv3 nuclei from a given area were used as input. The cross-area dendrogram was generated with a similar workflow but was downsampled to a maximum of 100 nuclei per cross-area cluster per area. The 3000 High-variance genes used for dendrogram construction were identified from the downsampled matrix containing Cv3 nuclei from all eight areas.

Cell-type comparisons across cortical areas

Differential gene expression

To identify subclass marker genes within an area, Cv3 datasets from each area were downsampled to a maximum of 100 nuclei per cluster per individual. Differentially expressed marker genes were then identified using the FindAllMarkers function from Seurat, using the Wilcoxon sum rank test on log-normalized matrices with a maximum of 500 nuclei per

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

group (subclass versus. all other nuclei as background). Statistical thresholds for markers are indicated in their respective figures. To identify area marker genes across subclasses, Cv3 datasets from each area were downsampled to a maximum of 50 nuclei per cluster per individual. Downsampled counts matrices were then grouped into pseudo-bulk replicates (area, individual, subclass) and the counts were summed per replicate. DESeq2 functionality was then used to perform a differential expression analysis between area pairs (or comparisons of interest) for each subclass using the Wald test statistic.

Transcriptomic entropy across areas

To quantify intercell transcriptomic heterogeneity across areas for each subclass we calculated the transcriptomic entropy in the observed data (structured) and compared against entropy in permuted data (unstructured). Transcriptomic heterogeneity is defined as the difference between the structured and unstructured entropy. To compute transcriptomic entropy, we followed these steps: (1) Randomly down-sample the cells within each subclass by taking 250 cells from each cross-area cell type. (2) Identify the highly variable genes in each area and take the union of genes as our set of interest. (3) Then, by following a recently reported computational approach to quantify transcriptomic heterogeneity (*66*), we computed the per-area transcriptomic entropy for each subclass.

Identifying changes in cell-type proportions across areas

Cell-type proportions are compositional, where the gain or loss of one population necessarily affects the proportions of the others, so we used scCODA (*25*) to determine which changes in cell class, subclass, and cell-type proportions across areas were statistically significant. We analyzed neuronal and non-neuronal populations separately because nuclei were sorted based on NeuN immunostaining to enrich for neurons. The proportion of each cell type was estimated using a Bayesian approach where proportion differences across individuals were used to estimate the posterior. All compositional and categorical analyses require a reference population to describe differences with respect to and, because we were uncertain which populations should be unchanged, we iteratively used each cell type and each area as a reference when computing abundance changes. To account for sex differences, we included it as a covariate when testing for abundance changes. Separately for neuronal and non-neuronal populations, we reported the effect size of each area for each cell type (table S10) and used a mean inclusion probability cutoff of 0.7 for calling a population consistently different.

Partitioning variation in gene expression across areas

Variation partitioning analysis was performed to prioritize the drivers of variation across areas within each subclass. Using linear mixed-effect models implemented in the variancePartitioning bioconductor package: http://bioconductor.org/packages/variancePartition (*66*) we identify genes whose variance is best explained along the M-S (anatomical left to right), R-C, and D-V axes as well as by cortical area and donor. The order of areas along these axes was defined based on the approximate x, y, and z coordinates of tissue samples based on a common coordinate framework of the adult human brain (*20*) (table S2). Genes were removed from the analysis based on the following criteria: (1) expressed in less than 10 cells, (2) greater than 80% dropout rate, (3) zero variance in expression, and (4) expression less than 1 CPM on average. The variance partitioning linear mixed-effect model was then defined as:

$$\text{Gene} \sim \text{medial\_lateral} + \text{rostral\_caudal} + \text{dorsal\_ventral} + (1|\text{area}) + (1|\text{donor})$$

and passed into the variancePartition function `fitVarPartModel()`. We determined the amount of variation explained per covariate for each gene from the `extractVarPart()` function.

In situ profiling of gene expression

Human postmortem frozen brain tissue was embedded in Optimum Cutting Temperature medium (VWR,25608-930) and sectioned on a Leica cryostat at −17°C at 10 μm onto Vizgen MERSCOPE coverslips (VIZGEN 2040003). These sections were then processed for MERSCOPE imaging according to the manufacturer's instructions. Briefly: sections were allowed to adhere to these coverslips at room temperature for 10 min prior to a 1 min wash in nuclease-free phosphate buffered saline (PBS) and fixation for 15 min in 4% paraformaldehyde in PBS. Fixation was followed by 3x5 min washes in PBS prior to a 1 min wash in 70% ethanol. Fixed sections were then stored in 70% ethanol at 4 C prior to use and for up to one month. Human sections were photobleached using a 150W LED array for 72 hours at 4°C prior to hybridization then washed in 5 ml Sample Prep Wash Buffer (VIZGEN 20300001) in a 5 cm petri dish. Sections were then incubated in 5 ml Formamide Wash Buffer (VIZGEN 20300002) at 37 C for 30 min. Sections were hybridized by placing 50 μl of VIZGEN-supplied Gene Panel Mix onto the section, covering with parafilm and incubating at 37°C for 36 to 48 hours in a humidified hybridization oven.

Following hybridization, sections were washed twice in 5 ml Formamide Wash Buffer for 30 min at 47°C. Sections were then embedded in acrylamide by polymerizing VIZGEN Embedding Premix (VIZGEN 20300004) according to the manufacturer's instructions. Sections were embedded by inverting sections onto 110 μl of Embedding Premix and 10% Ammonium Persulfate (Sigma A3678) and TEMED (BioRad 161-0800) solution applied to a Gel Slick (Lonza 50640) treated 2x3 glass slide. The coverslips were pressed gently onto the acrylamide solution and allowed to polymerize for 1.5 hours. Following embedding, sections were cleared for 24 to 48 hours with a mixture of VIZGEN Clearing Solution (VIZGEN 20300003) and Proteinase K (New England Biolabs P8107S) according to the Manufacturer's instructions. Following clearing, sections were washed twice for 5 min in Sample Prep Wash Buffer (PN 20300001). VIZGEN DAPI and PolyT Stain (PN 20300021) was applied to each section for 15 min followed by a 10 min wash in Formamide Wash Buffer. Formamide Wash Buffer was removed and replaced with Sample Prep Wash Buffer during MERSCOPE set up. 100 μl of RNAse Inhibitor (New England BioLabs M0314L) was added to 250 μl of Imaging Buffer Activator (PN 203000015) and this mixture was added through the cartridge activation port to a pre-thawed and mixed MERSCOPE Imaging cartridge (VIZGEN PN1040004). 15 ml mineral oil (Millipore-Sigma m5904-6X500ML) was added to the activation port and the MERSCOPE fluidics system was primed according to VIZGEN instructions. The flow chamber was assembled with the hybridized and cleared section coverslip according to VIZGEN specifications and the imaging session was initiated after collection of a 10X mosaic DAPI image and selection of the imaging area. For specimens that passed the minimum count threshold, imaging was initiated, and processing completed according to VIZGEN's proprietary protocol. Following processing and segmentation through MERSCOPE software, cells with fewer than 50 counts, or with an area outside the 100 to 300 μm^2 range were eliminated from the mapping process.

The 140 gene human cortical panel was selected using a combination of manual and algorithmic based strategies requiring a reference single cell/nucleus RNA-seq data set from the same tissue, in this case the human MTG snRNAseq dataset and resulting taxonomy (*11*). First, an initial set of high-confidence marker genes are selected through a combination of literature search and analysis of the reference data. These genes are used as input for a greedy algorithm (detailed below). Second, the reference RNA-seq data set is filtered to only include genes compatible with mFISH. Retained genes need to 1) be long enough to allow probe design (> 960 base pairs); 2) be expressed highly enough to be detected (FPKM >= 10), but not so high as to overcrowd the signal of other genes in a cell (FPKM < 500); 3) have low expression in off-target cells (FPKM < 50 in non-neuronal cells); and 4) be differentially expressed

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

between cell types (top 500 remaining genes by marker score20). To sample each cell type more evenly, the reference data set is also filtered to include a maximum of 50 cells per cluster.

The main step of gene selection uses a greedy algorithm to iteratively add genes to the initial set. To do this, each cell in the filtered reference data set is mapped to a cell type by taking the Pearson correlation of its expression levels with each cluster median using the initial gene set of size n, and the cluster corresponding to the maximum value is defined as the "mapped cluster". The "mapping distance" is then defined as the average cluster distance between the mapped cluster and the originally assigned cluster for each cell. In this case a weighted cluster distance, defined as one minus the Pearson correlation between cluster medians calculated across all filtered genes, is used to penalize cases where cells are mapped to very different types, but an unweighted distance, defined as the fraction of cells that do not map to their assigned cluster, could also be used. This mapping step is repeated for every possible n+1 gene set in the filtered reference data set, and the set with minimum cluster distance is retained as the new gene set. These steps are repeated using the new get set (of size n+1) until a gene panel of the desired size is attained. Code for reproducing this gene selection strategy is available as part of the mfishtools R library (https://github.com/AllenInstitute/mfishtools).

Cell-type mapping of MERSCOPE data

Any genes not matched across both the MERSCOPE gene panel and the snRNASeq mapping taxonomy were filtered from the snRNASeq dataset. We calculated the mean gene expression for each gene in each snRNAseq cluster. We assigned MERSCOPE cells to snRNAseq clusters by finding the nearest cluster to the mean expression vectors of the snRNASeq clusters using the cosine distance. All scripts and data used are available at: https://github.com/AllenInstitute/human_cross_areal.

GFAP Immunofluorescence

Tissue blocks from cortical areas of interest were removed from fresh-frozen tissue slabs as described above. Immediately after dissection, tissue blocks were drop-fixed in cold 4% paraformaldehyde overnight in a 4°C fridge. Tissue blocks were then rinsed in multiple washes of 1X PBS, cryoprotected in sequential 15% and 30% sucrose solutions, and embedded in OCT. Sections were cut free floating at 30 μm in the coronal plane on a Leica cryostat into 1X PBS and were stored at 4°C or at −20°C in cryoprotectant solution until the time of use. Sections were processed for immunofluorescence using a rabbit polyclonal anti-GFAP antibody (Agilent, Z0334) at a dilution of 1:1000 and mouse monoclonal anti-NeuN antibody (Millipore Sigma, MAB377) at a dilution of 1:1000. Primary antibodies were incubated overnight at 4°C, followed by incubation in Alexa Fluor conjugated secondary species-specific antibodies for 2 hours at room temperature. Sections were counterstained with DAPI and Neurotrace 500 fluorescent Nissl stain and were mounted in ProLong Gold Antifade Mountant (ThermoFisher Scientific). Sections were imaged on a Nikon TiE fluorescence microscope equipped with NIS-Elements Advanced Research imaging software (v4.20). GFAP processes were traced using the SNT plugin in the Fiji distribution of ImageJ.

REFERENCES AND NOTES

1. C. von Economo, G. N. Koskinas, L. C. Triarhou, *Atlas of Cytoarchitectonics of The Adult Human Cerebral Cortex* (Karger, 2008).
2. K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde: in ihren Prinzipien dargestellt auf Grund d. Zellenbaues* (Verlag, 1909).
3. C. Vogt, O. Vogt, *Allgemeine Ergebnisse unserer Hirnforschung* (Verlag, 1919).
4. K. Amunts, K. Zilles, Architectonic Mapping of the Human Brain beyond Brodmann. *Neuron* **88**, 1086–1107 (2015). doi: 10.1016/j.neuron.2015.12.001; pmid: 26687219
5. R. J. Douglas, K. A. C. Martin, Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004). doi: 10.1146/annurev.neuro.27.070203.144152; pmid: 15217339
6. J. Defelipe, H. Markram, K. S. Rockland, The neocortical column. *Front. Neuroanat.* **6**, 2012.00022 (2012). doi: 10.3389/fnana.2012.00022; pmid: 22745629
7. R. Benavides-Piccione, C. Rojo, A. Kastanauskaite, J. DeFelipe, Variation in Pyramidal Cell Morphology Across the Human Anterior Temporal Lobe. *Cereb. Cortex* **31**, 3592–3609 (2021). doi: 10.1093/cercor/bhab034; pmid: 33723567
8. G. N. Elston, R. Benavides-Piccione, A. Elston, P. R. Manger, J. Defelipe, Pyramidal cells in prefrontal cortex of primates: Marked differences in neuronal structure among species. *Front. Neuroanat.* **5**, 2011.00002 (2011). doi: 10.3389/fnana.2011.00002; pmid: 21347276
9. J. I. Luebke, Pyramidal Neurons Are Not Generalizable Building Blocks of Cortical Networks. *Front. Neuroanat.* **11**, 2017.00011 (2017) doi: 10.3389/fnana.2017.00011; pmid: 28326020
10. 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; pmid: 31435019
11. 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; pmid: 34616062
12. E. M. Callaway *et al*., A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021). doi: 10.1038/s41586-021-03950-0; pmid: 34616075
13. B. Tasic *et al*., Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016). doi: 10.1038/nn.4216; pmid: 26727548
14. 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; pmid: 30382198
15. Z. Yao *et al*., A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021). doi: 10.1038/s41586-021-03500-8; pmid: 34616066
16. Z. Yao *et al*., A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e26 (2021). doi: 10.1016/j.cell.2021.04.021; pmid: 34004146
17. M. Zhang *et al*., Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021). doi: 10.1038/s41586-021-03705-x; pmid: 34616063
18. F. Scala *et al*., Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021). doi: 10.1038/s41586-020-2907-3; pmid: 33184512
19. S. L. Ding *et al*., Comprehensive cellular-resolution atlas of the adult human brain. *J. Comp. Neurol.* **524**, 3127–3481 (2016). doi: 10.1002/cne.24080; pmid: 27418273
20. K. Amunts, H. Mohlberg, S. Bludau, K. Zilles, Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* **369**, 988–992 (2020). doi: 10.1126/science.abb4588; pmid: 32732281
21. M. J. Hawrylycz *et al*., An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012). doi: 10.1038/nature11405; pmid: 22996553
22. I. Yanai *et al*., Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005). doi: 10.1093/bioinformatics/bti042; pmid: 15388519
23. J. A. Miller *et al*., Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229
24. M. Büttner, J. Ostner, C. L. Müller, F. J. Theis, B. Schubert, scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 6876 (2021). doi: 10.1038/s41467-021-27150-6; pmid: 34824236
25. P. Balaram, N. A. Young, J. H. Kaas, Histological features of layers and sublayers in cortical visual areas V1 and V2 of chimpanzees, macaque monkeys, and humans. *Eye Brain*. **6**, 5–18 (2014). doi: 10.2147/EB.S51814; pmid: 25788835
26. N. H. Yabuta, E. M. Callaway, Functional streams and local connections of layer 4C neurons in primary visual cortex of the macaque monkey. *J. Neurosci.* **18**, 9489–9499 (1998). doi: 10.1523/JNEUROSCI.18-22-09489.1998; pmid: 9801386
27. J. C. Horton, E. T. Hedley-Whyte, T. N. Wiesel, Mapping of cytochrome oxidase patches and ocular dominance columns in human visual cortex. *Philos. Trans. R. Soc. B* **304**, 255–272 (1984). doi: 10.1098/rstb.1984.0022; pmid: 6142485
28. A. Bernard *et al*., Transcriptional architecture of the primate neocortex. *Neuron* **73**, 1083–1099 (2012). doi: 10.1016/j.neuron.2012.03.002; pmid: 22445337
29. C. R. Siu, J. L. Balsor, D. G. Jones, K. M. Murphy, Classic and Golli Myelin Basic Protein have distinct developmental trajectories in human visual cortex. *Front. Neurosci.* **9**, 2015.00138 (2015). doi: 10.3389/fnins.2015.00138; pmid: 25964736
30. H. L. Nhan, E. M. Callaway, Morphology of superior colliculus- and middle temporal area-projecting neurons in primate primary visual cortex. *J. Comp. Neurol.* **520**, 52–80 (2012). doi: 10.1002/cne.22685; pmid: 21674487
31. W. Fries, H. Distel, J. Z. Young, Large layer VI neurons of monkey striate cortex (Meynert cells) project to the superior colliculus. *Proc. R. Soc. B* **219**, 53–59 (1983). doi: 10.1098/rspb.1983.0058; pmid: 6137827
32. W. Fries, K. Keizer, H. G. J. M. Kuypers, Large layer VI cells in macaque striate cortex (Meynert cells) project to both superior colliculus and prestriate visual area V5. *Exp. Brain Res.* **58**, 613–616 (1985). doi: 10.1007/BF00235878; pmid: 3839191
33. R. D. Hodge *et al*., Transcriptomic evidence that von Economo neurons are regionally specialized extratelencephalic-projecting excitatory neurons. *Nat. Commun.* **11**, 1172 (2020). doi: 10.1038/s41467-020-14952-3; pmid: 32127543
34. S. R. Stevens *et al*., Ankyrin-R links Kv3.3 to the spectrin cytoskeleton and is required for purkinje neuron survival. *J. Neurosci.* **42**, 2–15 (2022). doi: 10.1523/JNEUROSCI.1132-21.2021; pmid: 34785580
35. A. C. Yang *et al*., A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022). doi: 10.1038/s41586-021-04369-3; pmid: 35165441
36. R. Fang *et al*., Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022). doi: 10.1126/science.abm1741; pmid: 35771910
37. M. F. Glasser, D. C. Van Essen, Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* **31**, 11597–11616 (2011). doi: 10.1523/JNEUROSCI.2180-11.2011; pmid: 21832190
38. M. F. Glasser *et al*., A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016). doi: 10.1038/nature18933; pmid: 27437579
39. K. Siletti *et al*., Transcriptomic diversity of cell types across the adult human brain. bioRxiv 2022.10.12.511898 [Preprint] (2022). doi: 10.1101/2022.10.12.511898v1
40. 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; pmid: 30096314

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

41. M. Y. Batiuk *et al*., Identification of region-specific astrocyte subtypes at single cell resolution. *Nat. Commun.* **11**, 1220 (2020). doi: 10.1038/s41467-019-14198-8; pmid: 32139688
42. N. A. Oberheim *et al*., Uniquely hominid features of adult human astrocytes. *J. Neurosci.* **29**, 3276–3287 (2009). doi: 10.1523/JNEUROSCI.4707-08.2009; pmid: 19279265
43. C. Falcone *et al*., Redefining varicose projection astrocytes in primates. *Glia* **70**, 145–154 (2022). doi: 10.1002/glia.24093; pmid: 34533866
44. C. Falcone *et al*., Cortical Interlaminar Astrocytes Are Generated Prenatally, Mature Postnatally, and Express Unique Markers in Human and Nonhuman Primates. *Cerebral Cortex.* **31**, 379–395 (2021). doi: 10.1002/cne.24605; pmid: 30552685
45. C. Falcone *et al*., Cortical interlaminar astrocytes across the therian mammal radiation. *J. Comp. Neurol.* **527**, 1654–1674 (2019). doi: 10.1002/cne.24605; pmid: 30552685
46. J. Berg *et al*., Human neocortical expansion involves glutamatergic neuron diversification. *Nature* **598**, 151–158 (2021). doi: 10.1038/s41586-021-03813-8; pmid: 34616067
47. M. Á. García-Cabezas, H. Barbas, Area 4 has layer IV in adult primates. *Eur. J. Neurosci.* **39**, 1824–1834 (2014). doi: 10.1111/ejn.12585; pmid: 24735460
48. H. Barbas, M. Á. García-Cabezas, Motor cortex layer 4: Less is more. *Trends Neurosci.* **38**, 259–261 (2015). doi: 10.1016/j.tins.2015.03.005; pmid: 25868984
49. N. Yamawaki, K. Borges, B. A. Suter, K. D. Harris, G. M. G. Shepherd, A genuine layer 4 in motor cortex with prototypical synaptic circuit connectivity. *eLife* **3**, e05422 (2014). doi: 10.7554/eLife.05422; pmid: 25525751
50. K. M. Bishop, J. L. R. Rubenstein, D. D. M. O'Leary, Distinct actions of Emx1, Emx2, and Pax6 in regulating the specification of areas in the developing neocortex. *J. Neurosci.* **22**, 7627–7638 (2002). doi: 10.1523/JNEUROSCI.22-17-07627.2002; pmid: 12196586
51. D. D. O'Leary, S. Sahara, Genetic regulation of arealization of the neocortex. *Curr. Opin. Neurobiol.* **18**, 90–100 (2008). doi: 10.1016/j.conb.2008.05.011; pmid: 18524571
52. G. N. Elston, R. Tweedale, M. G. P. Rosa, Cortical integration in the visual system of the macaque monkey: Large-scale morphological differences in the pyramidal neurons in the occipital, parietal and temporal lobes. *Proc. Biol. Sci.* **266**, 1367–1374 (1999). doi: 10.1098/rspb.1999.0789; pmid: 10445291
53. R. Hassler, in *Evolution of the Forebrain*, R. Hassler, H. Stephan, Eds. (Springer, 1966); pp. 419–434.
54. V. S. Sohal, J. L. R. Rubenstein, Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol. Psychiatry* **24**, 1248–1257 (2019). doi: 10.1038/s41380-019-0426-0; pmid: 31089192
55. J.-P. Hornung, N. De Tribolet, Distribution of GABA-containing neurons in human frontal cortex: A quantitative immunocytochemical study. *Anat. Embryol.* **189**, 139–145 (1994). doi: 10.1007/BF00185772; pmid: 8010412
56. C. Beaulieu, Z. Kisvarday, P. Somogyi, M. Cynader, A. Cowey, Quantitative distribution of GABA-immunopositive and -immunonegative neurons and synapses in the monkey striate cortex (area 17). *Cereb. Cortex* **2**, 295–309 (1992). doi: 10.1093/cercor/2.4.295; pmid: 1330121
57. S. H. Hendry, H. D. Schwark, E. G. Jones, J. Yan, Numbers and proportions of GABA-immunoreactive neurons in different areas of monkey cerebral cortex. *J. Neurosci.* **7**, 1503–1519 (1987). doi: 10.1523/JNEUROSCI.07-05-01503.1987; pmid: 3033170
58. S. Loomba *et al*., Connectomic comparison of mouse and human cortex. *Science* **377**, eabo0924 (2022). doi: 10.1126/science.abo0924; pmid: 35737810
59. T. Chartrand *et al*., Morpho-electric and transcriptomic divergence of the layer 1 interneuron repertoire in human versus mouse neocortex. bioRxiv 2022.10.24.511199 [Preprint] (2022); doi: 10.1101/2022.10.24.511199v1
60. E. Aronesty, Comparison of Sequencing Utility Programs. *Open Bioinform. J.* **7**, 1–8 (2013). doi: 10.2174/1875036201307010001
61. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
62. M. Lawrence *et al*., Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013). doi: 10.1371/journal.pcbi.1003118; pmid: 23950696
63. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019). doi: 10.1186/s13059-019-1874-1; pmid: 31870423
64. T. Stuart *et al*., Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019). doi: 10.1016/j.cell.2019.05.031; pmid: 31178118
65. G. E. Hoffman, E. E. Schadt, variancePartition: Interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016). doi: 10.1186/s12859-016-1323-z; pmid: 27884101

ACKNOWLEDGMENTS

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 postmortem and neurosurgical brain specimens; the technology team at the Allen Institute for assistance with data management; M. Vawter, J. Davis and the San Diego Medical Examiner's Office for assistance with postmortem tissue donations. This project made use of Connectome DB and Connectome Workbench, developed under the auspices of the Human Connectome Project at Washington University in St. Louis and associated consortium institutions (https://www.humanconnectome.org/). This publication was supported by and coordinated through the BICCN. This publication is part of the Human Cell Atlas- www.humancellatlas.org/publications/. Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Numbers U01MH114812 and U19MH114830. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank the founder of the Allen Institute, P. G. Allen, for his vision, encouragement, and support. **Funding:** This work was supported by the following: Knut and Alice Wallenberg Foundation 2018.0220 (to S.L.); Nancy and Buster Alvord Endowment (to C.D.K.); National Institutes of Health grant U01MH114812 (to A.M.Y., A.T., C.R., D.B., D.M., E.S.L., H.T., J.G., J.S., K.Si., K.Sm., M.T., N.D., N.L.J., R.D.H., S.D., S.L., T.C., T.E.B., and T.P.); National Institutes of Health grant U19MH117023 (to P.R.H.); National Institutes of Health grant U19MH114830 (to A.T., D.B., D.M., H.T., J.G., J.S., K.C., K.Sm., M.Kr., M.T., N.D., T.C., and T.P.) **Author contributions:** RNA data generation: A.M.Y., A.T., B.P.L., C.D.K., C.R., D.B., D.H., D.M., E.R.B., E.S.L., H.T., J.G., J.S., K.C., K.L., K.Si., K.Sm., K.K., M.Kr., M.T., N.D., N.S., R.D.H., S.D., S.L., S.Sh., T.C., T.E.B., and T.P. Spatial transcriptomic data generation: A.R., B.L., D.M., E.G., J.Ca., J.Cl., M.Ku., and N.M. Data archive / Infrastructure: J.G., S.So. Data analysis: B.L., E.G., E.S.L., J.Cl., J.G., K.J.T., K.Sm., N.J., N.L.J., R.D.H., and T.E.B. Data interpretation: E.M.C., E.S.L., J.Cl., J.G., K.J.T., N.J., N.L.J., P.P.M., P.R.H., R.D.H., S.D., T.E.B. Writing manuscript: E.S.L., J.Cl., K.J.T., N.L.J., P.R.H., R.D.H., T.E.B. **Competing interests:** From April 11, 2022, N.L.J. has been an employee of Genentech. **Data and materials availability:** Raw sequence data were produced as part of the BRAIN Initiative Cell Census Network (BICCN: RRID:SCR_015820) are available for download from the Neuroscience Multi-omics Archive (RRID:SCR_016152; https://assets.nemoarchive.org/dat-rg2rc5m) and the Brain Cell Data Center (RRID:SCR_017266; https://biccn.org/data). Code for analysis and generation of figures is available for download from https://github.com/AllenInstitute/human_cross_areal. MTG human SMARTseq v4 data (https://portal.brain-map.org/atlases-and-data/rnaseq/human-mtg-smart-seq, https://assets.nemoarchive.org/dat-swzf4kc). **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.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf6812
Figs. S1 to S12
Tables S1 to S13
Legends for tables S1 to S13

Submitted 6 November 2022; accepted 8 September 2023
10.1126/science.adf6812

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

Science
AAAS

Transcriptomic cytoarchitecture reveals principles of human neocortex organization

Nikolas L. Jorstad, Jennie Close, Nelson Johansen, Anna Marie Yanny, Eliza R. Barkan, Kyle J. Travaglini, Darren Bertagnolli, Jazmin Campos, Tamara Casper, Kirsten Crichton, Nick Dee, Song-Lin Ding, Emily Gelfand, Jeff Goldy, Daniel Hirschstein, Katelyn Kiick, Matthew Kroll, Michael Kunst, Kanan Lathia, Brian Long, Naomi Martin, Delissa McMillen, Trangthanh Pham, Christine Rimorin, Augustin Ruiz, Nadiya Shapovalova, Soraya Shehata, Kimberly Siletti, Saroja Somasundaram, Josef Sulc, Michael Tieu, Amy Torkelson, Herman Tung, Edward M. Callaway, Patrick R. Hof, C. Dirk Keene, Boaz P. Levi, Sten Linnarsson, Partha P. Mitra, Kimberly Smith, Rebecca D. Hodge, Trygve E. Bakken, and Ed S Lein

Science **382** (6667), eadf6812. DOI: 10.1126/science.adf6812

View the article online
https://www.science.org/doi/10.1126/science.adf6812
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at Cold Spring Harbor Laboratory on October 20, 2023

Use of this article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works