






# Cellular Diversity in Human Subgenual Anterior Cingulate and Dorsolateral Prefrontal Cortex by Single-Nucleus RNA-Sequencing

Billy Kim,<sup>1\*</sup>  Downon Kim,<sup>1\*</sup> Anton Schulmann,<sup>2\*</sup> Yash Patel,<sup>1\*</sup>  Carolina Caban-Rivera,<sup>1</sup> Paul Kim,<sup>1</sup> Ananya Jambhale,<sup>1</sup> Kory R. Johnson,<sup>3</sup> Ningping Feng,<sup>1</sup> Qing Xu,<sup>1</sup> Sun Jung Kang,<sup>4</sup> Ajeet Mandal,<sup>1</sup>  Michael Kelly,<sup>5</sup> Nirmala Akula,<sup>2</sup>  Francis J. McMahon,<sup>2</sup> Barbara Lipska,<sup>1</sup> Stefano Marengo,<sup>1\*\*</sup> and  Pavan K. Auluck<sup>1\*\*</sup>

<sup>1</sup>Human Brain Collection Core, National Institute of Mental Health-Intramural Research Program, Bethesda, Maryland 20892, <sup>2</sup>Human Genetics Branch, National Institute of Mental Health-Intramural Research Program, Bethesda, Maryland 20892, <sup>3</sup>Information Technology and Bioinformatics Program, National Institute of Neurological Disorders and Stroke-Intramural Research Program, Bethesda, Maryland 20892, <sup>4</sup>Genetic Epidemiology Research Branch, National Institute of Mental Health-Intramural Research Program, Bethesda, Maryland 20892, and <sup>5</sup>CCR Single Analysis Facility, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Bethesda, Maryland 20892

Regional cellular heterogeneity is a fundamental feature of the human neocortex; however, details of this heterogeneity are still undefined. We used single-nucleus RNA-sequencing to examine cell-specific transcriptional features in the dorsolateral PFC (DLPFC) and the subgenual anterior cingulate cortex (sgACC), regions implicated in major psychiatric disorders. Droplet-based nuclei-capture and library preparation were performed on replicate samples from 8 male donors without history of psychiatric or neurologic disorder. Unsupervised clustering identified major neural cell classes. Subsequent iterative clustering of neurons further revealed 20 excitatory and 22 inhibitory subclasses. Inhibitory cells were consistently more abundant in the sgACC and excitatory neuron subclusters exhibited considerable variability across brain regions. Excitatory cell subclasses also exhibited greater within-class transcriptional differences between the two regions. We used these molecular definitions to determine which cell classes might be enriched in loci carrying a genetic signal in genome-wide association studies or for differentially expressed genes in mental illness. We found that the heritable signals of psychiatric disorders were enriched in neurons and that, while the gene expression changes detected in bulk-RNA-sequencing studies were dominated by glial cells, some alterations could be identified in specific classes of excitatory and inhibitory neurons. Intriguingly, only two excitatory cell classes exhibited concomitant region-specific enrichment for both genome-wide association study loci and transcriptional dysregulation. In sum, by detailing the molecular and cellular diversity of the DLPFC and sgACC, we were able to generate hypotheses on regional and cell-specific dysfunctions that may contribute to the development of mental illness.

**Key words:** cingulate; dorsolateral prefrontal cortex; RNA-seq; schizophrenia; single-nucleus

Received Apr. 29, 2022; revised Feb. 27, 2023; accepted Mar. 15, 2023.

Author contributions: B.K., D.K., C.C.-R., P.K., N.F., Q.X., and M.K. performed research; B.K., A.S., Y.P., A.J., K.R.J., S.J.K., A.M., B.L., S.M., and P.K.A. analyzed data; B.K. and D.K. wrote the first draft of the paper; A.S., N.A., F.J.M., B.L., and P.K.A. edited the paper; M.K. and N.A. contributed unpublished reagents/analytic tools; F.J.M., S.M., and P.K.A. designed research; S.M. and P.K.A. wrote the paper.

This work was supported by the Intramural Research Program of the National Institute of Mental Health ZICMH002903-15. This work utilized the computational resources of the National Institutes of Health HPC Biowulf cluster (<http://hpc.nih.gov>). We thank the families of the deceased for tissue donations; the Offices of the Medical Examiners of Washington, D.C. and Virginia for the extraction of the tissue; the Single Cell Analysis Facility, National Cancer Institute Center for Cancer Research and the Sequencing Facility, and National Cancer Institute Center for Cancer Research for support in

the snRNA-seq protocol, which are funded through the Frederick National Laboratory for Cancer Research Contract 75N91019D00024; and Advanced Cell Diagnostics for support in the RNAscope protocol.

\*B.K., D.K., A.S., and Y.P. contributed equally to this work.

\*\*S.M. and P.K.A. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Stefano Marengo at [marencos@mail.nih.gov](mailto:marencos@mail.nih.gov) or Pavan K. Auluck at [pavan.auluck@nih.gov](mailto:pavan.auluck@nih.gov).

<https://doi.org/10.1523/JNEUROSCI.0830-22.2023>

Copyright © 2023 the authors

### Significance Statement

Dysfunction of the subgenual anterior cingulate cortex has been implicated in mood disorders, particularly major depressive disorder, and the dorsolateral PFC, a subsection of the PFC involved in executive functioning, has been implicated in schizophrenia. Understanding the cellular composition of these regions is critical to elucidating the neurobiology underlying psychiatric and neurologic disorders. We studied cell type diversity of the subgenual anterior cingulate cortex and dorsolateral PFC of humans with no neuropsychiatric illness using a clustering analysis of single-nuclei RNA-sequencing data. Defining the transcriptomic profile of cellular subpopulations in these cortical regions is a first step to demystifying the cellular and molecular pathways involved in psychiatric disorders.

### Introduction

The neurobiology underlying psychiatric and neurologic disorders is complex and poorly understood. Two brain regions that are repeatedly suspected to be dysfunctional in major psychiatric illnesses are the dorsolateral PFC (DLPFC) and the subgenual anterior cingulate cortex (sgACC).

The DLPFC is implicated in higher-order cognitive functions, including attention, working memory, abstract reasoning, and cognitive control (MacDonald et al., 2000; Egner and Hirsch, 2005; Hare et al., 2009). Accordingly, the DLPFC is considered a key region of dysfunction in various psychiatric illnesses, especially schizophrenia (SCZ) (Weinberger et al., 1986; Manoach et al., 1999; Potkin et al., 2009; Koenigs and Grafman, 2009; Price and Drevets, 2012). Consequently, the DLPFC has been the target of postmortem studies of SCZ (Fromer et al., 2016; Gandal et al., 2018b; Jaffe et al., 2018) and major depressive disorder (MDD) (Nagy et al., 2020). The less-studied sgACC is involved in linking emotional and cognitive systems, emotion processing, and autonomic and emotional arousal (Mayberg et al., 1999; Price and Drevets, 2012). It has been implicated in mood disorders, particularly MDD and bipolar disorder (BD), and has been the focus of imaging, postmortem, and treatment studies (Tripp et al., 2011; Scifo et al., 2018; Akula et al., 2021).

Detailed characterization of the cellular composition of these brain regions is essential because mechanisms of psychiatric diseases might act through distinct classes of cells. Single-nucleus RNA-sequencing (snRNA-seq) has recently emerged as a high-throughput method to characterize cell populations based on each individual nucleus's transcriptional profile (Darmanis et al., 2015; Habib et al., 2016, 2017; Krishnaswami et al., 2016). This unsupervised approach is advantageous for its ability to elucidate subtle gene expression differences on the individual cell level, undetectable at the bulk-tissue level. Single-cell transcriptomics has already proved useful in understanding various mental illnesses, such as the dysregulation of corticocortical projection neurons in autism spectrum disorder (ASD) (Velmeshev et al., 2019). While earlier snRNA-seq studies have largely used mouse brains, more recent studies have studied various cortical areas of the human brain, including the PFC (Lake et al., 2016, 2018; Habib et al., 2017). In the DLPFC, recent snRNA-seq studies have demonstrated cell subtype-specific abnormalities in MDD (Nagy et al., 2020) and SCZ (Ruzicka et al., 2020; Reiner et al., 2021). In the sgACC, there is only one study with snRNA-seq in 5 individuals (Tran et al., 2021). Additional studies using more subjects are necessary to create a consensus profile of cell types in these brain regions and elucidate how the molecular profiles of these two brain regions inform their functional roles.

This study seeks to map the molecular and cellular landscape of the DLPFC and sgACC using snRNA-seq in human postmortem brain. We profiled a total of 87,279 nuclei from both brain

regions in 8 individuals without a history of psychiatric or neurologic disorders. Our large dataset allowed us to identify conserved and unique cell populations between the two cortical regions. Using enrichment analyses, we localized psychiatric disorder-relevant transcriptional and heritability signals in the context of our cellular characterization. Overall, this study provides a resource that catalogs the transcriptomic heterogeneity of these key cortical regions at single-cell resolution.

### Materials and Methods

*Contact for reagent and resource sharing.* Further information and requests for resources and reagents should be directed to and will be fulfilled by S.M. (marencos@mail.nih.gov).

*Experimental design and statistical analysis.* snRNA-seq data were generated from the DLPFC and sgACC of 8 male subjects (age =  $44.8 \pm 11.2$  SD; sex = male; race = 5 White and 3 Black; postmortem interval =  $24.3 \pm 7$  h; pH =  $6.5 \pm 0.17$ ), each without a history of psychiatric (including substance use disorders) or neurologic disorders during their lifetime (for detailed demographics, see Extended Data Fig. 1-1). Technical replicates for all 16 unique specimens at the droplet-based nuclei capture step resulted in 32 total samples. Quality control steps (detailed in snRNA-seq data preprocessing) eliminated seven samples with low total cDNA yield resulting in 25 samples with 105,047 nuclei and 28,195 genes. Pearson correlations were performed to compare technical replicates.

Unsupervised clustering algorithms in the Seurat version 4.1.0 (Hao et al., 2021) pipeline were used to assign nuclei into general cell classes and to identify and eliminate doublets, clusters with high mitochondrial contamination, and artifact nuclei (for details, see Unsupervised clustering, quality control, and clustering) yielding a final dataset of 63,901 nuclei. Additional subclustering was performed on Ex and In cells using Seurat version 4.1.0 tools. The default nonparametric Wilcoxon rank sum test with Bonferroni correction was applied to identify cluster markers. Pearson correlation was used to compare our cell classification to published snRNA-seq datasets (for details, see Comparison with published human snRNA-seq datasets).

To define consensus cell class identities, our data were integrated with the Allen Cell Types Database Human Multiple Cortical Areas SMART-seq dataset via reciprocal PCA in Seurat version 4.1.0, and a nearest centroid classifier was applied to map each cell onto the cluster labels in the Allen dataset. Regional subpopulation proportional representation was assessed using the GLIMMIX (generalized linear mixed model) procedure with logit function and the option of  $\beta$  distribution with default covariance structure in SAS version 9.4 to test for effects of subpopulations, brain region, and their interaction (see Regional differences in subclusters). Differential expression between regions and subclusters was tested using a pseudo-bulk approach using the edgeR likelihood ratio test (for details, see Regional differences in subclusters).

Linkage disequilibrium score regression (LDSC) was implemented to examine cell type enrichment for GWAS signals from studies of ASD (Grove et al., 2019), attention-deficit/hyperactivity disorder (ADHD) (Demontis et al., 2019), BD (Stahl et al., 2019), MDD (Wray et al., 2018), and SCZ (Schizophrenia Working Group of the Psychiatric Genomics

Consortium, 2014) (for details, see Cell type-specific genes and psychiatric disorder enrichment).

**Postmortem human specimen collection and dissection.** Postmortem brains were collected at the Human Brain Collection Core (HBCC), National Institute of Mental Health Intramural Research Program, with permission from the legal next-of-kin according to the National Institutes of Health Institutional Review Board and ethical guidelines under protocol 17 M-N073. Cases were obtained from the Offices of the Chief Medical Examiner of the District of Columbia, Northern Virginia, and Central Virginia. Clinical, neuropathological, and toxicological characterization was performed as previously described (Deep-Soboslay et al., 2005; Lipska et al., 2006; Kunii et al., 2015).

Samples were dissected from frozen coronal slabs using a dental drill. For the DLPFC (BA9/46) dissections, gray matter tissue from the mid-point of the middle frontal gyrus was targeted. Subcortical white matter was carefully excluded. For the sgACC (BA24/25), the portion of the cingulum immediately ventral to the genu of the corpus callosum was dissected.

**snRNA-seq nuclei extraction and isolation.** The nuclei isolation protocol was adapted from Krishnaswami et al. (2016) with additional washes to reduce mitochondrial contamination. To maximize our ability to capture the entire cellular diversity in an anatomically defined brain region during dissection, we obtained two separate 50–80 mg aliquots from each frozen dissected block and combined them after nuclei isolation.

Frozen samples were physically dissociated in a homogenizing buffer using a Dounce homogenizer (DWK Life Sciences), five strokes with a narrow pestle followed by 15 strokes with a wide pestle. The resulting nuclear suspensions were filtered through a 35  $\mu$ m mesh cell strainer. Nuclear suspensions were washed 3 times, pelleting the nuclei with centrifugation at 1200  $\times$  g for 9 min before resuspending the pellets in 0.5 ml of wash buffer. Nuclear suspensions from the same frozen block (described above) were combined at this step and thoroughly mixed.

For nuclei counting, a 10  $\mu$ l aliquot of each nuclear suspension was stained 1:1 with 0.4% Trypan Blue. Nuclei were counted using a hemocytometer under a bright-field microscope. Concentrations ranged from 1 million to 1.5 million nuclei per milliliter. A final aliquot of 50,000 nuclei in 1 ml of storage buffer was taken for the 10 $\times$  Genomics nuclei capture.

**10 $\times$  genomics nuclei capture and library construction.** We followed the Chromium Single Cell 3' Reagent Kits User Guide version 2 Chemistry (CG00052) and used Chromium Single Cell 3' Library & Gel Bead Kit version 2 (PN-120237) to produce Single Cell 3' libraries ready for Illumina sequencing (<https://support.10xgenomics.com/>). Droplet-based nuclei-capture and library preparation experiments were conducted using materials and support from the Single Cell Analysis Facility, National Cancer Institute.

First, we created technical duplicates from the nuclei suspensions and loaded them onto the Chromium Controller microfluidic device. We loaded volumes provided in the Cell Suspension Volume Calculator Table of the 10 $\times$  user guide to target a recovery rate of 6000 nuclei per sample. The microfluidic device partitioned single nuclei into bar code-containing Gel Bead-In-Emulsions (GEMs). Samples were incubated, and poly-adenylated mRNAs were reverse-transcribed into cDNAs containing a unique molecular identifier (UMI) and a cell bar code unique to each GEM. Afterward, GEMs were broken, and the barcoded cDNAs were amplified by PCR. Next, a sample index, R2 (read 2 primer sequence), and standard P5 and P7 primers used in Illumina bridge amplification were added during library construction via End Repair, A-tailing, Adaptor Ligation, and PCR. The 16 bp 10 $\times$  bar code and 10 bp UMI were encoded at the start of Read 1 while the 8 bp sample index information was incorporated into the i7 index read, and cDNA insert sequenced as Read 2.

**Sequencing, bar code processing, and alignment.** Sequencing was performed by the Center for Cancer Research Sequencing Facility at the Frederick National Laboratory for Cancer Research. Thirty-two 10 $\times$  Genomics Chromium Single Cell 3' version 2 paired-end libraries were multiplexed and sequenced on one run on the HiSeq 4000. Base-calling was performed with Illumina instrument runtime analysis software, RTA version 2.4.11. The output raw base call (BCL) files were

demultiplexed using Bcl2fastq version 2.17, allowing one mismatch in barcodes, to generate FASTQ files. Sequence alignment and bar code processing were performed with 10 $\times$  Genomics's processing software, Cell Ranger version 2.1.1 (Zheng et al., 2017). A customized human reference genome, GRCh38-1.2.0\_premRNA (described below), was used for counting reads and the "expect-cells" parameter for estimating number of nuclei in each sample above a background expression level was set to 6000, consistent with the number of nuclei targeted during library preparation.

Because snRNA-seq captures precursor mRNA (pre-mRNA) transcripts that have not been alternatively spliced, these transcripts have retained intronic features. Recent RNA-seq and comparative studies of single nucleus and single-cell RNA-seq datasets have shown that including reads mapping to intronic regions as gene counts accurately estimates true gene expression (Gaidatzis et al., 2015; Lake et al., 2017; Bakken et al., 2018). Additionally, this method has an added benefit for downstream clustering analysis by increasing read depth for inherently low coverage snRNA-seq datasets. Therefore, a customized, pre-mRNA reference genome was created following recommendations from 10 $\times$  Genomics. "Gene\_biotype" tags in the original GTF file were modified such that Cell Ranger's count function included intronic features as gene counts.

**snRNA-seq data preprocessing.** Data processing, analysis, and handling were performed using the R package Seurat version 4.1.0 (Hao et al., 2021). Filtered gene count matrices were read into R using Seurat's Read10X function. After initial review of experimental and sequencing quality control (QC) characteristics, 7 of 32 total samples were identified as "dropout" samples and removed from the dataset. These samples displayed low total cDNA yield calculated from BioAnalyzer electropherograms. They contained high sequence duplication rate (>45%) calculated with the sequencing QC tool FastQC (Andrews, 2010) and high sequencing saturation rate (>85%) from Cellranger QC outputs. Together, these QC results suggested inadequate RNA capture and subsequent PCR overamplification of these "dropout" samples. Twenty-five samples (14 DLPFC, 11 sgACC) remained for downstream analysis and were assessed for further quality control. Detailed 10 $\times$  Genomics Library capture, sequencing, and alignment data for each sample of brain tissue is listed in Extended Data Fig. 1-2.

Filtering of noisy genes and outlier nuclei was performed; 5475 genes detected in fewer than three nuclei were filtered from the data. Nuclei with <200 genes do not contain enough transcriptional information for effective clustering and subsequent biological interpretation. This threshold is like those used in published snRNA-seq studies (Hu et al., 2017). No nuclei fell below this threshold in our data. Nuclei with large library sizes (>5000 genes detected and 15,000 UMIs) are likely to be "doublets" or "multiplets," which are expected technical artifacts of the microfluidic droplet technology (Habib et al., 2017; Hu et al., 2017; Lake et al., 2018). We excluded 227 nuclei whose values exceeded these thresholds (0.2% of total nuclei). Next, the fraction of reads mapping to mitochondrially encoded genes was calculated for each barcoded nucleus. Mitochondrially encoded mRNA transcripts represent false transcriptional signals and thus should not be present in snRNA-seq data. Nuclei with >20% of reads mapping to mitochondrially encoded genes were filtered out (25,273 nuclei). Subsequently, mitochondrially encoded genes were removed from the count matrix. After these preprocessing QC steps, 105,047 nuclei and 28,195 genes were retained for analysis.

**Correlation between technical replicates of samples.** To assess similarity between technical replicates, we calculated the Pearson correlation coefficient between average gene expression of replicates using the *cor* function in R (Extended Data Fig. 1-3). Correlations of average expression across all genes and then across only highly variable genes were assessed. Correlation of gene expression (Pearson  $r > 0.998$ ) and correlation of number of nuclei obtained (Pearson  $r = 0.992$ ) across the technical replicates was high (Extended Data Fig. 1-3). Given the high correlation coefficient values between technical replicates and the added clustering power that is achieved with larger sample size (number of nuclei), technical replicates were combined for subsequent downstream analyses. Technical replicates could be treated as additional samples of

the same subject because they were separate aliquots from nuclei suspensions, representing different nuclei from the same subject.

**Unsupervised clustering, quality control, and clustering.** Normalization of gene expression to account for library size and expression level was performed with the Seurat function *NormalizeData* using default settings. Next, highly variable genes (HVGs) in our dataset were identified using the Seurat function *FindVariableGenes* for unsupervised clustering. Briefly, genes were first binned into 20 bins based on their mean expressions. *Z* scores were calculated within each bin. This method of binning controlled for the correlation between gene expression and variability. After binning, genes with  $SD > 0.65$  and mean log normalized expression between 0.015 and 4 were selected as HVGs. Next, the Seurat function *ScaleData* was used to compute *z* scores for each gene in the list of HVGs. We calculated *z* scores by fitting gene expression to a negative binomial model because it more accurately resembles gene expression distributions. In this regression, the total number of UMIs detected and fraction of mitochondrial genome encoded gene counts per nucleus were regressed out to further eliminate library size effects on the downstream clustering analysis. Next, for linear dimensional reduction, a principal component analysis (PCA) was performed on the *z* scores of the HVGs using the Seurat function *RunPCA*. The top 50 principal components (PCs) were selected after assessment of PCs with an elbow plot and were used as input to Seurat's unsupervised clustering algorithm function *FindClusters*. In this algorithm, Euclidean distances between nuclei were calculated in the PCA space (top 50 PCs) and a K-nearest neighbor (KNN) graph-based approach was used to group nuclei into clusters. The resolution parameter, which represents the granularity of the clustering analysis, was set at 3, and all other parameters were set at default settings. For two-dimensional *t*-distributed stochastic neighbor embedding (tSNE) visualization, the *RunTSNE* function from Seurat was performed at default settings on the same principal components that were used for clustering.

After clustering, a variety of methods were used to identify and disqualify clusters. The distribution of number of genes, number of UMIs, fraction of mitochondrial gene counts, and fraction of ribosomal protein gene counts were visually assessed for each unbiased cluster using violin plots. To assess similarity of these clusters to each other, hierarchical clustering analysis on the average expression for each cluster was performed using R function *hclust*, a pairwise correlation heatmap between all clusters was generated with R function *cor* to view similar mean expression profiles, and distances of tSNE coordinates were visually assessed. Doublet or mixed clusters, identified by high numbers of genes and UMIs detected in addition to the combined expression of multiple canonical cell class marker genes, and clusters with very low numbers of detected genes, low UMI counts, and lack of distinguishing expression profiles, were discarded. The remaining subset of nuclei were reclustered using new HVGs calculated from the subset. Cluster removal and reclustering were repeated until all clusters that formed could be identified by their specific expression of canonical cell-type marker genes from existing literature and could not be removed by quality control as described above. A total of 87,729 of 105,047 nuclei remained after cluster curation.

Finally, to remove artifact nuclei, which include collections of ambient RNA encapsulated during the droplet experiment or doublets/multiplets, we curated each cluster by calculating cell-type scores for each nucleus (barnyard curation). Cell-type scores were calculated as a nucleus's average expression of the top five specific genes for each broad cell class, listed in Extended Data Figure 1-4. Top five specific genes were determined as the five highest-ranked genes when significant differentially expressed genes (DEGs) for each broad cell cluster were ranked by specificity (percent of in-group nuclei expressing the gene/percent of out-group nuclei expressing the gene). An additional criterion for top five specific genes for broad cell classes was that the genes were required to overlap with marker genes from at least one other single-cell or nucleus RNA-seq dataset (Darmanis et al., 2015; Lake et al., 2016; Habib et al., 2017; Hu et al., 2017; Sathyamurthy et al., 2018). By assessing each nucleus's score for each of the broad cell cluster, nuclei with "promiscuous" cellular identity (i.e., having scores greater in cell classes other than the cell class to which it belongs) were

filtered out from further analysis. A total of 63,901 of 87,279 nuclei remained after barnyard curation.

Following QC, 63,901 nuclear expression profiles remained from 25 samples, averaging 93 million reads per sample and 27,061 reads per nucleus, slightly above the range (19,000–26,000 reads per nucleus) typically required for sequencing saturation with 10× Genomics version 2 chemistry in human brain snRNA-seq (Habib et al., 2017). Using 10× Genomics' software Cell Ranger for read alignment and counting, 1641 median UMIs and 1100 median genes per nucleus were detected, which is consistent with other snRNA-seq studies in postmortem human brain samples (Habib et al., 2017; Lake et al., 2018). A modified human reference genome was used to incorporate intronic reads as gene counts (Gaidatzis et al., 2015; Lake et al., 2017; Bakken et al., 2018). Ninety percent of reads mapped confidently to the customized genome.

**Determination of cluster marker genes.** The *FindMarkers* and *FindAllMarkers* functions in Seurat were applied with default settings to identify marker genes for each cluster (i.e., genes enriched in a cluster compared with all other clusters). The default test used was a nonparametric Wilcoxon rank sum test. Only genes whose expression value were  $> 10\%$  of the nuclei in the in-group, whose Bonferroni corrected *p* values  $< 0.05$ , and whose log fold-change  $> 0.25$  were saved as marker genes. This approach was used to identify marker genes for broad cell clusters (Extended Data Fig. 1-5), and excitatory and inhibitory neuron subclusters (Extended Data Fig. 2-1).

**Subclustering inhibitory and excitatory neurons.** Transcriptional signals that differentiate neurons from glial cells may not be the same signals that differentiate subtypes of neurons, excitatory and inhibitory, from each other (Tasic et al., 2018). The signals that differentiate excitatory from inhibitory neurons, in turn, may also not be the same signals that differentiate canonical inhibitory neuron subtypes from each other. To effectively resolve the transcriptional diversity of inhibitory neuron subtypes without the influence of noninhibitory neurons, we performed subclustering on the restricted set of nuclei identified as inhibitory neurons. We implemented the same data-driven QC approaches described above for the broad cell class clustering analysis with an added step of merging nonunique clusters. From an initial 24 inhibitory neuron subclusters, we reassigned nuclei of one subcluster that did not have a uniquely identifiable expression signal ( $< 10$  unique DE genes) and one subcluster that mostly consisted of nuclei from one individual. This process is called "merging" and is performed by reassigning the identity of each nucleus individually to the subcluster with the highest average shared nearest neighbors. This process reassigned nuclei to subclusters that had very similar transcriptional profiles to its original cluster. Similarity of subclusters could be assessed by correlation heatmaps and relative position to one another in a hierarchical clustering tree. This resulted in 22 final inhibitory neuron subclusters. DE analysis was performed as described above for each subcluster against all other inhibitory neurons to identify potential marker genes. Subclustering was also performed for the excitatory neurons. From the initial 29 subclusters that formed, nine subclusters were identified as nonunique cell types and nuclei of these subclusters were reassigned to subclusters with highest average shared nearest neighbors. Merging these subclusters resulted in 20 final excitatory neuron subclusters represented by all 8 individuals. Statistics for the final list of subclusters are shown in Extended Data Figure 2-1.

**Comparison with published human snRNA-Seq datasets.** For comparison with published human brain snRNA-Seq datasets, we read in count matrices (Lake et al., 2016; Habib et al., 2017; Lake et al., 2017; Tasic et al., 2018; D. Wang et al., 2018) into R as Seurat objects. Average gene expression was computed for each cluster from each dataset. A Pearson correlation coefficient was calculated using R function, *cor*, on mean expression for all pairwise comparisons of clusters. We limited our genes for correlation comparisons to 3310 HVGs (disregarding 78 ribosomal protein genes) to reduce noise from gene expression signals that are less informative among brain cells and to avoid dataset-specific technical artifacts. All heatmaps were plotted with either the ggplot2 R package or the pheatmap R package with colors palettes generated from the RColorBrewer R package. Correlation heat maps are shown in Extended Data Figure 2-4.

**Dataset integration in Seurat and nearest centroid classification.** snRNA-seq count data from multiple cortical areas were downloaded from the Allen Cell Types Database (<https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq>) and imported into Seurat (version 4) along with HBCC data. Nuclei with a missing cluster label were removed. Inhibitory neurons, excitatory neurons, and non-neuronal cells were analyzed separately. For each major subclass, cells were log-normalized and the top 2000 most variable genes across both datasets were selected using the *SelectIntegrationFeatures* function in Seurat. PCA was run separately on both HBCC and Allen Institute data for the selected genes. Integration anchors were obtained via the *FindIntegrationAnchors* function using reciprocal PCA on the top 30 principal components and a k.anchor default value of 5. Datasets were integrated via the *IntegrateData* function with a k.weight default value of 100.

For classification of HBCC data based on Allen Institute labels, a nearest centroid classifier was implemented via the lolR library and was trained on the Allen cluster labels with the integrated gene expression values from 2000 genes as described above. This classifier was then applied to predict the Allen cluster label of each cell in the HBCC data. A contingency figure of the original HBCC cluster ID and predicted Allen label was created, and the fraction of HBCC cells mapping to each Allen label was visualized in a heatmap (see Fig. 2E,F).

**Regional differences in subclusters.** To establish whether particular neuronal subpopulations were unequally represented in the DLPFC or the sgACC, we calculated the proportion of neuronal subpopulations using all neurons as the denominator, then used the GLIMMIX (generalized linear mixed model) procedure with logit function and the option of  $\beta$  distribution with default covariance structure in SAS version 9.4 to test for effects of subpopulations, brain region, and their interaction. Subject (id) was used as random effect, assuming subjects participating in the study are a random sample of all possible subjects. This procedure was repeated on absolute counts of each neuronal subpopulation (omitting the  $\beta$  distribution).

Differential expression analysis between subclusters and brain regions was conducted using a conservative pseudo-bulk approach via edgeR likelihood ratio test (Robinson et al., 2010), while adjusting for donor status using the Libra R package (Squair et al., 2021). DEGs were defined as those that passed false discovery rate (FDR)-corrected  $p < 0.05$  and had an absolute  $\log_{2}FC > 0.1$ . To explore the relevance of DEGs, these genes were used as input into gene enrichment analysis with the Gene Ontology (GO) (Gene Ontology Consortium, 2001), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), and DisGeNET (Piñero et al., 2015) as implemented in the ClusterProfiler (Yu et al., 2012) and disgenet2r R packages (Piñero et al., 2020).

**Cell type-specific genes and psychiatric disorder enrichment.** To test the association between cell type-specific genes and the heritability for psychiatric disorders, we used LDSC, as implemented previously (Finucane et al., 2015; Skene et al., 2018; Bryois et al., 2020). In brief, single-nucleus RNA expression data were scaled to a total of 1 million UMIs for each cell type, separately for each cortical region. A specificity metric was calculated as a ratio of the expression of each gene in a given cell type cluster over the total expression of that gene across all cell types (Skene et al., 2018; Bryois et al., 2020). The top 10% most specific genes for each cell type were used as input for LD score regression. Genome Wide Association Study (GWAS) data for the following psychiatric disorders were used: ASD (Grove et al., 2019), ADHD (Demontis et al., 2019), BD (Stahl et al., 2019), MDD (Wray et al., 2018), and SCZ (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), accessed through the Psychiatric Genomics Consortium data sharing portal (<https://www.med.unc.edu/pgc/>). Gene coordinates were extended by 100 kb upstream and 100 kb downstream for each as done previously (Finucane et al., 2015; Bryois et al., 2020).  $p$  values were FDR-corrected for 7 broad cell types, and 47 subclusters (Benjamini and Hochberg, 1995).

The same top 10% of cell specific genes were used to test for enrichment with DEGs from a recently published large RNA seq study of ASD, SCZ, and BD samples (Gandal et al., 2018b). Hypergeometric tests were

used to assess the enrichment or overlap between gene sets, and the  $p$  values were corrected as described above.

**RNA ISH (RNAscope).** For RNAscope validation of cell types, 10- $\mu$ m-thin cryosections were taken from the DLPFC and sgACC dissected from postmortem brain samples (Extended Data Fig. 2-7). The sections underwent post-fixation in 10% NBF at 4°C for 1 h. Tissue was dehydrated and stained following the Advanced Cell Diagnostics protocol for RNAscope 2.5 Duplex Chromogenic Assay (Anderson et al., 2016) with the following modifications: Protease IV treatment reaction time was reduced to 15 min; amplification reaction time was optimized to obtain the best resolution of puncta for the different target probes labeled with HRP-based green; counterstaining was done with 25% Hematoxylin 560 MX (Leica 042817) diluted in dH<sub>2</sub>O, for 10 s; alternatively, Nissl staining was done with 1% thionine for 30 s, subsequent wash in dH<sub>2</sub>O for 30 s twice, and differentiated in acidified ethanol for 5-15 s. Whole slide image acquisition was performed on a NanoZoomer S60 Digital Slide Scanner (Hamamatsu).

**Code/software.** Raw FASTQ files, final Seurat R Object containing nuclei expression data, and metadata are publicly available for download on the National Institute of Mental Health Data Archive (<https://nda.nih.gov/experimentView.html?experimentId=1852&collectionId=3151>) and Synapse (<https://doi.org/10.7303/syn48374287.1>). R code for this project is available for download on our GitHub repository: <https://github.com/NIMH-HBCC>. A shareable interactive user interface using iSEE (Rue-Albrecht et al., 2018) and Binder 2.0 (<https://elifesciences.org/labs/8653a61d/introducing-binder-2-0-share-your-interactive-research-environment>) is also accessible from the GitHub HBCC repository. Datasets and gene sets from referenced literature were obtained online from supplementary data.

## Results

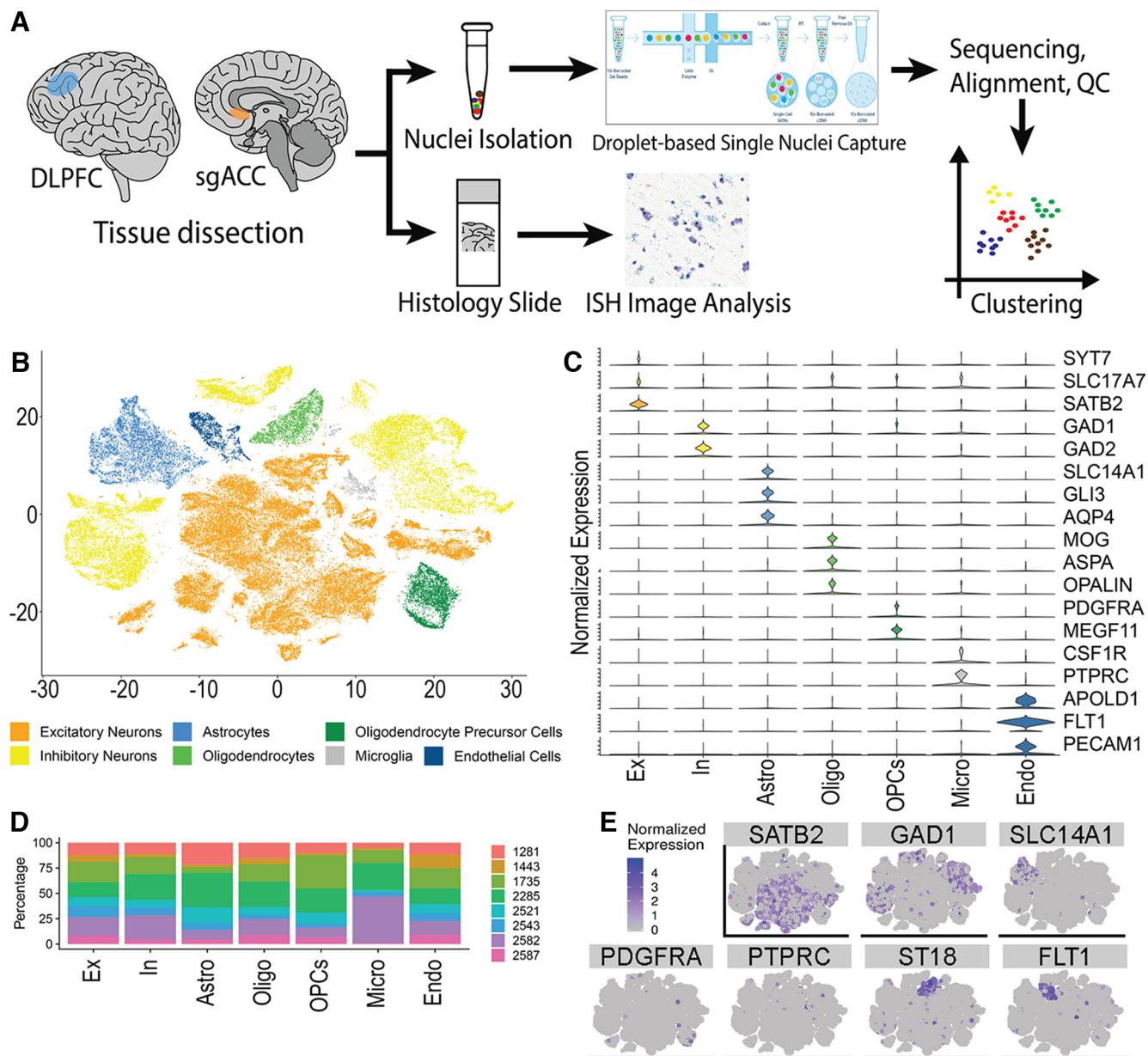
### Classification of nuclei into broad cell types

Unsupervised clustering of the nuclei that survived quality control and preprocessing steps (see Materials and Methods; Extended Data Figs. 1-1, 1-2, and 1-3) generated 47 initial clusters. Initial clusters of the same broad cell type were grouped together, which ultimately resulted in seven distinct broad clusters: excitatory neurons (Ex), inhibitory neurons (In), astrocytes (Astro), oligodendrocytes (Oligo), oligodendrocyte precursor cells (OPCs), endothelial cells (Endo), and microglia (Micro) (Fig. 1; Extended Data Figs. 1-4 and 1-5) (Habib et al., 2017; Lake et al., 2018; D. Wang et al., 2018). Neuronal nuclei contained a higher median number of genes and UMIs (1225 and 1875, respectively) than non-neuronal nuclei (674 and 915, respectively), as expected of the physically larger nuclear sizes and greater functional complexity of neurons.

### Transcriptional diversity in excitatory neurons

To investigate the diversity of neuronal subtypes, excitatory neuronal nuclei were extracted from the larger dataset and iteratively reclustered to yield 20 robust, transcriptionally distinct Ex subclusters (Fig. 2A,C). Each subcluster was given a simple assignment (e.g., Ex18), in which the number corresponds to the relative size of each subcluster in descending order (e.g., Ex01 is the largest group of excitatory cells). Differential gene expression was determined for each Ex subcluster versus all other Ex subclusters to identify enriched cell type-specific markers for each subcluster (Extended Data Fig. 2-1). Each subcluster contained at least one nucleus from all 8 subjects, thereby avoiding individual-biased clusters (Extended Data Fig. 2-1).

Ex subclusters were categorized into upper (L2-3), middle (L4-5), and lower (L6-6b) cortical layer groups based on the expression of known layer-specific marker genes: *CARTPT* (L2-3), *CUX2* (L2-4), *RORB* (L3-5), *FOXP2* (L6), *PCP4* (L5), *GRIN3A* (L5), *GRIK3* (L5-6), *TLE4* (L5-6), and *CTGF* (L6b)

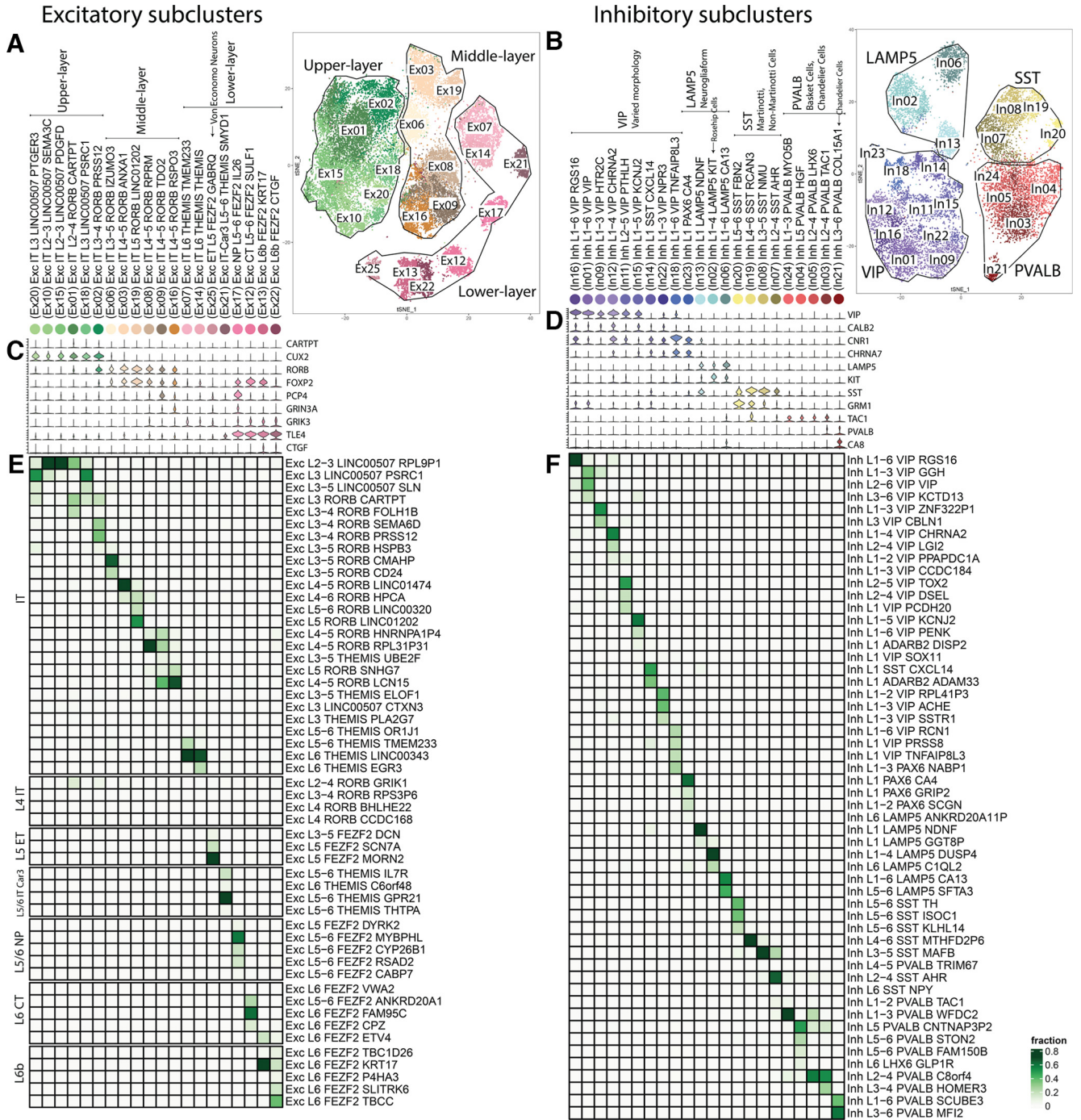


**Figure 1.** Molecularly defined broad cell types in the human DLPFC and sgACC from 8 male subjects (for donor details, see Extended Data Fig. 1-1). **A**, Workflow diagram of 10× Genomics snRNA-seq library capture, sequencing, and alignment (for sample metrics, see Extended Data Fig. 1-2), and RNAscope validation. Correlations between technical replicates can be found in Extended Data Figure 1-3. **B**, tSNE plot of broad cell clusters. **C**, Expression violin plots of canonical marker genes enriched in each broad cell type (see also Extended Data Figs. 1-4 and 1-5). **D**, Proportion of nuclei count from each of 8 donors in each broad cell cluster. **E**, Normalized expression of the most representative canonical marker gene for each broad cell type visualized in relation to the tSNE plot.

(Fig. 2) (Zeng et al., 2012; Hodge et al., 2019). Subcluster layer assignment was verified with comparison to the Allen Brain Institute's ISH atlas (Extended Data Fig. 2-3) (<https://human.brain-map.org/ish/search>). We further validated our subcluster classification by demonstrating strong overall correlation with excitatory neuronal clusters defined in other studies across various cortical regions in both mouse and human (Extended Data Fig. 2-5) (Habib et al., 2017; Lake et al., 2018; Tasic et al., 2018; D. Wang et al., 2018).

We next integrated the neuronal clusters identified here with the Allen Brain Institute's Human Multiple Cortical Areas SMART-seq dataset (Fig. 2E; Extended Data Fig. 2-4), which includes data from middle temporal gyrus, ACC, primary visual cortex, primary motor cortex, primary somatosensory cortex, and primary auditory cortex (<https://portal.brain-map.org/atlas-and-data/rnaseq>) and defines neuronal populations that are conserved across species with anatomically confirmed layer assignments and inferred connectivity of intratelencephalic (IT: all upper- and middle-layer subclasses), extratelencephalic (ET: Ex25), corticothalamic (CT: Ex12), layer 6b (Ex13 and Ex22), and near-projecting (NP: Ex17) cells (Hodge et al., 2019; Bakken et al., 2021). Notably, L4 IT and L3-5 IT neuron subtypes characteristic of primary visual, sensory, and motor cortices were not found. Other exceedingly rare L5, L5/6, and L6 classes were also not found in our dataset. Using this classification method, most HBCC subclusters mapped to several Allen clusters with one cluster being favored (Fig. 2E; Extended Data Fig. 2-4). Rarely, multiple HBCC subclusters would map to the same Allen cluster. A clear one-to-one correspondence for the lower-layer neuronal subpopulations

org/atlas-and-data/rnaseq) and defines neuronal populations that are conserved across species with anatomically confirmed layer assignments and inferred connectivity of intratelencephalic (IT: all upper- and middle-layer subclasses), extratelencephalic (ET: Ex25), corticothalamic (CT: Ex12), layer 6b (Ex13 and Ex22), and near-projecting (NP: Ex17) cells (Hodge et al., 2019; Bakken et al., 2021). Notably, L4 IT and L3-5 IT neuron subtypes characteristic of primary visual, sensory, and motor cortices were not found. Other exceedingly rare L5, L5/6, and L6 classes were also not found in our dataset. Using this classification method, most HBCC subclusters mapped to several Allen clusters with one cluster being favored (Fig. 2E; Extended Data Fig. 2-4). Rarely, multiple HBCC subclusters would map to the same Allen cluster. A clear one-to-one correspondence for the lower-layer neuronal subpopulations



**Figure 2.** Subclusters of excitatory (Ex) and inhibitory (In) neurons (for subcluster details, see Extended Data Fig. 2-1). **A, B**, Nomenclature and tSNE plot of Ex and In subclusters, respectively. Nomenclature format: Excitatory or versus inhibitory, projection type (if Ex), cortical layer, marker gene, unique differentially expressed marker gene; HBCC ID is written in parentheses (see Extended Data Fig. 2-2). tSNE plot shows subclassification of Ex subclusters to upper, middle, or lower cortical layers and subclassification of In subclusters to LAMP5, SST, VIP, and PVALB (see Extended Data Fig. 2-3). **C, D**, Expression violin plots for cortical layer marker genes for Ex subclusters and for ganglionic eminence marker genes for In subclusters, respectively (see Extended Data Fig. 2-3). **E, F**, Nearest centroid classification of Allen Cell Types Database and HBCC Ex and In subclusters, respectively. Ex Allen clusters are divided into Allen projection subclasses. See Extended Data Figure 2-3 for tSNE analysis comparing HBCC and Allen Brain Atlas subclusters and RNAscope validation of select markers and Extended Data Figure 2-4 for comparisons of subclusters by brain region. See Extended Data Figure 2-5 for correlation of HBCC subclusters with additional human snRNA-seq datasets. See Extended Data Figure 2-6 for correlation values of subclusters to each other and Extended Data Figure 2-7 for donor details of cases used for RNAscope studies.

was observed. Middle- and upper-layer subclusters were less well-matched to corresponding clusters. Using this data integration, we labeled our clusters following common cell type nomenclature guidelines (Miller et al., 2020). The labels included layer specificity, inferred connectivity, and major classifying markers, keeping the Allen nomenclature where

their classifying markers coincided with ours (Fig. 2E; Extended Data Figs. 2-2 and 2-4).

The smallest subcluster, Ex25 (Exc L5 ET *FEZF2 GABRQ*), integrated with multiple Allen L5 ET classes, but primarily with Exc L5 *FEZF2 MORN2*, a rare cell type found primarily in the ACC. Intriguingly, Ex25 displayed a gene expression signature

corresponding to a rare neuron type, the von Economo neuron (VEN), with enriched expression of VEN markers, such as *ADRA1A*, *GABRQ*, *VAT1L*, and *ADAMTS2* (Extended Data Fig. 2-5). VENS are morphologically defined by a unique spindle shape and are implicated in the pathogenesis of frontotemporal lobar dementia and SCZ (Seeley et al., 2006; L. Yang et al., 2019). The identification of this rare, well-characterized cell type reinforces the validity of our subcluster analysis and cell type classifications.

### Transcriptional diversity in inhibitory neurons

In parallel to our analysis of Ex subclusters, we subset the inhibitory neuron nuclei from the larger dataset and performed reclustering to obtain 22 transcriptionally distinct In subclusters (Fig. 2B,D). Except for In22 and In24, both of which consisted of at least one nucleus from 7 subjects, each subcluster contained at least one nucleus from all 8 subjects (Extended Data Fig. 2-2).

Differential gene expression was performed to identify specific markers for each subcluster (Extended Data Fig. 2-1). Our first observation was that In subclusters segregated along developmental lineages derived from caudal and medial ganglionic eminences (CGE and MGE) (Extended Data Fig. 2-3). CGE-derived interneurons accounted for 54% of all interneurons and expressed the transcription factors *NFIB* and *PROX1*. MGE-derived subclusters expressed *SOX6* and *LHX6*. Using RNAscope, we confirmed coexpression of the MGE marker *LHX6* in SST-expressing cells across all cortical layers and in PVALB-expressing cells in layers IV and above (Extended Data Fig. 2-3). The CGE marker *HTR2C* was coexpressed with *VIP* in all cortical layers (Extended Data Fig. 2-3). Intriguingly, the LAMP5 subcluster In06 expressed both CGE and MGE transcription factors (*NFIB*, *LHX6*, and *SOX6*).

We next observed that In subclusters mapped to the four major classes of inhibitory neurons: SST, PVALB, VIP, and LAMP5. In subclusters also matched robustly to corresponding Allen Institute clusters (Bakken et al., 2021). Integration of our data with the Allen Brain Institute's Human Multiple Cortical Areas SMART-seq dataset revealed distinct mapping of all In subclusters to few related Allen clusters with no overlap between In subclusters (Fig. 2F), emphasizing the robustness of our cluster classification. Few Allen clusters were not identified among our In subclusters: Inh L6 SST *NPY*, L1-2 PVALB *TAC1*, Inh L6 LAMP5 *ANKRD20A11P*, Inh L1 VIP *SOX11*, Inh L1 *ADARB2* *DISP2*, and Inh L1-3 VIP *CCDC184* classes, which are especially rare subtypes, and L4-5 PVALB *TRIM67*, which is primarily found in sensory and visual cortices. Similar to Ex subclusters, the In subclusters were assigned labels following common cell type nomenclature guidelines (Miller et al., 2020), which include broad neuronal class, canonical marker gene, and cell type-specific marker genes.

### Numerical differences in neuronal subclusters between DLPFC and sgACC

While there was no difference between the brain regions in the overall counts of excitatory cells (Extended Data Fig. 2-1), significant differences were found ( $p < 0.0001$ ) at the subcluster resolution (Fig. 3). Some Ex subclusters appeared almost exclusively in the DLPFC; these included three supragranular subclusters (Ex15 L2-3 *LINC00507* *PDGFD*, Ex01 L2-4 *RORB* *CARTPT*, and Ex02 L3-4 *RORB* *PRSS12*) and one middle-layer subcluster (Ex03 L4-5 *RORB* *ANXA1*) (Fig. 3B,C). Conversely, one supragranular subcluster (Ex20 L3 *LINC00507* *PTGER3*) and one middle-layer subcluster (Ex16 L4-5 *RORB* *RSPO3*)

exhibited enrichment in the sgACC (Fig. 3B,C). Notably, most deep-layer excitatory subclusters, consisting of pyramidal projection neurons, were present in both brain regions with similar frequencies.

Inhibitory neurons were more numerous in the sgACC ( $p < 0.0001$ ) than in the DLPFC (Fig. 3B,C); however, In subclusters exhibited less regional variability than Ex subclusters. Only In04 L5 PVALB *HGF* basket cells, In06 L1-6 LAMP5 *CA13* and In01 L1-6 VIP *VIP* subclusters were significantly enriched in the sgACC. Overall, we found that heterogeneity of supragranular Ex cells is the primary distinguishing feature of these cortical regions.

### Transcriptional differences between DLPFC and sgACC for neuronal subpopulations

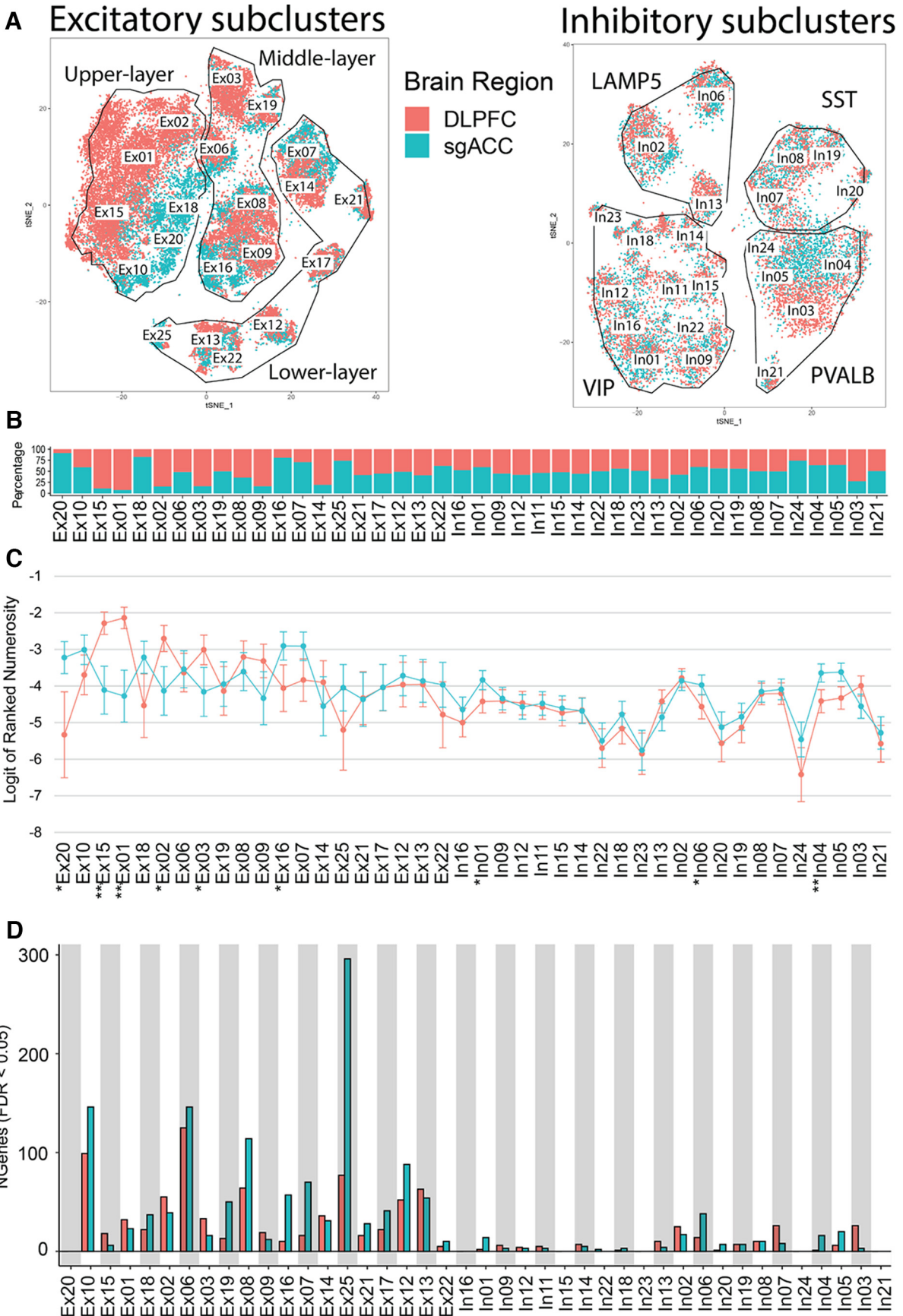
We next asked whether transcriptional diversity within subclusters further contributed to differences between these cortical areas by examining differential gene expression between DLPFC- and sgACC-derived nuclei within each subcluster (Fig. 3D). Very few DEGs were found for In subclusters with a median of 14 DEGs of 1083 median total genes detected (1.3%) (Fig. 3D). Ex subclusters had a median of 67 DEGs per cluster of 1640 median total genes detected (4.1%). In particular, six Ex subclusters had extremely high interregional variation in gene expression Ex06 (20.0%), Ex25 (17.4%), Ex10 (16.4%), Ex08 (12.6%), Ex12 (9.6%), Ex13 (6.3%), and Ex02 (5.2%) (Fig. 3D). GO analyses on these subclusters were performed to identify molecular processes that distinguish each brain region (Extended Data Fig. 3-2). The predominant differences included enrichment for synapse organization, cell-cell adhesion, trans-synaptic signaling, and other GO terms associated with functions of neurotransmission in the sgACC versus DLPFC (Extended Data Fig. 3-2).

Ex25 (Exc L5 ET *FEZF2* *GABRQ*) had the highest absolute number of DEGs of the neuronal subclusters. This was the only extratelencephalic projecting subcluster identified in our dataset. Upon closer inspection, we found that *GABRQ*, a marker gene for VENS (L. Yang et al., 2019; Hodge et al., 2020) was only expressed in sgACC-derived nuclei of the Ex25 subcluster (Fig. 4A), whereas other markers for VENS (*ADRA1A*, *VAT1L*, and *ADAMTS2*) were present in both brain regions. Despite the large number of interregional DEGs, this subcluster could not be further subdivided using our unbiased classification methods as clustering using a higher resolution would have yielded clusters without representation from all individuals. We therefore investigated whether this population of neurons simply had a lower level of *GABRQ* expression in the DLPFC not detected by our single-nucleus methodology or whether they were indeed distinct populations. Using RNAscope, we examined the coexpression of *GABRQ* and *ADRA1A* in the DLPFC and sgACC and found that (1) *GABRQ* expression was indeed restricted to the sgACC and was not present in *ADRA1A*-expressing neurons in the DLPFC and (2) that these cells exhibited distinct morphologies in the two brain regions. The *ADRA1A*/*GABRQ* double-positive neurons in layer V of the sgACC exhibited classic spindle-shaped morphology that typifies VENS, whereas the *ADRA1A*-positive neurons in the DLPFC exhibited the more generalized pyramidal morphology (Fig. 4B; Extended Data Fig. 4-1).

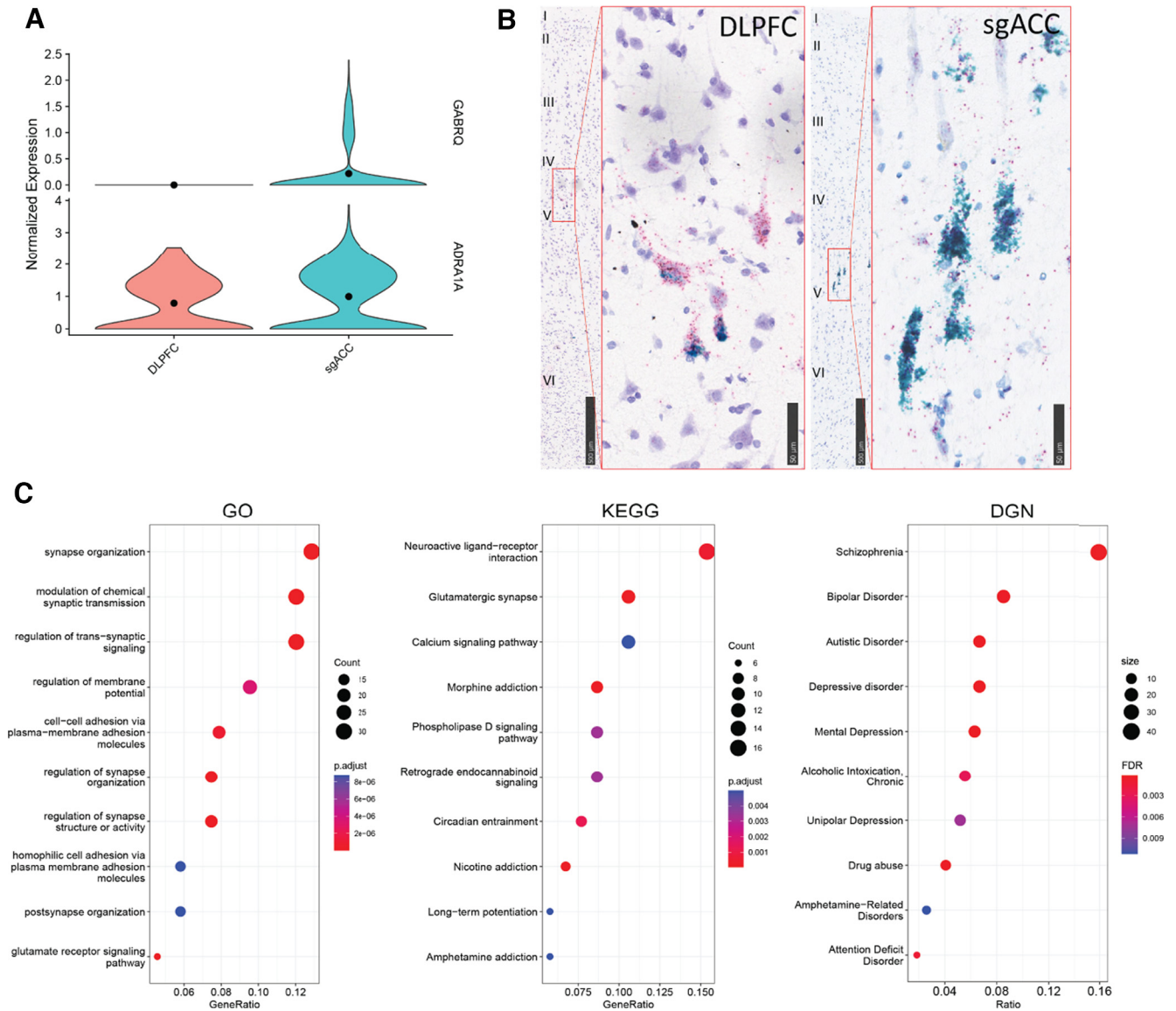
### Cellular enrichment of psychiatric disorder transcriptomic and GWAS signals

Last, we examined the enrichment of bulk-tissue RNA-seq studies of psychiatric disorders, SCZ, BD, and ASD (Fromer et al.,





**Figure 3.** Neuronal diversity between the DLPFC (salmon) and sgACC (cyan). **A**, tSNE plots showing excitatory (Ex) and inhibitory (In) neuronal subclusters colored by brain region. **B**, Percentage of cells in each brain region for each subcluster. **C**, Proportion of cells for each brain region for each subcluster, shown as logit values of ranked numerosity. Ex20, Ex15, Ex01, Ex02, Ex03, Ex16, In04, and In05 have nonoverlapping Cls. Wilcoxon rank sum test showed that Ex15, Ex01, and In04 have significant differences between DLPFC and sgACC after adjusted *p* values:



**Figure 4.** Putative VEN subcluster Ex25. **A**, Ex25 gene expression of VEN markers in the DLPFC and the sgACC. **B**, RNAscope histology images comparing VENs in human DLPFC and sgACC sections. ADRA1A and GABRQ transcripts were detected with chromogenic RNAscope probes, red and green, respectively. ADRA1A<sup>+</sup>/GABRQ<sup>+</sup> neurons were found in layer V in both the sgACC and DLPFC. Note the different morphology of the cell bodies in DLPFC and the apparent reduction in expression of GABRQ. For additional images, see Extended Data Figure 4-1. **C**, Enrichment analysis of genes with greater expression in the sgACC within Ex25 neuronal subpopulation, using GO (left), KEGG (middle), and DisGeNET database of gene-disease associations (right). Point size represents the number of genes that are differentially expressed and within a particular enrichment category (e.g., a GO term).

2016; Gandal et al., 2018a; Jaffe et al., 2018), in our transcriptomic profiles of neuronal subpopulations. To this end, we performed hypergeometric tests on data from cell type-specific enrichment of DEGs obtained from the largest postmortem RNA-expression meta-analysis to date (Gandal et al., 2018b). In parallel, we used LDSC on GWAS signals derived from large case-control studies (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Wray et al., 2018;

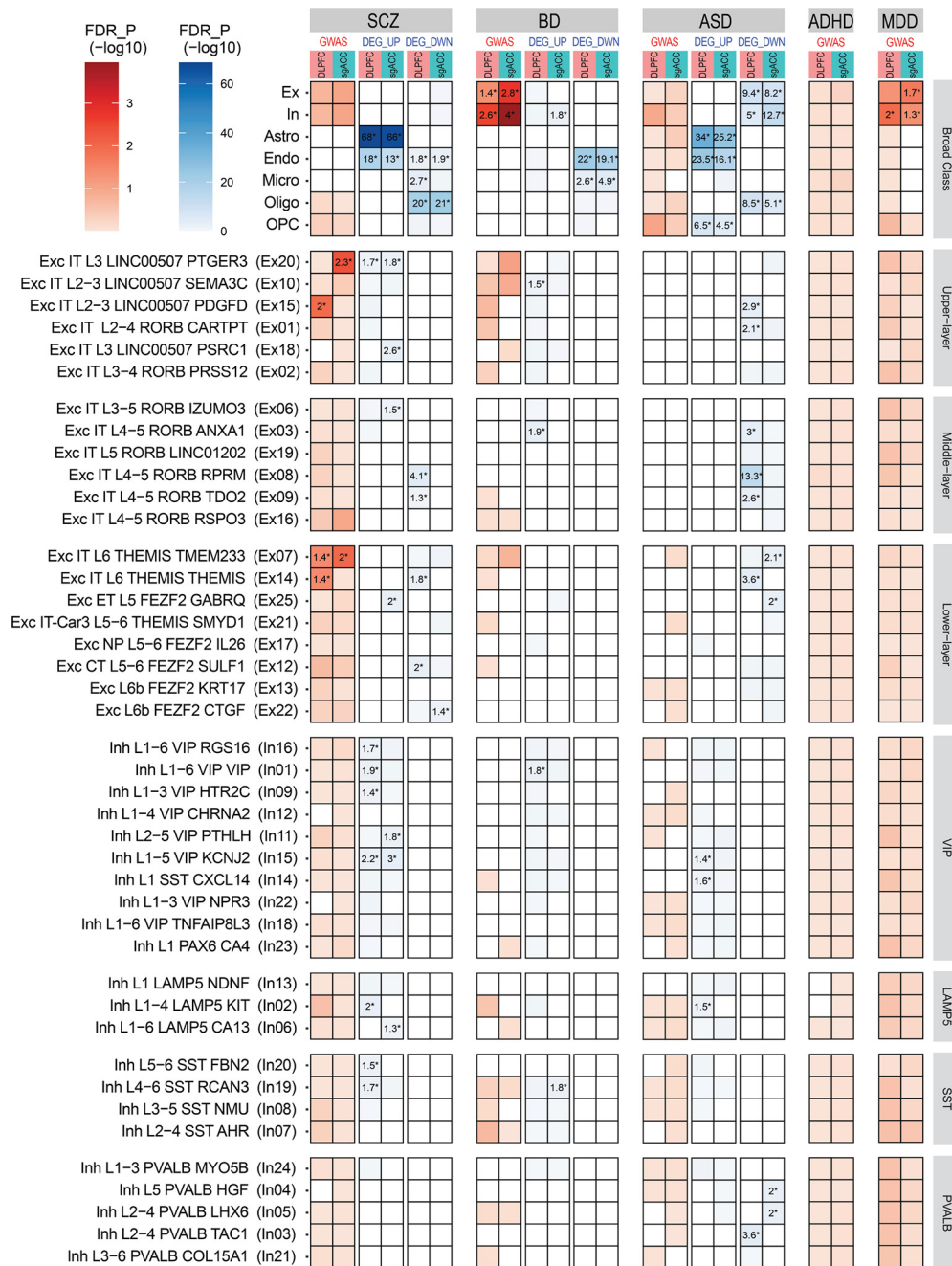
Demontis et al., 2019; Grove et al., 2019; Stahl et al., 2019) to examine the enrichment of heritable loci across neuronal subclusters for the same three neuropsychiatric disorders and two more: ADHD and MDD.

At the broad cell cluster level, glial cells exhibited the greatest enrichment for transcriptional changes across psychiatric disorders. Astrocytes were particularly enriched for DEGs upregulated in SCZ and ASD. Microglia and endothelial cells were highly enriched for DEGs downregulated in SCZ and BD. Neurons were enriched for DEGs downregulated in ASD and upregulated in BD but showed no enrichment in the other disorders (Fig. 5).

Additional DEG enrichment signals emerged when assessing neurons at the resolution of subclusters. Several Ex subclusters, primarily in the DLPFC, were associated with altered expression in ASD. Gene expression dysregulation in BD was associated with two In (In01 Inh L1-6 *VIP*, In19 Inh L4-6 *SST RCAN3*) and two Ex subclusters (Ex10 Exc IT L2-3 *LINC00507 SEMA3C*,

←

\*\**p* < 0.05. Ex20, Ex02, Ex03, Ex16, In01, and In06 also show meaningful differences between DLPFC and sgACC with unadjusted *p* values: \**p* < 0.005. **D**, Number of differential expressed genes for each brain region within each subcluster (FDR < 0.05, logFC > 0.1) (Extended Data Fig. 3-1). See Extended Data Figure 3-2 for subcluster enriched GO terms and Extended Data Figure 3-3 for the proportional contribution of individual donors to each subcluster.



**Figure 5.** Cellular enrichment of psychiatric disorders. Blue heatmaps represent transcriptomic cell type-specific enrichment of upregulated and downregulated DEGs for different psychiatric disorders (SCZ, BD, ASD). Red heatmaps represent genetic cell type-specific enrichment of GWAS loci for different psychiatric disorders (SCZ, BD, ASD, ADHD, MDD). Rows represent HBC cell types. Each cell type is divided by brain region. Scales represent  $-\log(\text{FDR } p \text{ value})$  of enrichment. Values that are statistically significant are shown and starred. Upregulated and downregulated DEGs were obtained from Gandal et al. (2018b). GWAS signals were obtained from the following studies: Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014); Wray et al. (2018); Demontis et al. (2019); Grove et al. (2019); and Stahl et al. (2019).

Ex03 Exc IT L4-5 *RORB ANXA1*). Eighteen In and Ex subclusters were enriched for transcriptional changes in SCZ in either or both the DLPPFC and sgACC (Fig. 5).

GWAS signals at the broad cell class level were generally enriched in neurons and OPCs, especially for BD and MDD, with no significant enrichment in astrocytes, oligodendrocytes, microglia, or endothelial cells across disorders. For SCZ, a nominal association for GWAS signals with broad neuronal classes was found; however, this did not survive FDR correction. Additional GWAS signal enrichment for specific neurons was found at the subcluster classification level, especially for SCZ,

where enrichment of SCZ genetic signals was identified in four Ex subclusters (Ex15 Exc IT L2-3 *LINC00507 PDGFD*, Ex7 Exc IT L6 *THEMIS TMEM233*, Ex14 Exc IT L6 *THEMIS THEMIS*, and Ex20 Exc IT L3 *LINC00507 PTGER3*). Intriguingly, two of these Ex subclusters also featured transcriptional alteration in SCZ: Ex20 in the sgACC and Ex14 in the DLPPFC. While no GWAS signals mapped to Ex25 Exc ET L5 *FEZF2 GABRQ*, this subcluster was enriched for DEGs upregulated in SCZ only in the sgACC. Only one In neuronal subcluster (In07 Inh L2-4 *SST AHR*) exhibited GWAS enrichment in BD, and no neuronal enrichment was found for ADHD and MDD (Fig. 5).

## Discussion

### Neuronal diversity in the DLPFC and sgACC

In this study, we sought to characterize the cellular composition of the DLPFC and sgACC, two brain regions implicated in various psychiatric disorders (Drevets et al., 2008; Guillozet-Bongaarts et al., 2014). Through snRNA-seq of these brain regions from 8 individuals with no history of mental illness, we were able to profile 63,901 nuclei, ultimately identifying 42 transcriptionally distinct neuronal populations using an unbiased, data-driven approach. The transcriptional identities of these subclusters proved to be quite robust. Most clusters corresponded tightly to those defined by the Allen Institute Multiple Cortical Areas SMART-seq dataset, which were achieved with layer-specific dissections and much deeper sequencing.

GABAergic interneurons have been classically defined by their derivation from either the MGE or CGE, their expression of the canonical markers SST, PVAlB, VIP, or LAMP5, and morphology. While the molecular definition of these cells remains unclear, a few populations could be clearly defined: In21 Inh L3-6 PVAlB COL15A1 are presumably Chandelier cells because of expression of TRPS1 and PTHLH and integration with Allen Inh L1-3 PVAlB MFI2 (Fig. 2F) (Marin and Muller, 2014); and In02 (Inh L1-4 LAMP5 KIT) are likely Rosehip neurons, with the molecular profile GAD1<sup>+</sup>/CCK<sup>+</sup>/CNRI<sup>-</sup>/SST<sup>-</sup>/CALB2<sup>-</sup>/PVAlB<sup>-</sup> (Boldog et al., 2018). We also identified a rare LAMP5 subtype (In06 Inh L1-6 LAMP5 CA13) with characteristics of both MGE and CGE derivation (Extended Data Fig. 2-2H). A similar population was identified in mouse (Tasic et al., 2018) and human (Luo et al., 2022), and a recent study using integrative cross-species alignment and ISH in mouse and marmoset hypothesizes that these LAMP5<sup>+</sup>/LHX6<sup>+</sup> cells may be MGE-derived (Krienen et al., 2020). The remaining populations of SST, VIP, PVAlB, and LAMP5 cells do not map unequivocally to specific morphologic or neurophysiological types, although recent works attempt to bridge this gap (Gouwens et al., 2019; Scala et al., 2021). An approach combining immunohistochemical markers, morphology, and RNA probes could help map molecular definitions to morphologic/functional GABAergic classes.

Unlike interneurons, Ex subclusters exhibited considerable diversity across cortical regions. Diversity in cell numbers was greatest in supragranular layers (Fig. 3C), while diversity in expression was distributed across different layers (Fig. 3D). The correspondence of our Ex subclusters with those from the Allen Institute was less precise for superficial layer classes than deeper-layer Ex subclasses. Three distinct cell types identified in our study (Ex10, Ex15, and Ex01) mapped to Exc L2-3 LINC00507 RPL9P17 in the Allen dataset, and Ex20 and Ex18 both mapped to Exc L3 LINC00507 PSRC1. This may reflect evolutionary expansion of supragranular layers in human (DeFelipe, 2011), resulting in regional specialization of cells within these superficial layers. Indeed, several superficial and middle-layer Ex subclasses tend to be enriched in one or few regions, whereas deeper-layer subclasses appear to be present across brain regions (Extended Data Fig. 3-1B,C). Interestingly, the Allen cell classes derived primarily from the ACC (Exc L3-5 RORB HSBP3, Exc L3-5 RORB CD24, Exc L3-5 LINC00507 SLN, and Exc L3-5 LINC00507 PSRC1) were found enriched in the sgACC, which is a contiguous brain region. Supragranular layers consist mostly of corticocortical projection neurons (Krienen et al., 2016), which are postulated to be more varied in higher-order cortical regions

to support specialized local information processing. Deeper-layer projection neurons responsible for communicating with other cortical and extracortical regions exhibited less regional variation in cell proportions.

These differences in cell types are also reflected in the cytoarchitectonics of the two regions, the major difference being the relative absence of granular layer 4 in the sgACC (Rajkowska and Goldman-Rakic, 1995; Palomero-Gallagher et al., 2008). Overall, layer 4 RORB-expressing Ex neurons were 2.6-fold overrepresented in DLPFC versus sgACC. Specifically, three abundant L4-enriched Ex subclasses (Ex01 IT L2-4 RORB CARTPT, Ex02 IT L3-4 RORB PRSS12, Ex03 IT L4-5 RORB ANXA1 in Fig. 3C) were significantly overrepresented in DLPFC. Only the rarer Ex16 IT L4-5 RORB RSPO3 was overrepresented in sgACC (Fig. 3B; Extended Data Fig. 2-1).

We also identified a rare, well-defined cell population, the VEN (Ex25 Exc ET L5 FEFZ2 GABRQ). VENs are extratelencephalic projecting excitatory neurons with morphologically characteristic spindle-shaped cell bodies primarily found in the fronto-insular, anterior cingulate, and medial fronto-polar cortices (González-Acosta et al., 2018; Banovac et al., 2021). In our dataset, Ex25 had the greatest number of intercortical DEGs and indeed expressed canonical VEN markers, such as ADRA1A and ADAMTS2. However, Ex25 in the DLPFC did not express GABRQ. Moreover, we noted that, in the DLPFC, these ADRA1A-positive cells exhibited pyramidal morphology as opposed to having a spindle or forked morphology. Hodge et al. (2020) similarly found that this population also mapped to both pyramidal and forked cell morphologies, and others have suggested the existence of VENs in the DLPFC (BA9) (Fajardo et al., 2008; González-Acosta et al., 2018). The fact that these populations of neurons consistently cocluster across studies and datasets suggests that L5 ET cells play similar functional roles across the two brain regions. Like Ex25, other intratelencephalic projecting cell types (Ex10 in L2-3 or Ex06 in L3-5) also exhibit marked intercortical differential gene expression. The sgACC has reciprocal connections with medial and orbital frontal cortical areas, entorhinal cortex and amygdala (Calderazzo et al., 2021; McHale et al., 2022), whereas the DLPFC connects to lateral frontal, parietal, insular, and superior temporal regions (Yeterian et al., 2012); it is possible that the large number of intercortical DEGs present in these long-range projection neurons reflect the underlying differences in connectivity.

### Enrichment of psychiatric genetic signals and transcriptional changes in specific neuronal populations

One key advantage of cell-type classification is the ability to interrogate the contribution of individual cell populations to the development and expression of psychiatric disorders. To this end, we examined the cell-specific enrichment for both heritability signals and alterations of gene expression. Our findings suggest: (1) that the genetic determinants of psychiatric disorders act primarily on neurons; (2) that the gene expression changes detected in bulk-RNA-seq studies are dominated by the response of glial cells to the changes in neurons or factors arising after illness onset; and (3) that gene expression changes in neuronal subpopulations are masked in bulk-RNA-seq studies.

Neurons exhibited enrichment for GWAS signals for BD and MDD at the broad class level; however, enrichment of specific neuronal subclusters was not evident for these disorders. However, neuronal subclassification revealed surprisingly strong enrichment for SCZ GWAS signals in four Ex

subclusters: Ex15 IT L2-3 *LINC00507 PDGFD* and Ex14 IT L6 *THEMIS THEMIS* in the DLPFC; Ex20 IT L3 *LINC00507 PTGER3* in the sgACC; and Ex7 IT L6 *THEMIS THEM233* in both brain regions. No cell type-specific enrichment was observed for GWAS variants associated with ASD, possibly reflecting (1) the smaller size of ASD GWAS studies (Grove et al., 2019) or (2) that rare *de novo* variants might contribute disproportionately to the genetics of ASD (Iossifov et al., 2014; Sestan and State, 2018; T. Wang et al., 2020). Others have found enrichment of GWAS signal for SCZ and BD in several excitatory and inhibitory subclasses (Skene et al., 2018; Tran et al., 2021) and for ASD in one inhibitory subtype, OPCs, oligodendrocytes, and astrocytes (Tran et al., 2021). Skene et al. (2018) associated GWAS signals for both SCZ and BD to many of the same inhibitory and excitatory cell types. Our study includes a larger number of nuclei (63,901 vs ~14,000) and finer resolution clustering for Ex and In subtypes, enabling resolution of differences in the neuronal association of SCZ and BD GWAS signals. Additional key differences include our use of human single nuclei versus mouse single cell data; our focus on two prefrontal cortical brain regions versus the inclusion of striatum, somatosensory, and hippocampus, and differences in the multiple comparison correction procedures used. Regarding Tran et al. (2021), differing results may be explained by: (1) our use of LDSC versus MAGMA to determine relevant GWAS signals; (2) the resolution of the cell types resolved; and (3) the threshold for cell type gene specificity selected for analysis. Nevertheless, the convergent finding is that the heritable factors contributing to the development of a psychiatric disorder act primarily through neurons.

Differential gene expression paints a different picture. DEGs for SCZ, BD, and ASD were enriched primarily in glia, microglia, and endothelial cells with less significant enrichment scattered throughout several In and Ex subclusters. For BD, DEGs derived from studies of the DLPFC (Gandal et al., 2018b) were minimally enriched in neuronal subclusters. However, DEGs identified in a separate study of sgACC (Akula et al., 2021) were not enriched in any cell type, most likely because this study contained only 39 cases with BD (data not shown). Analysis of a more recent RNA-seq study of 138 cases with BD in the sgACC (Zandi et al., 2022) confirmed the findings based on transcriptomics in the DLPFC of BD (Gandal et al., 2018b) (data not shown). With respect to ASD, DEG enrichment primarily localized to upper- and middle-layer Ex subclusters in the DLPFC, corroborating findings by others that upper-layer excitatory neurons (Velmeshv et al., 2019) and the PFC is most affected in ASD (Courchesne et al., 2011; Parikshak et al., 2016; Brumback et al., 2018; Schwede et al., 2018). Two PVALB In subclusters, In04 and In05, were enriched for genes downregulated in ASD. This result aligns with the finding of decreased numbers of PVALB<sup>+</sup> interneurons in the PFC of ASD brains (Hashemi et al., 2017).

For SCZ, astrocytes exhibited the strongest enrichment for genes with increased expression in disease and a few Ex subclusters exhibited enrichment for genes with either increased expression in superficial layer subclusters (Ex20 IT L3 *LINC00507 PTGER3*, Ex18 IT L3 *LINC00507 PSRC1*, and Ex06 IT L3 – 5 *RORB IZUMO3*) or reduced expression in middle- or lower-layer subclusters (Ex08 IT L4 – 5 *RORB RPRM*, Ex09 IT L4 – 5 *RORB TDO2*, Ex14 IT L6 *THEMIS THEMIS*, and Ex22 L6b *FEZF2 CTGF*). Recent snRNA-seq studies of SCZ also found enrichment of DEGs in similar excitatory cell classes (Ruzicka et al., 2020; Reiner et al., 2021; Batiuk et al., 2022). We noted that Ex25 L5-ET L5 *FEZF2 GABRQ* (VENs) in the sgACC, but not DLPFC,

exhibited modest enrichment for genes upregulated in SCZ and ASD, in line with other studies that suggest VENs may be selectively vulnerable in SCZ, ASD, and frontotemporal lobar dementia (Uppal et al., 2014; Krause et al., 2017; Y. Yang et al., 2017; Hodge et al., 2020). Regarding inhibitory cells, VIP In subclusters were enriched for genes with increased expression in SCZ; and while some enrichment was found for SST and LAMP5 In subclusters, no enrichment was noted for any PVALB In subclusters. This finding stands in contrast to findings of presynaptic and postsynaptic abnormalities in PVALB cells (for review, see Lewis et al., 2012), possibly because some within-subcluster alterations may go undetected in homogenate RNA-seq studies, which are more likely to capture global changes across multiple cell classes or changes from a dominant cell population.

Only for SCZ did two cell types demonstrate enrichment for both GWASs and differential expression signals: Ex20 IT L3 *LINC00507 PTGER3* and Ex14 IT L6 *THEMIS THEMIS*. This finding is tantalizing as it suggests a possible connection between heritable features and transcriptional phenotypes within these cell types. Given that more cell types did not exhibit convergent signals suggests that these excitatory neuronal cell types may exhibit the strongest dysregulation in SCZ and may be the primary driving defect underlying the disorder. In this hypothetical framework, the transcriptional changes identified in glia and In subtypes are a compensatory/regulatory response to Ex cell dysfunction. Still, it is possible that (1) some transcriptional changes in other individual neuronal subclusters are obscured in homogenate RNA-seq studies; and that (2) GWAS signals that act during development may not be expressed in the appropriate cell type in adulthood and remain undetected.

Our study was limited by: (1) small sample size of only male individuals; (2) technical differences between the 10× Genomics and SMART-seq platforms resulting in reduced detection of low-expressed genes; and (3) nuclei isolation biased toward collection of neuronal nuclei over glial nuclei, making detection of rare cell types difficult. We were unable to identify the Allen Brain SST CHODL cluster (Inh L6 *SST NPY*), which is a rare SST interneuron with long-range projections. Additionally, our requirement that each subcluster be present in all or almost all the subjects, while conservative, also increases the likelihood of missing rare cell types. While it is possible that clusters might represent distinct transcriptional states of a similar cell subtype (Kharchenko, 2021), the correspondence of the cell types identified here to those from the Allen Institute supports the validity of our clustering methods.

In conclusion, we analyzed snRNA-sequencing data from the DLPFC and sgACC, identified 42 robust neuronal subclasses, elucidated cellular and molecular differences between the regions, and mapped genetic risk and expression changes associated with psychiatric disorders onto specific cell types. These findings reveal new insights into the functional differences between the sgACC and DLPFC and identify hypothetical mechanisms by which these functions might be altered in SCZ, MDD, BP, and ASD.

## References

- Akula N, et al. (2021) Deep transcriptome sequencing of subgenual anterior cingulate cortex reveals cross-diagnostic and diagnosis-specific RNA expression changes in major psychiatric disorders. *Neuropsychopharmacology* 46:1364–1372.
- Anderson CM, Zhang B, Miller M, Butko E, Wu X, Laver T, Kernag C, Kim J, Luo Y, Lamparski H, Park E, Su N, Ma X (2016) Fully automated RNAscope in situ hybridization assays for formalin-fixed paraffin-embedded cells and tissues. *J Cell Biochem* 117:2201–2208.

- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Bakken TE, et al. (2018) Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* 13:e0209648.
- Bakken TE, et al. (2021) Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* 598:111–119.
- Banovac I, Sedmak D, Judaš M, Petanjek Z (2021) von Economo neurons: primate-specific or commonplace in the mammalian brain? *Front Neural Circuits* 15:714611.
- Batiuk MY, Tyler T, Dragicevic K, Mei S, Rydbirk R, Petukhov V, Deviatarov R, Sedmak D, Frank E, Feher V, Habek N, Hu Q, Iqolkina A, Roszik L, Pfisterer U, Garcia-Gonzalez D, Petanjek Z, Adorjan I, Kharchenko PV, Khodosevich K (2022) Upper cortical layer-driven network impairment in schizophrenia. *Sci Adv* 8:eabn8367.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.
- Boldog E, et al. (2018) Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat Neurosci* 21:1185–1195.
- Brumback AC, Ellwood IT, Kjaerby C, Iafraji J, Robinson S, Lee AT, Patel T, Nagaraj S, Davatolhagh F, Sohal VS (2018) Identifying specific prefrontal neurons that contribute to autism-associated abnormalities in physiology and social behavior. *Mol Psychiatry* 23:2078–2089.
- Broyois J, et al. (2020) Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat Genet* 52:482–493.
- Calderazzo SM, Busch SE, Moore TL, Rosene DL, Medalla M (2021) Distribution and overlap of entorhinal, premotor, and amygdalar connections in the monkey anterior cingulate cortex. *J Comp Neurol* 529:885–904.
- Gene Ontology Consortium (2001) Creating the gene ontology resource: design and implementation. *Genome Res* 11:1425–1433.
- Courchesne E, Mouton PR, Calhoun ME, Semendeferi K, Ahrens-Barbeau C, Hallet MJ, Barnes CC, Pierce K (2011) Neuron number and size in prefrontal cortex of children with autism. *JAMA* 306:2001–2010.
- Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, Quake SR (2015) A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci USA* 112:7285–7290.
- Deep-Soboslay A, Akil M, Martin CE, Bigelow LB, Herman MM, Hyde TM, Kleinman JE (2005) Reliability of psychiatric diagnosis in postmortem research. *Biol Psychiatry* 57:96–101.
- DeFelipe J (2011) The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front Neuroanat* 5:29.
- Demontis D, et al. (2019) Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 51:63–75.
- Drevets WC, Savitz J, Trimble M (2008) The subgenual anterior cingulate cortex in mood disorders. *CNS Spectr* 13:663–681.
- Egner T, Hirsch J (2005) Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat Neurosci* 8:1784–1790.
- Fajardo C, Escobar MI, Buriticá E, Arteaga G, Umbarila J, Casanova MF, Pimienta H (2008) von Economo neurons are present in the dorsolateral (dysgranular) prefrontal cortex of humans. *Neurosci Lett* 435:215–218.
- Finucane HK, et al., RACI Consortium (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 47:1228–1235.
- Fromer M, et al. (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 19:1442–1453.
- Gaidatzis D, Burger L, Florescu M, Stadler MB (2015) Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat Biotechnol* 33:722–729.
- Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, Schork AJ, Appadurai V, Buil A, Werge TM, Liu C, White KP; CommonMind Consortium; PsychENCODE Consortium; iPSYCH-BROAD Working Group; Horvath S, Geschwind DH (2018a) Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 359:693–697.
- Gandal MJ, et al. (2018b) Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 362:eaat8127.
- González-Acosta CA, Escobar MI, Casanova MF, Pimienta H, Buriticá E (2018) von Economo neurons in the human medial frontopolar cortex. *Front Neuroanat* 12:64.
- Gouwens NW, et al. (2019) Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nat Neurosci* 22:1182–1195.
- Grove J, et al. (2019) Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* 51:431–444.
- Guillozet-Bongaerts AL, Hyde TM, Dalley RA, Hawrylycz MJ, Henry A, Hof PR, Hohmann J, Jones AR, Kuan CL, Royall J, Shen E, Swanson B, Zeng H, Kleinman JE (2014) Altered gene expression in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol Psychiatry* 19:478–485.
- Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi I, Trombetta JJ, Hession C, Zhang F, Regev A (2016) Div-Seq: single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* 353:925–928.
- Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F, Regev A (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* 14:955–958.
- Hao Y, et al. (2021) Integrated analysis of multimodal single-cell data. *Cell* 184:3573–3587.e29.
- Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324:646–648.
- Hashemi E, Ariza J, Rogers H, Noctor SC, Martínez-Cerdeño V (2017) The number of parvalbumin-expressing interneurons is decreased in the prefrontal cortex in autism. *Cereb Cortex* 27:1931–1943.
- Hodge RD, et al. (2019) Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573:61–68.
- Hodge RD, et al. (2020) Transcriptomic evidence that von Economo neurons are regionally specialized extralencephalic-projecting excitatory neurons. *Nat Commun* 11:1172.
- Hu P, Fabyanic E, Kwon DY, Tang S, Zhou Z, Wu H (2017) Dissecting cell-type composition and activity-dependent transcriptional state in mammalian brains by massively parallel single-nucleus RNA-Seq. *Mol Cell* 68:1006–1015.e1007.
- Iossifov I, et al. (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515:216–221.
- Jaffe AE, et al. (2018) Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci* 21:1117–1125.
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30.
- Kharchenko PV (2021) The triumphs and limitations of computational methods for scRNA-seq. *Nat Methods* 18:723–732.
- Koenigs M, Grafman J (2009) The functional neuroanatomy of depression: distinct roles for ventromedial and dorsolateral prefrontal cortex. *Behav Brain Res* 201:239–243.
- Krause M, Theiss C, Brüne M (2017) Ultrastructural alterations of von Economo neurons in the anterior cingulate cortex in schizophrenia. *Anat Rec (Hoboken)* 300:2017–2024.
- Krienen FM, Yeo BT, Ge T, Buckner RL, Sherwood CC (2016) Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proc Natl Acad Sci USA* 113:E469–E478.
- Krienen FM, et al. (2020) Innovations present in the primate interneuron repertoire. *Nature* 586:262–269.
- Krishnaswami SR, et al. (2016) Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* 11:499–524.
- Kunii Y, Zhang W, Xu Q, Hyde TM, McFadden W, Shin JH, Deep-Soboslay A, Ye T, Li C, Kleinman JE, Wang KH, Lipska BK (2015) CHRNA7 and CHRFA7A mRNAs: co-localized and their expression levels altered in the postmortem dorsolateral prefrontal cortex in major psychiatric disorders. *Am J Psychiatry* 172:1122–1130.
- Lake BB, et al. (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 352:1586–1590.
- Lake BB, Codoluppi S, Yung YC, Gao D, Chun J, Kharchenko PV, Linnarsson S, Zhang K (2017) A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci Rep* 7:6031.

- Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, Duong TE, Gao D, Chun J, Kharchenko PV, Zhang K (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* 36:70–80.
- Lewis DA, Curley AA, Glausier JR, Volk DW (2012) Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends Neurosci* 35:57–67.
- Lipska BK, Deep-Soboslay A, Weickert CS, Hyde TM, Martin CE, Herman MM, Kleinman JE (2006) Critical factors in gene expression in postmortem human brain: focus on studies in schizophrenia. *Biol Psychiatry* 60:650–658.
- Luo C, et al. (2022) Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genom* 2:100107.
- MacDonald AW, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288:1835–1838.
- Manoach DS, Press DZ, Thangaraj V, Searl MM, Goff DC, Halpern E, Saper CB, Warach S (1999) Schizophrenic subjects activate dorsolateral prefrontal cortex during a working memory task, as measured by fMRI. *Biol Psychiatry* 45:1128–1137.
- Marin O, Muller U (2014) Lineage origins of GABAergic versus glutamatergic neurons in the neocortex. *Curr Opin Neurobiol* 26:132–141.
- Mayberg HS, Liotti M, Brannan SK, McGinnis S, Mahurin RK, Jerabek PA, Silva JA, Tekell JL, Martin CC, Lancaster JL, Fox PT (1999) Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am J Psychiatry* 156:675–682.
- McHale AC, Cho YT, Fudge JL (2022) Cortical granularity shapes the organization of afferent paths to the amygdala and its striatal targets in nonhuman primate. *J Neurosci* 42:1436–1453.
- Miller JA, Gouwens NW, Tasic B, Collman F, van Velthoven CT, Bakken TE, Hawrylycz MJ, Zeng H, Lein ES, Bernard A (2020) Common cell type nomenclature for the mammalian brain. *Elife* 9:e59928.
- Nagy C, Maitra M, Tanti A, Suderman M, Thérroux JF, Davoli MA, Perlman K, Yerko V, Wang YC, Tripathy SJ, Pavlidis P, Mechawar N, Ragoussis J, Turecki G (2020) Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat Neurosci* 23:771–781.
- Palomero-Gallagher N, Mohlberg H, Zilles K, Vogt B (2008) Cytology and receptor architecture of human anterior cingulate cortex. *J Comp Neurol* 508:906–926.
- Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, Hartl C, Leppa V, Ubieta LT, Huang J, Lowe JK, Blencowe BJ, Horvath S, Geschwind DH (2016) Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540:423–427.
- Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford)* 2015:bav028.
- Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI (2020) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 48:D845–D855.
- Potkin SG, Turner JA, Brown GG, McCarthy G, Greve DN, Glover GH, Manoach DS, Belger A, Diaz M, Wible CG, Ford JM, Mathalon DH, Gollub R, Lauriello J, O’Leary D, van Erp TG, Toga AW, Preda A, Lim KO, FBIRN (2009) Working memory and DLPFC inefficiency in schizophrenia: the FBIRN study. *Schizophr Bull* 35:19–31.
- Price JL, Drevets WC (2012) Neural circuits underlying the pathophysiology of mood disorders. *Trends Cogn Sci* 16:61–71.
- Rajkowska G, Goldman-Rakic PS (1995) Cytoarchitectonic definition of prefrontal areas in the normal human cortex: I. Remapping of areas 9 and 46 using quantitative criteria. *Cereb Cortex* 5:307–322.
- Reiner BC, Crist RC, Stein LM, Weller AE, Doyle GA, Arauco-Shapiro G, Turecki G, Ferraro TN, Hayes MR, Berrettini WH (2021) Single-nuclei transcriptomics of schizophrenia prefrontal cortex primarily implicates neuronal subtypes. *bioRxiv* 227355. <https://doi.org/10.1101/2020.07.29.227355>.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427.
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Rue-Albrecht K, Marini F, Sonesson C, Lun AT (2018) iSEE: interactive SummarizedExperiment Explorer. *F1000Res* 7:741.
- Ruzicka WB, Mohammadi S, Davila-Velderrain J, Subburaju S, Tso DR, Hourihan M, Kellis M (2020) Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. *medRxiv* 22279406. <https://doi.org/10.1101/2022.08.31.22279406>.
- Sathyamurthy A, Johnson KR, Matson KJ, Dobrott CI, Li L, Ryba AR, Bergman TB, Kelly MC, Kelley MW, Levine AJ (2018) Massively parallel single nucleus transcriptional profiling defines spinal cord neurons and their activity during behavior. *Cell Rep* 22:2216–2225.
- Scala F, Kobak D, Bernabucci M, Bernaerts Y, Cadwell CR, Castro JR, Hartmanis L, Jiang X, Laturnus S, Miranda E, Mulherkar S, Tan ZH, Yao Z, Zeng H, Sandberg R, Berens P, Tolias AS (2021) Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* 598:144–150.
- Schwede M, Nagpal S, Gandal MJ, Parikshak NN, Mirmics K, Geschwind DH, Morrow EM (2018) Strong correlation of downregulated genes related to synaptic transmission and mitochondria in post-mortem autism cerebral cortex. *J Neurodev Disord* 10:18.
- Scifo E, Pabba M, Kapadia F, Ma T, Lewis DA, Tseng GC, Sibille E (2018) Sustained molecular pathology across episodes and remission in major depressive disorder. *Biol Psychiatry* 83:81–89.
- Seeley WW, Carlin DA, Allman JM, Macedo MN, Bush C, Miller BL, Dearmond SJ (2006) Early frontotemporal dementia targets neurons unique to apes and humans. *Ann Neurol* 60:660–667.
- Sestan N, State MW (2018) Lost in translation: traversing the complex path from genomics to therapeutics in autism spectrum disorder. *Neuron* 100:406–423.
- Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA, Giusti-Rodríguez P, Hodge RD, Miller JA, Muñoz-Manchado AB, O’Donovan MC, Owen MJ, Pardiñas AF, Ryge J, Walters JT, Linnarsson S, Lein ES, Sullivan PF, Hjerling-Lefler J, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (2018) Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* 50:825–833.
- Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, Hudelle R, Qaiser T, Matson KJ, Barraud Q, Levine AJ, La Manno G, Skinnider MA, Courtine G (2021) Confronting false discoveries in single-cell differential expression. *Nat Commun* 12:5692.
- Stahl EA, et al. (2019) Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 51:793–803.
- Tasic B, et al. (2018) Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563:72–78.
- Tran MN, Maynard KR, Spangler A, Huuki LA, Montgomery KD, Sadashivaiah V, Tippani M, Barry BK, Hancock DB, Hicks SC, Kleinman JE, Hyde TM, Collado-Torres L, Jaffe AE, Martinowich K (2021) Single-nucleus transcriptome analysis reveals cell-type-specific molecular signatures across reward circuitry in the human brain. *Neuron* 109:3088–3103. e3085.
- Tripp A, Kota RS, Lewis DA, Sibille E (2011) Reduced somatostatin in subgenual anterior cingulate cortex in major depression. *Neurobiol Dis* 42:116–124.
- Uppal N, Wicinski B, Buxbaum JD, Heinsen H, Schmitz C, Hof PR (2014) Neuropathology of the anterior midcingulate cortex in young children with autism. *J Neuropathol Exp Neurol* 73:891–902.
- Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, Bhaduri A, Goyal N, Rowitch DH, Kriegstein AR (2019) Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* 364:685–689.
- Wang D, et al. (2018) Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362:eaat8464.
- Wang T, et al. (2020) Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nat Commun* 11:4932.
- Weinberger DR, Berman KF, Zec RF (1986) Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia: I. Regional cerebral blood flow evidence. *Arch Gen Psychiatry* 43:114–124.

- Wray NR, et al. (2018) Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 50:668–681.
- Yang L, Yang Y, Yuan J, Sun Y, Dai J, Su B (2019) Transcriptomic landscape of von Economo neurons in human anterior cingulate cortex revealed by microdissected-cell RNA sequencing. *Cereb Cortex* 29:838–851.
- Yang Y, Halliday GM, Hodges JR, Tan RH (2017) von Economo neuron density and thalamus volumes in behavioral deficits in frontotemporal dementia cases with and without a C9ORF72 repeat expansion. *J Alzheimers Dis* 58:701–709.
- Yeterian EH, Pandya DN, Tomaiuolo F, Petrides M (2012) The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48:58–81.
- Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16:284–287.
- Zandi PP, Jaffe AE, Goes FS, Burke EE, Collado-Torres L, Huuki-Myers L, Seyedian A, Lin Y, Seifuddin F, Pirooznia M, Ross CA, Kleinman JE, Weinberger DR, Hyde TM (2022) Amygdala and anterior cingulate transcriptomes from individuals with bipolar disorder reveal downregulated neuroimmune and synaptic pathways. *Nat Neurosci* 25:381–389.
- Zeng H, et al. (2012) Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* 149:483–496.
- Zheng GX, et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8:14049.