

RESEARCH ARTICLE SUMMARY

PSYCHENCODE2

Single-cell genomics and regulatory networks for 388 human brains

Prashant S. Emani *et al.*

INTRODUCTION: Single-cell genomics offers a powerful method to understand how variants influence gene expression, especially across the numerous cell types in the human brain. Moreover, it can potentially refine our understanding of the regulatory mechanisms underlying brain-related traits. However, population-scale cohorts with a wide range of brain phenotypes are needed to infer key associations among variants and to develop models of regulation at the single-cell scale.

RATIONALE: To address this, the PsychENCODE Consortium performed many single-cell experiments [single-nucleus RNA sequencing (snRNA-seq), snATAC-seq (ATAC, assay for transposase-accessible chromatin), and snMultiome plus genotyping] and computational analyses on prefrontal-cortex samples of adults with a range of brain-related disorders such as schizophrenia, autism spectrum disorder, bipolar disorder, and Alzheimer's disease, as well as controls.

RESULTS: We developed a uniformly processed resource comprising >2.8 million nuclei from 388 individuals (brainscope.psychencode.org).

The resource is based on harmonized cell typing, with 28 neuronal and non-neuronal cell types (registered against BICCN). Partitioning the expression variation within these types revealed higher cell-type variability than inter-individual variability; this pattern was amplified in neurotransmitter and neurorelated drug-target genes such as *CNR1*.

Integration of expression and genotype data revealed >1.4 million single-cell expression quantitative trait loci (eQTLs), many of which were not seen in bulk gene-expression datasets and a subset of which involved variants related to brain disorders. Moreover, we found that expression patterns across cell types recapitulated the spatial relationships of excitatory neurons across cortical layers and enabled the identification of “dynamic eQTLs,” with smooth changes in regulatory effect across cortical layers. The chromatin datasets in the resource allowed for identification of >550,000 single-cell cis-regulatory elements, which were enriched at loci linked to brain-related traits. Combining expression, chromatin, and eQTL datasets, we built cell type-specific gene regulatory networks. In these, information-flow bottleneck

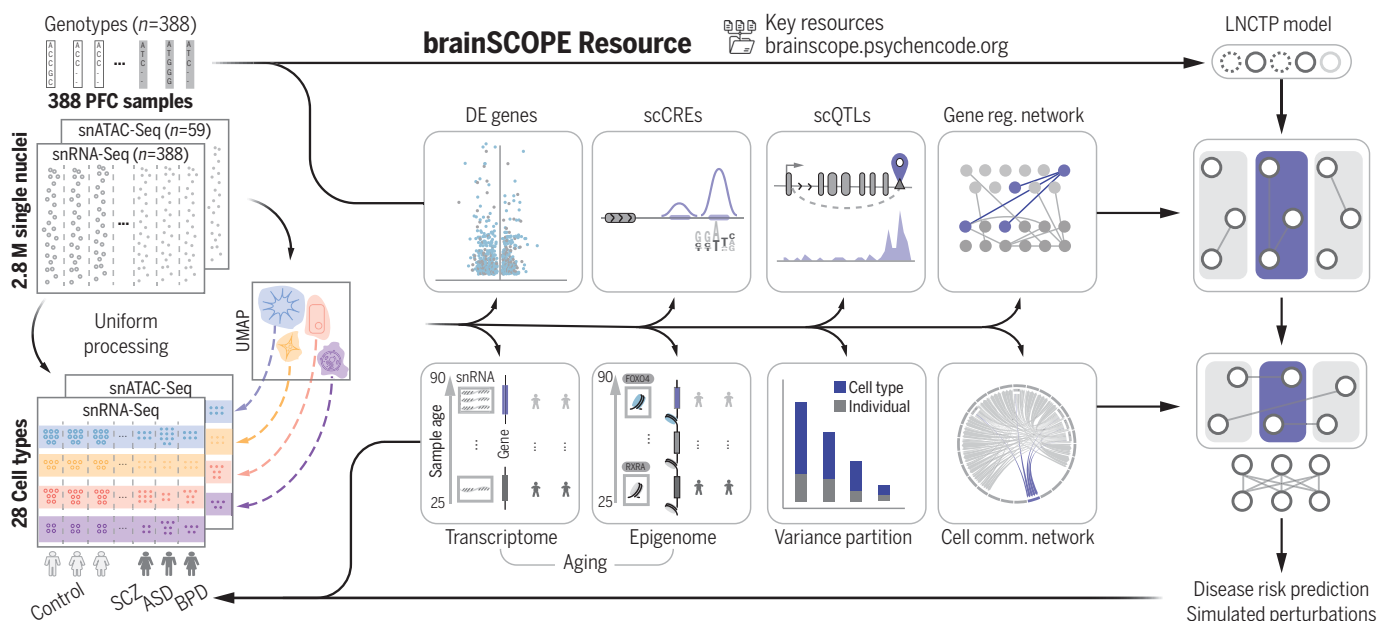
genes tended to be specific to particular cell types, in contrast to hubs. We also developed cell-to-cell communication networks, which highlighted differences in signaling pathways in disorders, including altered Wnt signaling in schizophrenia and bipolar disorder.

We developed an integrative deep-learning model with embedded layers for genotypes, eQTLs, and regulatory and cell-to-cell communications networks. The model allowed for accurate imputation of cell type-specific expression and phenotype from genotype. It prioritized >250 risk genes and drug targets for brain-related disorders along with associated cell types. Simulated perturbation of individual genes led to predicted expression changes mirroring those for disease cases, suggesting drug targets. Lastly, we constructed predictive models for aging and Alzheimer's disease, showing, for instance, that expression and chromatin in specific neurons were highly predictive of an individual's age.

CONCLUSION: Our population-scale single-cell resource for the human brain can help facilitate precision-medicine approaches for neuropsychiatric disorders, especially by prioritizing follow-up genes and drug targets linked to cell types. ■

The list of author affiliations is available in the full article online.
*Corresponding authors: Matthew J. Girgenti (matthew.girgenti@yale.edu); Jing Zhang (zhang.jing@uci.edu); Daifeng Wang (daifeng.wang@wisc.edu); Mark Gerstein (mark@gersteinlab.org)
Cite this article as P. S. Emani *et al.*, *Science* **384**, eadi5199 (2024). DOI: 10.1126/science.adi5199

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.adi5199>



brainSCOPE resource. snRNA-seq and snATAC-seq from 388 individuals allowed assessment of regulatory elements (scCREs), single-cell eQTLs (scQTLs), and gene regulatory networks across cell types. These were integrated into a model (LNCTP, Linear Network of Cell Type Phenotypes) to predict phenotypes and prioritize genes and cell types.

RESEARCH ARTICLE

PSYCHENCODE2

Single-cell genomics and regulatory networks for 388 human brains

Prashant S. Emani^{1,2,†}, Jason J. Liu^{1,2,†}, Declan Clarke^{1,2,†}, Matthew Jensen^{1,2,†}, Jonathan Warrell^{1,2,†}, Chirag Gupta^{3,4,†}, Ran Meng^{1,2,†}, Che Yu Lee^{5,†}, Siwei Xu^{5,†}, Gagatay Dursun^{1,2,†}, Shaoke Lou^{1,2,†}, Yuhang Chen^{1,2}, Zhiyuan Chu¹, Timur Galeev^{1,2}, Ahyeon Hwang^{5,6}, Yunyong Li^{2,7}, Pengyu Ni^{1,2}, Xiao Zhou^{1,2}, PsychENCODE Consortium[†], Trygve E. Bakken⁸, Jaroslav Bendl^{9,10,11,12}, Lucy Bicks¹³, Tanima Chatterjee^{1,2}, Lijun Cheng¹⁴, Yuyan Cheng^{13,15}, Yi Dai⁵, Ziheng Duan⁵, Mary Flaherty¹⁴, John F. Fullard^{9,10,11,12}, Michael Gancz^{1,2}, Diego Garrido-Martín¹⁶, Sophia Gaynor-Gillett^{14,17}, Jennifer Grundman¹³, Natalie Hawken¹³, Ella Henry^{1,2}, Gabriel E. Hoffman^{9,10,11,12,18,19}, Ao Huang¹, Yunzhe Jiang^{1,2}, Ting Jin^{3,4}, Nikolas L. Jorstad⁸, Riki Kawaguchi^{13,20}, Saniya Khullar^{3,4}, Jianyin Liu¹³, Junhao Liu⁵, Shuang Liu⁴, Shaojie Ma^{21,22}, Michael Margolis¹³, Samantha Mazariegos¹³, Jill Moore²³, Jennifer R. Moran¹⁴, Eric Nguyen^{1,2}, Nishigandha Phalke²³, Milos Pjanic^{9,10,11,12}, Henry Pratt²³, Diana Quintero¹³, Ananya S. Rajagopalan⁷, Tiernon R. Riesenmy²⁴, Nicole Shedd²³, Manman Shi¹⁴, Megan Spector¹⁴, Rosemarie Terwilliger²⁵, Kyle J. Travaglini⁸, Brie Wamsley¹³, Gaoyuan Wang^{1,2}, Yan Xia^{1,2}, Shaohua Xiao¹³, Andrew C. Yang^{1,2}, Suchen Zheng^{1,2}, Michael J. Gandal^{26,27,28,29,30}, Donghoon Lee^{9,10,11,12}, Ed S. Lein^{8,31,32}, Panos Roussos^{9,10,11,12,18,19}, Nenad Sestan²¹, Zhiping Weng²³, Kevin P. White³³, Hyejung Won³⁴, Matthew J. Girgenti^{25,35,36*}, Jing Zhang^{5*}, Daifeng Wang^{3,4,37*}, Daniel Geschwind^{13,20,27,28,38*}, Mark Gerstein^{1,2,7,24,39*}

Single-cell genomics is a powerful tool for studying heterogeneous tissues such as the brain. Yet little is understood about how genetic variants influence cell-level gene expression. Addressing this, we uniformly processed single-nuclei, multiomics datasets into a resource comprising >2.8 million nuclei from the prefrontal cortex across 388 individuals. For 28 cell types, we assessed population-level variation in expression and chromatin across gene families and drug targets. We identified >550,000 cell type-specific regulatory elements and >1.4 million single-cell expression quantitative trait loci, which we used to build cell-type regulatory and cell-to-cell communication networks. These networks manifest cellular changes in aging and neuropsychiatric disorders. We further constructed an integrative model accurately imputing single-cell expression and simulating perturbations; the model prioritized ~250 disease-risk genes and drug targets with associated cell types.

Genetic variants linked to neuropsychiatric disorders affect brain functions on multiple levels, from gene expression in individual cells to complex brain circuits between cells (1–3). At every level, they manifest themselves differently depending

on the cell type in question. Previously, groups such as GTEx (Genotype-Tissue Expression), the PsychENCODE Consortium, and ROSMAP (Religious Orders Study and Memory and Aging Project) assembled cohorts large enough to link variants to their effects on gene expression in

bulk tissue, generating comprehensive eQTL (expression quantitative trait locus) catalogs for the brain (4–6). Although useful, these tissue-level results do not reflect the specific cell types involved; moreover, they do not provide strong evidence that eQTLs act in cell type-specific fashion (7–10).

Recently, dramatic technological advances have allowed the measurement of gene expression and chromatin accessibility at the single-cell level (11–13). The resulting datasets have shown that the brain has a particularly large number of distinct cell types; cell-type complexity is one of the brain's distinguishing features (12). Many brain cell types have been rigorously defined, particularly by the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative Cell Census Network (BICCN) (12, 14, 15). Using these, we can potentially refine our understanding of how variants and gene regulation affect brain phenotypes, including neuropsychiatric disorders (16). However, up to now we have not had sufficiently large cohorts, including a wide enough range of brain phenotypes, to make statistically meaningful associations between variants, regulatory elements, and expression and to develop comprehensive models of brain gene regulation at the single-cell level.

To address this gap, the PsychENCODE Consortium generated single-cell sequencing data from adult brains with multiple neuropsychiatric disorders in the human prefrontal cortex, using single-nucleus (sn) assays for RNA sequencing (snRNA-seq), snATAC-seq (ATAC, assay for transposase-accessible chromatin), and snMultiome. Leveraging these data and integrating them with other published studies (12, 17–19), we created a uniformly processed single-cell resource at the population level. This resource, which we call brainSCOPE (brain single-cell omics for PsychENCODE), comprises >2.8 million nuclei from 388 individual brains, including 333 newly generated

¹Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA. ²Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. ³Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53706, USA. ⁴Waisman Center, University of Wisconsin-Madison, Madison, WI 53705, USA. ⁵Department of Computer Science, University of California, Irvine, CA 92697, USA. ⁶Mathematical, Computational and Systems Biology, University of California, Irvine, CA 92697, USA. ⁷Department of Computer Science, Yale University, New Haven, CT 06520, USA. ⁸Allen Institute for Brain Science, Seattle, WA 98109, USA. ⁹Center for Disease Neurogenetics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ¹⁰Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ¹¹Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ¹²Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ¹³Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA. ¹⁴Tempus Labs, Chicago, IL 60654, USA. ¹⁵Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. ¹⁶Department of Genetics, Microbiology and Statistics, Universitat de Barcelona, Barcelona 08028, Spain. ¹⁷Department of Biology, Cornell College, Mount Vernon, IA 52314, USA. ¹⁸Mental Illness Research Education and Clinical Center, James J. Peters VA Medical Center, Bronx, NY 10468, USA. ¹⁹Center for Precision Medicine and Translational Therapeutics, James J. Peters VA Medical Center, Bronx, NY 10468, USA. ²⁰Center for Autism Research and Treatment, Semel Institute, University of California, Los Angeles, CA 90095, USA. ²¹Department of Neuroscience, Yale University, New Haven, CT 06510, USA. ²²Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. ²³Department of Genomics and Computational Biology, UMass Chan Medical School, Worcester, MA 01605, USA. ²⁴Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA. ²⁵Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06520, USA. ²⁶Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90095, USA. ²⁷Department of Psychiatry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. ²⁸Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. ²⁹Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. ³⁰Lifespan Brain Institute, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. ³¹Department of Neurological Surgery, University of Washington, Seattle, WA 98195, USA. ³²Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA 98195, USA. ³³Yong Loo Lin School of Medicine, National University of Singapore, 117597 Singapore. ³⁴Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. ³⁵Wu Tsai Institute, Yale University, New Haven, CT 06520, USA. ³⁶Clinical Neuroscience Division, National Center for Posttraumatic Stress Disorder, Veterans Affairs Connecticut Healthcare System, West Haven, CT 06516, USA. ³⁷Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA. ³⁸Institute for Precision Health, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA.

*Corresponding author. Email: matthew.girgenti@yale.edu (M.J.G.); zhang.jing@uci.edu (J.Z.); daifeng.wang@uisc.edu (D.W.); dhg@mednet.ucla.edu (D.G.); mark@gersteinlab.org (M.G.)

†These authors contributed equally to this work.

‡PsychENCODE Consortium authors and affiliations are listed in the supplementary materials.

samples and 55 from external sources (figs. S1 and S2). It enables us to assess 28 distinct brain cell types that can be registered against previously identified canonical cell types (12, 19). Using the resource, we identified an average of ~85,000 cis-eQTLs per cell type and ~550,000 cell type-specific cis-regulatory elements, which were enriched for variants associated with brain-related disorders. Using our regulatory elements and eQTLs, we inferred cell type-specific gene regulatory networks (which show great changes across cell types) as well as cell-to-cell communication networks. Moreover, we precisely quantified expression variation in the population, finding, for instance, that common neurorelated drug targets such as *CNRI* demonstrate a high degree of cell-type variability and low interindividual variability and that the transcriptomes of specific neurons are highly predictive of an individual's age. Lastly, we developed an integrative model to impute cell type-specific functional genomic information for individuals from genotype data alone. Using this model, we prioritized many known and some additional disease genes, now with information about their specific cell type of action. We further associated this prioritization with potential drug targets and simulated the effects of perturbing the expression of particular genes.

All sequencing data, derived analysis files, and computer codes are available from the brainSCOPE resource portal (brainscope.psychencode.org; figs. S3 to S5); these data include gene-expression matrices from snRNA-seq data, regulatory regions from snATAC-seq data, variability metrics for all genes, single-cell QTL callsets, regulatory and cell-to-cell communication networks, and the integrative model and its prioritization outputs.

Constructing a single-cell genomic resource for 388 individuals

We compiled and analyzed population-scale single-cell multiomics data from the human prefrontal cortex (PFC) for a cohort consisting of 388 adults. The individuals in our cohort are diverse in terms of biological sex, ancestry, and age, and include 182 healthy controls as well as individuals with schizophrenia, bipolar disorder, autism spectrum disorder (ASD), and Alzheimer's disease (AD) (Fig. 1A, fig. S1, data S1 and S2, and table S1) (20). We used various filters on the total cohort of 388 for different downstream analyses (fig. S2 and data S3). In total, to build the resource, we uniformly processed 447 snRNA-seq, snATAC-seq, and snMultiome datasets from within PsychENCODE and external studies with >2.8 million total nuclei (after quality control and filtering from a raw number of nearly 4 million; figs. S6 and S7 and table S2) (20). Our processing required harmonizing datasets derived from different technologies and modalities; for instance, we generated uniform genotypes, includ-

ing structural variants, from combining whole-genome sequencing (WGS), single-nucleotide polymorphism (SNP) array, and snRNA-seq data (figs. S1 and S8) (20). We also generated custom datasets to bridge studies, in particular, snMultiome sequencing of controls (20).

We developed a cell-type annotation scheme that harmonizes the BICCN reference atlas (12) and published analyses specifically focusing on the PFC [labeled "Ma-Sestan" here (19); Fig. 1B and figs. S9 to S11 (20)]. In particular, we leveraged the deep sampling of neurons from BICCN and of non-neuronal cells from Ma-Sestan. This resulted in a set of 28 cell subclasses, which we will hereafter refer to as "cell types," most of which are robustly represented across all cohorts (tables S3 and S4). For select downstream analyses that require increased power, we grouped excitatory and inhibitory neuron types into larger "excitatory" and "inhibitory" classes to yield seven major cell groupings. Overall, we assessed a total of 2,557,291 high-quality annotated nuclei from the snRNA-seq data (table S2). We validated our annotation scheme by assessing the expression of key marker genes (Fig. 1C).

Using these datasets, we first calculated cell-type fractions in each sample (figs. S12 to S14 and data S4) (20). Fractions based on raw cell counts in snRNA-seq show great consistency with those inferred from bulk RNA-seq using deconvolution (fig. S12 and data S5 and S6). We further found that some cell types demonstrate cell-fraction differences in neuropsychiatric traits (fig. S13). For example, as previously suggested, the Sst cell fraction is different in individuals with bipolar compared with that found in controls (21, 22) [false discovery rate (FDR) < 0.05, two-sided Welch's *t* test]. To more broadly quantify differences relevant to population-wide traits, we computed lists of cell type-specific differentially expressed (DE) genes for each disorder on the basis of established approaches (23) (figs. S15 to S18 and data S7) (20). Figure 1D shows a representative plot for DE genes in schizophrenia, highlighting many previously known risk genes in a cell type-specific context (24, 25). We also found that individuals with schizophrenia differ from controls with respect to the number of aging DE genes, which may reflect the increased expression variability in schizophrenia patients (Fig. 1E and fig. S19).

Our snRNA-seq data also recapitulates the spatial relationships among cell types in the PFC. Figure 1F shows a cell-trajectory analysis (26, 27) across four subclasses of excitatory neurons in controls. We found smoothed patterns of gene expression variation along the cortical-depth axis [specifically for layers 2 and 3 (L2/3), L4, L5, and L6 intratelencephalic (IT); figs. S20 to S22] (20). These findings expand on previous MERFISH-based results for 258 genes in the mouse motor cortex (MERFISH, multiplexed error-robust fluorescence in situ

hybridization), now showing that cortical depth is related to gene-expression variation for thousands of genes (28, 29). Overall, we found 76 genes with significant variation (FDR < 0.05, Wald test) across cortical layers, including several genes involved in neural development, such as *SEMA6A*, *RUNX2*, *SOX6*, and *PROX1* (figs. S20 to S22, list in table S5, and data S8).

Determining regulatory elements for cell types from snATAC-seq

In addition to snRNA-seq data, our resource contains 59 samples with snATAC-seq data, including 40 snMultiome datasets. After strict quality control, we extracted 273,502 deeply sequenced nuclei, allowing us to learn cell embeddings simultaneously from transcriptomic and epigenetic information (table S1) (20). As a result, we recovered 28 distinct PFC cell types consistent with the snRNA-seq annotation and validated these with the chromatin accessibility of marker genes (Fig. 2, A and B, and figs. S23 and S24). Further, uniform snATAC-seq processing identified a total of 562,098 open-chromatin regions across all datasets, representing a much larger number of regions than those identified in previous brain studies [Fig. 2C and (20)] (2, 30). Following the ENCODE (Encyclopedia of DNA Elements) convention (31), we call these "scCREs" (single-cell candidate cis-regulatory elements). About half of these are cell type-specific and located distal to genes (fig. S25). We validated the functionality of select scCREs using self-transcribing active regulatory region sequencing (STARR-seq) (Fig. 2D) (20, 32).

Using bulk data, we also developed a reference set of >400,000 open-chromatin regions, representing brain-tissue candidate cis-regulatory elements (b-cCREs) (20). The b-cCREs were generated in a fashion comparable to ENCODE cCREs, which are not tissue-specific (31). As expected, they show strong overlap with scCREs (Fig. 2C).

To identify how our cell type-specific regulatory elements relate to genetic associations, we performed a linkage-disequilibrium score regression (LDSC) analysis (20, 33). In general, we found stronger LDSC enrichment for brain phenotypes in b-cCREs compared with those in cCREs (Fig. 2E, fig. S26, data S9 and S10, and table S6). Furthermore, we found additional enrichment when comparing cell type-specific scCREs in excitatory neurons to b-cCREs, highlighting how snATAC-seq allows for better linkage between regulatory regions and brain phenotypes (Fig. 2E) (34–37).

Next, we explored transcription factor (TF) usage across major brain cell types (fig. S27) (20). Figure 2F shows that major brain cell types clearly use distinct TFs. For instance, *CUX1*, *NEUROG1*, and *PAX3* are mostly active in excitatory neurons, whereas *SPL1* and *SPI1* are specific to microglia. We further observed differences between proximal and distal regulation,

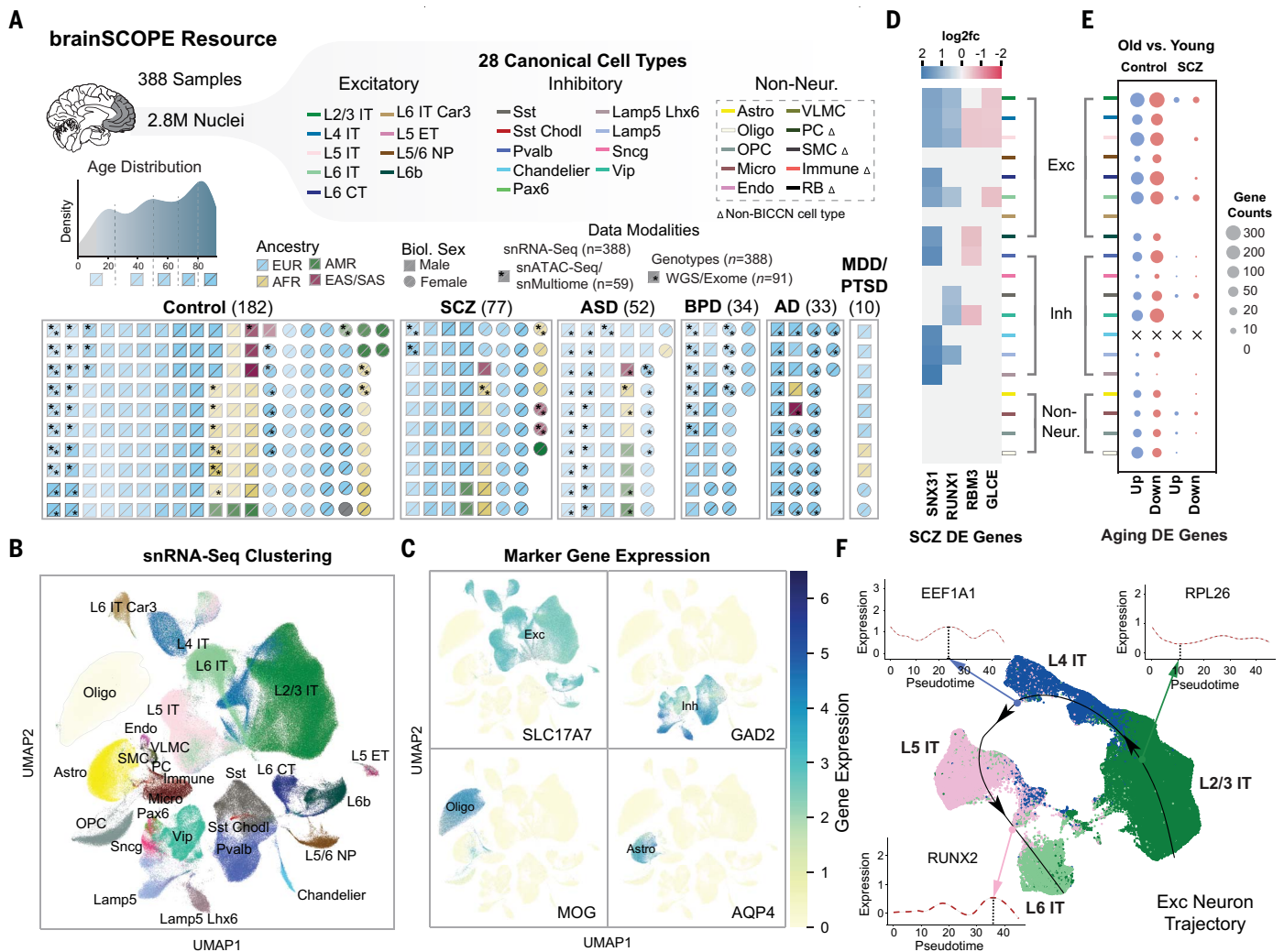


Fig. 1. Constructing a single-cell genomic resource for 388 individuals.

(A) Overview of the integrative single-cell analysis performed on 388 adult prefrontal cortex samples. (Top) Schematic for 28 cell types grouped by excitatory (Exc), inhibitory (Inh), and non-neuronal cell types (table S3); color labels for each subclass are used consistently throughout all figures (table S4). Dashed box indicates cell types defined with Ma-Sestan marker genes (19), with Δ indicating cell types specific to Ma-Sestan. (Bottom) Schematic showing all samples labeled by disease (SCZ, schizophrenia, ASD, autism spectrum disorder, BPD, bipolar disorder, AD, Alzheimer's disease), biological sex, ancestry (EUR, Europeans, AFR, Africans, AMR, Admixed Americans, EAS, East Asians, SAS, South Asians), age, and available data modalities, including a distribution plot for sample ages (gray indicates pediatric samples excluded from most analyses). Astro, astrocytes; Oligo, oligodendrocytes; OPC, oligodendrocyte progenitor cells; Micro, microglial cells; Endo, endothelial cells; VLMC, vascular leptomeningeal cells; PC, pericytes, SMC, smooth muscle cells; RB, red blood lineage cells. (B) Uniform manifold approximation and projection (UMAP) plot for clustering of

28 harmonized cell types from snRNA-seq data derived from 72 samples in the SZBDMulti-seq cohort (using this study as an example of pan-cohort cell typing; see fig. S10 for UMAPs of other studies). (C) UMAP plots highlighting expression of key marker genes in four broad cell types (excitatory: *SLC17A7*; inhibitory: *GAD2*; oligodendrocytes: *MOG*; and astrocytes: *AQP4*). (D) Differential expression (\log_2 -fold change) of four schizophrenia-related genes across cell types in samples from individuals with schizophrenia (blue for up-regulation, red for down-regulation). (E) Numbers of DE genes up-regulated (blue) and down-regulated (red) in older (≥ 70 years) control (left) and schizophrenia (right) individuals per cell type when compared with younger individuals (< 70 years). "X" indicates no DE genes were compared for a particular cell type. (F) UMAP plot showing predicted trajectory for excitatory IT neurons in adult control samples from the SZBDMulti-seq cohort. The predicted trajectory proceeds along the cortical layer dimension from L2/3 to L6 in the prefrontal cortex. (Insets) Log-normalized gene expression is highlighted in cells along the pseudotime axis for three genes.

for example, in *ELF1* (Fig. 2G and data S11). We were able to validate many TF activities with footprinting (38) (Fig. 2H and fig. S27).

Measuring transcriptome and epigenome variation across the cohort at the single-cell level

Single-cell data across a large cohort offers a distinctive opportunity to study the sources

of expression variation in the brain (Fig. 3A and figs. S28 to S30) (20, 39). We partitioned the variation in expression of each gene according to the relative contribution of individual and cell-type variability while correcting for covariates (data S12 and S13). This allowed us to determine relative contributions to variability that were based on the function of each

gene. For example, brain-specific genes, such as those associated with central nervous system (CNS) morphogenesis and neurotransmitter reuptake, demonstrate a high degree of cell-type variability and a lower interindividual variation (Fig. 3, B and C; fig. S31; and data S14). Conversely, genes associated with common molecular or cellular processes tend to

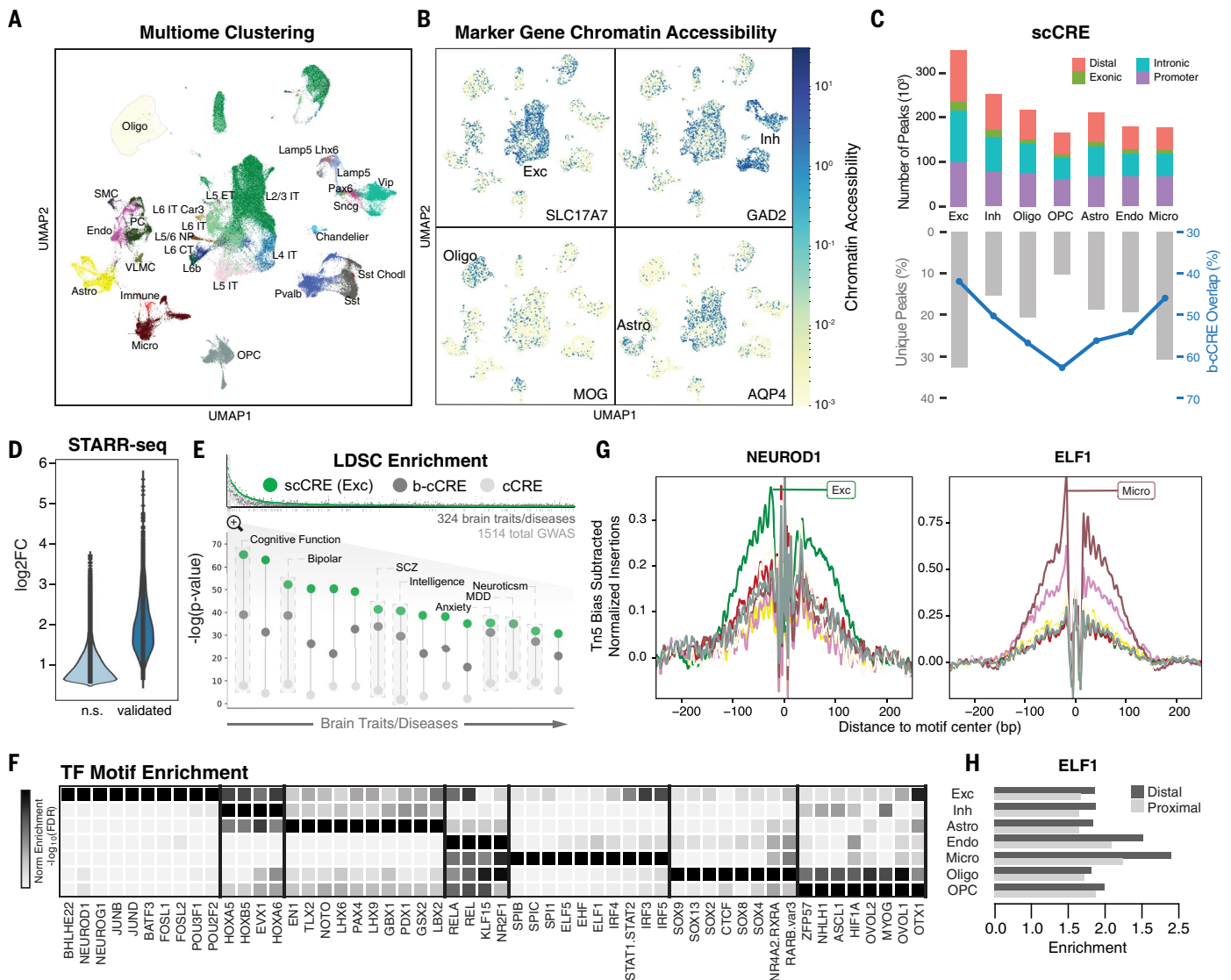


Fig. 2. Determining regulatory elements for cell types from snATAC-seq.

(A) UMAP plot for clustering of 28 harmonized cell types from snMultiome data derived from 40 individuals. (B) UMAP plots highlighting chromatin accessibility of key marker genes for four broad cell types (Fig. 1C). (C) (Top) Counts of open-chromatin regions from combined snATAC-seq and snMultiome peaks across cohorts by gene context (promoter, intronic, exonic, or distal). (Bottom) Percentage of unique ATAC peaks found in each cell type. Blue line indicates the percentage of ATAC peaks that overlap with b-cCREs derived from bulk data. (D) Change in enhancer activity among open-chromatin regions determined using STARR-seq assays of predicted enhancers, comparing the log₂-fold expression change of validated regions with nonvalidated regions (n.s., not

significant). (E) (Top) LDSC enrichment across GWAS summary statistics for UK BioBank traits and diseases, including brain-related traits (gray bars), cCREs (white circles), b-cCREs (gray circles), and snATAC-seq peaks in excitatory neurons (scCREs, green circles). (Bottom) LDSC enrichment [log-scaled *P* values for LDSC analysis as explained in (20)] for select brain traits and disorders. Trait names are listed in table S6. MDD, major depressive disorder. (F) Enrichment (log-scaled FDR) of TF binding motifs among cell type-specific (CTS) snATAC-seq peaks. (G) Differential activity of *ELF1* in proximal and distal regions across cell types. (H) Cell type-specific location of TF binding for *NEUROD1* (left) and *ELF1* (right) across cell types (colors defined in Fig. 1A), based on snATAC-seq footprinting analysis.

have lower cell-type variation and higher individual variation [for instance, carbohydrate homeostasis and adenosine triphosphate (ATP) generation] (Fig. 3B). Furthermore, within families of CNS-specific genes, some neurotransmitter families manifest higher interindividual variation than others (for example, glutamate versus serotonin, *P* value = 3.7×10^{-6} , one-sided *t* test) (Fig. 3C and fig. S31). We also identified a few outliers with very large inter-

individual variation, such as *ARL17B*, likely resulting from copy number variation (40, 41).

An additional application of quantifying expression variability is the characterization of drug-target genes. In particular, we selected 280 common CNS-related drug-target genes and showed that, overall, they have high cell-type variability and low individual-level variability (Fig. 3C and fig. S32A) (42). However, some of the 280 exhibit much higher inter-

individual variation than others; *HSPA5* and *CNR1* provide a good illustration (Fig. 3, C and E, and fig. S32B). Also, two adrenergic receptor family genes, *ADRA1A* and *ADRA1B*, demonstrate high cell-type variation but distinctly different cell-type expression patterns (fig. S32C).

Next, we found that genes with lower expression variability have higher sequence conservation (Fig. 3F and figs. S33 and S34) (20). However, some genes not following this trend

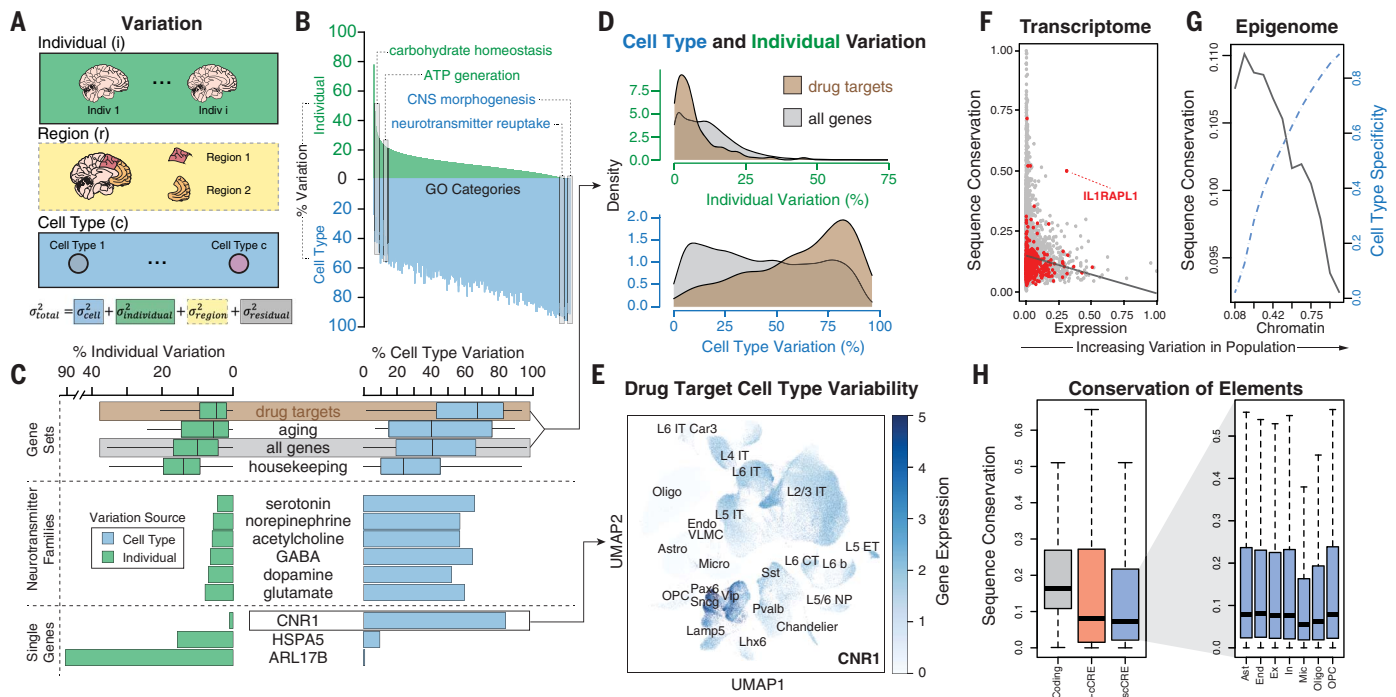


Fig. 3. Measuring transcriptome and epigenome variation across the cohort at the single-cell level. (A) Schematic for the calculation of overall gene-level variance partition by integrating individual, brain region, and cell type-specific variation. Variation analysis using different brain regions (denoted with a dashed yellow box) was performed on a subset of individuals (shown in fig. S28). (B) Percent expression variation attributed to individuals (green) and cell types (blue) for Gene Ontology (GO) categories, with select GO categories highlighted. (C) Percent interindividual and cell-type variation for specific genes and gene sets, including neurotransmitter families and drug targets. (D) Distribution of individual variation and cell-type variation in

drug-target genes versus all genes. (E) UMAP plot of example drug target, *CNR1*, demonstrating cell type-specific expression patterns that contribute to high cell-type variability. We also assessed other genes such as serotonin receptor genes in fig. S31C. (F) Comparison of observed expression variation of individual genes with predicted conservation scores (phastCons). Red dots indicate outlier genes. The black line shows a trend of decreasing conservation as expression variation increases. (G) Increased cell-type specificity (dashed blue line) and decreased conservation (black line) observed as the population variability of scCREs increases. (H) (Left) Conservation of protein-coding regions, b-cCREs, and scCREs. (Right) Conservation of scCREs by cell type.

serve as interesting exceptions (that is, highly conserved genes with high expression variance). The gene deviating most from the trend is *IL1RAPL1* (Fig. 3F and fig. S34B), an interleukin-1 receptor-family gene inhibiting neurotransmitter release (43); *IL1RAPL1* is highly expressed in the brain and has been implicated in intellectual disability and ASD (44).

We also leveraged our snATAC-seq profiles to deconvolve population-scale chromatin data (fig. S33) (20). Similar to the transcriptome, open-chromatin regions with higher sequence conservation have less variability in their chromatin openness (Fig. 3G and fig. S35). Furthermore, an increase in variability is concurrently observed with an increase in cell-type specificity. These patterns held when we jointly considered a gene and its linked upstream regulatory region: A more variably expressed gene is associated with a more variable upstream chromatin region, and both of these are less conserved at the sequence level. (fig. S34A) (20). Lastly, we found that microglia scCREs exhibit the least sequence conservation, which is consistent with previous studies (Fig. 3H) (19, 45, 46).

Determining cell type-specific eQTLs from single-cell data

To evaluate cell-type expression variation in more detail, we used our processed snRNA-seq data to identify single-cell cis-eQTLs (hereafter referred to as “scQTLs”). We followed the same general procedure used by GTEx (5), including conservative filtering at the cell-type level when generating pseudobulk data (20). We used this set of scQTLs as our “core callset,” with the objective of facilitating consistent comparisons with those from existing datasets (such as GTEx and PsychENCODE bulk data) (data S15). The sparsity intrinsic to snRNA-seq data reduces power, particularly for rarer cell types (fig. S36 and table S7) (20, 47). To ameliorate the low power, we developed a Bayesian linear mixed-effects model to identify more scQTLs for rare cell types as an additional callset (Fig. 4A, figs. S36C and S37, and table S8) (20). We also generated further alternative callsets and a merge of results from all approaches (figs. S38 and S39). These callsets include results based on linkage-disequilibrium pruning (table S9) (20), regression across pseudotime trajectories [below and (20)], and conditional analysis (giving rise

to ~1 signal per eGene, where an eGene is a gene involved in an eQTL) (20). Lastly, we identified a limited number of cell type-specific isoform-usage QTLs (iso-QTLs), taking into account limitations in isoform identification from short-read snRNA-seq data (~134,000 candidate iso-QTLs with 1389 associated “isoGenes” (figs. S40 to S42 and data S16) (20).

Overall, we identified an average of ~85K scQTLs and ~690 eGenes per cell type in our core set, resulting in ~1.4 million scQTLs when totaled over cell types (Fig. 4A, figs. S36A and S43, and table S7) (20). Many of the scQTLs are solely found in a single cell type, but ~47% appear in more than one cell type (Fig. 4A and fig. S36). About 30% of the scQTLs overlap with bulk cis-eQTLs (4). Among these “overlappers,” the direction of effect is consistent (Fig. 4B) but the magnitude of the scQTL effect size is greater than that of the matched bulk eQTL (Fig. 4C, fig. S44, and table S10). We posit a “dilution effect” as an explanation, wherein scQTL effect sizes may be diluted in bulk data when they occur only in a relatively small number of cell types. This line of reasoning is supported by comparing scQTLs

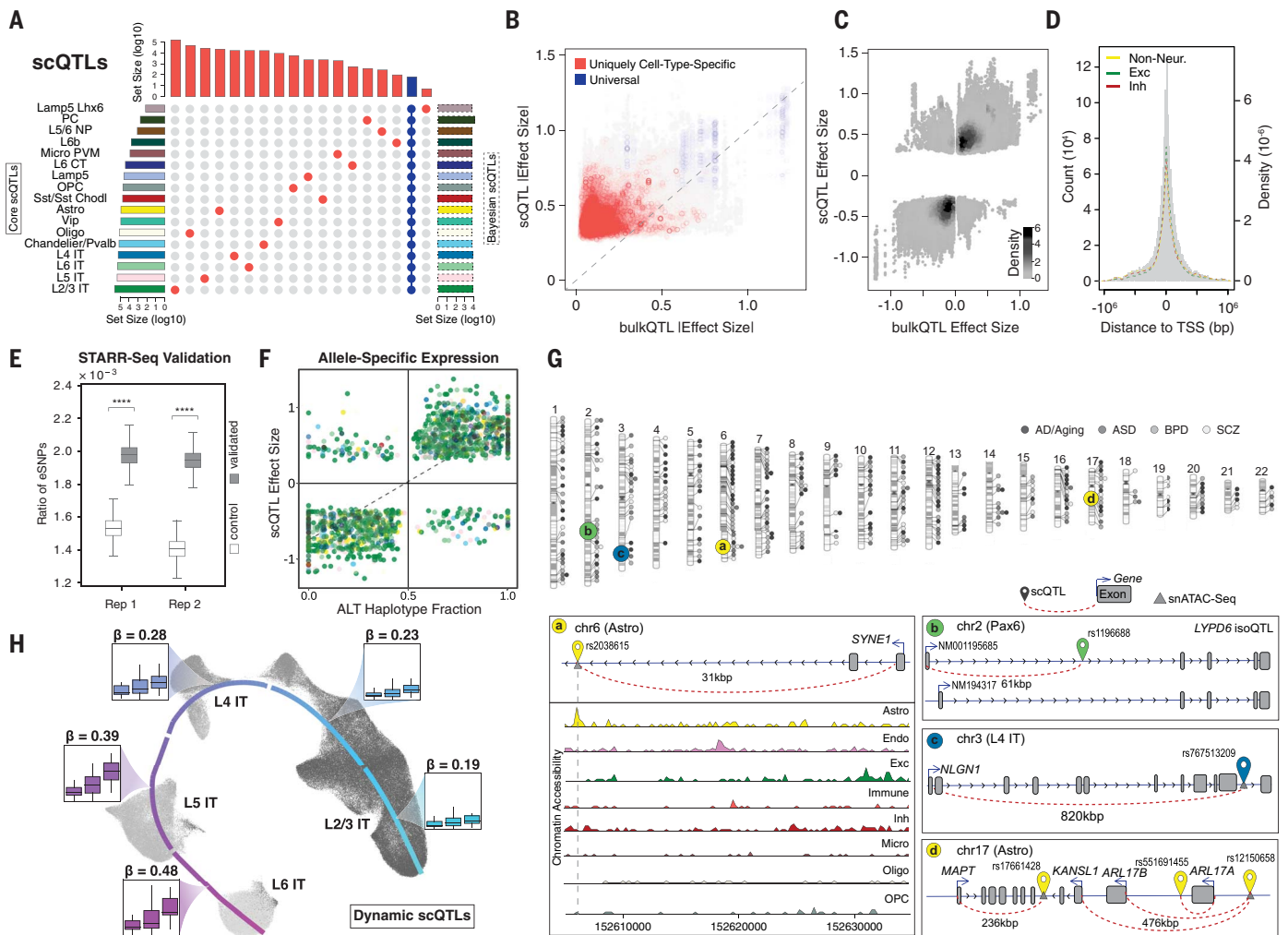


Fig. 4. Determining cell type-specific eQTLs from single-cell data. (A) Partial UpSet plot with identified scQTLs (from the core analysis) that are unique to individual cell types (red) or present across all cell types (blue). Left histogram summarizes the log-scaled total number of core scQTLs per cell type. Right histogram summarizes the log-scaled total number of Bayesian scQTLs per cell type. More complete plots are presented in figs. S36C and S42. (B) Scatter plots comparing absolute eQTL effect sizes between single-cell and bulk RNA datasets, highlighting QTLs shared across >14 cell types (blue) and unique to one cell type (red). (C) Density plot comparing eQTL effect sizes between single-cell and bulk RNA datasets. (D) Histogram with the distribution of scQTLs by distance from eGene transcription start site (TSS), with normalized distributions highlighted for the union of scQTLs across excitatory, inhibitory, and non-neuronal cell types. (E) Box plot showing a significantly higher enrichment of eQTL SNPs (eSNPs) in active STARR-seq peaks compared with the control group ($P < 1.0 \times 10^{-4}$, Mann-Whitney U rank sum test). Two replicates are shown. (F) Scatter plot comparing scQTL effect sizes

with allelic ratios of ASE eGenes, or the fraction of ASE gene reads originating from the haplotype with the scQTL alternative allele. ASE genes were identified in 21 MultiomeBrain cohort samples; cell types are represented with the color scheme used in (A). (G) (Top) Chromosome ideogram for the location of eGenes in all cell types related to four brain disorders. (Bottom) Schematics for specific instances of scQTLs for disease-related eGenes. Left schematic (a) shows astrocyte-specific eSNP for *SYNE1* along with chromatin accessibility (snATAC-seq) tracks for eight cell types. Top-right schematic (b) shows the isoform QTL (iso-QTL) for *LYPD6* in Pax6 inhibitory neurons, leading to altered expression of isoforms with different start codons. Middle (c) and bottom-right (d) schematics show SNP-gene pairs for scQTLs associated with *NLGN1* in L4 IT neurons and *MAPT* in astrocytes, respectively. (H) UMAP plot for predicted trajectory of excitatory neurons in samples from the SZBDMulti-seq cohort. Box plots highlight the expression of *EFCAB13*, stratified by eSNP genotype in each sample, for cell types in each cortical layer; effect size (β) values for the eSNP increase over pseudotime. Additional information is shown in fig. S53.

appearing in a few cell types with those observed in many (Fig. 4B and fig. S36A). Overall, we found cell type-specific QTLs were likely difficult to detect in bulk measurements, which is borne out by the fact that more than two-thirds of our scQTLs are not found in bulk despite much larger sample sizes available in bulk.

Our scQTLs are strongly enriched in narrow regions around the transcription start sites

(Fig. 4D and figs. S45 and S46). We validated some of our core scQTLs by comparing them with functional elements identified by STARR-seq, mutation STARR-seq, and massively parallel reporter assays (MPRA) (Fig. 4E and figs. S47 and S48) (20, 32). As further validation, we were able to identify allele-specific expression (ASE) at the single-cell level in samples with WGS-based phased variants (Fig. 4F and fig. S49) (20). Determination of single-cell ASE is particularly

challenging owing to the sparsity of the data (48–52). In this work, we compared the magnitude of the ASE effect at a single-nucleotide variant (SNV) with the corresponding effect size of the scQTL involving the same SNV, finding significant correlation as expected (Fig. 4F and fig. S49; $P < 2.0 \times 10^{-16}$, Fisher's exact test).

Overall, we identified 330 scQTLs for eGenes related to brain disorders (Fig. 4G, figs. S50 and S51, and data S17). For example, we found

scQTLs for *SYNE1*, a candidate autism and schizophrenia gene (53, 54), and *NLGN1*, a candidate gene for multiple brain disorders encoding a ligand for neuroligin signaling (55). We also found multiple scQTLs within the complex 17q21.31 locus related to brain disorders, including an astrocyte-specific scQTL for the Tau protein gene *MAPT* and a multicell-type scQTL for the neurodegenerative-disorder risk gene *KANS1* (Fig. 4G) (40). We also highlighted an iso-QTL for *LYPD6*, which inhibits acetylcholine-receptor activity in Pax6-type inhibitory neurons (56) (Fig. 4G).

Lastly, we developed a Poisson-regression model that incorporates a continuous trajectory and a pseudotime-genotype interaction term to further expand our scQTLs, allowing for the calculation of “dynamic scQTLs” that exhibit a changing effect size along the pseudotime trajectory (figs. S52 and S53 and data S18

and S19) (20, 57). In particular, for 1692 of the 6255 distinct eGenes in four types of excitatory neurons, we found a corresponding dynamic scQTL (with a nonzero interaction term); an example is shown in Fig. 4H and fig. S52. Moreover, many of these dynamic scQTLs imply widespread QTL effects in cell types where we do not discover a scQTL with our core approach (fig. S52).

Building a gene regulatory network for each cell type

By integrating multiple data modalities, including scQTLs, snATAC-seq, TF-binding sites, and gene coexpression, we constructed gene-regulatory networks (GRNs) for PFC cell types (Fig. 5A, figs. S54 to S58, and data S20) (20). In particular, we linked TFs to potential target genes based on their coexpression relationships from snRNA-seq data (58, 59) and mapped

scQTLs to connect promoters and enhancers (data S21). We have made these networks available in a variety of easy-to-use formats (20). For instance, we applied a network-diffusion method that provides the key regulators of a given target gene—specifically, the aggregate regulatory score of each TF for that target (figs. S59 and S60).

We experimentally validated a subset of these linkages using CRISPR knockouts (Fig. 5B, fig. S61, and data S22) (20). Overall, we found that TF expression in the GRNs explain an average of 52% of the variation in expression of target genes, with merged networks explaining more variance than just the promoter or enhancer connections (Fig. 5C and fig. S62). Additionally, mapping loss-of-function mutations in individuals to select TFs provided further validation by showing the expected change in expression of their target genes in a cell type—

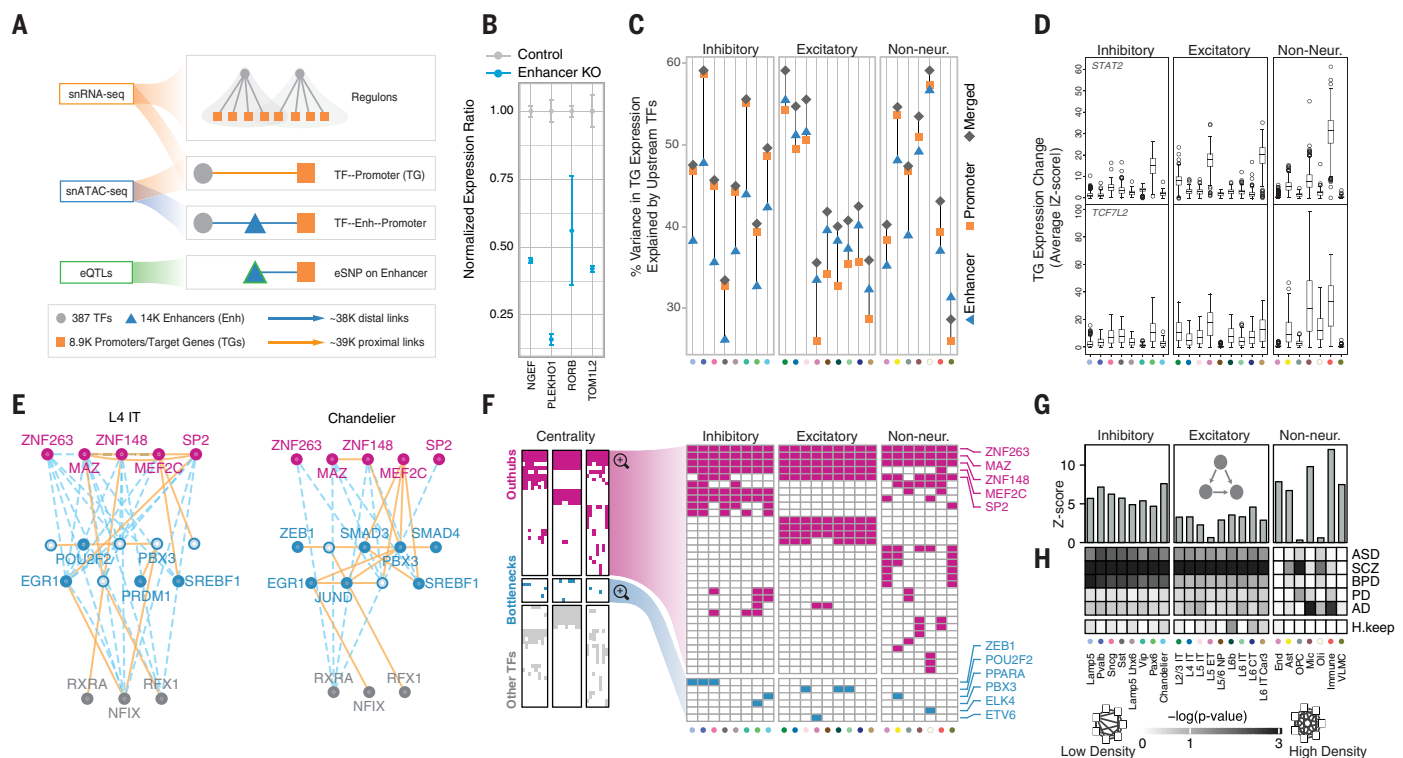


Fig. 5. Building a gene regulatory network for each cell type. (A) Schematic for the construction of cell type-specific GRNs based on snRNA-seq, snATAC-seq, and scQTL datasets. (B) Change in expression of four genes after CRISPR-mediated deletion (KO, knockout) of enhancers identified in cell type-specific GRNs (blue bars) compared with control samples (gray bars). (C) Percent variance in target gene expression explained by the networks. Orange squares, blue triangles, and gray diamonds indicate variance explained by promoter, enhancer, and merged GRNs, respectively. (D) Changes in expression (average z-score) of target genes in cell type-specific regulons among samples with loss-of-function variants that disrupt the TFs *TCF7L2* and *STAT2*. (E) Network graphs depicting a subset of the excitatory (L4 IT) and inhibitory (Chandelier) GRNs that show differential usage of enhancers and promoters. Nodes (TFs) are colored in pink, blue, or gray to represent out-hubs, bottlenecks, and in-hubs, respectively. Nodes without blue fill represent TFs that are absent as bottlenecks in that cell

type. Solid orange lines indicate proximal links; distal links are indicated by dashed blue lines. (F) Panel representing the full set of TFs (y axis) that act as hubs or bottlenecks in different cell types (x axis). Cells are colored if the TF is found to be a pure hub (magenta) or bottleneck (cyan) in the corresponding cell type. (Hubs here are out-hubs.) The right panel zooms in to highlight hubs (top) and bottlenecks (bottom). (G) Motif enrichment analysis bar plots showing a stronger enrichment of transcriptional feed-forward loops (illustrated in inset) in inhibitory neurons (left) and most non-neuronal cell types (right) compared with excitatory neurons (center). (H) Co-regulatory network changes of disease gene sets across cell types. The white-to-black gradient shows low to high probability (log *P* value obtained by random sampling, *n* = 10,000) of a disease gene set or housekeeping genes (H.keep, y axis) forming a dense subnetwork in the corresponding cell type (x axis) (20). Cell types on the x axis in (C), (D), and (F) are individually colored according to names in (G).

specific manner (fig. S63) (20). Overall, 77% of TFs with LOF variants, including *TCF7L2* and *STAT2*, lead to the expected expression alteration within their cell type-specific regulons (Fig. 5D and fig. S63).

Our analyses of GRNs uncovered complex network rewiring across the cell types (Fig. 5E, figs. S64 to S66, and data S23) (20). In particular, the most highly connected TFs (hubs) are largely shared across cell types, suggesting their involvement in common machinery used by all brain cells (Fig. 5F). By contrast, bottlenecks (key connector TFs) have much more cell type-specific activity (Fig. 5F and fig. S67A). Furthermore, the targets of bottleneck TFs are enriched for cell type-specific functions, such as myelination and axon ensheathment for oligodendrocytes (60) (fig. S67B and data S24). Additionally, cell type-specific GRNs greatly differ in the usage of network motifs, such as feed-forward loops (Fig. 5G). These particular motifs, which are thought of as a noise-filtering mechanism (61), are notably enriched in certain non-neuronal cell types.

Lastly, disease genes for a particular disorder tend to be co-regulated in a cell type-specific manner (Fig. 5H and figs. S65 and S68) (20). For instance, gene sets related to schizophrenia form relatively dense subnetworks in neurons, whereas the AD subnetwork is actively co-regulated just in microglia and immune cells (fig. S69) (37, 62, 63).

Constructing a cell-to-cell communication network

To further understand cellular signaling and regulation, we leveraged publicly available ligand-receptor pairs (64) in combination with our snRNA-seq data to construct a cell-to-cell communication network (Fig. 6A, tables S11 and S12, and fig. S70) (20). As expected, we observed three broad ligand-receptor usage patterns among excitatory, inhibitory, and glial cell types, indicating that these cell types use distinct signaling pathways in their communication. For instance, in both incoming and outgoing communication, we observed that all nine glial cell types are grouped together according to their ligand-receptor interactions, with growth-factor genes as some of the top contributing ligand-receptor pairs (65–67) (Fig. 6B).

We next explored how cell-cell communication patterns are altered in individuals with neuropsychiatric disorders, finding that they are greatly changed for schizophrenia and bipolar disorder (Fig. 6B; fig. S71, A and B, and fig. S72; and data S25 and S26). Notable intermixings occur among the three broad patterns of ligand-receptor usage. For instance, in bipolar disorder, the excitatory pattern (inferred from controls) now also contains oligodendrocyte precursor cells (OPCs) and some inhibitory neurons (Pvalb and Sst Chodl). In individuals with schizophrenia (compared with controls),

we also found that excitatory neurons received less incoming signaling, whereas inhibitory neurons received more (Fig. 6C).

To further highlight network perturbations in disease, we assessed signaling-pathway changes for bipolar disorder and schizophrenia (Fig. 6D). In bipolar, we observed down-regulation of the Wnt pathway, which is consistent with previous findings (Fig. 6D) (68–71). Mechanistically, this down-regulation could result in the over-activity of the lithium-targeted GSK3 β enzyme in neurons (72, 73). In schizophrenia, the Wnt pathway is down-regulated as expected, but we also found increased sender communication strength for L6 IT Car3 neurons, different from what is found in bipolar (74). We further found down-regulation of pleiotrophin (PTN) pathway interactions from glial cells to neurons, which is consistent with previous studies (75–77), and a decrease in signaling to glial cells involving various growth factors (fibroblast, epidermal, and insulin) (fig. S71, C and E). These findings support the “glial cell hypothesis,” which posits that deleterious effects on glial cells cascade to neurons (78).

Lastly, we extended our extracellular cell-to-cell communication analysis by considering related disruptions to intracellular signaling pathways (Fig. 6E and fig. S73) (20, 79). By utilizing disease-risk genes and setting support cells (non-neurons) as the senders and neurons as the receivers, we identified ligand-receptor links connecting known risk genes to potential upstream effectors. For instance, we linked *FOXPI* and its ligand *EBI3* in bipolar disorder and *MECP2* and its ligand *PDGFB* in schizophrenia (80, 81).

Assessing cell type-specific transcriptomic and epigenetic changes in aging

We used our population-scale single-cell data to systematically highlight transcriptomic and epigenetic changes due to aging. First, we assessed cell-fraction changes according to deconvolution of bulk data using our single-cell profiles and found that Chandelier and OPC cell types decrease with age, as in previous reports (FDR < 0.05, two-sided *t* test; Fig. 7A and data S27) (82, 83). This result is consistent with findings from raw cell counts in the single-cell data (FDR < 0.05, two-sided *t* test; Fig. 7A and data S27) (20). Next, we identified a list of aging DE genes across cell types (Fig. 7B, fig. S74, and data S28) (20). This list shows, for instance, that *HSPBI*, which encodes a heat-shock protein and has been previously implicated in longevity, is up-regulated in multiple cell types in older individuals (84, 85).

To further explore the relationship between the transcriptome and aging, we constructed a model to predict an individual’s age from their single-cell expression data (Fig. 7C and fig. S75, A and B) (20). The model shows that the transcriptomes of six cell types (L2/3 IT,

L4 IT, L5 IT, L6 IT, oligodendrocytes, and OPCs) have strong predictive value (Fig. 7C and fig. S75C). It also shows that many individual genes contribute to the model, highlighting broad transcriptome changes in aging. From these, we selected two particularly predictive genes previously associated with aging, *FKBP5* and *MKRN3*, and observed a clear correlation between their expression and aging (Fig. 7C and fig. S76) (86–88).

We also investigated the effects of age on the epigenome using our scCREs to deconvolve bulk chromatin accessibility for 628 individuals into those for specific cell types (Fig. 7D and fig. S77). The resulting scCRE activity patterns in certain cell types, particularly microglia, cluster individuals into distinct age groups (Fig. 7D and fig. S77) (20). We further expanded our analysis to highlight how patterns of enriched TF motifs in active scCREs change with age in a cell type-specific fashion (Fig. 7E and fig. S78) (20). Some TFs demonstrate consistent patterns across cell types (*FOXO4* and *RXR α*), whereas others exhibit more cell type-specific patterns (*NEUROG1*).

Lastly, we extended our analysis to identify cell type-specific changes in neurodegenerative disease. We obtained cell-type fractions by using our single-cell expression profiles to deconvolve 638 bulk RNA-seq samples, containing AD cases and controls (fig. S79A) (20, 89). Certain glial fractions show a significant increase in AD ($P < 0.005$, *t* test), whereas several neuronal fractions decrease, especially Sst, Pvalb, and L2/3 IT, in line with previous studies (90) (Fig. 7F). We compared this result with that from directly comparing cell type-specific gene expression and methylation signatures to determine case-control status (91), finding that the fractions and signatures capture independent information (fig. S79B and data S29) (20).

Imputing gene expression and prioritizing disease genes across cell types with an integrative model

We incorporated many of the preceding single-cell datasets and derived networks into an integrative framework to model and interpret the connections between genotype and phenotype. We term our modeling framework a Linear Network of Cell Type Phenotypes (LNCTP; Fig. 8A) (20). This framework serves four tasks: (i) to impute cell type-specific and bulk-tissue gene expression from genotype, (ii) to predict the risk of disorders on the basis of input genotypes, (iii) to highlight genes and pathways contributing to particular phenotypes in their specific cell type of action, and (iv) to simulate perturbations of select genes and quantify their impact on overall gene expression or trait propensity. The LNCTP has several visible layers associated with components of the resource described above, including genotypes at scQTL

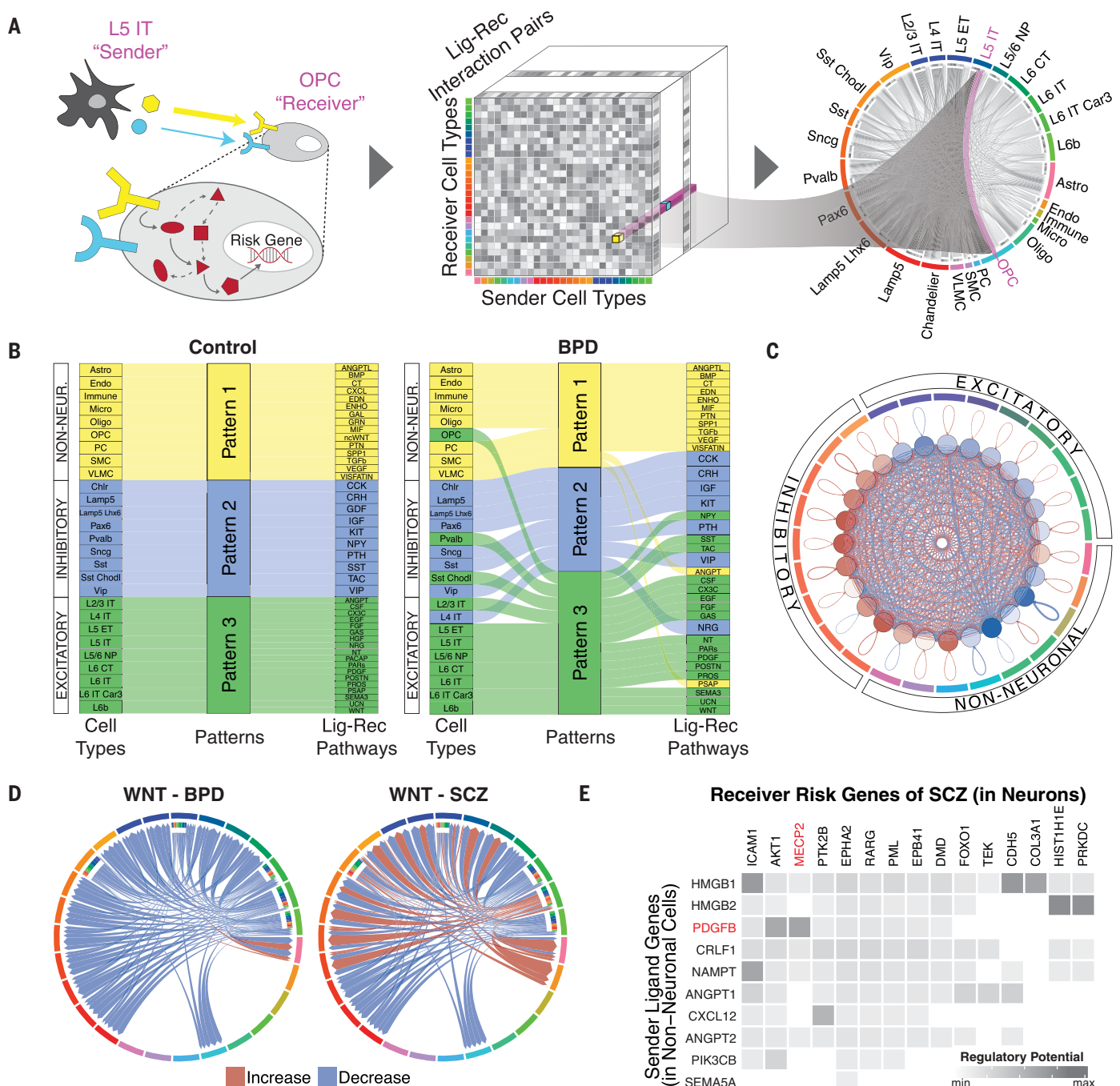
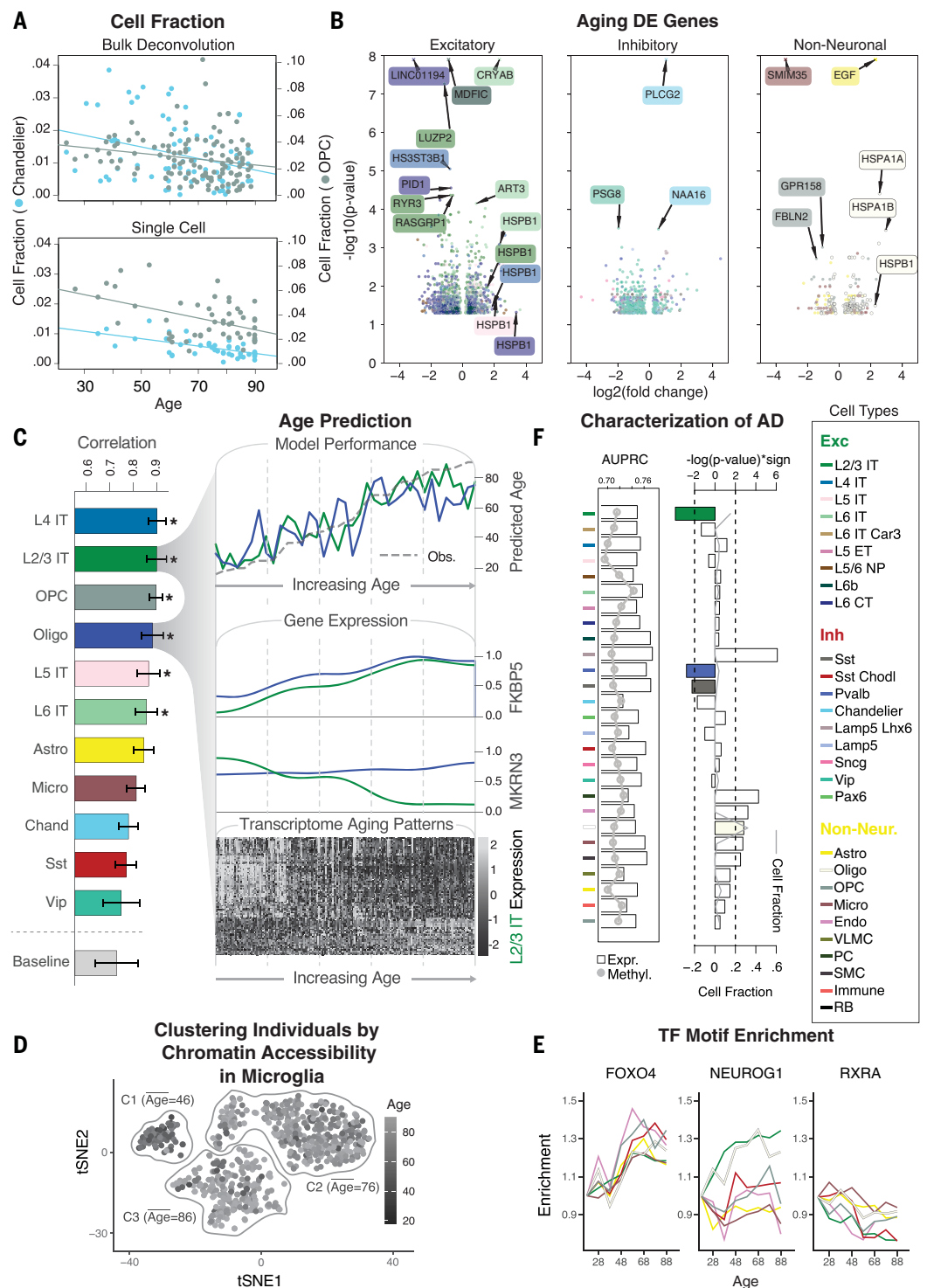


Fig. 6. Constructing a cell-to-cell communication network. (A) Schematic for the construction of cell-to-cell communication networks, based on a matrix of coexpressed ligand-receptor (Lig-Rec) gene pairs in signaling pathways between sender and receiver cell types. (Right) Circos plot shows the strength of all identified cell-to-cell interactions, highlighting L5 IT to OPC cell types as an example. This model does not consider the synaptic connectivity between neurons. (B) Sankey plots for differential clustering of incoming interactions in receiver cells across cell types and ligand-receptor signaling pathways for control (left) and bipolar disorder (right) samples. For example, inhibitory Sst and Sst Chodl cell types were assigned to Pattern 2 in controls, along with the SST-SSTR signaling pathway. This makes sense because these cell types are predominantly

characterized by the SST gene. However, in BPD samples, Sst Chodl cells switched from Pattern 2 to 3, along with the SST-SSTR signaling pathway. (C) Circos plot showing differential strength of all cell-to-cell interactions between individuals with schizophrenia and control individuals. Red edges indicate increased interaction strength in schizophrenia samples; blue edges indicate weaker interaction strength. (D) Circos plots showing changes in cell-to-cell interaction strengths for ligand-receptor genes in the Wnt signaling pathway between individuals with bipolar disorder (left) and schizophrenia (right) compared with control individuals. (E) Predicted likelihoods that ligand genes in non-neuronal cells (y axis) regulate schizophrenia-associated risk genes (x axis) in neuronal cell types, with the neurological risk gene *MECP2* highlighted in red.

Fig. 7. Assessing cell type-specific transcriptomic and epigenetic changes in aging.

(A) Normalized changes in the fraction of OPC (gray) and Chandelier cells (blue) by age, based on bulk RNA-seq deconvolution (top) and single-cell annotation (bottom), with best-fitted lines. **(B)** Log₂-fold changes and *P* values from DESeq2 (20) for DE genes in older versus younger individuals (≥70 years versus <70 years, respectively) among excitatory, inhibitory, and non-neuronal cell types. Values with $-\log(P) > 8$ are shown as crosses. **(C)** (Left) Pearson correlation values between model prediction of age and observed age for each cell type and baseline model (covariates). (Top right) Predicted and observed age for oligodendrocytes and L2/3 IT neurons along the age spectrum. (Middle right) Transcriptomic profiles along the age spectrum of two key genes (*MKRN3* and *FKBP5*) related to aging. (Bottom right) Genes demonstrate an increase (light gray) or decrease (dark gray) in expression along the age spectrum. **(D)** t-distributed stochastic neighbor embedding (tSNE) plot of chromatin peaks showing how chromatin patterns in microglia stratify younger and older individuals into three distinct clusters. **(E)** Examples of TF binding motifs that display distinct enrichment patterns across cell types and age. **(F)** (Left) Predictive accuracy [area under the precision-recall curve (AUPRC)] of cell type-specific expression (bars) and methylation signatures (gray line) toward AD status. (Right) Enrichment of cell-fraction changes among individuals with AD. L2/3 IT, *Pvalb*, and *Sst* (colored bars) are significantly associated with a decreased cell fraction in AD (log-*P* value, *t* test). Gray line shows the overall median cell fraction of each cell type in AD individuals.



and bulk eQTL sites, cell type-specific and bulk tissue-based GRNs, cell-type fractions, cell-to-cell communication networks, gene coexpression modules, and sample covariates (20).

The LNCTP was trained as a conditional energy-based model that represents the joint distribution of the above “visible” variables conditioned on genotype, with additional latent layers (Fig. 8A) (20). It imputes cell type-

specific gene expression from genotype with high cross-validated accuracy: The mean correlation between the imputed and experimentally observed expression profiles is 69% across major cell types and ~78% in excitatory and inhibitory neurons (Fig. 8B). This corresponds to explaining 38% of the variance in cell-type gene expression (or, equivalently, estimating the heritability of cell-type gene expression

h^2), compared with a 34% baseline achieved by combining prior methods for bulk imputation and cell-type deconvolution (20, 92, 93). The baseline does not include our derived GRNs and cell-to-cell networks, so the improvement represents the additional predictive performance possible with these networks (Fig. 8C). Moreover, the inclusion of imputed single-cell gene-expression data also improves the overall

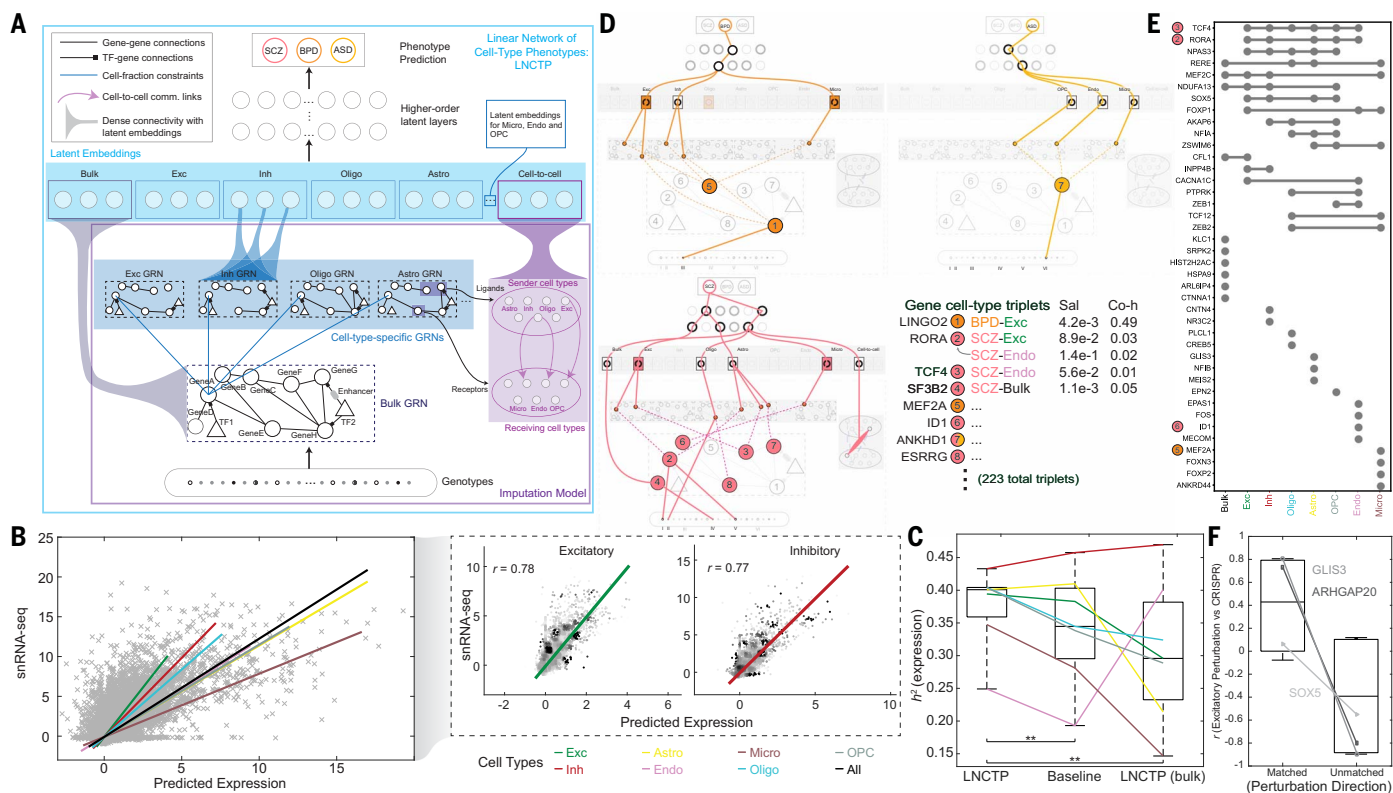


Fig. 8. Imputing gene expression and prioritizing disease genes across cell types with an integrative model. (A) LNCTP schematic. Bulk and cell-type gene-expression levels were imputed from genotype by using a conditional energy-based model incorporating GRNs and cell-to-cell networks. Cell type-specific nodes with dense connectivity were then incorporated into a deep linear model to predict phenotypes in each sample and prioritize cell types and genes for each trait. (B) (Left) Imputed single-cell expression values from LNCTP compared with observed snRNA-seq expression values, with best-fit lines for all cells and individual cell types. (Right) Correlations among imputed expression values for genes in excitatory and inhibitory neurons with best-fit lines. (C) Comparison of explained variance in gene expression from the LNCTP model with a baseline model, using deconvolved, imputed bulk expression data and a model that includes only bulk expression data. Colored lines indicate the performance of individual cell types in each model (** $P < 0.01$, two-tailed paired t test over gene:cell-type pairs). (D) Schematic for LNCTP model interpretation, showing relationships between prioritized intermediate phenotypes for schizophrenia (SCZ, pink), bipolar disorder

(BPD, orange) and ASD (light orange). Gene:cell-type:disease triplets are associated with salience (Sal) and coheritability (Co-h) values ($*P < 0.05$, $**P < 0.01$, two-tailed t test; data S30). Significant cell type, GRN, and cell-to-cell associations are shown at the latent-embedding layers ($P < 0.05$, two-tailed t test; data S31 and S32). Tree structures connect representative subgraphs (feature combinations) in each model (figs. S80 and S81). The schematic also highlights QTL variants linked to the prioritized genes: (I) eQTL (bulk) chr15:60578052, (II) scQTL (Oligo) chr1:216891970, (III) eQTL (bulk) chr9:27902874, (IV) scQTL (Oligo) chr11:66017740, (V) eQTL (bulk) chr15:61553688, and (VI) scQTL (Astro) chr3:158668177. (We shorten the readthrough transcript ANKHD1-EIF4EBP3 to ANKHD1 in ASD) chr, chromosome. (E) UpSet plot for SCZ showing overlap between genes with the highest saliency per cell type or bulk expression, including four genes highlighted in (D) (colored circles). (F) Pearson correlations of LNCTP (excitatory neurons in SCZ) and CRISPR perturbation vectors for three example genes, when perturbation directions are matched versus unmatched, and correlations are calculated across imputed genes ($*P < 0.05$, one-tailed t test; fig. S88).

prediction accuracy of disorders (discussed below) and accounts for a larger fraction of common-SNP heritability of these disorders beyond predictions based solely on bulk expression or polygenic risk scores (94) (table S13).

We exploited the ability of the LNCTP to impute missing data for discovery of cell type-specific molecular phenotypes important for neuropsychiatric disorders. Doing so allowed us to link variants with their “intermediate” functional genomic activities, such as cell type-specific gene expression, pathway activity, and cell-cell communication. We used a hierarchical linear architecture for the trait-prediction portion of the LNCTP, which performed comparably to or better than nonlinear architectures (tables S14 and S15) (20). Moreover, the

LNCTP generates a model that is directly interpretable at multiple scales, avoiding many of the difficulties arising in the interpretation of deep neural networks, while maintaining a hierarchical structure. Our linear architecture allowed us to prioritize intermediate phenotypes by both gradient-based saliency, a metric directly derived from weights in the model, and co-heritability, which directly compares the genetic components of two traits. For instance, we can use the LNCTP to calculate the coheritability of the genetic component of a particular gene’s cell type-specific expression with respect to schizophrenia or other disorders (fig. S80) (20).

Figure 8D and fig. S81 provide an overview of key prioritized genes, cell types, and cell-to-

cell interactions in various disorders (full lists are provided in data S30 to S32). We found 64, 51, 108, and 34 pairs of genes and cell types for schizophrenia, bipolar disorder, ASD, and AD, respectively (20). In particular, *TCF4*, the first identified cross-psychiatric disorder locus (95), is important for neurons in schizophrenia (96); *LINGO2* is important for excitatory neurons in bipolar disorder; and *ANKHD1* is highly weighted in ASD, supporting current hypotheses (97, 98). Figure 8E shows the associated cell types for the most highly prioritized genes. For example, *RORA* is important in many cell types for schizophrenia (but is, nevertheless, not prioritized in the bulk data) (20). It is associated with retinoic-acid signaling, which has been proposed to be an important determinant

of schizophrenia and bipolar risk (99). Furthermore, the retinoic-acid signaling-associated gene *ESRRG* is prioritized in oligodendrocytes (Fig. 8D).

Overall, prioritized genes associated with bulk expression exhibit only a modest overlap with the prioritized cell type-specific genes, indicating that integration of single-cell data in the LNCTP permits the prioritization of distinct genes compared with those found with bulk data alone (Fig. 8E). Moreover, as expected, the prioritized genes are enriched for cell type-specific scQTLs, disease DE genes, and brain-related functional categories (figs. S82 and S83). They are also enriched for prior GWAS and literature support as well as bottleneck locations in the regulatory network (figs. S84 and S85 and data S33). However, several genes specifically prioritized by the LNCTP are not differentially expressed for their respective disorders, including *MEF2A* and *IDI*, perhaps highlighting that they act through network effects (figs. S83 and S85) (100).

In terms of cell types, excitatory neurons and microglia are prioritized in schizophrenia and bipolar disorder, supporting their importance for conferring genetic risk (101), with oligodendrocytes also prioritized in bipolar disorder. Moreover, in schizophrenia, we observed an increase in cell-to-cell interactions between excitatory neurons and microglia as well as a decrease between microglia and oligodendrocytes, consistent with the known glial dysregulation in the disease (Fig. 8D) (78).

We further used the LNCTP to perform in silico perturbation analysis, in which we perturbed a specific gene's expression and observed the induced expression changes in other genes (and the ensuing changes in trait propensity). Perturbations of our prioritized genes as well as known drug targets [retrieved through DrugBank (102)] induce overall expression changes strongly characteristic of case status (fig. S86, A and B). As expected, the induced changes more strongly impact genes in close proximity to the perturbed gene in the GRNs (fig. S87 and table S16). We synthesized the perturbations into a workflow to suggest potential drugs for repurposing with CLUE (42) by matching a perturbation's effects to drugs inducing changes potentially complementary to those found in a particular disorder (fig. S86C and table S17) (20).

To independently validate the results of our simulated perturbation analysis, we used data from CRISPR perturbations (CRISPRi and CRISPRa) applied to specific genes in glutamatergic neurons (103). Induced gene-expression changes resulting from the CRISPR perturbations are more highly correlated with those resulting from LNCTP perturbations when the direction of the perturbation is matched (versus not matched; Fig. 8F, figs. S88 and S89, and table S18) (20). Furthermore, they are more aligned with the direction of case-control DE

for LNCTP-prioritized genes than for non-prioritized ones (fig. S90). Although more comprehensive validation is essential, these results offer promising indications that LNCTP can find verifiable prioritizations of pairs of genes and cell types.

Discussion

In this study, we used population-scale multiomic data to build a comprehensive single-cell functional genomics resource (brainSCOPE) for investigating brain disorders in adults (figs. S3 to S5) (20). The resource can be summarized at multiple levels: (i) raw data and metadata with a harmonized identifier system for each of the individuals; (ii) quantifications of single-cell gene expression (count matrices) with a BICCN-compatible cell-typing system for the PFC; (iii) lists of DE genes and differential cell fractions for various phenotypes; (iv) snATAC-seq signal tracks for various cell types and ENCODE-compatible regulatory elements (b-cCREs and scCREs), including lists of validated ones; (v) the variability for each gene and functional category (by individual, cell type, and brain region) and the associated sequence conservation of genes and regulatory elements; (vi) a core set of GTEx-compatible scQTLs and other additional sets of QTLs (such as dynamic eQTLs); (vii) full GRNs for each cell type, including enhancer-to-gene and TF-to-regulatory element links, and associated files relating each downstream gene to its most important upstream regulators; (viii) cell-to-cell communication networks (expressed as ligand-receptor-by-cell-type matrices); (ix) integrative models with code for imputation, perturbation, and prioritization of cell type-specific functional genomics in brain disease; and (x) the resulting prioritized genes, cell types, and cell-to-cell linkages. The brainSCOPE portal also includes visualizer tools for many of the data types (fig. S4).

The resource allows for several important observations. These include the robustness of cell typing to population variation in 388 individuals and the identification, through shared scQTLs and dynamic scQTLs, of common regulatory programs between cell types. Moreover, by partitioning the observed expression variation, we identified certain drug targets demonstrating high variability between cell types but low variation across individuals (e.g., *CNRI*), a fact that is perhaps key to their therapeutic efficacy. We also found that gene-expression changes in certain neurons and glial cells can accurately predict the age of an individual.

Moreover, a key outcome of our work is providing a set of promising targets for experimental validation. We see these falling into three classes. Class 1 comprises genes that are prioritized by the LNCTP model but not found by traditional DE analysis. This class is ideal for CRISPR assays seeking to test predicted cell-type and phenotypic effects. Other intriguing

candidates are genes that have impacts on cell-to-cell communication spanning multiple cell types (class 2) and genes prioritized in disorders by the LNCTP with further support from DE analysis but lacking prior literature support (class 3). Overall, the LNCTP prioritized gene targets that are consistent with previous findings and also suggested new avenues for investigation. We further used the LNCTP to simulate perturbations and make predictions regarding the effects of known drug-gene interactions on resulting phenotypes—for instance, by perturbing drug-target expression levels. This application will potentially allow for assessing combinations of drugs for targeting multiple genes.

A few limitations should be noted regarding the data used in this study. First, a number of recent works have demonstrated that RNA expression does not completely correlate with protein abundance, and this observation can be even more pronounced in the context of subregions within the brain (104–106). Another related complication is the uncertainty in the extent to which expression in postmortem tissues accurately reflects the expression patterns in live ones (107).

Future efforts could potentially address these limitations. They can also expand our analyses beyond the PFC and integrate functional genomic data from other connecting brain regions (such as the anterior cingulate cortex) to create a comprehensive brain-wide functional genomic atlas. This work could include the incorporation of developmental data as well as experimentally tractable models (such as those from cortical organoids); regulatory network changes over time can then be imputed across developmental axes toward fully mature brain GRNs. We could also incorporate imaging into our integrative model to improve our predictions of brain-associated phenotypes. In addition, more extensive validation of our results would be valuable, such as through targeted CRISPR assays.

Overall, the brainSCOPE resource has the potential to facilitate precision medicine by linking variants to specific cell types and their cell type-specific impacts—for example, to help identify the cell type of action for potential therapies. Through our integrative analyses, we provide an extensive collection of inferences and predictions for neuroscientists to verify in new cohorts, populations, assays, and experimental conditions.

Materials and methods summary

The materials and methods for each section of the main text are available in the supplementary materials, which is organized with the same section headings as in the main text. These include a detailed description of the individuals and datasets assessed in the integrative analysis, protocols used for generating

additional sequencing data and replication experiments for the analysis, and all computational and statistical analysis performed for each part of the integrative analysis.

REFERENCES AND NOTES

- P. F. Sullivan, D. H. Geschwind, Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162–183 (2019). doi: [10.1016/j.cell.2019.01.015](https://doi.org/10.1016/j.cell.2019.01.015); pmid: 30901538
- A. E. Ashley-Koch et al., Revealing the brain's molecular architecture. *Science* **362**, 1262–1263 (2018). doi: [10.1126/science.362.6420.1262](https://doi.org/10.1126/science.362.6420.1262); pmid: 30545881
- M. J. Gandal, V. Leppa, H. Won, N. N. Parikshak, D. H. Geschwind, The road to precision psychiatry: Translating genetics into disease mechanisms. *Nat. Neurosci.* **19**, 1397–1407 (2016). doi: [10.1038/nn.4409](https://doi.org/10.1038/nn.4409); pmid: 27786179
- D. Wang et al., Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018). doi: [10.1126/science.aat8464](https://doi.org/10.1126/science.aat8464); pmid: 30545857
- GTEX Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi: [10.1126/science.aaz1776](https://doi.org/10.1126/science.aaz1776); pmid: 32913098
- B. Ng et al., An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017). doi: [10.1038/nn.4632](https://doi.org/10.1038/nn.4632); pmid: 28869584
- S. Liu et al., Illuminating links between cis-regulators and trans-acting variants in the human prefrontal cortex. *Genome Med.* **14**, 133 (2022). doi: [10.1186/s13073-022-01133-8](https://doi.org/10.1186/s13073-022-01133-8); pmid: 36424644
- J. Bryois et al., Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.* **25**, 1104–1112 (2022). doi: [10.1038/s41593-022-01128-z](https://doi.org/10.1038/s41593-022-01128-z); pmid: 35915177
- S. Kim-Hellmuth et al., Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020). doi: [10.1126/science.aaz8528](https://doi.org/10.1126/science.aaz8528); pmid: 32913075
- B. Zeng et al., Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat. Genet.* **54**, 161–169 (2022). doi: [10.1038/s41588-021-00987-9](https://doi.org/10.1038/s41588-021-00987-9); pmid: 35058635
- K. Zhang et al., A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021). doi: [10.1016/j.cell.2021.10.024](https://doi.org/10.1016/j.cell.2021.10.024); pmid: 34774128
- BRAIN Initiative Cell Census Network (BICCN), A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021). doi: [10.1038/s41586-021-03950-0](https://doi.org/10.1038/s41586-021-03950-0); pmid: 34616075
- C. Luo et al., Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genomics* **2**, 100107 (2022). doi: [10.1016/j.xgen.2022.100107](https://doi.org/10.1016/j.xgen.2022.100107); pmid: 35419551
- H. Zeng, What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022). doi: [10.1016/j.cell.2022.06.031](https://doi.org/10.1016/j.cell.2022.06.031); pmid: 35868277
- G. La Manno et al., Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021). doi: [10.1038/s41586-021-03775-x](https://doi.org/10.1038/s41586-021-03775-x); pmid: 34321664
- M. Song et al., Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* **51**, 1252–1262 (2019). doi: [10.1038/s41588-019-0472-1](https://doi.org/10.1038/s41588-019-0472-1); pmid: 31367015
- P. L. De Jager et al., A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018). doi: [10.1038/sdata.2018.142](https://doi.org/10.1038/sdata.2018.142); pmid: 30084846
- D. Velmeshev et al., Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019). doi: [10.1126/science.aav8130](https://doi.org/10.1126/science.aav8130); pmid: 31097668
- S. Ma et al., Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eaab07257 (2022). doi: [10.1126/science.aba07257](https://doi.org/10.1126/science.aba07257); pmid: 36007006
- Materials and methods are available as supplementary materials.
- H. Pantazopoulos, J. T. Wiseman, M. Markota, L. Ehrenfeld, S. Berretta, Decreased Numbers of Somatostatin-Expressing Neurons in the Amygdala of Subjects With Bipolar Disorder or Schizophrenia: Relationship to Circadian Rhythms. *Biol. Psychiatry* **81**, 536–547 (2017). doi: [10.1016/j.biopsych.2016.04.006](https://doi.org/10.1016/j.biopsych.2016.04.006); pmid: 27259817
- L.-C. Lin, E. Sibille, Reduced brain somatostatin in mood disorders: A common pathophysiological substrate and drug target? *Front. Pharmacol.* **4**, 110 (2013). doi: [10.3389/fphar.2013.00110](https://doi.org/10.3389/fphar.2013.00110); pmid: 24058344
- M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8); pmid: 25516281
- W. B. Ruzicka et al., Single-cell multi-cohort dissection of the schizophrenia transcriptome. *Science* **384**, eadg5136 (2024).
- P. Karpiński, J. Samochowicz, M. M. Szaśniadek, L. Łaczmański, B. Misiak, Analysis of global gene expression at seven brain regions of patients with schizophrenia. *Schizophr. Res.* **223**, 119–127 (2020). doi: [10.1016/j.schres.2020.06.032](https://doi.org/10.1016/j.schres.2020.06.032); pmid: 32631700
- K. Street et al., Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018). doi: [10.1186/s12864-018-4772-0](https://doi.org/10.1186/s12864-018-4772-0); pmid: 29914354
- K. Van den Berge et al., Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020). doi: [10.1038/s41467-020-14766-3](https://doi.org/10.1038/s41467-020-14766-3); pmid: 32139671
- M. Zhang et al., Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021). doi: [10.1038/s41586-021-03705-x](https://doi.org/10.1038/s41586-021-03705-x); pmid: 34616063
- R. Fang et al., Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022). doi: [10.1126/science.abm1741](https://doi.org/10.1126/science.abm1741); pmid: 35771910
- J. Bryois et al., Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, 3121 (2018). doi: [10.1038/s41467-018-05379-y](https://doi.org/10.1038/s41467-018-05379-y); pmid: 30087329
- ENCODE Project Consortium et al., Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020). doi: [10.1038/s41586-020-2493-4](https://doi.org/10.1038/s41586-020-2493-4); pmid: 32728249
- S. C. Gaynor et al., Validation of enhancer regions in primary human neural progenitor cells using capture STARR-seq. NIMH Data Archive (2023); <https://doi.org/10.7303/SYN509003021>.
- C. Sudlow et al., UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015). doi: [10.1371/journal.pmed.1001779](https://doi.org/10.1371/journal.pmed.1001779); pmid: 25826379
- D. Polioudakis et al., A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785–801.e8 (2019). doi: [10.1016/j.neuron.2019.06.011](https://doi.org/10.1016/j.neuron.2019.06.011); pmid: 31303374
- E. K. Ruzzo et al., Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850–866.e26 (2019). doi: [10.1016/j.cell.2019.07.015](https://doi.org/10.1016/j.cell.2019.07.015); pmid: 31398340
- C. L. Hartl et al., Coexpression network architecture reveals the brain-wide and multi-regional basis of disease susceptibility. *Nat. Neurosci.* **24**, 1313–1323 (2021). doi: [10.1038/s41593-021-00887-5](https://doi.org/10.1038/s41593-021-00887-5); pmid: 34294919
- B. Hu et al., Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat. Commun.* **12**, 3968 (2021). doi: [10.1038/s41467-021-24243-0](https://doi.org/10.1038/s41467-021-24243-0); pmid: 34172755
- J. M. Granja et al., ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021). doi: [10.1038/s41588-021-00790-6](https://doi.org/10.1038/s41588-021-00790-6); pmid: 33633365
- N. Johansen et al., Interindividual variation in human cortical cell type abundance and expression. *Science* **382**, eadfc359 (2023). doi: [10.1126/science.adf359](https://doi.org/10.1126/science.adf359); pmid: 37824649
- Y. A. Cooper et al., Functional regulatory variants implicate distinct transcriptional networks in dementia. *Science* **377**, eabi8654 (2022). doi: [10.1126/science.abi8654](https://doi.org/10.1126/science.abi8654); pmid: 35981026
- N. L. Jorstad et al., Comparative transcriptomics reveals human-specific cortical features. *Science* **382**, eade9516 (2023). doi: [10.1126/science.ade9516](https://doi.org/10.1126/science.ade9516); pmid: 37824638
- A. Subramanian et al., A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452.e17 (2017). doi: [10.1016/j.cell.2017.10.049](https://doi.org/10.1016/j.cell.2017.10.049); pmid: 29195078
- F. Gambino et al., IL1-receptor accessory protein-like 1 (ILIRAP1), a protein involved in cognitive functions, regulates N-type Ca²⁺-channel and neurite elongation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9063–9068 (2007). doi: [10.1073/pnas.0701133104](https://doi.org/10.1073/pnas.0701133104); pmid: 17502602
- C. Montani et al., The X-Linked Intellectual Disability Protein ILIRAP1 Regulates Dendrite Complexity. *J. Neurosci.* **37**, 6606–6627 (2017). doi: [10.1523/JNEUROSCI.3775-16.2017](https://doi.org/10.1523/JNEUROSCI.3775-16.2017); pmid: 28576939
- W. G. Pembroke, C. L. Hartl, D. H. Geschwind, Evolutionary conservation and divergence of the human brain transcriptome. *Genome Biol.* **22**, 52 (2021). doi: [10.1186/s13059-020-02257-z](https://doi.org/10.1186/s13059-020-02257-z); pmid: 33514394
- K. J. Miller et al., Cortical activity during motor execution, motor imagery, and imagery-exome feedback. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4430–4435 (2010). doi: [10.1073/pnas.0913697107](https://doi.org/10.1073/pnas.0913697107); pmid: 20160084
- M. Maria, N. Pouyanfar, T. Örd, M. U. Kaikkonen, The Power of Single-Cell RNA Sequencing in eQTL Discovery. *Genes* **13**, 502 (2022). doi: [10.3390/genes13030502](https://doi.org/10.3390/genes13030502); pmid: 35328055
- Q. Deng, D. Ramsköld, B. Reinius, R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014). doi: [10.1126/science.1245316](https://doi.org/10.1126/science.1245316); pmid: 24408435
- C. Borel et al., Biased allelic expression in human primary fibroblast single cells. *Am. J. Hum. Genet.* **96**, 70–80 (2015). doi: [10.1016/j.ajhg.2014.12.001](https://doi.org/10.1016/j.ajhg.2014.12.001); pmid: 25557783
- W. Mu et al., Airport: Interpretable statistical models for analyzing allelic imbalance in single-cell datasets. *Bioinformatics* **38**, 2773–2780 (2022). doi: [10.1093/bioinformatics/btac212](https://doi.org/10.1093/bioinformatics/btac212); pmid: 35561168
- K. Choi, N. Raghupathy, G. A. Churchill, A Bayesian mixture model for the analysis of allelic expression in single cells. *Nat. Commun.* **10**, 5188 (2019). doi: [10.1038/s41467-019-13099-0](https://doi.org/10.1038/s41467-019-13099-0); pmid: 31729374
- Y. Jiang, N. R. Zhang, M. Li, SCALE: Modeling allele-specific gene expression by single-cell RNA sequencing. *Genome Biol.* **18**, 74 (2017). doi: [10.1186/s13059-017-1200-8](https://doi.org/10.1186/s13059-017-1200-8); pmid: 28446220
- T. W. Yu et al., Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**, 259–273 (2013). doi: [10.1016/j.neuron.2012.11.002](https://doi.org/10.1016/j.neuron.2012.11.002); pmid: 23352163
- S. A. Sloan, B. A. Barres, Mechanisms of astrocyte development and their contributions to neurodevelopmental disorders. *Curr. Opin. Neurobiol.* **27**, 75–81 (2014). doi: [10.1016/j.conb.2014.03.005](https://doi.org/10.1016/j.conb.2014.03.005); pmid: 24694749
- A. M. Craig, Y. Kang, Neurexin-neurologin signaling in synapse development. *Curr. Opin. Neurobiol.* **17**, 43–52 (2007). doi: [10.1016/j.conb.2007.01.011](https://doi.org/10.1016/j.conb.2007.01.011); pmid: 17275284
- D. Kulbatskii et al., Human Three-Finger Protein Lypd6 Is a Negative Modulator of the Cholinergic System in the Brain. *Front. Cell Dev. Biol.* **9**, 662227 (2021). doi: [10.3389/fcell.2021.662227](https://doi.org/10.3389/fcell.2021.662227); pmid: 34631692
- A. Nathan et al., Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120–128 (2022). doi: [10.1038/s41586-022-04713-1](https://doi.org/10.1038/s41586-022-04713-1); pmid: 35545678
- S. Albar et al., SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017). doi: [10.1038/nmeth.4463](https://doi.org/10.1038/nmeth.4463); pmid: 28991892
- T. Jin et al., scGRNomic: A computational pipeline of integrative multi-omics analyses for predicting cell-type disease genes and regulatory networks. *Genome Med.* **13**, 95 (2021). doi: [10.1186/s13073-021-00908-9](https://doi.org/10.1186/s13073-021-00908-9); pmid: 34044854
- I. D. Duncan et al., The adult oligodendrocyte can participate in remyelination. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11807–E11816 (2018). doi: [10.1073/pnas.1808064115](https://doi.org/10.1073/pnas.1808064115); pmid: 30487224
- U. Alon, Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007). doi: [10.1038/nrg2102](https://doi.org/10.1038/nrg2102); pmid: 17510665
- I. E. Jansen et al., Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019). doi: [10.1038/s41588-018-0311-9](https://doi.org/10.1038/s41588-018-0311-9); pmid: 30617256
- H. Keren-Shaul et al., A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276–1290.e17 (2017). doi: [10.1016/j.cell.2017.05.018](https://doi.org/10.1016/j.cell.2017.05.018); pmid: 28602351
- S. Jin et al., Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021). doi: [10.1038/s41467-021-21246-9](https://doi.org/10.1038/s41467-021-21246-9); pmid: 33597522
- E. Savchenko et al., FGF family members differentially regulate maturation and proliferation of stem cell-derived astrocytes. *Sci. Rep.* **9**, 9610 (2019). doi: [10.1038/s41598-019-46110-1](https://doi.org/10.1038/s41598-019-46110-1); pmid: 31270389
- C. J. Carter, Multiple genes and factors associated with bipolar disorder converge on growth factor and stress

- activated kinase pathways controlling translation initiation: Implications for oligodendrocyte viability. *Neurochem. Int.* **50**, 461–490 (2007). doi: [10.1016/j.neuint.2006.11.009](https://doi.org/10.1016/j.neuint.2006.11.009); pmid: [17239488](https://pubmed.ncbi.nlm.nih.gov/17239488/)
67. Q.-L. Cui, W.-H. Zheng, R. Quirion, G. Almazan, Inhibition of Src-like kinases reveals Akt-dependent and -independent pathways in insulin-like growth factor I-mediated oligodendrocyte progenitor survival. *J. Biol. Chem.* **280**, 8918–8928 (2005). doi: [10.1074/jbc.M414267200](https://doi.org/10.1074/jbc.M414267200); pmid: [15632127](https://pubmed.ncbi.nlm.nih.gov/15632127/)
68. J. C. McAfee et al., Systematic investigation of allelic regulatory activity of schizophrenia-associated common variants. *Cell Genomics* **3**, 100404 (2023). doi: [10.1016/j.xgen.2023.100404](https://doi.org/10.1016/j.xgen.2023.100404); pmid: [37868037](https://pubmed.ncbi.nlm.nih.gov/37868037/)
69. A. Muneer, Wnt and GSK3 Signaling Pathways in Bipolar Disorder: Clinical and Therapeutic Implications. *Clin. Psychopharmacol. Neurosci.* **15**, 100–114 (2017). doi: [10.9758/cpn.2017.15.2.100](https://doi.org/10.9758/cpn.2017.15.2.100); pmid: [28449557](https://pubmed.ncbi.nlm.nih.gov/28449557/)
70. R. Santos et al., Deficient LEF1 expression is associated with lithium resistance and hyperexcitability in neurons derived from bipolar disorder patients. *Mol. Psychiatry* **26**, 2440–2456 (2021). doi: [10.1038/s41380-020-00981-3](https://doi.org/10.1038/s41380-020-00981-3); pmid: [33398088](https://pubmed.ncbi.nlm.nih.gov/33398088/)
71. E. M. Wexler, D. H. Geschwind, T. D. Palmer, Lithium regulates adult hippocampal progenitor development through canonical Wnt pathway activation. *Mol. Psychiatry* **13**, 285–292 (2008). doi: [10.1038/sj.mp.4002093](https://doi.org/10.1038/sj.mp.4002093); pmid: [17968353](https://pubmed.ncbi.nlm.nih.gov/17968353/)
72. E. Z. Hoseth et al., Exploring the Wnt signaling pathway in schizophrenia and bipolar disorder. *Transl. Psychiatry* **8**, 55 (2018). doi: [10.1038/s41398-018-0102-1](https://doi.org/10.1038/s41398-018-0102-1); pmid: [29507296](https://pubmed.ncbi.nlm.nih.gov/29507296/)
73. A. J. Valvezan, P. S. Klein, GSK-3 and Wnt Signaling in Neurogenesis and Bipolar Disorder. *Front. Mol. Neurosci.* **5**, 1 (2012). doi: [10.3389/fnmol.2012.00001](https://doi.org/10.3389/fnmol.2012.00001); pmid: [22319467](https://pubmed.ncbi.nlm.nih.gov/22319467/)
74. S. Lovestone, R. Killick, M. Di Forti, R. Murray, Schizophrenia as a GSK-3 dysregulation disorder. *Trends Neurosci.* **30**, 142–149 (2007). doi: [10.1016/j.tins.2007.02.002](https://doi.org/10.1016/j.tins.2007.02.002); pmid: [17324475](https://pubmed.ncbi.nlm.nih.gov/17324475/)
75. I. Panaccione et al., Neurodevelopment in schizophrenia: The role of the wnt pathways. *Curr. Neuropharmacol.* **11**, 535–558 (2013). doi: [10.2174/1570159X113119990037](https://doi.org/10.2174/1570159X113119990037); pmid: [24403877](https://pubmed.ncbi.nlm.nih.gov/24403877/)
76. R. D. McCurdy et al., Cell cycle alterations in biopsied olfactory neuroepithelium in schizophrenia and bipolar I disorder using cell culture and gene expression analyses. *Schizophr. Res.* **82**, 163–173 (2006). doi: [10.1016/j.schres.2005.10.012](https://doi.org/10.1016/j.schres.2005.10.012); pmid: [16406496](https://pubmed.ncbi.nlm.nih.gov/16406496/)
77. J. Xu et al., RNA-Seq analysis implicates dysregulation of the immune system in schizophrenia. *BMC Genomics* **13** (Suppl 8), S2 (2012). doi: [10.1186/1471-2164-13-S8-S2](https://doi.org/10.1186/1471-2164-13-S8-S2); pmid: [23282246](https://pubmed.ncbi.nlm.nih.gov/23282246/)
78. A. F. Terwisscha van Scheltinga, S. C. Bakker, R. S. Kahn, Fibroblast growth factors in schizophrenia. *Schizophr. Bull.* **36**, 1157–1166 (2010). doi: [10.1093/schbul/sbp033](https://doi.org/10.1093/schbul/sbp033); pmid: [19429845](https://pubmed.ncbi.nlm.nih.gov/19429845/)
79. R. Browaeys, W. Saelens, Y. Saeys, NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020). doi: [10.1038/s41592-019-0667-5](https://doi.org/10.1038/s41592-019-0667-5); pmid: [31819264](https://pubmed.ncbi.nlm.nih.gov/31819264/)
80. R. E. Amir et al., Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999). doi: [10.1038/13810](https://doi.org/10.1038/13810); pmid: [10508514](https://pubmed.ncbi.nlm.nih.gov/10508514/)
81. C. Bacon, G. A. Rappold, The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Hum. Genet.* **131**, 1687–1698 (2012). doi: [10.1007/s00439-012-1193-z](https://doi.org/10.1007/s00439-012-1193-z); pmid: [22736078](https://pubmed.ncbi.nlm.nih.gov/22736078/)
82. M. C. Inda, J. Defelipe, A. Muñoz, The distribution of chandelier cell axon terminals that express the GABA plasma membrane transporter GAT-1 in the human neocortex. *Cereb. Cortex* **17**, 2060–2071 (2007). doi: [10.1093/cercor/bhl114](https://doi.org/10.1093/cercor/bhl114); pmid: [17099065](https://pubmed.ncbi.nlm.nih.gov/17099065/)
83. W. E. Allen, T. R. Blosser, Z. A. Sullivan, C. Dulac, X. Zhuang, Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**, 194–208.e18 (2023). doi: [10.1016/j.cell.2022.12.010](https://doi.org/10.1016/j.cell.2022.12.010); pmid: [36580914](https://pubmed.ncbi.nlm.nih.gov/36580914/)
84. C. R. Gomez, Role of heat shock proteins in aging and chronic inflammatory diseases. *Geroscience* **43**, 2515–2532 (2021). doi: [10.1007/s11357-021-00394-2](https://doi.org/10.1007/s11357-021-00394-2); pmid: [34241808](https://pubmed.ncbi.nlm.nih.gov/34241808/)
85. C. Schultz et al., Expression of stress proteins alpha B-crystallin, ubiquitin, and hsp27 in pallido-nigral spheroids of aged rhesus monkeys. *Neurobiol. Aging* **22**, 677–682 (2001). doi: [10.1016/S0197-4580\(01\)00229-9](https://doi.org/10.1016/S0197-4580(01)00229-9); pmid: [11445268](https://pubmed.ncbi.nlm.nih.gov/11445268/)
86. A. P. Abreu et al., Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N. Engl. J. Med.* **368**, 2467–2475 (2013). doi: [10.1056/NEJMoaj302160](https://doi.org/10.1056/NEJMoaj302160); pmid: [23738509](https://pubmed.ncbi.nlm.nih.gov/23738509/)
87. A. S. Zannas et al., Epigenetic upregulation of FKBP5 by aging and stress contributes to NF- κ B-driven inflammation and cardiovascular risk. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11370–11379 (2019). doi: [10.1073/pnas.1816847116](https://doi.org/10.1073/pnas.1816847116); pmid: [31113877](https://pubmed.ncbi.nlm.nih.gov/31113877/)
88. A. S. Zannas, T. Wiechmann, N. C. Gassen, E. B. Binder, Gene-Stress-Epigenetic Regulation of FKBP5: Clinical and Translational Implications. *Neuropsychopharmacology* **41**, 261–274 (2016). doi: [10.1038/npp.2015.235](https://doi.org/10.1038/npp.2015.235); pmid: [26250598](https://pubmed.ncbi.nlm.nih.gov/26250598/)
89. D. A. Bennett et al., Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers Dis.* **64**, S161–S189 (2018). doi: [10.3233/JAD-179939](https://doi.org/10.3233/JAD-179939); pmid: [29865057](https://pubmed.ncbi.nlm.nih.gov/29865057/)
90. C. A. Saura, A. Deprada, M. D. Capilla-López, A. Parra-Damas, Revealing cell vulnerability in Alzheimer's disease by single-cell transcriptomics. *Semin. Cell Dev. Biol.* **139**, 73–83 (2023). doi: [10.1016/j.semdb.2022.05.007](https://doi.org/10.1016/j.semdb.2022.05.007); pmid: [35623983](https://pubmed.ncbi.nlm.nih.gov/35623983/)
91. J. Wang, K. Roeder, B. Devlin, Bayesian estimation of cell type-specific gene expression with prior derived from single-cell data. *Genome Res.* **31**, 1807–1818 (2021). doi: [10.1101/gr.268722.120](https://doi.org/10.1101/gr.268722.120); pmid: [33837133](https://pubmed.ncbi.nlm.nih.gov/33837133/)
92. E. R. Gamazon et al., A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015). doi: [10.1038/ng.3367](https://doi.org/10.1038/ng.3367); pmid: [26258848](https://pubmed.ncbi.nlm.nih.gov/26258848/)
93. Y. H. Wang et al., A CIBERSORTx-based immune cell scoring system could independently predict the prognosis of patients with myelodysplastic syndromes. *Blood Adv.* **5**, 4535–4548 (2021). doi: [10.1182/bloodadvances.2021005141](https://doi.org/10.1182/bloodadvances.2021005141); pmid: [34614508](https://pubmed.ncbi.nlm.nih.gov/34614508/)
94. Brainstorm Consortium et al., Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018). doi: [10.1126/science.aap8757](https://doi.org/10.1126/science.aap8757); pmid: [29931010](https://pubmed.ncbi.nlm.nih.gov/29931010/)
95. Cross-Disorder Group of the Psychiatric Genomics Consortium, Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* **381**, 1371–1379 (2013). doi: [10.1016/S0140-6736\(12\)61219-1](https://doi.org/10.1016/S0140-6736(12)61219-1); pmid: [23453885](https://pubmed.ncbi.nlm.nih.gov/23453885/)
96. J. Gelernter et al., Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. *Nat. Neurosci.* **22**, 1394–1401 (2019). doi: [10.1038/s41593-019-0447-7](https://doi.org/10.1038/s41593-019-0447-7); pmid: [31358989](https://pubmed.ncbi.nlm.nih.gov/31358989/)
97. C. Pisanu et al., Evidence that genes involved in hedgehog signaling are associated with both bipolar disorder and high BMI. *Transl. Psychiatry* **9**, 315 (2019). doi: [10.1038/s41398-019-0652-x](https://doi.org/10.1038/s41398-019-0652-x); pmid: [31754094](https://pubmed.ncbi.nlm.nih.gov/31754094/)
98. M. Chopra et al., Heterozygous ANKRD17 loss-of-function variants cause a syndrome with intellectual disability, speech delay, and dysmorphism. *Am. J. Hum. Genet.* **108**, 1138–1150 (2021). doi: [10.1016/j.ajhg.2021.04.007](https://doi.org/10.1016/j.ajhg.2021.04.007); pmid: [33909992](https://pubmed.ncbi.nlm.nih.gov/33909992/)
99. W. R. Reay et al., Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. *Mol. Psychiatry* **25**, 719–731 (2020). doi: [10.1038/s41380-018-0305-0](https://doi.org/10.1038/s41380-018-0305-0); pmid: [30532020](https://pubmed.ncbi.nlm.nih.gov/30532020/)
100. M. J. Gandal et al., Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018). doi: [10.1126/science.aat8127](https://doi.org/10.1126/science.aat8127); pmid: [30545856](https://pubmed.ncbi.nlm.nih.gov/30545856/)
101. M. Wang, L. Zhang, F. H. Gage, Microglia, complement and schizophrenia. *Nat. Neurosci.* **22**, 333–334 (2019). doi: [10.1038/s41593-019-0343-1](https://doi.org/10.1038/s41593-019-0343-1); pmid: [30796420](https://pubmed.ncbi.nlm.nih.gov/30796420/)
102. D. S. Wishart et al., DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006). doi: [10.1093/nar/gkj067](https://doi.org/10.1093/nar/gkj067); pmid: [16381955](https://pubmed.ncbi.nlm.nih.gov/16381955/)
103. R. Tian et al., Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. *Nat. Neurosci.* **24**, 1020–1034 (2021). doi: [10.1038/s41593-021-00862-0](https://doi.org/10.1038/s41593-021-00862-0); pmid: [34031600](https://pubmed.ncbi.nlm.nih.gov/34031600/)
104. A. L. Bauernfeind, C. C. Babbitt, The predictive nature of transcript expression levels on protein expression in adult human brain. *BMC Genomics* **18**, 322 (2017). doi: [10.1186/s12864-017-3674-x](https://doi.org/10.1186/s12864-017-3674-x); pmid: [28438116](https://pubmed.ncbi.nlm.nih.gov/28438116/)
105. C. P. Moritz, T. Mühlhaus, S. Tenzer, T. Schulenburg, E. Friauf, Poor transcript-protein correlation in the brain: Negatively correlating gene products reveal neuronal polarity as a potential cause. *J. Neurochem.* **149**, 582–604 (2019). doi: [10.1111/jnc.14664](https://doi.org/10.1111/jnc.14664); pmid: [30664243](https://pubmed.ncbi.nlm.nih.gov/30664243/)
106. B. C. Carlyle et al., A multiregional proteomic survey of the postnatal human brain. *Nat. Neurosci.* **20**, 1787–1795 (2017). doi: [10.1038/s41593-017-0011-2](https://doi.org/10.1038/s41593-017-0011-2); pmid: [29184206](https://pubmed.ncbi.nlm.nih.gov/29184206/)
107. L. E. Liharska et al., A study of gene expression in the living human brain. medRxiv [Preprint] (2023). <https://doi.org/10.1101/2023.04.21.23288916>
108. PsychENCODE Consortium (PEC), PsychENCODE Consortium (PEC) Capstone II Cross-study Harmonized Data, version 5. Synapse (2023). <https://doi.org/10.7303/SYNS1111084.1>
109. PsychENCODE Consortium (PEC), PEC Integrative Analysis Processing of ROSMAP data, version 1, Synapse (2024); <https://doi.org/10.7303/SYNS53479857.1>
110. D. Clarke, et al., Single-cell genomics & regulatory networks for 388 human brains, NCBI GEO (2024); <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE261983>
111. R. E. Rodin et al., The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat. Neurosci.* **24**, 176–185 (2021). doi: [10.1038/s41593-020-00765-6](https://doi.org/10.1038/s41593-020-00765-6); pmid: [33432195](https://pubmed.ncbi.nlm.nih.gov/33432195/)
112. P. Emani et al., gersteinlab/PsychENCODE_SingleCell_Integrative: v1.0.0, Zenodo (2024); <https://doi.org/10.5281/ZENODO.10849968>

ACKNOWLEDGMENTS

The authors thank the founder of the Allen Institute, P. G. Allen, for his vision, encouragement, and support. R.T. and M. J. Girgenti thank Keck Microarray Shared Resource (KMSR) and Yale Center for Genome Analysis (YCGA) at Yale University for their assistance with 10x Genomics single-cell RNA-seq services. **Funding:** Data were generated as part of the PsychENCODE Consortium, supported by U01DA048279, U01MH103339, U01MH103340, U01MH103346, U01MH103365, U01MH103392, U01MH116438, U01MH116441, U01MH116442, U01MH116488, U01MH116489, U01MH116492, U01MH122590, U01MH122591, U01MH122592, U01MH122849, U01MH122678, U01MH122681, U01MH116487, U01MH122509, R01MH094714, R01MH105472, R01MH105898, R01MH109677, R01MH109715, R01MH110905, R01MH110920, R01MH110921, R01MH110926, R01MH110927, R01MH110928, R01MH111721, R01MH117291, R01MH117292, R01MH117293, R21MH102791, R21MH103877, R21MH105853, R21MH105881, R21MH109956, R56MH114899, R56MH114901, R56MH114911, R01MH125516, R01MH126459, R01MH129301, R01MH126393, R01MH121521, R01MH116529, R01MH129817, R01MH117406, and P50MH106934 awarded to authors and collaborators A. Abzyov, N. Ahituv, S. Akbarian, K. Brennand, A. Chess, G. Cooper, G. Crawford, S. Dracheva, P. Farnham, M.J.Ga., M.Ge., D.Ge., F. Goes, J. F. Hallmayer, V. Haroutunian, T. M. Hyde, A. Jaffe, P. Jin, M. Kellis, J. Kleinman, J. A. Knowles, A. Kriegstein, C. Liu, C. E. Mason, K. Martinovich, E. Mukamel, R. Myers, C. Nemeroff, M. Peters, D. Pinto, K. Pollard, K. Ressler, P.R., S. Sanders, N.Se., P. Sklar, M. P. Snyder, M. State, J. Stein, P. Sullivan, A. E. Urban, F. Vaccarino, S. Warren, D. Weinberger, S. Weissman, Z.W., K.P.W., A. J. Willsey, H. Won, and P. Zandi. Additional data were provided to the PsychENCODE Consortium, supported by 2015 and 2018 NARSAD Young Investigator grants from Brain & Behavior Research Foundation awarded to collaborator N. Daskalakis. Additionally, P.R. (Icahn School of Medicine at Mount Sinai and James J. Peters VA Medical Center) and J.B. (Icahn School of Medicine at Mount Sinai) were supported by the National Institute of Mental Health (NIMH), NIH grants RFI-MH128970, R01-MH125246, and R01-MH109897, as well as the National Institute on Aging, NIH grants R01-AG050986, R01-AG067025, and R01-AG065582 and by Veterans Affairs Merit grant BX002395. S.G.-G. was supported by grant 5U01MH116489. J.Z. was supported by grants R01HG012572 and R01NS128523. M.J.Ga. was supported by NIMH grant R01-MH123922 as well as NIH/NIMH grants R01MH121521, R01MH124018, and R01MH125252. M.Ge. was supported by NIH NIMH grant U01-MH16492. D.W. and D.W. were supported by NIH NIMH grant U01-MH16492. D.W. was also supported by NIH grants R01AG067025, R1MH128695, R2INS127432, R2INS128761, and P50HD105353 (Waisman Center), National Science Foundation Career Award 2144475, and a Simons Foundation Autism Research Initiative pilot grant 971316. This work was also supported by NIH grant U01MH114812 (E.S.L., N.L.J., and T.E.B.), and by the National Institute on Aging grant U19AG060909 (E.S.L., K.J.T.). The research reported here was supported by the Department of Veterans Affairs, Veteran Health Administration, VISNI Career Development Award, a Brain and Behavior Research Foundation Young Investigator Award, and an American Foundation for Suicide Prevention Young Investigator Award to M.J.G. This work was funded in part by the State of Connecticut, Department of Mental Health and Addiction Services. The views expressed here are those of the authors and do not necessarily reflect the position or policy of the US Department of Veterans Affairs (VA) or the US government or the views of the Department of Mental Health and Addiction Services or the State of Connecticut. **Author contributions:** The following 11 co-first authors contributed equally to this work: P.S.E., J.J.L., D.C., M.J., J.W., C.G., R.M., C.L., S.Xu, C.D., and S.Lou. The following seven co-second authors contributed equally to this work: Y.Chen, Z.C.,

T.G., A.Hw., Y.L., P.N., and X.Z. The following authors contributed substantially to the paper through either data generation or analysis. Data generation: T.E.B., J.B., L.B., L.C., Y.Cheng, M.F., J.F.F., S.G., J.G., N.H., N.L.J., R.K., Ji.L., S.Ma, M.M., S.Maz, J.R.M., D.Q., M.Sh., M.Sp., R.T., K.J.T., B.W., S.Xi., M.J.Ga., E.S.L., P.R., N.Se., K.P.W., and M.J.Gi. Data analysis: P.S.E., J.J.L., D.C., M.J., J.W., C.G., R.M., C.L., S.Xu, C.D., S.Lou, Y.Chen, Z.C., T.G., A.Hw., Y.L., P.N., X.Z., T.E.B., T.C., L.C., Y.D., Z.D., M.Ga., D.Ga., S.G., E.H., G.E.H., A.Hu., Y.J., T.J., N.L.J., S.K., Ju.L., S.Li., J.M., E.N., N.P., M.P., H.P., A.S.R., T.R.R., N.Sh., K.J.T., G.W., Y.X., A.C.Y., S.Z., D.L., E.S.L., P.R., Z.W., H.W., J.Z., D.W., D.Ge., and M.Ge. The following seven senior authors played instrumental roles in this study: M.J.Ga., D.L., E.S.L., P.R., N.Se., Z.W., K.P.W., and H.W. The following five corresponding authors co-led the analysis: M.J.Gi., J.Z., D.W., D.Ge., and M.Ge. **Competing interests:** Z.W. (UMass Chan Medical School) co-founded and serves as a scientific advisor for Rgenta Therapeutics, N.L.J. (Allen Institute for Brain Science) has been an employee of Genentech since 11April 2022. K.P.W. (National University of Singapore) is a shareholder in Tempus AI and

Provaxus Inc. S.Xi. is currently employed by Denali Therapeutics. The other authors declare that they have no competing interests. **Data and materials availability:** The brainSCOPE resource was developed from raw sequencing data (snRNA-Seq, snATAC-Seq, snMultiome, and genotype) derived from 12 individual cohorts, including eight PsychENCODE cohorts and four external cohorts. Raw datasets for the PsychENCODE cohorts, as well as protected-access integrated datasets such as imputed genotypes, are available at the PsychENCODE portal (108). For the external cohorts, Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) Project raw datasets and imputed genotypes are available at the AD Knowledge Portal (109). Girgenti-snMultiome datasets are deposited at NCBI GEO (GSE261983) (110). Ma-Sestan and Velmeshev datasets are available from their respective publications (18, 19, 111). Other key resources and additional datasets used in the integrative analysis are available in the supplementary materials (for smaller datasets) or through the brainSCOPE portal at <http://brainscope.psychencode.org> (for larger datasets) (20). Code used in this manuscript is deposited on GitHub

and linked from the brainSCOPE portal (112). **License information:** Copyright © 2024 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi5199

Consortium Authors and Affiliations
Materials and Methods

Figs. S1 to S90

Tables S1 to S18

References (113–249)

Data S1 to S33

MDAR Reproducibility Checklist

Submitted 6 May 2023; accepted 5 April 2024
10.1126/science.adi5199



Single-cell genomics and regulatory networks for 388 human brains

Prashant S. Emani, Jason J. Liu, Declan Clarke, Matthew Jensen, Jonathan Warrell, Chirag Gupta, Ran Meng, Che Yu Lee, Siwei Xu, Cagatay Dursun, Shaoke Lou, Yuhang Chen, Zhiyuan Chu, Timur Galeev, Ahyeon Hwang, Yunyang Li, Pengyu Ni, Xiao Zhou, PsychENCODE Consortium, Trygve E. Bakken, Jaroslav Bendl, Lucy Bicks, Tanim Chatterjee, Lijun Cheng, Yuyan Cheng, Yi Dai, Ziheng Duan, Mary Flaherty, John F. Fullard, Michael Gancz, Diego Garrido-Martín, Sophia Gaynor-Gillett, Jennifer Grundman, Natalie Hawken, Ella Henry, Gabriel E. Hoffman, Ao Huang, Yunzhe Jiang, Ting Jin, Nikolas L. Jorstad, Riki Kawaguchi, Saniya Khullar, Jianyin Liu, Junhao Liu, Shuang Liu, Shaojie Ma, Michael Margolis, Samantha Mazariegos, Jill Moore, Jennifer R. Moran, Eric Nguyen, Nishigandha Phalke, Milos Pjanic, Henry Pratt, Diana Quintero, Ananya S. Rajagopalan, Tiernon R. Riesenmy, Nicole Shedd, Manman Shi, Megan Spector, Rosemarie Terwilliger, Kyle J. Travaglini, Brie Wamsley, Gaoyuan Wang, Yan Xia, Shaohua Xiao, Andrew C. Yang, Suchen Zheng, Michael J. Gandal, Donghoon Lee, Ed S. Lein, Panos Roussos, Nenad Sestan, Zhiping Weng, Kevin P. White, Hyejung Won, Matthew J. Girgenti, Jing Zhang, Daifeng Wang, Daniel Geschwind, Mark Gerstein, Schahram Akbarian, Alexej Abyzov, Nadav Ahituv, Dhivya Arasappan, Jose Juan Almagro Armenteros, Brian J. Beliveau, Sabina Berretta, Rahul A. Bharadwaj, Arjun Bhattacharya, Kristen Brennand, Davide Caputo, Frances A. Champagne, Chris Chatzinakos, H. Isaac Chen, Lijun Cheng, Andrew Chess, Jo-fan Chien, Ashley Clement, Leonardo Collado-Torres, Gregory M. Cooper, Gregory E. Crawford, Rujia Dai, Nikolaos P. Daskalakis, Jose Davila-Velderrain, Amy Deep-Soboslay, Chengyu Deng, Christopher P. DiPietro, Stella Dracheva, Shiron Drusinsky, Duc Duong, Nicholas J. Eagles, Jonathan Edelstein, Kiki Galani, Kiran Girdhar, Fernando S. Goes, William Greenleaf, Hanmin Guo, Qiuyu Guo, Yoav Hadas, Joachim Hallmayer, Xikun Han, Vahram Haroutunian, Chuan He, Stephanie C. Hicks, Marcus Ho, Li-Lun Ho, Yiling Huang, Louise A. Huuki-Myers, Thomas M. Hyde, Artemis Iatrou, Fumitaka Inoue, Aarti Jajoo, Lihua Jiang, Peng Jin, Connor Jops, Alexandre Jourdon, Manolis Kellis, Joel E. Kleinman, Steven P. Kleopoulos, Alex Kozlenkov, Arnold Kriegstein, Anshul Kundaje, Soumya Kundu, Junhao Li, Mingfeng Li, Xiao Lin, Shuang Liu, Chunyu Liu, Jacob M. Loupe, Dan Lu, Liang Ma, Jessica Mariani, Keri Martinowich, Kristen R. Maynard, Richard M. Myers, Courtney Micallef, Tatiana Mikhailova, Guo-li Ming, Shahin Mohammadi, Emma Monte, Kelsey S. Montgomery, Eran A. Mukamel, Angus C. Nairn, Charles B. Nemeroff, Scott Norton, Tomasz Nowakowski, Larsson Omberg, Stephanie C. Page, Saejeong Park, Ashok Patowary, Reenal Pattni, Geo Perteau, Mette A. Peters, Dalila Pinto, Sirisha Pochareddy, Katherine S. Pollard, Alex Pollen, Pawel F. Przytycki, Carolin Purmann, Zhaohui S. Qin, Ping-Ping Qu, Towfique Raj, Sarah Reach, Thomas Reimann, Kerry J. Ressler, Deanna Ross, Joel Rozowsky, Misir Ruth, W. Brad Ruzicka, Stephan J. Sanders, Juliane M. Schneider, Soraya Scuderi, Robert Sebra, Nicholas Seyfried, Zhiping Shao, Annie W. Shieh, Joo Heon Shin, Mario Skarica, Clara Snijders, Hongjun Song, Matthew W. State, Jason Stein, Marilyn Steyert, Sivan Subburaju, Thomas Sudhof, Michael Snyder, Ran Tao, Karen Therrien, Li-Huei Tsai, Alexander E. Urban, Flora M. Vaccarino, Harm van Bakel, Daniel Vo, Georgios Voloudakis, Tao Wang, Sidney H. Wang, Yifan Wang, Yu Wei, Annika K. Weimer, Daniel R. Weinberger, Cindy Wen, Sean Whalen, A. Jeremy Willsey, Wing Wong, Hao Wu, Feinan Wu, Stefan Wuchty, Dennis Wylie, Chloe X. Yap, Biao Zeng, Pan Zhang, Chunling Zhang, Bin Zhang, Yanqiong Zhang, Ryan Ziffra, Zane R. Zeier, and Trisha M. Zintel

Science **384** (6698), eadi5199. DOI: 10.1126/science.adi5199

View the article online

<https://www.science.org/doi/10.1126/science.adi5199>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

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

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works