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, Tanima Chatterjee, Lijun Cheng, Yuyan Cheng, Yi Dai, Ziheng Duan, Mary Flaherty, John F. Fullard, Michael Gancz, Diego Garrido-Martín, Sophie 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*

† Co-first authors, these authors contributed equally to this work.
* Corresponding authors.

Correspondence to: matthew.girgenti@yale.edu, zhang.jing@uci.edu, daifeng.wang@wisc.edu, dhg@mednet.ucla.edu, pi@gersteinlab.org

Abstract

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, multi-omics datasets into a resource comprising >2.8M 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 >550K cell-type-specific regulatory elements and >1.4M 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.

Introduction

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), PsychENCODE, and ROSMAP (Religious Orders Study/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*). While 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, in fact, is one of the brain's distinguishing features (*12*). Many brain cell types have been rigorously defined, particularly by the 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, with 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 such as snRNA-seq, snATAC-seq, 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 M nuclei from 388 individual brains, including 333 newly generated samples and 55 from external sources (figs. S1-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 ~85K cis-eQTLs per cell type and ~550K 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 neuro-related drug targets like *CNR1* demonstrate a high degree of cell-type variability and low inter-individual variability and that the transcriptomes of specific neurons are highly predictive of an individual's age. Finally, 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-S5; (*20*)); these 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-S2; table S1; (*20*)). We used various filters on the total cohort of 388 for different downstream analyses (fig. S2; 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 >2.8M total nuclei (after QC and filtering from a raw number of nearly 4M; figs. S6, S7; table S2; (*20*)). Our processing required harmonizing datasets derived from different technologies and modalities; for instance, we generated uniform genotypes, including SVs, from combining whole-genome sequencing (WGS), SNP array, and snRNA-seq data (figs. S1, 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; figs. S9-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-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-S14; 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; data S5-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 to controls (*21–23*) (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 based on established approaches (*24*) (figs. S15-S18; data S7; (*20*)). Fig. 1D shows a representative plot for DE genes in schizophrenia, highlighting many previously known risk genes in a cell-type-specific context (*25*, *26*). 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; fig. S19).

Our snRNA-seq data also recapitulates the spatial relationships among cell types in the PFC. Fig. 1F shows a cell-trajectory analysis (*27*, *28*) across four subclasses of excitatory neurons in controls. We found smoothed patterns of gene-expression variation along the cortical-depth axis (specifically for L2/3, L4, L5, and L6 IT; figs. S20-S22; (*20*)). These findings expand on previous MERFISH-based results for 258 genes in the mouse motor cortex, now showing that cortical depth is related to gene expression variation for thousands of genes (*29*, *30*). 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-S22; list in table S5; 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 (Figs. 2A-2B; figs. S23-S24). Further, uniform snATAC-seq processing identifies 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; (*20*)) (*2*, *31*). Following the ENCODE (Encyclopedia of DNA Elements) convention (*32*), 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 targeted STARR-seq (Fig. 2D; (*20*, *33*)).

Using bulk data, we also developed a reference set of >400K open-chromatin regions, representing brain-tissue candidate cis-Regulatory Elements (b-cCREs; (*20*)). The b-cCREs were generated in a comparable fashion to ENCODE cCREs, which are not tissue-specific (*32*). 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 LDSC (linkage-disequilibrium score regression) analysis (*20*, *34*). In general, we found stronger LDSC enrichment for brain phenotypes in b-cCREs compared to cCREs (Fig. 2E; fig. S26; data S9-S10; 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) (*35*–*38*).

Next, we explored transcription factor (TF) usage across major brain cell types (fig. S27; (*20*)). Fig. 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, for example, in *ELF1* (Fig. 2G; data S11). We were able to validate many TF activities with footprinting (*39*) (Fig. 2H; fig. S27).

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

Single-cell data across a large cohort offers a unique opportunity to study the sources of expression variation in the brain (Fig. 3A; figs. S28-S30; (*20*)) (*40*). We partitioned the variation in expression of each gene based on the relative contribution of individual and cell-type variability while correcting for covariates (data S12-S13). This allowed us to determine relative contributions to variability 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 inter-individual variation (Figs. 3B-3C; fig. S31; data S14). Conversely, genes associated with common molecular or cellular processes, tend to have lower cell-type variation and higher individual variation (for instance, carbohydrate homeostasis and ATP generation; Fig. 3B). Furthermore, within families of CNS-specific genes, some neurotransmitter families manifest higher inter-individual variation compared to others (for example, glutamate vs serotonin, *p*-value=$3.7 \times 10^{-6,}$ one-sided t-test; Fig. 3C; fig. S31). We also identified a few outliers with very large inter-individual variation such as *ARL17B*, likely resulting from copy-number variation (*41*, *42*).

An additional application of quantifying expression variability is characterizing 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; fig. S32A) (*43*). That said, some of the 280 exhibit much higher inter-individual variation than others; *HSPA5* and *CNR1* provide a good illustration (Figs. 3C-3E; 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; figs. S33-S34; (*20*)). However, some genes not following this trend serve as interesting exceptions (that is, highly conserved genes with high expression variance). The gene deviating most from the trend is *IL1RAPL1* (Fig. 3F; fig. S34B), an interleukin-1 receptor-family gene inhibiting neurotransmitter release (*44*); *IL1RAPL1* is highly expressed in the brain and has been implicated in intellectual disability and ASD (*45*).

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; 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; that is, 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*)). Finally, we found that microglia scCREs exhibit the least sequence conservation, consistent with previous studies (Fig. 3H) (*19*, *46*, *47*).

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). Note the sparsity intrinsic to snRNA-seq data reduces power, particularly for rarer cell types (fig. S36; table S7; (*20*)) (*48*). 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, S37; table S8; (*20*)). We also generated further alternative callsets and a merge of results from all approaches (figs. S38-S39). These callsets include results based on linkage-disequilibrium pruning (table S9; (*20*)), regression across pseudo-time 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*)). Finally, 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 (~134K candidate iso-QTLs with 1389 associated "isoGenes"; figs. S40-S42; data S16; (*20*)).

Overall, we identified an average of ~85K scQTLs and ~690 eGenes per cell type in our core set, resulting in ~1.4M scQTLs when totaled over cell types (Fig. 4A; fig. S36A, S43; table S7; (*20*)). Many of the scQTLs are uniquely cell-type-specific (i.e. not in any other cell type), but ~47% appear in more than one cell type (Fig. 4A; 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; 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 appearing in a few cell types to those observed in many (Fig. 4B; 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; figs. S45-S46). We validated some of our core scQTLs by comparing them with functional elements identified by STARR-seq, mut-STARR-seq, and massively parallel reporter assays (MPRA) (Fig. 4E; figs. S47-S48; (*20*, *33*)). 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; fig. S49; (*20*)). Determination of single-cell ASE is particularly challenging due the sparsity of the data (*49–53*). Here, we compared the magnitude of the ASE effect at an SNV with the corresponding effect size of the scQTL involving the same SNV, finding significant correlation as expected (Fig. 4F; 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-S51; data S17). For example, we found scQTLs for *SYNE1*, a candidate autism and schizophrenia gene (*54*, *55*), and *NLGN1*, a candidate gene for multiple brain disorders encoding a ligand for neurexin signaling (*56*). 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 multi-cell type scQTL for the neurodegenerative-disorder risk gene *KANSL1* (Fig. 4G) (*41*). We further highlight an iso-QTL for *LYPD6*, which inhibits

acetylcholine-receptor activity in Pax6-type inhibitory neurons (*57*) (Fig. 4G).

Finally, 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-S53; data S18-S19; (*20*)) (*58*). In particular, for 1692 of the 6255 unique eGenes in four types of excitatory neurons, we found a corresponding dynamic scQTL (with a non-zero interaction term); Fig. 4H and fig. S52 show examples. 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 co-expression, we constructed gene-regulatory networks (GRNs) for PFC cell types (Fig. 5A; figs. S54-S58; data S20; (*20*)). In particular, we linked TFs to potential target genes based on their co-expression relationships from snRNA-seq data (*59*, *60*), and mapped scQTLs to connect promoters and enhancers (data S21). We make 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-S60).

We experimentally validated a subset of these linkages using CRISPR knockouts (Fig. 5B; fig. S61; 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; fig. S62). Additionally, mapping loss-of-function (LOF) mutations in individuals to select TFs ("natural knock-outs") provided further validation by showing the expected change in expression of their target genes in a cell-type-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; fig. S63).

Our analyses of GRNs uncovered complex network rewiring across the cell types (Fig. 5E; figs. S64-S66; 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). In contrast, bottlenecks (key connector TFs) have much more cell-type-specific activity (Fig. 5F; fig. S67A). Furthermore, the targets of bottleneck TFs are enriched for cell-type-specific functions, such as myelination and axon ensheathment for oligodendrocytes (*61*) (fig. S67B; 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 (*62*), are notably enriched in certain non-neuronal cell types.

Finally, disease genes for a particular disorder tend to be co-regulated in a cell-type-specific manner (Fig 5H; figs. S65,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) (*38*, *63*, *64*).

Constructing a cell-to-cell communication network

To further understand cellular signaling and regulation, we leveraged publicly available ligand-receptor pairs (*65*) in combination with our snRNA-seq data to construct a cell-to-cell communication network (Fig. 6A; table S11-S12, 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 based on their ligand-receptor interactions, with growth-factor genes as some of the top contributing ligand-receptor pairs (*66–68*) (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; figs. S71A-B, S72; data S25-S26). In fact, notable inter-mixings occur among the three broad patterns of ligand-receptor usage. For instance, in bipolar disorder, the excitatory pattern (inferred from controls) now also contains OPCs and some inhibitory neurons (Pvalb and Sst Chodl). In individuals with schizophrenia (compared to controls), we also found that excitatory neurons received less incoming signaling, while 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 downregulation of the Wnt pathway, consistent with previous findings (Fig. 6D) (*69–72*). Mechanistically, this downregulation could result in the overactivity of the lithium-targeted GSK3β enzyme in neurons (*73*, *74*). In schizophrenia, the Wnt pathway is downregulated as expected, but we also found increased sender communication strength for L6 IT Car3 neurons, different from bipolar (*75*). We further found downregulation of PTN pathway interactions from glial cells to neurons, consistent with previous studies (*76–78*), and a decrease in signaling to glial cells involving various growth factors (fibroblast, epidermal and insulin) (figs. S71C-E). These findings support the “glial cell hypothesis,” which posits that deleterious effects on glial cells cascade to neurons (*79*).

Lastly, we extended our extracellular cell-to-cell communication analysis by considering related disruptions to intracellular signaling pathways (Fig. 6E; fig. S73; (*20*)) (*80*). 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 *FOXP1* and its ligand *EBI3* in bipolar disorder and *MECP2* and its ligand *PDGFB* in schizophrenia (*81*, *82*).

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 based on 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, data S27) (*83*, *84*). This result is consistent with findings from raw cell counts in the single-cell data

(FDR<0.05, two-sided t-test; Fig. 7A; data S27; (*20*)). Next, we identified a list of aging DE genes across cell types (Fig. 7B; fig. S74; data S28; (*20*)). This list shows, for instance, that *HSPB1*, which encodes a heat-shock protein and has been previously implicated in longevity, is upregulated in multiple cell types in older individuals (*85*, *86*).

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; figs. S75A-B; (*20*)). The model shows that the transcriptomes of six cell types (L2/3 IT, L4 IT, L5 IT, L6 IT, Oligodendrocytes, and OPC) have strong predictive value (Fig. 7C; 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; fig. S76) (*87*–*89*).

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; fig. S77). The resulting scCRE activity patterns in certain cell types, particularly microglia, cluster individuals into distinct age groups (Fig. 7D; 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; fig. S78; (*20*)). Some TFs demonstrate consistent patterns across cell types (*FOXO4* and *RXRA)*, while others exhibit more cell-type-specific patterns (*NEUROG1*).

Finally, we extended our analysis to identify cell-type-specific changes in neurodegenerative disease. We obtain 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*)) (*90*). Certain glial fractions show a significant increase in AD (p<0.005, t-test), while several neuronal fractions decrease, especially Sst, Pvalb, and L2/3 IT, in line with previous studies (*91*) (Fig. 7F). We compared this result with that from directly comparing cell-type-specific gene-expression and methylation signatures to determine case-control status (*92*), finding that the fractions and signatures capture independent information (fig. S79B; 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: (1) to impute cell-type-specific and bulk tissue gene expression from genotype; (2) to predict the risk of disorders based on input genotypes; (3) to highlight genes and pathways contributing to particular phenotypes in their specific cell type of action; and (4) 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 and bulk eQTL sites, cell-type-specific and bulk tissue-based GRNs, cell-type fractions, cell-to-cell communication networks, gene co-expression 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 to a 34% baseline achieved by combining prior methods for bulk-imputation and cell-type deconvolution (*20*, *93*, *94*). 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 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 (*95*) (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 non-linear architectures (table S14-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 co-heritability of the genetic component of a particular gene's cell-type-specific expression with respect to schizophrenia or other disorders (fig. S80; (*20*)).

Fig. 8D and fig. S81 provide an overview of key prioritized genes, cell types, and cell-to-cell interactions in various disorders (full lists in data S30-S32). We found 64, 51, 108, and 34 gene/cell-type pairs for schizophrenia, bipolar disorder, ASD, and AD, respectively (*20*). In particular, *TCF4*, the first identified cross-psychiatric disorder locus (*96*), is important for neurons in schizophrenia (*97*), *LINGO2* is important for excitatory neurons in bipolar disorder, and *ANKHD1* is highly weighted in ASD, supporting current hypotheses (*98*, *99*). Fig. 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 (*100*). Further, we note 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 to 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-S83). They are also enriched for prior GWAS and literature support as well as bottleneck locations in the

regulatory network (figs. S84-S85; data S33). However, several genes specifically prioritized by the LNCTP are not differentially expressed for their respective disorders, including *MEF2A* and *ID1,* perhaps highlighting that they act through network effects (figs. S83, S85) (*23*).

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) (*79*).

We further used the LNCTP to perform in silico perturbation analysis, where 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 via DrugBank (*102*)) both induce overall expression changes strongly characteristic of case status (figs. S86A-B). As expected, the induced changes more strongly impact genes in close proximity to the perturbed gene in the GRNs (fig. S87; table S16). We synthesized the perturbations into a workflow to suggest potential drugs for repurposing with CLUE (*43*) by matching a perturbation's effects to drugs inducing changes potentially complementary to those found in a particular disorder (fig. S86C; table S17; (*20*)).

Finally, 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-S89; 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). While more comprehensive validation is essential, these results offer promising indications that LNCTP can find verifiable prioritizations of gene/cell-type pairs.

Discussion

Here, we used population-scale multi-omic data to build a comprehensive single-cell functional genomics resource (brainSCOPE) for investigating brain disorders in adults (Figs. S3-S5; (*20*)). The resource can be summarized at multiple levels: (1) raw data and metadata with a harmonized identifier system for each of the individuals; (2) quantifications of single-cell gene expression (count matrices) with a BICCN-compatible cell-typing system for the PFC; (3) lists of DE genes and differential cell-fractions for various phenotypes; (4) snATAC-seq signal tracks for various cell types and ENCODE-compatible regulatory elements (b-cCREs and scCREs), including lists of validated ones; (5) 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; (6) a core set of GTEx-compatible scQTLs and other additional sets of QTLs (such as dynamic eQTLs); (7) 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 significant upstream regulators; (8) cell-to-cell communication networks (expressed as ligand-receptor-by-cell-type matrices); (9) integrative models with code for imputation, perturbation and prioritization of cell-type-specific functional genomics in brain disease; and (10)

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, via 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., *CNR1*), 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.

Finally, 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 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. Firstly, 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 sub-regions 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. Finally, more extensive validation of our results would be valuable, such as via 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 (*20*), which is organized using 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

1. P. F. Sullivan, D. H. Geschwind, Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162–183 (2019).

2. PsychENCODE Consortium, Revealing the brain's molecular architecture. *Science* **362**, 1262–1263 (2018).

3. 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).

4. D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, *et al.*, Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).

5. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

6. B. Ng, C. C. White, H.-U. Klein, S. K. Sieberts, C. McCabe, E. Patrick, *et al.*, An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).

7. S. Liu, H. Won, D. Clarke, N. Matoba, S. Khullar, Y. Mu, *et al.*, Illuminating links between cis-regulators and trans-acting variants in the human prefrontal cortex. *Genome Med.* **14**, 133 (2022).

8. J. Bryois, D. Calini, W. Macnair, L. Foo, E. Urich, W. Ortmann, *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).

9. S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Muñoz-Aguirre, S. Kasela, V. Wucher, *et al.*, Cell type–specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).

10. B. Zeng, J. Bendl, R. Kosoy, J. F. Fullard, G. E. Hoffman, P. Roussos, Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat. Genet.* **54**, 161–169 (2022).

11. K. Zhang, J. D. Hocker, M. Miller, X. Hou, J. Chiou, O. B. Poirion, *et al.*, A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001.e19 (2021).

12. BRAIN Initiative Cell Census Network (BICCN), A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).

13. C. Luo, H. Liu, F. Xie, E. J. Armand, K. Siletti, T. E. Bakken, *et al.*, Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genomics* **2**, 100107 (2022).

14. H. Zeng, What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).

15. G. La Manno, K. Siletti, A. Furlan, D. Gyllborg, E. Vinsland, A. Mossi Albiach, *et al.*, Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021).

16. M. Song, X. Yang, X. Ren, L. Maliskova, B. Li, I. R. Jones, *et al.*, Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* **51**, 1252–1262 (2019).

17. P. L. De Jager, Y. Ma, C. McCabe, J. Xu, B. N. Vardarajan, D. Felsky, *et al.*, A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data* **5**, 180142 (2018).

18. D. Velmeshev, L. Schirmer, D. Jung, M. Haeussler, Y. Perez, S. Mayer, *et al.*, Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).

19. S. Ma, M. Skarica, Q. Li, C. Xu, R. D. Risgaard, A. T. N. Tebbenkamp, *et al.*, Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022).

20. Materials and methods are available as supplementary materials.

21. 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).

22. L.-C. Lin, E. Sibille, Reduced brain somatostatin in mood disorders: a common pathophysiological substrate and drug target? Front. Pharmacol. 4, 110 (2013).

23. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, *et al.*, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).

24. 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).

25. W. B. Ruzicka, S. Mohammadi, J. F. Fullard, J. Davila-Velderrain, S. Subburaju, D. R. Tso, *et al.*, "Single-cell multi-cohort dissection of the schizophrenia transcriptome"

(preprint, Psychiatry and Clinical Psychology, 2022); https://doi.org/10.1101/2022.08.31.22279406.

26. P. Karpiński, J. Samochowiec, M. M. Sąsiadek, Ł. Łaczmański, B. Misiak, Analysis of global gene expression at seven brain regions of patients with schizophrenia. Schizophr. Res. 223, 119–127 (2020).

27. K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, *et al.*, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

28. K. Van den Berge, H. Roux de Bezieux, K. Street, W. Saelens, R. Cannoodt, Y. Saeys, *et al.*, Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**, 1201 (2020).

29. M. Zhang, S. W. Eichhorn, B. Zingg, Z. Yao, K. Cotter, H. Zeng, *et al.*, Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).

30. R. Fang, C. Xia, J. L. Close, M. Zhang, J. He, Z. Huang, *et al.*, Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022).

31. J. Bryois, M. E. Garrett, L. Song, A. Safi, P. Giusti-Rodriguez, G. D. Johnson, *et al.*, Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, 3121 (2018).

32. E. P. Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, *et al.*, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

33. Validation of enhancer regions in primary human neural progenitor cells using capture STARR-seq, (2023); https://doi.org/10.7303/SYN50900302.1.

34. C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, *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).

35. D. Polioudakis, L. de la Torre-Ubieta, J. Langerman, A. G. Elkins, X. Shi, J. L. Stein, *et al.*, A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785-801.e8 (2019).

36. E. K. Ruzzo, L. Pérez-Cano, J.-Y. Jung, L.-K. Wang, D. Kashef-Haghighi, C. Hartl, *et al.*, Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850-866.e26 (2019).

37. C. L. Hartl, G. Ramaswami, W. G. Pembroke, S. Muller, G. Pintacuda, A. Saha, *et al.*, Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nat. Neurosci.* **24**, 1313–1323 (2021).

38. B. Hu, H. Won, W. Mah, R. B. Park, B. Kassim, K. Spiess, *et al.*, Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat. Commun.* **12**, 3968 (2021).

39. J. M. Granja, M. R. Corces, S. E. Pierce, S. T. Bagdatli, H. Choudhry, H. Y. Chang, *et al.*, ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

40. N. Johansen, S. Somasundaram, K. J. Travaglini, A. M. Yanny, M. Shumyatcher, T. Casper, *et al.*, Interindividual variation in human cortical cell type abundance and expression. *Science* **382**, eadf2359 (2023).

41. Y. A. Cooper, N. Teyssier, N. M. Dräger, Q. Guo, J. E. Davis, S. M. Sattler, *et al.*, Functional regulatory variants implicate distinct transcriptional networks in dementia. *Science* **377**, eabi8654 (2022).

42. N. L. Jorstad, J. H. T. Song, D. Exposito-Alonso, H. Suresh, N. Castro-Pacheco, F. M. Krienen, *et al.*, Comparative transcriptomics reveals human-specific cortical features. *Science* **382**, eade9516 (2023).

43. A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, *et al.*, A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).

44. F. Gambino, A. Pavlowsky, A. Béglé, J.-L. Dupont, N. Bahi, R. Courjaret, *et al.*, IL1-receptor accessory protein-like 1 (IL1RAPL1), 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).

45. C. Montani, M. Ramos-Brossier, L. Ponzoni, L. Gritti, A. W. Cwetsch, D. Braida, *et al.*, The X-Linked Intellectual Disability Protein IL1RAPL1 Regulates Dendrite Complexity. *J. Neurosci. Off. J. Soc. Neurosci.* **37**, 6606–6627 (2017).

46. W. G. Pembroke, C. L. Hartl, D. H. Geschwind, Evolutionary conservation and divergence of the human brain transcriptome. *Genome Biol.* **22**, 52 (2021).

47. K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, R. P. N. Rao, Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4430–4435 (2010).

48. M. Maria, N. Pouyanfar, T. Örd, M. U. Kaikkonen, The Power of Single-Cell RNA Sequencing in eQTL Discovery. *Genes* **13**, 502 (2022).

49. 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).

50. C. Borel, P. G. Ferreira, F. Santoni, O. Delaneau, A. Fort, K. Y. Popadin, *et al.*, Biased allelic expression in human primary fibroblast single cells. *Am. J. Hum. Genet.* **96**, 70–80 (2015).

51. W. Mu, H. Sarkar, A. Srivastava, K. Choi, R. Patro, M. I. Love, Airpart: interpretable statistical models for analyzing allelic imbalance in single-cell datasets. *Bioinformatics* **38**, 2773–2780 (2022).

52. 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).

53. Y. Jiang, N. R. Zhang, M. Li, SCALE: modeling allele-specific gene expression by single-cell RNA sequencing. *Genome Biol.* **18**, 74 (2017).

54. T. W. Yu, M. H. Chahrour, M. E. Coulter, S. Jiralerspong, K. Okamura-Ikeda, B. Ataman, *et al.*, Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**, 259–273 (2013).

55. S. A. Sloan, B. A. Barres, Mechanisms of astrocyte development and their contributions to neurodevelopmental disorders. *Curr. Opin. Neurobiol.* **27**, 75–81 (2014).

56. A. M. Craig, Y. Kang, Neurexin–neuroligin signaling in synapse development. *Curr. Opin. Neurobiol.* **17**, 43–52 (2007).

57. D. Kulbatskii, Z. Shenkarev, M. Bychkov, E. Loktyushov, M. Shulepko, S. Koshelev, *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).

58. A. Nathan, S. Asgari, K. Ishigaki, C. Valencia, T. Amariuta, Y. Luo, *et al.*, Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120–128 (2022).

59. S. Aibar, C. B. Gonzalez-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, *et al.*, SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).

60. T. Jin, P. Rehani, M. Ying, J. Huang, S. Liu, P. Roussos, *et al.*, scGRNom: a computational pipeline of integrative multi-omics analyses for predicting cell-type disease genes and regulatory networks. *Genome Med.* **13**, 95 (2021).

61. I. D. Duncan, A. B. Radcliff, M. Heidari, G. Kidd, B. K. August, L. A. Wierenga, The adult oligodendrocyte can participate in remyelination. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11807–E11816 (2018).

62. U. Alon, Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).

63. I. E. Jansen, J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, *et al.*, Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

64. H. Keren-Shaul, A. Spinrad, A. Weiner, O. Matcovitch-Natan, R. Dvir-Szternfeld, T. K. Ulland, *et al.*, A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e17 (2017).

65. S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C.-H. Kuan, *et al.*, Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

66. E. Savchenko, G. N. Teku, A. Boza-Serrano, K. Russ, M. Berns, T. Deierborg, *et al.*, FGF family members differentially regulate maturation and proliferation of stem cell-derived astrocytes. *Sci. Rep.* **9**, 9610 (2019).

67. 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).

68. 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).

69. J. C. McAfee, S. Lee, J. Lee, J. L. Bell, O. Krupa, J. Davis, *et al.*, Systematic investigation of allelic regulatory activity of schizophrenia-associated common variants. *Cell Genomics* **3**, 100404 (2023).

70. A. Muneer, Wnt and GSK3 Signaling Pathways in Bipolar Disorder: Clinical and Therapeutic Implications. *Clin Psychopharmacol Neurosci* **15**, 100–114 (2017).

71. R. Santos, S. B. Linker, S. Stern, A. P. D. Mendes, M. N. Shokhirev, G. Erikson, *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).

72. 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).

73. E. Z. Hoseth, F. Krull, I. Dieset, R. H. Morch, S. Hope, E. S. Gardsjord, *et al.*, Exploring the Wnt signaling pathway in schizophrenia and bipolar disorder. *Transl Psychiatry* **8**, 55 (2018).

74. A. J. Valvezan, P. S. Klein, GSK-3 and Wnt Signaling in Neurogenesis and Bipolar Disorder. *Front Mol Neurosci* **5**, 1 (2012).

75. S. Lovestone, R. Killick, M. Di Forti, R. Murray, Schizophrenia as a GSK-3 dysregulation disorder. *Trends Neurosci.* **30**, 142–149 (2007).

76. I. Panaccione, F. Napoletano, A. M. Forte, G. D. Kotzalidis, A. Del Casale, C. Rapinesi, *et al.*, Neurodevelopment in schizophrenia: the role of the wnt pathways. *Curr Neuropharmacol* **11**, 535–58 (2013).

77. R. D. McCurdy, F. Feron, C. Perry, D. C. Chant, D. McLean, N. Matigian, *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**, 163–73 (2006).

78. J. Xu, J. Sun, J. Chen, L. Wang, A. Li, M. Helm, *et al.*, RNA-Seq analysis implicates dysregulation of the immune system in schizophrenia. *BMC Genomics* **13 Suppl 8**, S2 (2012).

79. A. F. Terwisscha van Scheltinga, S. C. Bakker, R. S. Kahn, Fibroblast growth factors in schizophrenia. *Schizophr. Bull.* **36**, 1157–1166 (2010).

80. R. Browaeys, W. Saelens, Y. Saeys, NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**, 159–162 (2020).

81. R. E. Amir, I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, H. Y. Zoghbi, Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).

82. C. Bacon, G. A. Rappold, The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Hum. Genet.* **131**, 1687–1698 (2012).

83. 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 N. Y. N 1991* **17**, 2060–2071 (2007).

84. 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).

85. C. R. Gomez, Role of heat shock proteins in aging and chronic inflammatory diseases. *GeroScience* **43**, 2515–2532 (2021).

86. C. Schultz, E. J. Dick, A. B. Cox, G. B. Hubbard, E. Braak, H. Braak, Expression of stress proteins alpha B-crystallin, ubiquitin, and hsp27 in pallido-nigral spheroids of aged rhesus monkeys. *Neurobiol. Aging* **22**, 677–682 (2001).

87. A. P. Abreu, A. Dauber, D. B. Macedo, S. D. Noel, V. N. Brito, J. C. Gill, *et al.*, Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N. Engl. J. Med.* **368**, 2467–2475 (2013).

88. A. S. Zannas, M. Jia, K. Hafner, J. Baumert, T. Wiechmann, J. C. Pape, *et al.*, Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-κB-driven inflammation and cardiovascular risk. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11370–11379 (2019).

89. A. S. Zannas, T. Wiechmann, N. C. Gassen, E. B. Binder, Gene-Stress-Epigenetic Regulation of FKBP5: Clinical and Translational Implications. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **41**, 261–274 (2016).

90. D. A. Bennett, A. S. Buchman, P. A. Boyle, L. L. Barnes, R. S. Wilson, J. A. Schneider, Religious Orders Study and Rush Memory and Aging Project. *J Alzheimers Dis* **64**, S161–S189 (2018).

91. C. A. Saura, A. Deprada, M. D. Capilla-Lopez, A. Parra-Damas, Revealing cell vulnerability in Alzheimer's disease by single-cell transcriptomics. *Semin Cell Dev Biol* **139**, 73–83 (2023).

92. 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).

93. E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, *et al.*, A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091–8 (2015).

94. Y. H. Wang, H. A. Hou, C. C. Lin, Y. Y. Kuo, C. Y. Yao, C. L. Hsu, *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).

95. Brainstorm Consortium, V. Anttila, B. Bulik-Sullivan, H. K. Finucane, R. K. Walters, J. Bras, *et al.*, Analysis of shared heritability in common disorders of the brain. *Science* **360** (2018).

96. 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 Lond. Engl.* **381**, 1371–1379 (2013).

97. J. Gelernter, N. Sun, R. Polimanti, R. Pietrzak, D. F. Levey, J. Bryois, *et al.*, Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. *Nat Neurosci* **22**, 1394–1401 (2019).

98. C. Pisanu, M. J. Williams, D. M. Ciuculete, G. Olivo, M. Del Zompo, A. Squassina, *et al.*, Evidence that genes involved in hedgehog signaling are associated with both bipolar disorder and high BMI. *Transl. Psychiatry* **9**, 315 (2019).

99. M. Chopra, M. McEntagart, J. Clayton-Smith, K. Platzer, A. Shukla, K. M. Girisha, *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).

100. W. R. Reay, J. R. Atkins, Y. Quidé, V. J. Carr, M. J. Green, M. J. Cairns, Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. *Mol. Psychiatry* **25**, 719–731 (2020).

101. M. Wang, L. Zhang, F. H. Gage, Microglia, complement and schizophrenia. *Nat. Neurosci.* **22**, 333–334 (2019).

102. D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, *et al.*, DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668-672 (2006).

103. R. Tian, A. Abarientos, J. Hong, S. H. Hashemi, R. Yan, N. Dräger, *et al.*, Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. *Nat. Neurosci.* **24**, 1020–1034 (2021).

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).

105. C. P. Moritz, T. Mühlhaus, S. Tenzer, T. Schulenborg, 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).

106. B. C. Carlyle, R. R. Kitchen, J. E. Kanyo, E. Z. Voss, M. Pletikos, A. M. M. Sousa, *et al.*, A multiregional proteomic survey of the postnatal human brain. *Nat. Neurosci.* **20**, 1787–1795 (2017).

107. L. E. Liharska, Y. J. Park, K. Ziafat, L. Wilkins, H. Silk, L. M. Linares, *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), PyschENCODE Consortium (PEC) Capstone II Cross-study Harmonized Data, Synapse (2023); https://doi.org/10.7303/SYN51111084.1.

109. PEC Integrative Analysis Processing of ROSMAP data, (2024); https://doi.org/10.7303/SYN53479857.1.

110. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE216270.

111. R. E. Rodin, Y. Dou, M. Kwon, M. A. Sherman, A. M. D'Gama, R. N. Doan, *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).

112. S. Akbarian, C. Liu, J. A. Knowles, F. M. Vaccarino, P. J. Farnham, G. E. Crawford, *et al.*, The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).

113. M. Stoeckius, S. Zheng, B. Houck-Loomis, S. Hao, B. Z. Yeung, W. M. Mauck, *et al.*, Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).

114. H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, *et al.*, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

115. A. K. Greenwood, K. S. Montgomery, N. Kauer, K. H. Woo, Z. J. Leanza, W. L. Poehlman, *et al.*, The AD Knowledge Portal: A Repository for Multi-Omic Data on Alzheimer's Disease and Aging. *Curr. Protoc. Hum. Genet.* **108**, e105 (2020).

116. M. Freund, A. Taylor, C. Ng, A. R. Little, The NIH NeuroBioBank: creating opportunities for human brain research. *Handb. Clin. Neurol.* **150**, 41–48 (2018).

117. B. Li, J. Gould, Y. Yang, S. Sarkizova, M. Tabaka, O. Ashenberg, *et al.*, Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nat. Methods* **17**, 793–798 (2020).

118. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

119. C. S. McGinnis, D. M. Patterson, J. Winkler, D. N. Conrad, M. Y. Hein, V. Srivastava, *et al.*, MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).

120. P. Roelli, Bbimber, B. Flynn, Santiagorevale, Gege Gui, Hoohm/CITE-seq-Count: 1.4.2, version 1.4.2, Zenodo (2019); https://doi.org/10.5281/ZENODO.2585469.

121. S. J. Fleming, M. D. Chaffin, A. Arduini, A.-D. Akkad, E. Banks, J. C. Marioni, *et al.*, Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).

122. S. Rath, R. Sharma, R. Gupta, T. Ast, C. Chan, T. J. Durham, *et al.*, MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* **49**, D1541–D1547 (2021).

123. R. D. Hodge, T. E. Bakken, J. A. Miller, K. A. Smith, E. R. Barkan, L. T. Graybuck, *et al.*, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).

124. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).

125. A. Gayoso, J. Shor, JonathanShor/DoubletDetection: doubletdetection v4.2, Zenodo (2022); https://doi.org/10.5281/zenodo.6349517.

126. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

127. N. L. Jorstad, J. Close, N. Johansen, A. M. Yanny, E. R. Barkan, K. J. Travaglini, *et al.*, Transcriptomic cytoarchitecture reveals principles of human neocortex organization. *Science* **382**, eadf6812 (2023).

128. T. E. Bakken, N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, *et al.*, Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).

129. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

130. R. Poplin, V. Ruano-Rubio, M. A. DePristo, T. J. Fennell, M. O. Carneiro, G. A. Van der Auwera, *et al.*, “Scaling accurate genetic variant discovery to tens of thousands of samples” (preprint, Genomics, 2017); https://doi.org/10.1101/201178.

131. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).

132. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

133. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

134. S. Das, L. Forer, S. Schonherr, C. Sidore, A. E. Locke, A. Kwong, *et al.*, Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287 (2016).

135. B. S. Pedersen, A. R. Quinlan, Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *Am J Hum Genet* **100**, 406–413 (2017).

136. G. Gursoy, P. Emani, C. M. Brannon, O. A. Jolanki, A. Harmanci, J. S. Strattan, *et al.*, Data Sanitization to Reduce Private Information Leakage from Functional Genomics. *Cell* **183**, 905-917 e16 (2020).

137. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

138. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alfoldi, Q. Wang, *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

139. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894 (2019).

140. J. Ebler, P. Ebert, W. E. Clarke, T. Rausch, P. A. Audano, T. Houwaart, *et al.*, Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes. *Nat. Genet.* **54**, 518–525 (2022).

141. P. Ebert, P. A. Audano, Q. Zhu, B. Rodriguez-Martin, D. Porubsky, M. J. Bonder, *et al.*, Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372** (2021).

142. X. Li, Y. Kim, E. K. Tsang, J. R. Davis, F. N. Damani, C. Chiang, *et al.*, The impact of rare variation on gene expression across tissues. *Nature* **550**, 239–243 (2017).

143. B. Jew, M. Alvarez, E. Rahmani, Z. Miao, A. Ko, K. M. Garske, *et al.*, Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).

144. G. Hoffman, D. Lee, J. Bendl, P. Fnu, A. Hong, C. Casey, *et al.*, Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. *Res. Sq.*, rs.3.rs-2705625 (2023).

145. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

146. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

147. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

148. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

149. Gene Ontology Consortium, The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).

150. M. Li, G. Santpere, Y. Imamura Kawasawa, O. V. Evgrafov, F. O. Gulden, S. Pochareddy, *et al.*, Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).

151. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

152. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

153. O. Fornes, J. A. Castro-Mondragon, A. Khan, R. van der Lee, X. Zhang, P. A. Richmond, *et al.*, JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).

154. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, *et al.*, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

155. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, *et al.*, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

156. S. Köhler, M. Gargano, N. Matentzoglu, L. C. Carmody, D. Lewis-Smith, N. A. Vasilevsky, *et al.*, The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).

157. S. Gazal, H. K. Finucane, N. A. Furlotte, P.-R. Loh, P. F. Palamara, X. Liu, *et al.*, Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **179**, 1421–1427 (2017).

158. J. L. Stein, L. de la Torre-Ubieta, Y. Tian, N. N. Parikshak, I. A. Hernández, M. C. Marchetto, *et al.*, A Quantitative Framework to Evaluate Modeling of Cortical Development by Neural Stem Cells. *Neuron* **83**, 69–86 (2014).

159. L. de la Torre-Ubieta, J. L. Stein, H. Won, C. K. Opland, D. Liang, D. Lu, *et al.*, The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289-304.e18 (2018).

160. A. E. Trevino, N. Sinnott-Armstrong, J. Andersen, S.-J. Yoon, N. Huber, J. K. Pritchard, *et al.*, Chromatin accessibility dynamics in a model of human forebrain development. *Science* **367**, eaay1645 (2020).

161. R. L. Walker, G. Ramaswami, C. Hartl, N. Mancuso, M. J. Gandal, L. de la Torre-Ubieta, *et al.*, Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).

162. G. E. Hoffman, E. E. Schadt, variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).

163. M. J. Gandal, J. R. Haney, B. Wamsley, C. X. Yap, S. Parhami, P. S. Emani, *et al.*, Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).

164. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, *et al.*, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

165. X. Dong, X. Li, T. W. Chang, C. R. Scherzer, S. T. Weiss, W. Qiu, powerEQTL: An R package and shiny application for sample size and power calculation of bulk tissue and single-cell eQTL analysis. *Bioinformatics* **37**, 4269–71 (2021).

166. O. Delaneau, H. Ongen, A. A. Brown, A. Fort, N. I. Panousis, E. T. Dermitzakis, A complete tool set for molecular QTL discovery and analysis. *Nat Commun* **8**, 15452 (2017).

167. H. Ongen, A. Buil, A. A. Brown, E. T. Dermitzakis, O. Delaneau, Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).

168. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**, 9440–9445 (2003).

169. U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, *et al.*, g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

170. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, *et al.*, SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).

171. P. Jia, G. Han, J. Zhao, P. Lu, Z. Zhao, SZGR 2.0: a one-stop shop of schizophrenia candidate genes. *Nucleic Acids Res.* **45**, D915–D924 (2017).

172. C. Franklin, D. S. Dwyer, Candidate risk genes for bipolar disorder are highly conserved during evolution and highly interconnected. *Bipolar Disord.* **23**, 400–408 (2021).

173. Y.-S. Hu, J. Xin, Y. Hu, L. Zhang, J. Wang, Analyzing the genes related to Alzheimer's disease via a network and pathway-based approach. *Alzheimers Res. Ther.* **9**, 29 (2017).

174. R. Tacutu, D. Thornton, E. Johnson, A. Budovsky, D. Barardo, T. Craig, *et al.*, Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).

175. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, *et al.*, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

176. G. Casella, R. L. Berger, *Statistical Inference* (Duxbury, Pacific Grove, Calif, 2. ed., 2002).

177. An Introduction to Statistical Learning: with Applications in R | SpringerLink. https://link.springer.com/book/10.1007/978-1-4614-7138-7.

178. P. D. Hoff, *A First Course in Bayesian Statistical Methods* (Springer, New York, NY, 2009; http://link.springer.com/10.1007/978-0-387-92407-6)*Springer Texts in Statistics*.

179. L. Xiong, K. Tian, Y. Li, W. Ning, X. Gao, Q. C. Zhang, Online single-cell data integration through projecting heterogeneous datasets into a common cell-embedding space. *Nat. Commun.* **13**, 6118 (2022).

180. L. Pan, H. Q. Dinh, Y. Pawitan, T. N. Vu, Isoform-level quantification for single-cell RNA sequencing. *Bioinformatics* **38**, 1287–1294 (2022).

181. D. Garrido-Martin, B. Borsari, M. Calvo, F. Reverter, R. Guigo, Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. *Nat Commun* **12**, 727 (2021).

182. D. Garrido-Martín, E. Palumbo, R. Guigó, A. Breschi, ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput. Biol.* **14**, e1006360 (2018).

183. J. Rozowsky, A. Abyzov, J. Wang, P. Alves, D. Raha, A. Harmanci, *et al.*, AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* **7**, 522 (2011).

184. J. Chen, J. Rozowsky, T. R. Galeev, A. Harmanci, R. Kitchen, J. Bedford, *et al.*, A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat. Commun.* **7**, 11101 (2016).

185. J. Rozowsky, J. Gao, B. Borsari, Y. T. Yang, T. Galeev, G. Gürsoy, *et al.*, The EN-TEx resource of multi-tissue personal epigenomes & variant-impact models. *Cell* **186**, 1493-1511.e40 (2023).

186. G. Konopka, E. Wexler, E. Rosen, Z. Mukamel, G. E. Osborn, L. Chen, *et al.*, Modeling the functional genomics of autism using human neurons. *Mol. Psychiatry* **17**, 202–214 (2012).

187. D. Lee, M. Shi, J. Moran, M. Wall, J. Zhang, J. Liu, *et al.*, STARRPeaker: uniform processing and accurate identification of STARR-seq active regions. *Genome Biol.* **21**, 298 (2020).

188. C. Deng, S. Whalen, M. Steyert, R. Ziffra, P. F. Przytycki, F. Inoue, *et al.*, "Massively parallel characterization of psychiatric disorder-associated and cell-type-specific regulatory elements in the developing human cortex" (preprint, Genomics, 2023); https://doi.org/10.1101/2023.02.15.528663.

189. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

190. Y. Baran, A. Bercovich, A. Sebe-Pedros, Y. Lubling, A. Giladi, E. Chomsky, *et al.*, MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol* **20**, 206 (2019).

191. T. Moerman, S. Aibar Santos, C. Bravo Gonzalez-Blas, J. Simm, Y. Moreau, J. Aerts, *et al.*, GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).

192. S. Suo, Q. Zhu, A. Saadatpour, L. Fei, G. Guo, G. C. Yuan, Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas. *Cell Rep* **25**, 1436-1445 e3 (2018).

193. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, *et al.*, Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

194. C. Gupta, J. Xu, T. Jin, S. Khullar, X. Liu, S. Alatkar, *et al.*, Single-cell network biology characterizes cell type gene regulation for drug repurposing and phenotype prediction in Alzheimer's disease. *PLoS Comput. Biol.* **18**, e1010287 (2022).

195. L. Garcia-Alonso, C. H. Holland, M. M. Ibrahim, D. Turei, J. Saez-Rodriguez, Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).

196. D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, *et al.*, Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife* **6** (2017).

197. N. Kashtan, S. Itzkovitz, R. Milo, U. Alon, Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20**, 1746–58 (2004).

198. G. Liu, L. Wong, H. N. Chua, Complex discovery from weighted PPI networks. *Bioinformatics* **25**, 1891–7 (2009).

199. A. Baptista, A. Gonzalez, A. Baudot, Universal multilayer network exploration by random walk with restart. *Commun. Phys.* **5**, 170 (2022).

200. J.-P. Brunet, P. Tamayo, T. R. Golub, J. P. Mesirov, Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4164–4169 (2004).

201. R. Gaujoux, C. Seoighe, A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).

202. L. Huuki-Myers, A. Spangler, N. Eagles, K. D. Montgomery, S. H. Kwon, B. Guo, *et al.*, "Integrated single cell and unsupervised spatial transcriptomic analysis defines molecular anatomy of the human dorsolateral prefrontal cortex" (preprint, Neuroscience, 2023); https://doi.org/10.1101/2023.02.15.528722.

203. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

204. J. Ernst, Z. Bar-Joseph, STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**, 191 (2006).

205. M. Dai, L. Zhao, Z. Li, X. Li, B. You, S. Zhu, *et al.*, The Transcriptional Differences of Avian CD4+CD8+ Double-Positive T Cells and CD8+ T Cells From Peripheral Blood of ALV-J Infected Chickens Revealed by Smart-Seq2. *Front. Cell. Infect. Microbiol.* **11**, 747094 (2021).

206. W. Wei, C. Jiang, X. Chai, J. Zhang, C.-C. Zhang, W. Miao, *et al.*, RNA Interference by Cyanobacterial Feeding Demonstrates the SCSG1 Gene Is Essential for Ciliogenesis during Oral Apparatus Regeneration in Stentor. *Microorganisms* **9**, 176 (2021).

207. Q. Song, J. Wang, Z. Bar-Joseph, scSTEM: clustering pseudotime ordered single-cell data. *Genome Biol.* **23**, 150 (2022).

208. O. Ben-Kiki, A. Bercovich, A. Lifshitz, A. Tanay, Metacell-2: a divide-and-conquer metacell algorithm for scalable scRNA-seq analysis. *Genome Biol.* **23**, 100 (2022).

209. M.-A. Schulz, B. T. T. Yeo, J. T. Vogelstein, J. Mourao-Miranada, J. N. Kather, K. Kording, *et al.*, Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nat. Commun.* **11**, 4238 (2020).

210. S. Arora, N. Cohen, W. Hu, Y. Luo, Implicit Regularization in Deep Matrix Factorization. arXiv arXiv:1905.13655 [Preprint] (2019). https://doi.org/10.48550/arXiv.1905.13655.

211. M. J. Wainwright, M. I. Jordan, Graphical Models, Exponential Families, and Variational Inference. *Found. Trends® Mach. Learn.* **1**, 1–305 (2007).

212. A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, *et al.*, Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

213. S. W. Choi, T. S.-H. Mak, P. F. O'Reilly, Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).

214. J. Grove, S. Ripke, T. D. Als, M. Mattheisen, R. K. Walters, H. Won, *et al.*, Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

215. N. Mullins, A. J. Forstner, K. S. O'Connell, B. Coombes, J. R. I. Coleman, Z. Qiao, *et al.*, Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

216. V. Trubetskoy, A. F. Pardiñas, T. Qi, G. Panagiotaropoulou, S. Awasthi, T. B. Bigdeli, *et al.*, Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

217. D. P. Wightman, I. E. Jansen, J. E. Savage, A. A. Shadrin, S. Bahrami, D. Holland, *et al.*, A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).

218. F. Privé, J. Arbel, B. J. Vilhjálmsson, LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).

219. F. K. Satterstrom, R. K. Walters, T. Singh, E. M. Wigdor, F. Lescai, D. Demontis, *et al.*, Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat. Neurosci.* **22**, 1961–1965 (2019).

220. J. Kaplanis, N. Akawi, G. Gallone, J. F. McRae, E. Prigmore, C. F. Wright, *et al.*, Exome-wide assessment of the functional impact and pathogenicity of multinucleotide mutations. *Genome Res.* **29**, 1047–1056 (2019).

221. Y. Xia, R. Dai, K. Wang, C. Jiao, C. Zhang, Y. Xu, *et al.*, Sex-differential DNA methylation and associated regulation networks in human brain implicated in the sex-biased risks of psychiatric disorders. *Mol. Psychiatry* **26**, 835–848 (2021).

222. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

223. M. Reale, E. Costantini, N. H. Greig, Cytokine Imbalance in Schizophrenia. From Research to Clinic: Potential Implications for Treatment. *Front. Psychiatry* **12**, 536257 (2021).

224. A.-M. Tsimberidou, A. Skliris, A. Valentine, J. Shaw, U. Hering, H. H. Vo, *et al.*, AKT inhibition in the central nervous system induces signaling defects resulting in psychiatric symptomatology. *Cell Biosci.* **12**, 56 (2022).

225. L. A. Farrelly, S. Zheng, N. Schrode, A. Topol, N. V. Bhanu, R. M. Bastle, *et al.*, Chromatin profiling in human neurons reveals aberrant roles for histone acetylation and BET family proteins in schizophrenia. *Nat. Commun.* **13**, 2195 (2022).

226. S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw. ISDN Syst.* **30**, 107–117 (1998).

227. G. Csárdi, T. Nepusz, K. Müller, S. Horvát, V. Traag, F. Zanini, *et al.*, igraph for R: R interface of the igraph library for graph theory and network analysis, version v2.0.2 (2024); https://doi.org/10.5281/ZENODO.7682609.

228. D. B. West, *Introduction to Graph Theory* (Prentice Hall, Upper Saddle River, 1996).

229. J. Wang, S. Vasaikar, Z. Shi, M. Greer, B. Zhang, WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).

230. J. Jourquin, D. Duncan, Z. Shi, B. Zhang, GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics* **13**, S20 (2012).

231. F. Privé, C. Albiñana, J. Arbel, B. Pasaniuc, B. J. Vilhjálmsson, Inferring disease architecture and predictive ability with LDpred2-auto. *Am. J. Hum. Genet.* **110**, 2042–2055 (2023).

232. C. Park, J. Ha, S. Park, Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset. *Expert Syst. Appl.* **140**, 112873 (2020).

233. T. Lee, H. Lee, Prediction of Alzheimer's disease using blood gene expression data. *Sci. Rep.* **10**, 3485 (2020).

234. R. Sims, M. Hill, J. Williams, The multiplex model of the genetics of Alzheimer's disease. *Nat. Neurosci.* **23**, 311–322 (2020).

235. H. Chen, Y. He, J. Ji, Y. Shi, A Machine Learning Method for Identifying Critical Interactions Between Gene Pairs in Alzheimer's Disease Prediction. *Front. Neurol.* **10** (2019).

236. J. Li, T. Cai, Y. Jiang, H. Chen, X. He, C. Chen, *et al.*, Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Mol. Psychiatry* **21**, 298 (2016).

237. J. C. Darnell, S. J. Van Driesche, C. Zhang, K. Y. S. Hung, A. Mele, C. E. Fraser, *et al.*, FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* **146**, 247–261 (2011).

238. S. N. Basu, R. Kollu, S. Banerjee-Basu, AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* **37**, D832–D836 (2009).

239. M. J. Gandal, J. R. Haney, N. N. Parikshak, V. Leppa, G. Ramaswami, C. Hartl, *et al.*, Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).

240. N. N. Parikshak, V. Swarup, T. G. Belgard, M. Irimia, G. Ramaswami, M. J. Gandal, *et al.*, Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).

241. S. Gupta, S. E. Ellis, F. N. Ashar, A. Moes, J. S. Bader, J. Zhan, *et al.*, Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* **5**, 5748 (2014).

242. I. Voineagu, X. Wang, P. Johnston, J. K. Lowe, Y. Tian, S. Horvath, *et al.*, Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).

243. International Schizophrenia Consortium, Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).

244. M. Ayalew, H. Le-Niculescu, D. F. Levey, N. Jain, B. Changala, S. D. Patel, *et al.*, Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol. Psychiatry* **17**, 887–905 (2012).

245. C. M. Lewis, D. F. Levinson, L. H. Wise, L. E. DeLisi, R. E. Straub, I. Hovatta, *et al.*, Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am. J. Hum. Genet.* **73**, 34–48 (2003).

246. X. He, C. K. Fuller, Y. Song, Q. Meng, B. Zhang, X. Yang, *et al.*, Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am. J. Hum. Genet.* **92**, 667–680 (2013).

247. M. Y. M. Ng, D. F. Levinson, S. V. Faraone, B. K. Suarez, L. E. DeLisi, T. Arinami, *et al.*, Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol. Psychiatry* **14**, 774–785 (2009).

248. C. Chen, L. Cheng, K. Grennan, F. Pibiri, C. Zhang, J. A. Badner, *et al.*, Two gene co-expression modules differentiate psychotics and controls. *Mol. Psychiatry* **18**, 1308–1314 (2013).

Funding and Acknowledgments

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: Alexej Abyzov, Nadav Ahituv, Schahram Akbarian, Kristen Brennand, Andrew Chess, Gregory Cooper, Gregory Crawford, Stella Dracheva, Peggy Farnham, Michael Gandal, Mark Gerstein, Daniel Geschwind, Fernando Goes, Joachim F. Hallmayer, Vahram Haroutunian, Thomas M. Hyde, Andrew Jaffe, Peng Jin, Manolis Kellis, Joel Kleinman, James A. Knowles, Arnold Kriegstein, Chunyu Liu, Christopher E. Mason, Keri Martinowich, Eran Mukamel, Richard Myers, Charles Nemeroff, Mette Peters, Dalila Pinto, Katherine Pollard, Kerry Ressler, Panos Roussos, Stephan Sanders, Nenad Sestan, Pamela Sklar, Michael P. Snyder, Matthew State, Jason Stein, Patrick Sullivan, Alexander E. Urban, Flora Vaccarino, Stephen Warren, Daniel Weinberger, Sherman Weissman, Zhiping Weng, Kevin White, A. Jeremy Willsey, Hyejung Won, and Peter 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: Nikolaos Daskalakis. Additionally, Daifeng Wang was supported by R01AG067025, RF1MH128695, R21NS127432, R21NS128761, P50HD105353 (Waisman Center), National Science Foundation Career Award 2144475, and a Simons Foundation Autism Research Initiative pilot grant 971316, and Panos Roussos (Icahn School of Medicine at Mount Sinai and James J. Peters VA Medical Center) and Jaroslav Bendl (Icahn School of Medicine at Mount Sinai) were supported by the National Institute of Mental Health, NIH grants, RF1-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. Sophie Gaynor-Gillett was supported by 5U01MH116489. Jing Zhang was supported by R01HG012572 and R01NS128523. Michael J. Gandal was supported by NIMH R01-MH123922. This work was also supported by National Institutes of Health grant U01MH114812 (Ed S. Lein, Nikolas L. Jorstad, T Trygve E. Bakken), and by the National Institute on Aging grant U19AG060909 (Ed S. Lein, Kyle J. Travaglini). The authors thank the founder of the Allen Institute, P. G. Allen, for his vision, encouragement, and support. Rosemarie Terwilliger and Matthew 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. The research reported here was supported by the Department of Veterans Affairs, Veteran Health Administration, VISN1 Career Development Award, a Brain and Behavior Research Foundation Young Investigator Award, and an American Foundation for Suicide Prevention Young Investigator Award to Matthew J. Girgenti. 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 U.S. government or the views of the Department of Mental Health and Addiction Services or the State of Connecticut.

Author contributions

All individually named 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.; and 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 eleven 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 five corresponding authors co-led the analysis.

Competing interests

Z. Weng (UMass Chan Medical School) co-founded and serves as a scientific advisor for Rgenta Inc. From April 11, 2022, N.L. Jorstad (Allen Institute for Brain Science) has been an employee of Genentech. K.P.W. (National University of Singapore) is a shareholder in Tempus AI and Provaxus Inc. The other authors declare 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 for approved researchers at the PsychENCODE Knowledge portal (*108*). For the external cohorts, AMP-AD raw datasets and imputed genotypes are available for approved researchers at the AD Knowledge Portal (*109*). Girgenti-snMultiome datasets are deposited at NCBI GEO and will be available from the PIs upon reasonable request (*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 on 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.

List of supplementary materials

Materials and Methods
Figs. S1 to S90
Tables S1 to S18
Data S1 to S33
References (*112-248*)

Figure legends

Figure 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 unique to Ma-Sestan (Bottom) Schematic showing all samples labeled by disease, biological sex, ancestry, age, and available data modalities, including a distribution plot for sample ages (gray indicates pediatric samples excluded from most analyses). **(B)** 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 upregulation, red for downregulation). **(E)** Numbers of DE genes upregulated (blue) and downregulated (red) in older (>70 years) control (left) and schizophrenia (right) individuals per cell type when compared with younger individuals in each group (<70 years). “X” indicates no DE genes were observed 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. Inset highlights log-normalized gene expression in cells along the pseudotime axis for three genes.

Figure 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 (see 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 using STARR-seq assays of predicted

enhancers, comparing the $\log_2$-fold expression change of validated regions to non-validated regions (n.s.). **(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. **(F)** Enrichment (log-scaled FDR) of TF binding motifs among cell-type-specific 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.

Figure 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 gray 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 GO categories, with select GO categories highlighted. **(C)** Percent inter-individual 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. 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.

Figure 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, with normalized distributions highlighted for the union of scQTLs across excitatory, inhibitory, and non-neuronal cell types. **(E)** Boxplot showing a significantly higher enrichment of eSNPs in active STARR-seq peaks compared to the control group ($p<1.0\times10^{-4}$, Mann-Whitney U 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 shows astrocyte-specific eSNP for *SYNE1* along with chromatin accessibility (snATAC-seq) tracks for eight cell types. Top-right schematic shows the isoQTL for *LYPD6* in Pax6 inhibitory neurons, leading to altered expression of isoforms with different start codons. Middle and bottom-right 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 (□) values for the eSNP increase over pseudotime. Additional information is shown in fig. S53.

Figure 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 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. (Note 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 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 panels C, D, and F are colored uniquely according to names in panel G.

Figure 6. Constructing a cell-to-cell communication network. (A) Schematic for the construction of cell-to-cell communication networks, based on a matrix of co-expressed ligand-receptor gene pairs in signaling pathways between sender and receiver cell types. Circos plot on the right shows the strength of all identified cell-to-cell interactions, highlighting L5 IT to OPC cell types as an example. Note that 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, as 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, while 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.

Figure 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)** Log_2-fold changes and p-values from DESeq2 (*20*) for differentially expressed genes in older vs. younger individuals (±70 years) 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)** 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 (AUPRC) of cell-type-specific expression (bars) and methylation signatures (gray line) towards 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.

Figure 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 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-S32). Tree structures connect representative subgraphs (feature combinations) in each model (figs. S80-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. (Note, we shorten the readthrough transcript ANKHD1-EIF4EBP3 to ANKHD1 in ASD.) **(E)** UpSet plot for SCZ showing overlap between genes with the highest saliency per cell type or bulk expression, including four genes highlighted in panel 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 vs. unmatched, and correlations are calculated across imputed genes (*$p<0.05$, one-tailed t-test; fig. S88).

Figure 1

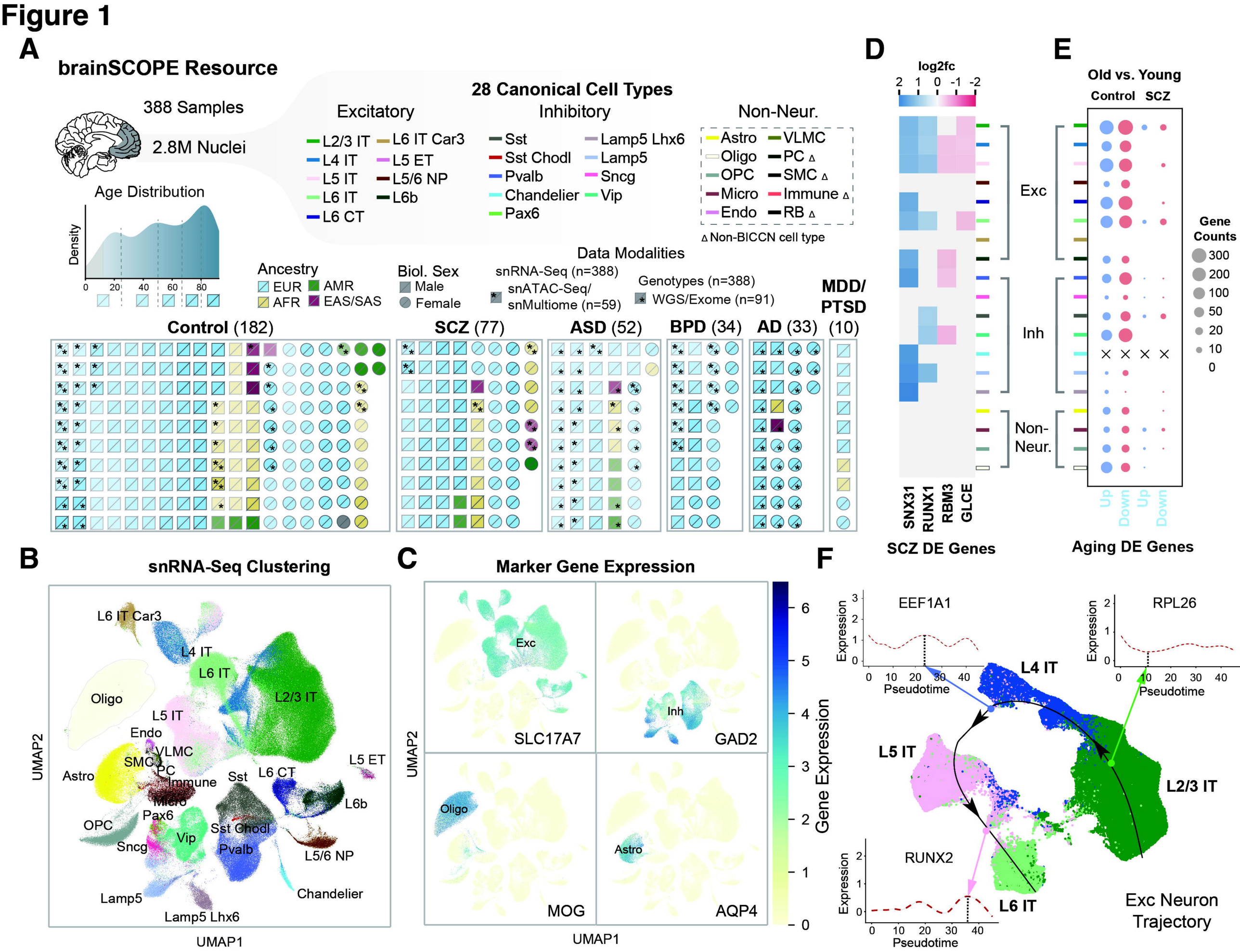

A
brainSCOPE Resource
388 Samples
2.8M Nuclei
Age Distribution
Density
0 20 40 60 80
28 Canonical Cell Types
Excitatory
L2/3 IT
L4 IT
L5 IT
L6 IT
L6 CT
L6 IT Car3
L5 ET
L5/6 NP
L6b
Inhibitory
Sst
Sst Chodl
Pvalb
Chandelier
Pax6
Lamp5 Lhx6
Lamp5
Sncg
Vip
Non-Neur.
Astro
Oligo
OPC
Micro
Endo
VLMC
PC Δ
SMC Δ
Immune Δ
RB Δ
Δ Non-BICCN cell type
Ancestry
EUR
AFR
AMR
EAS/SAS
Biol. Sex
Male
Female
Data Modalities
snRNA-Seq (n=388)
snATAC-Seq/ snMultiome (n=59)
Genotypes (n=388)
WGS/Exome (n=91)
Control (182)
SCZ (77)
ASD (52)
BPD (34)
AD (33)
MDD/ PTSD (10)
D
log2fc
2 1 0 -1 -2
Exc
Inh
Non-Neur.
SNX31
RUNX1
RBM3
GLCE
SCZ DE Genes
E
Old vs. Young
Control
SCZ
Gene Counts
300
200
100
50
20
10
0
Up
Down
Up
Down
Aging DE Genes
B
snRNA-Seq Clustering
UMAP2
UMAP1
L6 IT Car3
L4 IT
L6 IT
L2/3 IT
Oligo
L5 IT
Endo
VLMC
SMC
PC
Astro
Immune
Sst
L6 CT
L5 ET
Micro
L6b
Pax6
OPC
Sst Chodl
Vip
Sncg
Pvalb
L5/6 NP
Chandelier
Lamp5
Lamp5 Lhx6
C
Marker Gene Expression
Exc
SLC17A7
Inh
GAD2
Oligo
MOG
Astro
AQP4
Gene Expression
0 1 2 3 4 5 6
F
EEF1A1
RPL26
RUNX2
Expression
Pseudotime
0 10 20 30 40
L4 IT
L5 IT
L2/3 IT
L6 IT
Exc Neuron Trajectory

Figure 2

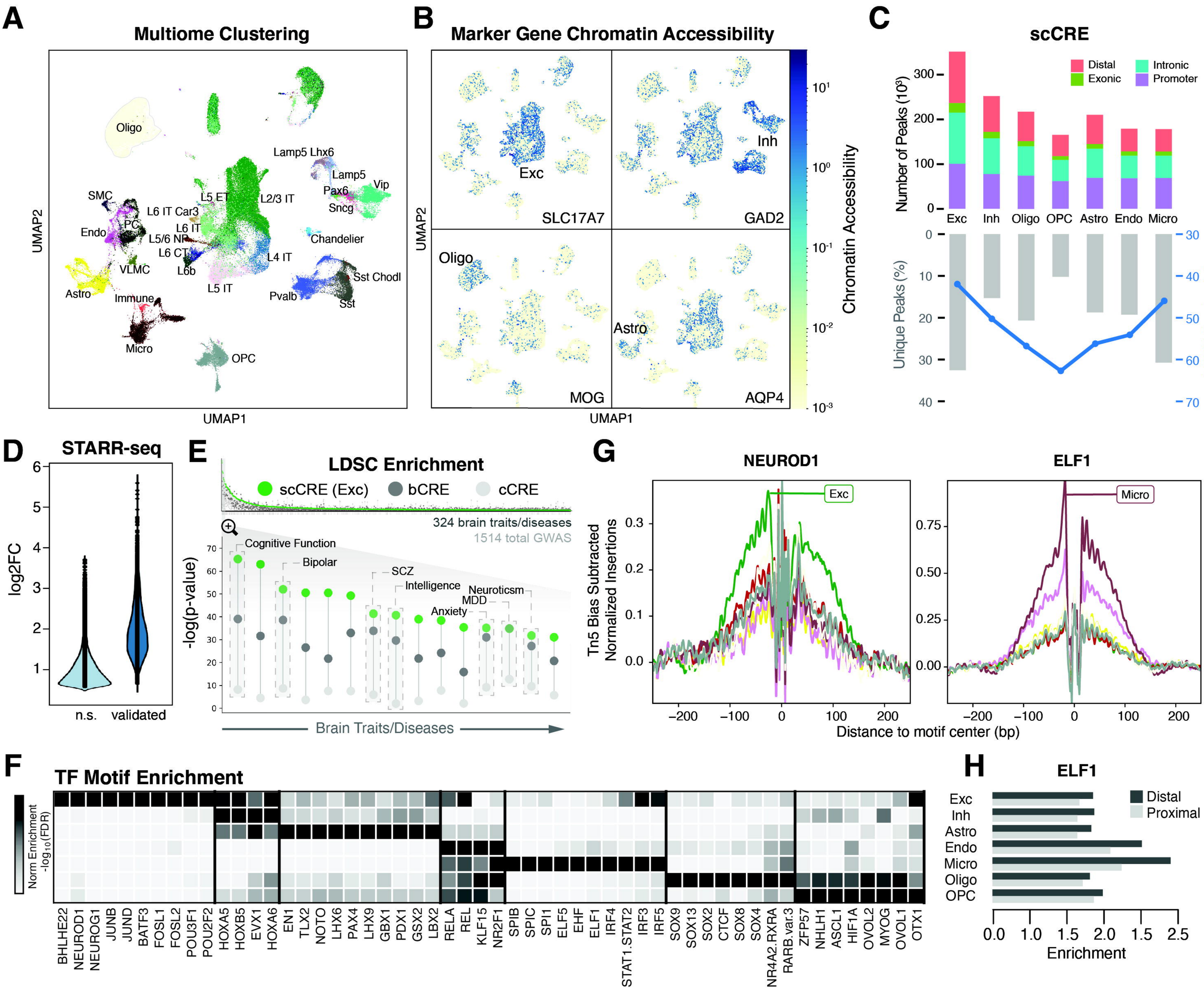

A
Multiome Clustering
UMAP2
UMAP1
Oligo
Lamp5 Lhx6
Lamp5
Vip
Pax6
Sncg
SMC
L5 ET
L2/3 IT
L6 IT Car3
PC
Endo
L6 IT
L5/6 NP
L6 CT
L6b
L4 IT
Chandelier
VLMC
Sst Chodl
Astro
Immune
L5 IT
Pvalb
Sst
Micro
OPC
B
Marker Gene Chromatin Accessibility
Exc
SLC17A7
Inh
GAD2
Oligo
MOG
Astro
AQP4
Chromatin Accessibility
C
scCRE
Distal
Exonic
Intronic
Promoter
Number of Peaks (10³)
Exc
Inh
Oligo
OPC
Astro
Endo
Micro
Unique Peaks (%)
bCRE Overlap (%)
D
STARR-seq
log2FC
n.s.
validated
E
LDSC Enrichment
scCRE (Exc)
bCRE
cCRE
324 brain traits/diseases
1514 total GWAS
-log(p-value)
Cognitive Function
Bipolar
SCZ
Intelligence
Neuroticsm
MDD
Anxiety
Brain Traits/Diseases
G
NEUROD1
ELF1
Exc
Micro
Tn5 Bias Subtracted Normalized Insertions
Distance to motif center (bp)
F
TF Motif Enrichment
Norm Enrichment -log10(FDR)
BHLHE22
NEUROD1
NEUROG1
JUNB
JUND
BATF3
FOSL1
FOSL2
POU3F1
POU2F2
HOXA5
HOXB5
EVX1
HOXA6
EN1
TLX2
NOTO
LHX6
PAX4
LHX9
GBX1
PDX1
GSX2
LBX2
RELA
REL
KLF15
NR2F1
SPIB
SPIC
SPI1
ELF5
EHF
ELF1
IRF4
STAT1.STAT2
IRF3
IRF5
SOX9
SOX13
SOX2
CTCF
SOX8
SOX4
NR4A2.RXRA
RARB.var.3
ZFP57
NHLH1
ASCL1
HIF1A
OVOL2
MYOG
OVOL1
OTX1
H
ELF1
Distal
Proximal
Exc
Inh
Astro
Endo
Micro
Oligo
OPC
Enrichment

Figure 3

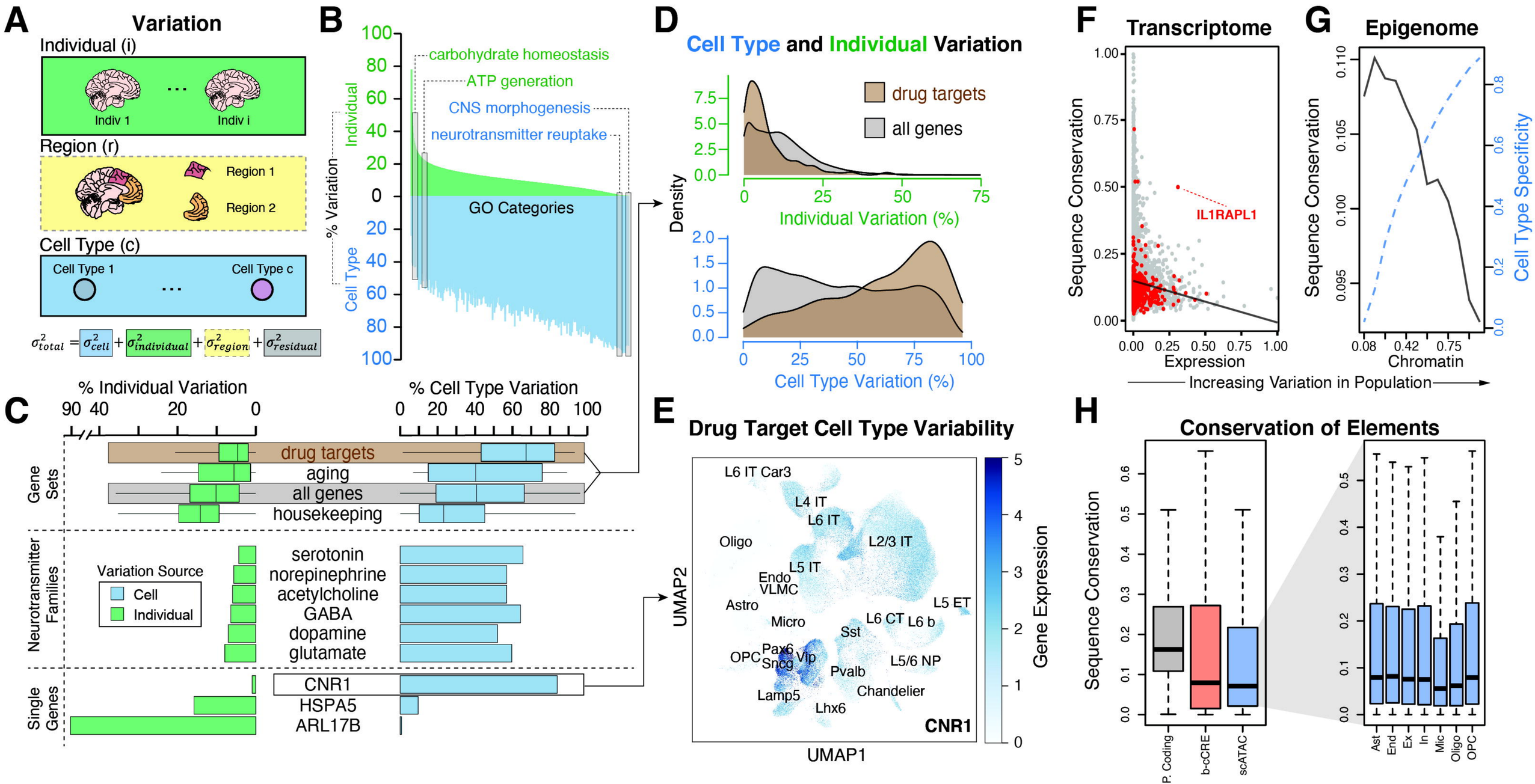

A
Variation
Individual (i)
Indiv 1
Indiv i
Region (r)
Region 1
Region 2
Cell Type (c)
Cell Type 1
Cell Type c
$\sigma^2_{total} = \sigma^2_{cell} + \sigma^2_{individual} + \sigma^2_{region} + \sigma^2_{residual}$
B
% Variation
Individual
Cell Type
carbohydrate homeostasis
ATP generation
CNS morphogenesis
neurotransmitter reuptake
GO Categories
C
% Individual Variation
% Cell Type Variation
Gene Sets
drug targets
aging
all genes
housekeeping
Neurotransmitter Families
serotonin
norepinephrine
acetylcholine
GABA
dopamine
glutamate
Variation Source
Cell
Individual
Single Genes
CNR1
HSPA5
ARL17B
D
Cell Type and Individual Variation
drug targets
all genes
Density
Individual Variation (%)
Cell Type Variation (%)
E
Drug Target Cell Type Variability
L6 IT Car3
L4 IT
L6 IT
Oligo
L2/3 IT
L5 IT
Endo
VLMC
Astro
L5 ET
Micro
L6 CT
L6 b
Sst
OPC
Pax6
Sncg
Vip
Pvalb
L5/6 NP
Lamp5
Chandelier
Lhx6
CNR1
UMAP1
UMAP2
Gene Expression
F
Transcriptome
IL1RAPL1
Sequence Conservation
Expression
G
Epigenome
Sequence Conservation
Chromatin
Cell Type Specificity
Increasing Variation in Population
H
Conservation of Elements
Sequence Conservation
P. Coding
b-cCRE
scATAC
Ast
End
Ex
In
Mic
Oligo
OPC

Figure 4

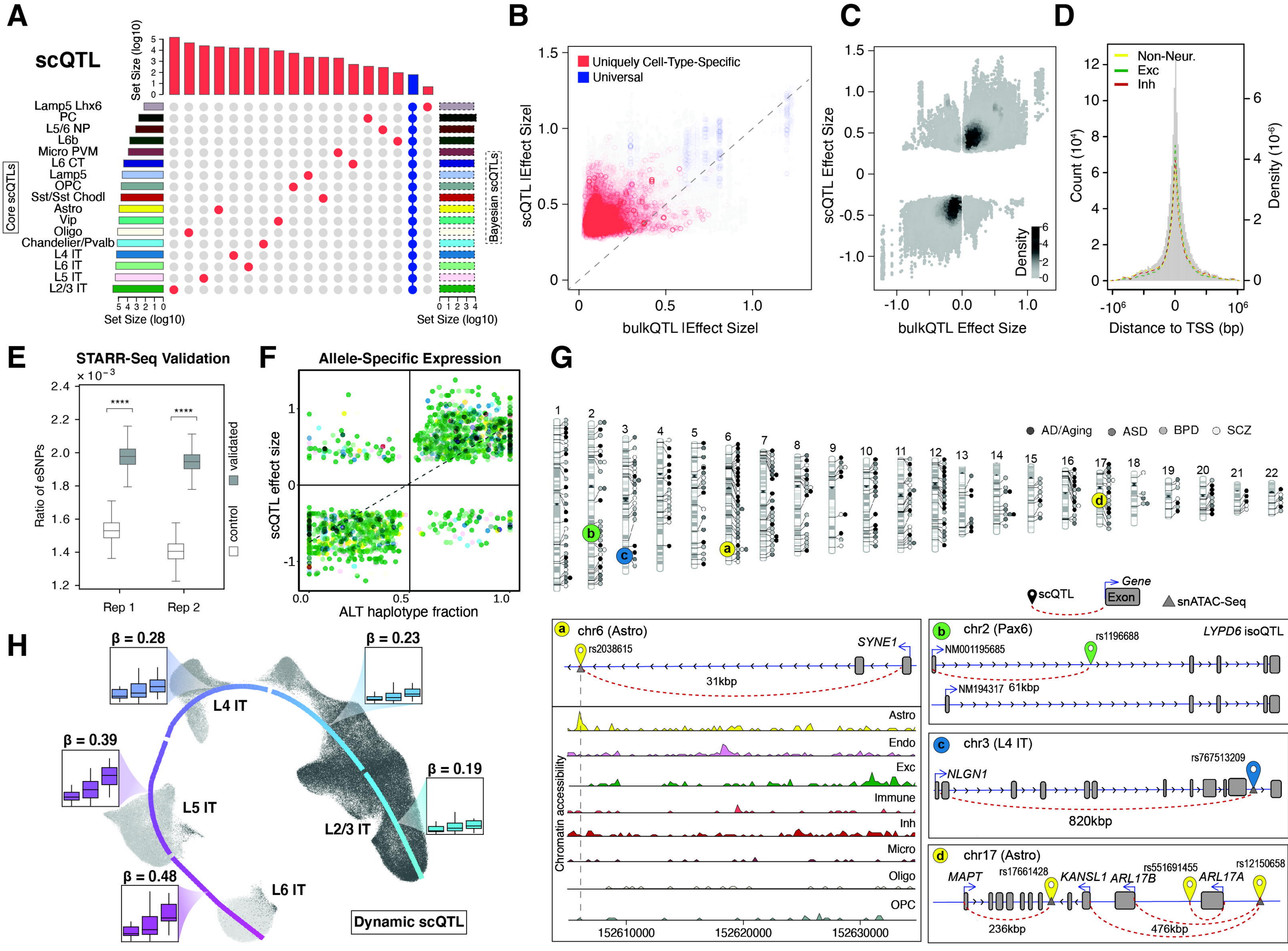

A
scQTL
Set Size (log10)
Lamp5 Lhx6
PC
L5/6 NP
L6b
Micro PVM
L6 CT
Lamp5
OPC
Sst/Sst Chodl
Astro
Vip
Oligo
Chandelier/Pvalb
L4 IT
L6 IT
L5 IT
L2/3 IT
Core scQTLs
Bayesian scQTLs
Set Size (log10)
Set Size (log10)
B
Uniquely Cell-Type-Specific
Universal
scQTL |Effect Size|
bulkQTL |Effect Size|
C
scQTL Effect Size
bulkQTL Effect Size
Density
D
Non-Neur.
Exc
Inh
Count (10⁴)
Density (10⁻⁶)
Distance to TSS (bp)
E
STARR-Seq Validation
Ratio of eSNPs
validated
control
Rep 1
Rep 2
F
Allele-Specific Expression
scQTL effect size
ALT haplotype fraction
G
AD/Aging
ASD
BPD
SCZ
scQTL
Gene
Exon
snATAC-Seq
a chr6 (Astro)
rs2038615
SYNE1
31kbp
Astro
Endo
Exc
Immune
Inh
Micro
Oligo
OPC
Chromatin accessibility
152610000
152620000
152630000
b chr2 (Pax6)
rs1196688
LYPD6 isoQTL
NM001195685
NM194317
61kbp
c chr3 (L4 IT)
rs767513209
NLGN1
820kbp
d chr17 (Astro)
MAPT
rs17661428
KANSL1
ARL17B
rs551691455
ARL17A
rs12150658
236kbp
476kbp
H
β = 0.28
β = 0.23
β = 0.39
β = 0.19
β = 0.48
L4 IT
L5 IT
L2/3 IT
L6 IT
Dynamic scQTL

Figure 5

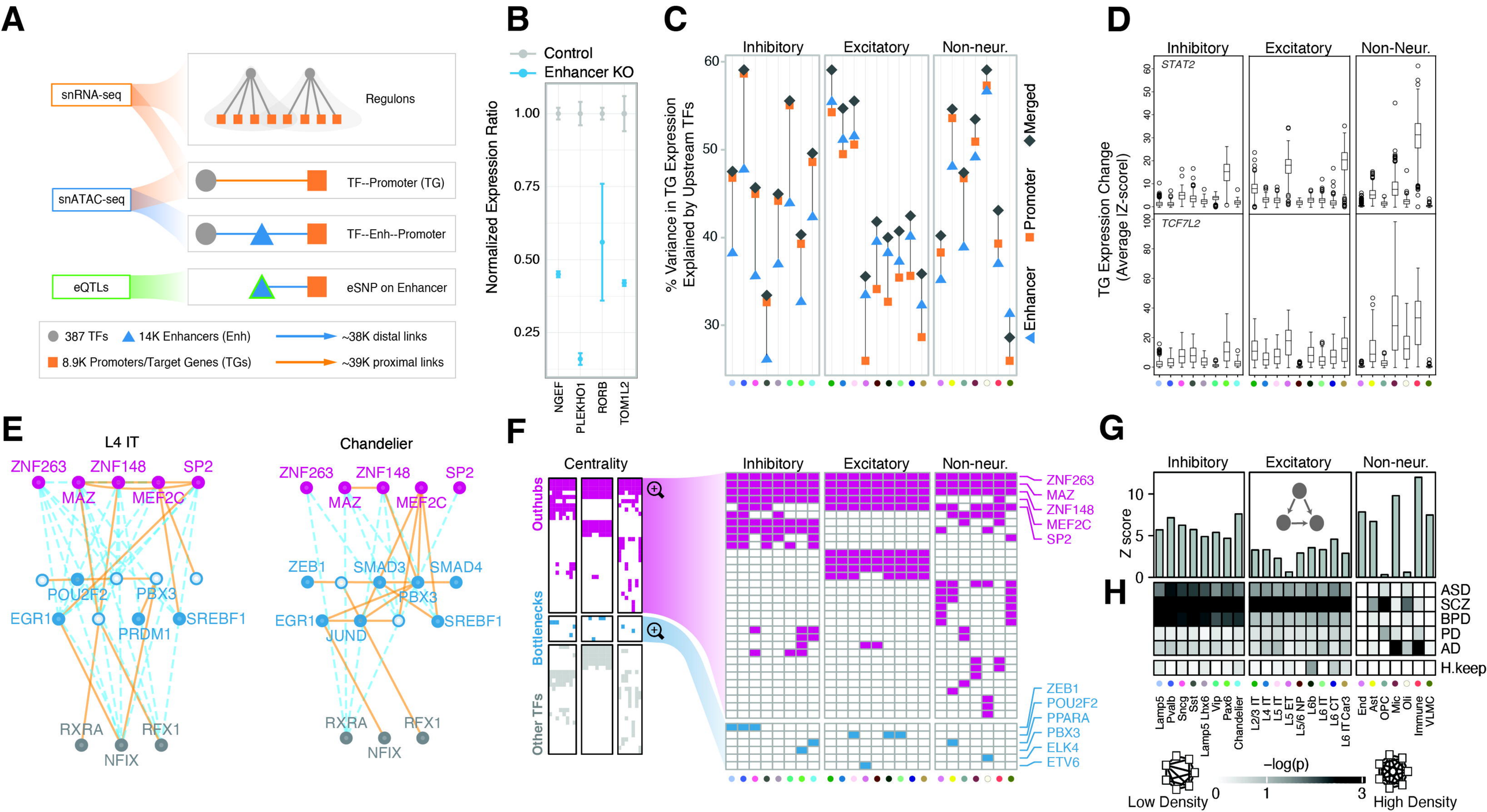

A
snRNA-seq
Regulons
snATAC-seq
TF--Promoter (TG)
TF--Enh--Promoter
eQTLs
eSNP on Enhancer
387 TFs
14K Enhancers (Enh)
~38K distal links
8.9K Promoters/Target Genes (TGs)
~39K proximal links
B
Control
Enhancer KO
Normalized Expression Ratio
1.00
0.75
0.50
0.25
NGEF
PLEKHO1
RORB
TOM1L2
C
Inhibitory
Excitatory
Non-neur.
% Variance in TG Expression Explained by Upstream TFs
60
50
40
30
Merged
Promoter
Enhancer
D
Inhibitory
Excitatory
Non-Neur.
STAT2
TCF7L2
TG Expression Change (Average |Z-score|)
E
L4 IT
ZNF263
ZNF148
SP2
MAZ
MEF2C
POU2F2
PBX3
EGR1
SREBF1
PRDM1
RXRA
RFX1
NFIX
Chandelier
ZNF263
ZNF148
SP2
MAZ
MEF2C
ZEB1
SMAD3
SMAD4
PBX3
EGR1
SREBF1
JUND
RXRA
RFX1
NFIX
F
Centrality
Outhubs
Bottlenecks
Other TFs
Inhibitory
Excitatory
Non-neur.
ZNF263
MAZ
ZNF148
MEF2C
SP2
ZEB1
POU2F2
PPARA
PBX3
ELK4
ETV6
G
Inhibitory
Excitatory
Non-neur.
Z score
10
5
0
H
ASD
SCZ
BPD
PD
AD
H.keep
Lamp5
Pvalb
Sncg
Sst
Lamp5 Lhx6
Vip
Pax6
Chandelier
L2/3 IT
L4 IT
L5 IT
L5 ET
L5/6 NP
L6b
L6 IT
L6 CT
L6 IT Car3
End
Ast
OPC
Mic
Oli
Immune
VLMC
Low Density
−log(p)
0
1
3
High Density

Figure 6

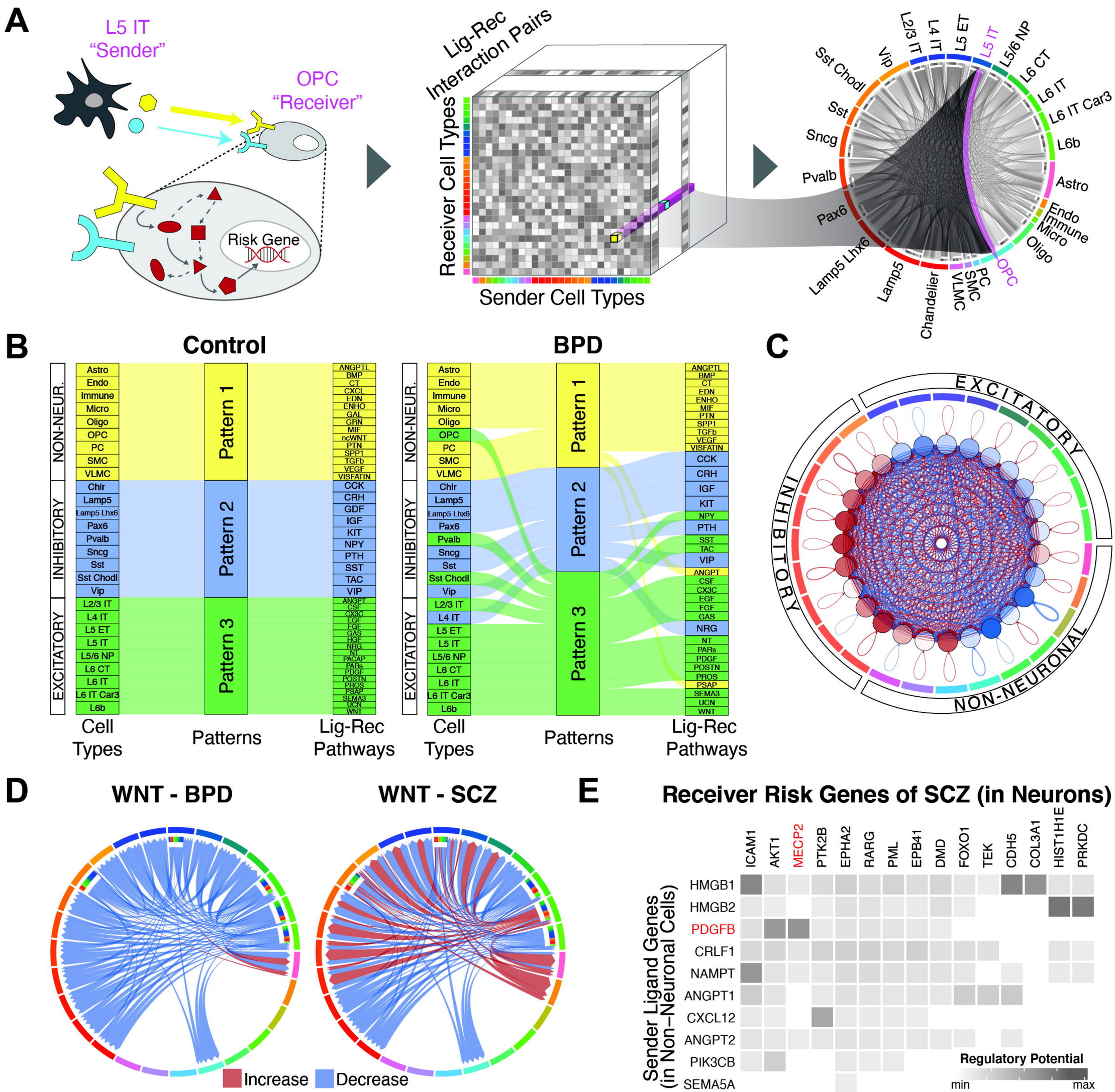
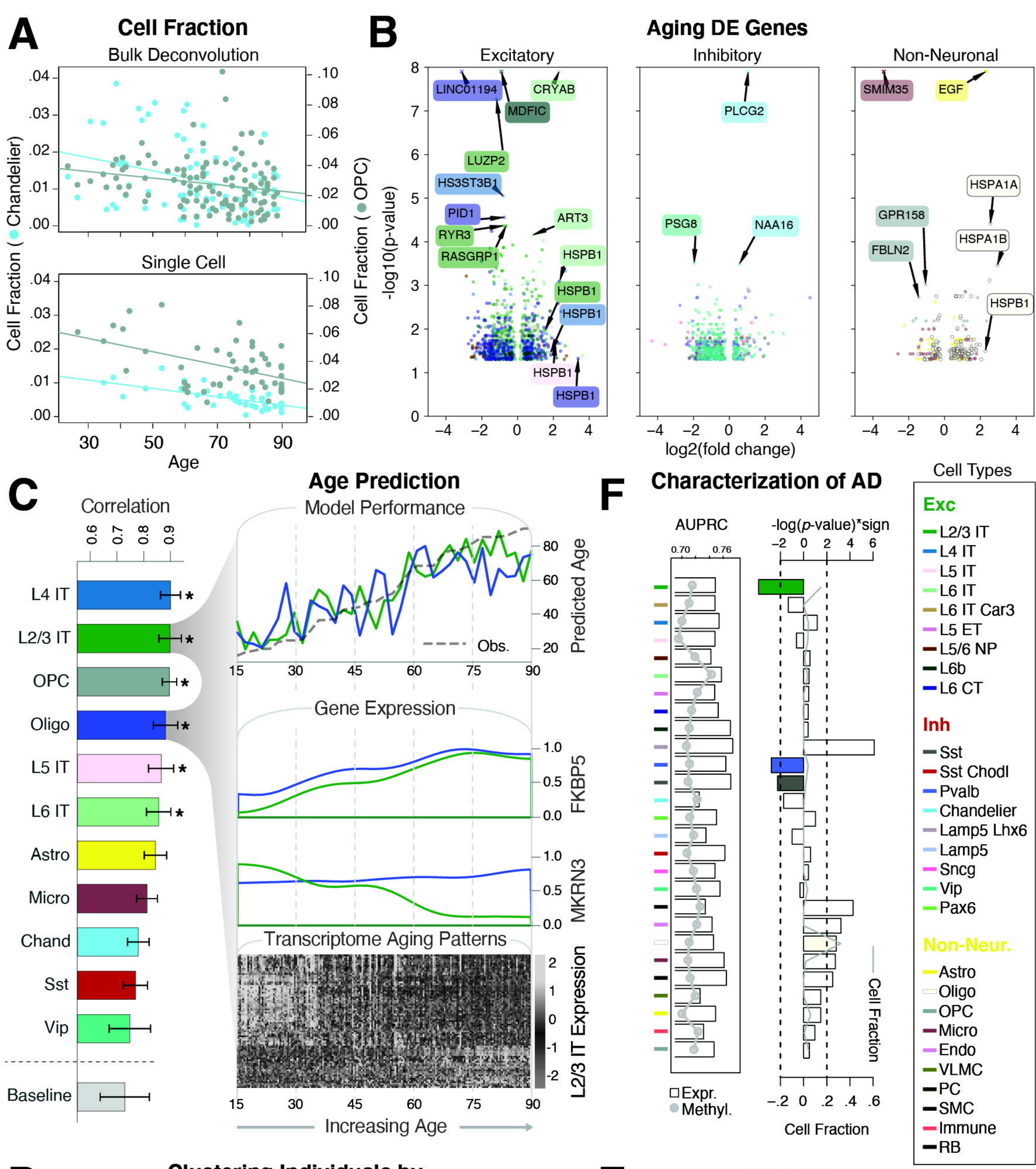

A
L5 IT
"Sender"
OPC
"Receiver"
Risk Gene
Lig-Rec
Interaction Pairs
Receiver Cell Types
Sender Cell Types
Sst Chodl
Sst
Sncg
Pvalb
Pax6
Lamp5 Lhx6
Lamp5
Chandelier
VLMC
SMC
PC
OPC
Oligo
Micro
Immune
Endo
Astro
L6b
L6 IT Car3
L6 IT
L6 CT
L5/6 NP
L5 IT
L5 ET
L4 IT
L2/3 IT
Vip
B
Control
BPD
NON-NEUR.
INHIBITORY
EXCITATORY
Astro
Endo
Immune
Micro
Oligo
OPC
PC
SMC
VLMC
Chlr
Lamp5
Lamp5 Lhx6
Pax6
Pvalb
Sncg
Sst
Sst Chodl
Vip
L2/3 IT
L4 IT
L5 ET
L5 IT
L5/6 NP
L6 CT
L6 IT
L6 IT Car3
L6b
Pattern 1
Pattern 2
Pattern 3
ANGPTL
BMP
CT
CXCL
EDN
ENHO
GAL
GRN
MIF
ncWNT
PTN
SPP1
TGFb
VEGF
VISFATIN
CCK
CRH
GDF
IGF
KIT
NPY
PTH
SST
TAC
VIP
ANGPT
CSF
CX3C
EGF
FGF
GAS
HGF
NRG
NT
PACAP
PARs
PDGF
POSTN
PROS
PSAP
SEMA3
UCN
WNT
Cell Types
Patterns
Lig-Rec Pathways
C
EXCITATORY
INHIBITORY
NON-NEURONAL
D
WNT - BPD
WNT - SCZ
Increase
Decrease
E
Receiver Risk Genes of SCZ (in Neurons)
ICAM1
AKT1
MECP2
PTK2B
EPHA2
RARG
PML
EPB41
DMD
FOXO1
TEK
CDH5
COL3A1
HIST1H1E
PRKDC
Sender Ligand Genes (in Non-Neuronal Cells)
HMGB1
HMGB2
PDGFB
CRLF1
NAMPT
ANGPT1
CXCL12
ANGPT2
PIK3CB
SEMA5A
Regulatory Potential
min
max

Figure 7


A
Cell Fraction
Bulk Deconvolution
Single Cell
Cell Fraction (● Chandelier)
Cell Fraction (● OPC)
Age
B
Aging DE Genes
Excitatory
Inhibitory
Non-Neuronal
-log10(p-value)
log2(fold change)
LINC01194
CRYAB
MDFIC
LUZP2
HS3ST3B1
PID1
RYR3
RASGRP1
ART3
HSPB1
PLCG2
PSG8
NAA16
SMIM35
EGF
HSPA1A
GPR158
HSPA1B
FBLN2
C
Correlation
L4 IT
L2/3 IT
OPC
Oligo
L5 IT
L6 IT
Astro
Micro
Chand
Sst
Vip
Baseline
Age Prediction
Model Performance
Predicted Age
Obs.
Gene Expression
FKBP5
MKRN3
Transcriptome Aging Patterns
L2/3 IT Expression
Increasing Age
F
Characterization of AD
AUPRC
-log(p-value)*sign
Expr.
Methyl.
Cell Fraction
Cell Types
Exc
L2/3 IT
L4 IT
L5 IT
L6 IT
L6 IT Car3
L5 ET
L5/6 NP
L6b
L6 CT
Inh
Sst
Sst Chodl
Pvalb
Chandelier
Lamp5 Lhx6
Lamp5
Sncg
Vip
Pax6
Non-Neur.
Astro
Oligo
OPC
Micro
Endo
VLMC
PC
SMC
Immune
RB

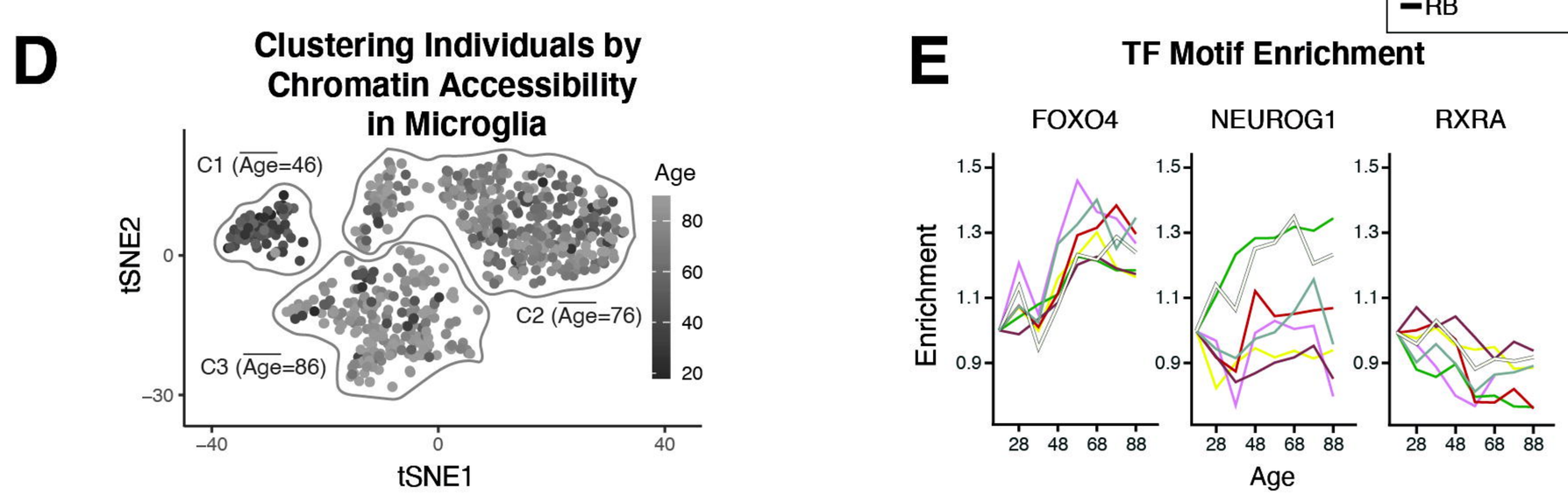
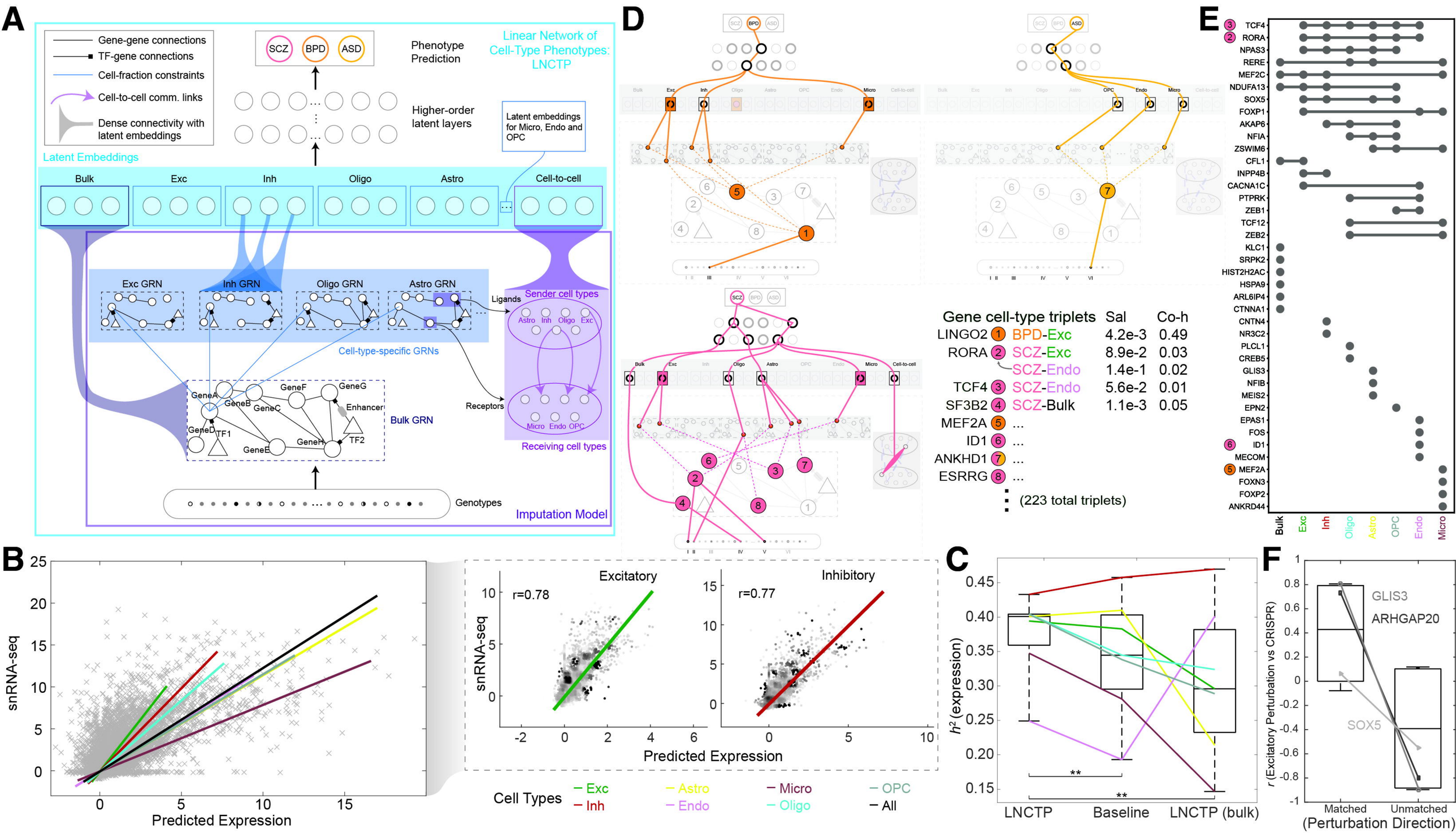

D
Clustering Individuals by Chromatin Accessibility in Microglia
C1 (Age=46)
C2 (Age=76)
C3 (Age=86)
Age
tSNE1
tSNE2
E
TF Motif Enrichment
FOXO4
NEUROG1
RXRA
Enrichment
Age

Figure 8


A
Gene-gene connections
TF-gene connections
Cell-fraction constraints
Cell-to-cell comm. links
Dense connectivity with latent embeddings
SCZ
BPD
ASD
Phenotype Prediction
Linear Network of Cell-Type Phenotypes: LNCTP
Higher-order latent layers
Latent embeddings for Micro, Endo and OPC
Latent Embeddings
Bulk
Exc
Inh
Oligo
Astro
Cell-to-cell
Exc GRN
Inh GRN
Oligo GRN
Astro GRN
Ligands
Sender cell types
Astro Inh Oligo Exc
Cell-type-specific GRNs
GeneA
GeneB
GeneC
GeneD
GeneE
GeneF
GeneG
GeneH
Enhancer
TF1
TF2
Bulk GRN
Receptors
Micro Endo OPC
Receiving cell types
Genotypes
Imputation Model
D
Gene cell-type triplets Sal Co-h
LINGO2 1 BPD-Exc 4.2e-3 0.49
RORA 2 SCZ-Exc 8.9e-2 0.03
SCZ-Endo 1.4e-1 0.02
TCF4 3 SCZ-Endo 5.6e-2 0.01
SF3B2 4 SCZ-Bulk 1.1e-3 0.05
MEF2A 5 ...
ID1 6 ...
ANKHD1 7 ...
ESRRG 8 ...
(223 total triplets)
E
TCF4
RORA
NPAS3
RERE
MEF2C
NDUFA13
SOX5
FOXP1
AKAP6
NFIA
ZSWIM6
CFL1
INPP4B
CACNA1C
PTPRK
ZEB1
TCF12
ZEB2
KLC1
SRPK2
HIST2H2AC
HSPA9
ARL6IP4
CTNNA1
CNTN4
NR3C2
PLCL1
CREB5
GLIS3
NFIB
MEIS2
EPN2
EPAS1
FOS
ID1
MECOM
MEF2A
FOXN3
FOXP2
ANKRD44
Bulk Exc Inh Oligo Astro OPC Endo Micro
B
snRNA-seq
Predicted Expression
Excitatory
r=0.78
Inhibitory
r=0.77
Cell Types
Exc
Inh
Astro
Endo
Micro
Oligo
OPC
All
C
h^2 (expression)
LNCTP
Baseline
LNCTP (bulk)
F
r (Excitatory Perturbation vs CRISPR)
GLIS3
ARHGAP20
SOX5
Matched
Unmatched
(Perturbation Direction)