HHS Public Access

Single-cell genomics and regulatory networks for 388 human brains

A full list of authors and affiliations appears at the end of the article.

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

[*] Corresponding authors.

[†] Co-first authors, these authors contributed equally to this work.

Author contributions

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. 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 seven senior authors played instrumental roles in this study: M.J.Ga., D.L., E.S.L., P.R., N.Se., Z.W., K.P.W., and H.W. The following five corresponding authors co-led the analysis: M.J.Gi., J.Z., D.W., D.Ge., and M.Ge.

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, 22) (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 (23) (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 (24, 25). We also found that individuals with schizophrenia differ from controls with respect to the number of aging DE genes, which may reflect the increased expression variability in schizophrenia patients (Fig. 1E; fig. S19).

Our snRNA-seq data also recapitulates the spatial relationships among cell types in the PFC. Fig. 1F shows a cell-trajectory analysis (26, 27) across four subclasses of excitatory neurons in controls. We found smoothed patterns of gene-expression variation along the cortical-depth axis (specifically for 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 (28, 29). Overall, we found 76 genes with significant variation (FDR<0.05, Wald test) across cortical layers, including several genes involved in neural development such as *SEMA6A*, *RUNX2*, *SOX6*, and *PROX1* (figs. S20–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, 30). Following the ENCODE (Encyclopedia of DNA Elements) convention (31), we call these scCREs (single-cell candidate cis-Regulatory Elements). About half of these are cell-type-specific and located distal to genes (fig. S25). We validated the functionality of select scCREs using targeted STARR-seq (Fig. 2D; (20, 32)).

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 (31). As expected, they show strong overlap with scCREs (Fig. 2C).

To identify how our cell-type-specific regulatory elements relate to genetic associations, we performed a LDSC (linkage-disequilibrium score regression) analysis (20, 33). 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) (34–37).

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 (38) (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)) (39). 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=3.7×10^{-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 (40, 41).

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) (42). 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 (43); *IL1RAPL1* is highly expressed in the brain and has been implicated in intellectual disability and ASD (44).

We also leveraged our snATAC-seq profiles to deconvolve population-scale chromatin data (fig. S33; (20)). Similar to the transcriptome, open chromatin regions with higher sequence conservation have less variability in their chromatin openness (Fig. 3G; 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, 45, 46).

Determining cell-type-specific eQTLs from single-cell data

To evaluate cell-type expression variation in more detail, we used our processed snRNA-seq data to identify single-cell cis-eQTLs (hereafter referred to as "scQTLs"). We followed the same general procedure used by GTEx (5), including conservative filtering at the cell-type level when generating pseudobulk data (20). We used this set of scQTLs as our "core callset," with the objective of facilitating consistent comparisons with those from existing datasets (such as GTEx and PsychENCODE bulk data) (data S15). Note the sparsity intrinsic to snRNA-seq data reduces power, particularly for rarer cell types (fig. S36; table S7; (20)) (47). To ameliorate the low power, we developed a Bayesian linear mixed-effects model to

identify more scQTLs for rare cell types as an additional callset (Fig. 4A; figs. S36C, 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, 32)). As further validation, we were able to identify allele-specific expression (ASE) at the single-cell level in samples with WGS-based phased variants (Fig. 4F; fig. S49; (20)). Determination of single-cell ASE is particularly challenging due the sparsity of the data (48–52). 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 (53, 54), and *NLGN1*, a candidate gene for multiple brain disorders encoding a ligand for neurexin signaling (55). We also found multiple scQTLs within the complex 17q21.31 locus related to brain disorders, including an astrocyte-specific scQTL for the Tau protein gene *MAPT* and a multi-cell type scQTL for the neurodegenerative-disorder risk gene *KANSL1* (Fig. 4G) (40). We further highlight an iso-QTL for *LYPD6*, which inhibits acetylcholine-receptor activity in Pax6-type inhibitory neurons (56) (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)) (57). 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 (58, 59), 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 (60) (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 (61), 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) (37, 62, 63).

Constructing a cell-to-cell communication network

To further understand cellular signaling and regulation, we leveraged publicly available ligand-receptor pairs (64) in combination with our snRNA-seq data to construct a cell-to-cell

communication network (Fig. 6A; 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 (65–67) (Fig. 6B).

We next explored how cell-cell communication patterns are altered in individuals with neuropsychiatric disorders, finding that they are greatly changed for schizophrenia and bipolar disorder (Fig. 6B; 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) (68–71). Mechanistically, this downregulation could result in the overactivity of the lithium-targeted GSK3β enzyme in neurons (72, 73). 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 (74). We further found downregulation of PTN pathway interactions from glial cells to neurons, consistent with previous studies (75–77), 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 (78).

Lastly, we extended our extracellular cell-to-cell communication analysis by considering related disruptions to intracellular signaling pathways (Fig. 6E; fig. S73; (20)) (79). By utilizing disease-risk genes and setting support cells (non-neurons) as the senders and neurons as the receivers, we identified ligand-receptor links connecting known risk genes to potential upstream effectors. For instance, we linked *FOXP1* and its ligand *EBI3* in bipolar disorder and *MECP2* and its ligand *PDGFB* in schizophrenia (80, 81).

Assessing cell-type-specific transcriptomic and epigenetic changes in aging

We used our population-scale single-cell data to systematically highlight transcriptomic and epigenetic changes due to aging. First, we assessed cell-fraction changes 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) (82, 83). This result is consistent with findings from raw cell counts in the single-cell data (FDR<0.05, two-sided t-test; Fig. 7A; 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 (84, 85).

To further explore the relationship between the transcriptome and aging, we constructed a model to predict an individual's age from their single-cell expression data (Fig. 7C; 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) (86–88).

We also investigated the effects of age on the epigenome using our scCREs to deconvolve bulk chromatin accessibility for 628 individuals into those for specific cell types (Fig. 7D; 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)) (89). 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 (90) (Fig. 7F). We compared this result with that from directly comparing cell-type-specific gene-expression and methylation signatures to determine case-control status (91), finding that the fractions and signatures capture independent information (fig. S79B; 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, 92, 93). The baseline does not include our derived GRNs and cell-to-cell networks, so the improvement represents the additional predictive performance possible with these networks (Fig. 8C). Moreover, the inclusion of imputed single-cell gene expression data also improves the overall prediction accuracy of disorders (discussed below) and accounts for a larger fraction of common-SNP heritability of these disorders beyond predictions based solely on bulk expression or polygenic risk scores (94) (table S13).

We exploited the ability of the LNCTP to impute missing data for discovery of cell-type-specific molecular phenotypes important for neuropsychiatric disorders. Doing so allowed us to link variants with their "intermediate" functional genomic activities, such as cell-type-specific gene expression, pathway activity, and cell-cell communication. We used a hierarchical linear architecture for the trait-prediction portion of the LNCTP, which performed comparably to or better than 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 (95), is important for neurons in schizophrenia (96), *LINGO2* is important for excitatory neurons in bipolar disorder, and *ANKHD1* is highly weighted in ASD, supporting current hypotheses (97, 98). 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 (99). 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) (100).

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

We further used the LNCTP to perform in silico perturbation analysis, 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 (42) 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Prashant S. Emani[1,2,†], Jason J. Liu[1,2,†], Declan Clarke[1,2,†], Matthew Jensen[1,2,†], Jonathan Warrell[1,2,†], Chirag Gupta[3,4,†], Ran Meng[1,2,†], Che Yu Lee[5,†], Siwei Xu[5,†], Cagatay Dursun[1,2,†], Shaoke Lou[1,2,†], Yuhang Chen[1,2], Zhiyuan Chu[1], Timur Galeev[1,2], Ahyeon Hwang[5,6], Yunyang Li[2,7], Pengyu Ni[1,2], Xiao Zhou[1,2], PsychENCODE Consortium, Trygve E. Bakken[8], Jaroslav Bendl[9,10,11,12], Lucy Bicks[13], Tanima Chatterjee[1,2], Lijun Cheng[14], Yuyan Cheng[13,15], Yi Dai[5], Ziheng Duan[5], Mary Flaherty[14], John F. Fullard[9,10,11,12], Michael Gancz[1,2], Diego Garrido-Martín[16], Sophia Gaynor-Gillett[14,17], Jennifer Grundman[13], Natalie Hawken[13], Ella Henry[1,2], Gabriel E. Hoffman[9,10,11,12,18,19], Ao Huang[1], Yunzhe Jiang[1,2], Ting Jin[3,4], Nikolas L. Jorstad[8], Riki Kawaguchi[13,20], Saniya Khullar[3,4], Jianyin Liu[13], Junhao Liu[5], Shuang Liu[4], Shaojie Ma[21,22], Michael Margolis[13], Samantha Mazariegos[13], Jill Moore[23], Jennifer R. Moran[14], Eric Nguyen[1,2], Nishigandha Phalke[23], Milos Pjanic[9,10,11,12], Henry Pratt[23], Diana Quintero[13], Ananya S. Rajagopalan[7], Tiernon R. Riesenmy[24], Nicole Shedd[23], Manman Shi[14], Megan Spector[14], Rosemarie Terwilliger[25], Kyle J. Travaglini[8], Brie Wamsley[13], Gaoyuan Wang[1,2], Yan Xia[1,2], Shaohua Xiao[13], Andrew C. Yang[1,2], Suchen Zheng[1,2], Michael J. Gandal[26,27,28,29,30], Donghoon Lee[9,10,11,12], Ed S. Lein[8,31,32], Panos Roussos[9,10,11,12,18,19], Nenad Sestan[21], Zhiping Weng[23], Kevin P. White[33], Hyejung Won[34], Matthew J. Girgenti[25,35,36,*], Jing Zhang[5,*], Daifeng Wang[3,4,37,*], Daniel Geschwind[13,20,27,28,38,*], Mark Gerstein[1,2,7,24,39,*]

Affiliations

[1]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, 06520, USA.

[2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 06520, USA.

[3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, 53706, USA.

[4]Waisman Center, University of Wisconsin-Madison, Madison, WI, 53705, USA.

[5]Department of Computer Science, University of California, Irvine, CA, 92697, USA.

[6]Mathematical, Computational and Systems Biology, University of California, Irvine, CA, 92697, USA.

[7]Department of Computer Science, Yale University, New Haven, CT, 06520, USA.

[8]Allen Institute for Brain Science, Seattle, WA, 98109, USA.

[9]Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.

[10]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.

[11]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.

[12]Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA.

[13]Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA.

[14]Tempus Labs, Inc., Chicago, IL, 60654, USA.

[15]Department of Opthalmology, Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.

[16]Department of Genetics, Microbiology and Statistics, Universitat de Barcelona, Barcelona, 08028, Spain.

[17]Department of Biology, Cornell College, Mount Vernon, IA, 52314, USA.

[18]Mental Illness Research Education and Clinical Center, James J. Peters VA Medical Center, Bronx, NY, 10468, USA.

[19]Center for Precision Medicine and Translational Therapeutics, James J. Peters VA Medical Center, Bronx, NY, 10468, USA.

[20]Center for Autism Research and Treatment, Semel Institute, University of California, Los Angeles, CA, 90095, USA.

[21]Department of Neuroscience, Yale University, New Haven, CT, 06510, USA.

[22]Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, 200031, China.

[23]Department of Genomics and Computational Biology, UMass Chan Medical School, Worcester, MA, 01605, USA.

[24]Department of Statistics & Data Science, Yale University, New Haven, CT, 06520, USA.

[25]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, 06520, USA.

[26]Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA, 90095, USA.

[27]Department of Psychiatry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, 90095, USA.

[28]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, 90095, USA.

[29]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.

[30]Lifespan Brain Institute, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.

[31]Department of Neurological Surgery, University of Washington, Seattle, WA, 98195, USA.

[32]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, 98195, USA.

[33]Yong Loo Lin School of Medicine, National University of Singapore, 117597, Singapore.

[34]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA.

[35]Wu Tsai Institute, Yale University, New Haven, CT, 06520, USA.

[36]Clinical Neuroscience Division, National Center for Posttraumatic Stress Disorder, Veterans Affairs Connecticut Healthcare System, West Haven, CT, 06516, USA.

[37]Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, 53706, USA.

[38]Institute for Precision Health, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA.

[39]Department of Biomedical Informatics & Data Science, Yale University, New Haven, CT, 06520, USA.

Acknowledgments

Data and materials availability

The brainSCOPE resource was developed from raw sequencing data (snRNA-Seq, snATAC-Seq, snMultiome, and genotype) derived from 12 individual cohorts, including eight PsychENCODE cohorts and four external cohorts. Raw datasets for the PsychENCODE cohorts, as well as protected-access integrated datasets such as imputed genotypes, are available at the PsychENCODE Knowledge portal (108). For the external cohorts, AMP-AD raw datasets and imputed genotypes are available at the AD Knowledge Portal (109). Girgenti-snMultiome datasets are deposited at NCBI GEO (GSE261983) (110). Ma-Sestan and Velmeshev datasets are available from their respective publications (18, 19, 111). Other key resources and additional datasets used in the integrative analysis are available in the supplementary materials (for smaller datasets) or 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 (112).

References

1. Sullivan PF, Geschwind DH, Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. Cell 177, 162–183 (2019). [PubMed: 30901538]

2. PsychENCODE Consortium, Revealing the brain's molecular architecture. Science 362, 1262–1263 (2018). [PubMed: 30545881]

3. Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH, The road to precision psychiatry: translating genetics into disease mechanisms. Nat. Neurosci. 19, 1397–1407 (2016). [PubMed: 27786179]

4. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, 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). [PubMed: 32913098]

6. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, et al. , An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat. Neurosci. 20, 1418–1426 (2017). [PubMed: 28869584]

7. Liu S, Won H, Clarke D, Matoba N, Khullar S, Mu Y, et al. , Illuminating links between cis-regulators and trans-acting variants in the human prefrontal cortex. Genome Med. 14, 133 (2022). [PubMed: 36424644]

8. Bryois J, Calini D, Macnair W, Foo L, Urich E, Ortmann W, 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). [PubMed: 35915177]

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

10. Zeng B, Bendl J, Kosoy R, Fullard JF, Hoffman GE, Roussos P, Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. Nat. Genet. 54, 161–169 (2022). [PubMed: 35058635]

11. Zhang K, Hocker JD, Miller M, Hou X, Chiou J, Poirion OB, et al. , A single-cell atlas of chromatin accessibility in the human genome. Cell 184, 5985–6001.e19 (2021). [PubMed: 34774128]

12. BRAIN Initiative Cell Census Network (BICCN), A multimodal cell census and atlas of the mammalian primary motor cortex. Nature 598, 86–102 (2021). [PubMed: 34616075]

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

14. Zeng H, What is a cell type and how to define it? Cell 185, 2739–2755 (2022). [PubMed: 35868277]

15. La Manno G, Siletti K, Furlan A, Gyllborg D, Vinsland E, Mossi Albiach A, et al. , Molecular architecture of the developing mouse brain. Nature 596, 92–96 (2021). [PubMed: 34321664]

16. Song M, Yang X, Ren X, Maliskova L, Li B, Jones IR, et al. , Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. Nat. Genet. 51, 1252–1262 (2019). [PubMed: 31367015]

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

18. Velmeshev D, Schirmer L, Jung D, Haeussler M, Perez Y, Mayer S, et al. , Single-cell genomics identifies cell type-specific molecular changes in autism. Science 364, 685–689 (2019). [PubMed: 31097668]

19. Ma S, Skarica M, Li Q, Xu C, Risgaard RD, Tebbenkamp ATN, 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. Pantazopoulos H, Wiseman JT, Markota M, Ehrenfeld L, Berretta S, 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). [PubMed: 27259817]

22. Lin L-C, Sibille E, Reduced brain somatostatin in mood disorders: a common pathophysiological substrate and drug target? Front. Pharmacol. 4, 110 (2013). [PubMed: 24058344]

23. Love MI, Huber W, Anders S, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]

24. Ruzicka WB, Mohammadi S, Fullard JF, Davila-Velderrain J, Subburaju S, Tso DR, et al. , "Single-cell multi-cohort dissection of the schizophrenia transcriptome" (preprint, Psychiatry and Clinical Psychology, 2022); 10.1101/2022.08.31.22279406.

25. Karpi ski P, Samochowiec J, S siadek MM, Łaczma ski Ł, Misiak B, Analysis of global gene expression at seven brain regions of patients with schizophrenia. Schizophr. Res. 223, 119–127 (2020). [PubMed: 32631700]

26. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. , Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics 19, 477 (2018). [PubMed: 29914354]

27. Van den Berge K, Roux de Bezieux H, Street K, Saelens W, Cannoodt R, Saeys Y, et al. , Trajectory-based differential expression analysis for single-cell sequencing data. Nat Commun 11, 1201 (2020). [PubMed: 32139671]

28. Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, et al. , Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. Nature 598, 137–143 (2021). [PubMed: 34616063]

29. Fang R, Xia C, Close JL, Zhang M, He J, Huang Z, et al. , Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. Science 377, 56–62 (2022). [PubMed: 35771910]

30. Bryois J, Garrett ME, Song L, Safi A, Giusti-Rodriguez P, Johnson GD, et al. , Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. Nat. Commun. 9, 3121 (2018). [PubMed: 30087329]

31. Consortium EP, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, et al. , Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710 (2020). [PubMed: 32728249]

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

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

34. Polioudakis D, de la Torre-Ubieta L, Langerman J, Elkins AG, Shi X, Stein JL, et al. , A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron 103, 785–801.e8 (2019). [PubMed: 31303374]

35. Ruzzo EK, Pérez-Cano L, Jung J-Y, Wang L-K, Kashef-Haghighi D, Hartl C, et al. , Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. Cell 178, 850–866.e26 (2019). [PubMed: 31398340]

36. Hartl CL, Ramaswami G, Pembroke WG, Muller S, Pintacuda G, Saha A, et al. , Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. Nat. Neurosci. 24, 1313–1323 (2021). [PubMed: 34294919]

37. Hu B, Won H, Mah W, Park RB, Kassim B, Spiess K, et al. , Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. Nat. Commun. 12, 3968 (2021). [PubMed: 34172755]

38. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. , ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. 53, 403–411 (2021). [PubMed: 33633365]

39. Johansen N, Somasundaram S, Travaglini KJ, Yanny AM, Shumyatcher M, Casper T, et al. , Interindividual variation in human cortical cell type abundance and expression. Science 382, eadf2359 (2023).

40. Cooper YA, Teyssier N, Dräger NM, Guo Q, Davis JE, Sattler SM, et al. , Functional regulatory variants implicate distinct transcriptional networks in dementia. Science 377, eabi8654 (2022).

41. Jorstad NL, Song JHT, Exposito-Alonso D, Suresh H, Castro-Pacheco N, Krienen FM, et al. , Comparative transcriptomics reveals human-specific cortical features. Science 382, eade9516 (2023).

42. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. , A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 171, 1437–1452.e17 (2017). [PubMed: 29195078]

43. Gambino F, Pavlowsky A, Béglé A, Dupont J-L, Bahi N, Courjaret R, 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). [PubMed: 17502602]

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

45. Pembroke WG, Hartl CL, Geschwind DH, Evolutionary conservation and divergence of the human brain transcriptome. Genome Biol. 22, 52 (2021). [PubMed: 33514394]

46. Miller KJ, Schalk G, Fetz EE, den Nijs M, Ojemann JG, Rao RPN, Cortical activity during motor execution, motor imagery, and imagery-based online feedback. Proc. Natl. Acad. Sci. U. S. A. 107, 4430–4435 (2010). [PubMed: 20160084]

47. Maria M, Pouyanfar N, Örd T, Kaikkonen MU, The Power of Single-Cell RNA Sequencing in eQTL Discovery. Genes 13, 502 (2022). [PubMed: 35328055]

48. Deng Q, Ramsköld D, Reinius B, Sandberg R, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196 (2014). [PubMed: 24408435]

49. Borel C, Ferreira PG, Santoni F, Delaneau O, Fort A, Popadin KY, et al. , Biased allelic expression in human primary fibroblast single cells. Am. J. Hum. Genet. 96, 70–80 (2015). [PubMed: 25557783]

50. Mu W, Sarkar H, Srivastava A, Choi K, Patro R, Love MI, Airpart: interpretable statistical models for analyzing allelic imbalance in single-cell datasets. Bioinformatics 38, 2773–2780 (2022). [PubMed: 35561168]

51. Choi K, Raghupathy N, Churchill GA, A Bayesian mixture model for the analysis of allelic expression in single cells. Nat. Commun. 10, 5188 (2019). [PubMed: 31729374]

52. Jiang Y, Zhang NR, Li M, SCALE: modeling allele-specific gene expression by single-cell RNA sequencing. Genome Biol. 18, 74 (2017). [PubMed: 28446220]

53. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, et al. , Using whole-exome sequencing to identify inherited causes of autism. Neuron 77, 259–273 (2013). [PubMed: 23352163]

54. Sloan SA, Barres BA, Mechanisms of astrocyte development and their contributions to neurodevelopmental disorders. Curr. Opin. Neurobiol. 27, 75–81 (2014). [PubMed: 24694749]

55. Craig AM, Kang Y, Neurexin–neuroligin signaling in synapse development. Curr. Opin. Neurobiol. 17, 43–52 (2007). [PubMed: 17275284]

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

57. Nathan A, Asgari S, Ishigaki K, Valencia C, Amariuta T, Luo Y, et al. , Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. Nature 606, 120–128 (2022). [PubMed: 35545678]

58. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. , SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14, 1083–1086 (2017). [PubMed: 28991892]

59. Jin T, Rehani P, Ying M, Huang J, Liu S, Roussos P, 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). [PubMed: 34044854]

60. Duncan ID, Radcliff AB, Heidari M, Kidd G, August BK, Wierenga LA, The adult oligodendrocyte can participate in remyelination. Proc. Natl. Acad. Sci. U. S. A. 115, E11807–E11816 (2018). [PubMed: 30487224]

61. Alon U, Network motifs: theory and experimental approaches. Nat. Rev. Genet. 8, 450–461 (2007). [PubMed: 17510665]

62. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. , Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. 51, 404–413 (2019). [PubMed: 30617256]

63. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. , A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell 169, 1276–1290.e17 (2017). [PubMed: 28602351]

64. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. , Inference and analysis of cell-cell communication using CellChat. Nat. Commun. 12, 1088 (2021). [PubMed: 33597522]

65. Savchenko E, Teku GN, Boza-Serrano A, Russ K, Berns M, Deierborg T, et al. , FGF family members differentially regulate maturation and proliferation of stem cell-derived astrocytes. Sci. Rep. 9, 9610 (2019). [PubMed: 31270389]

66. Carter CJ, 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). [PubMed: 17239488]

67. Cui Q-L, Zheng W-H, Quirion R, Almazan G, 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). [PubMed: 15632127]

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

69. Muneer A, Wnt and GSK3 Signaling Pathways in Bipolar Disorder: Clinical and Therapeutic Implications. Clin Psychopharmacol Neurosci 15, 100–114 (2017). [PubMed: 28449557]

70. Santos R, Linker SB, Stern S, Mendes APD, Shokhirev MN, Erikson G, 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). [PubMed: 33398088]

71. Wexler EM, Geschwind DH, Palmer TD, Lithium regulates adult hippocampal progenitor development through canonical Wnt pathway activation. Mol. Psychiatry 13, 285–292 (2008). [PubMed: 17968353]

72. Hoseth EZ, Krull F, Dieset I, Morch RH, Hope S, Gardsjord ES, et al. , Exploring the Wnt signaling pathway in schizophrenia and bipolar disorder. Transl Psychiatry 8, 55 (2018). [PubMed: 29507296]

73. Valvezan AJ, Klein PS, GSK-3 and Wnt Signaling in Neurogenesis and Bipolar Disorder. Front Mol Neurosci 5, 1 (2012). [PubMed: 22319467]

74. Lovestone S, Killick R, Di Forti M, Murray R, Schizophrenia as a GSK-3 dysregulation disorder. Trends Neurosci. 30, 142–149 (2007). [PubMed: 17324475]

75. Panaccione I, Napoletano F, Forte AM, Kotzalidis GD, Del Casale A, Rapinesi C, et al. , Neurodevelopment in schizophrenia: the role of the wnt pathways. Curr Neuropharmacol 11, 535–58 (2013). [PubMed: 24403877]

76. McCurdy RD, Feron F, Perry C, Chant DC, McLean D, Matigian N, 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). [PubMed: 16406496]

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

78. van Scheltinga A. F. Terwisscha, Bakker SC, Kahn RS, Fibroblast growth factors in schizophrenia. Schizophr. Bull. 36, 1157–1166 (2010). [PubMed: 19429845]

79. Browaeys R, Saelens W, Saeys Y, NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods 17, 159–162 (2020). [PubMed: 31819264]

80. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY, Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. 23, 185–188 (1999). [PubMed: 10508514]

81. Bacon C, Rappold GA, The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. Hum. Genet. 131, 1687–1698 (2012). [PubMed: 22736078]

82. Inda MC, Defelipe J, Muñoz A, 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).

83. Allen WE, Blosser TR, Sullivan ZA, Dulac C, Zhuang X, Molecular and spatial signatures of mouse brain aging at single-cell resolution. Cell 186, 194–208.e18 (2023). [PubMed: 36580914]

84. Gomez CR, Role of heat shock proteins in aging and chronic inflammatory diseases. GeroScience 43, 2515–2532 (2021). [PubMed: 34241808]

85. Schultz C, Dick EJ, Cox AB, Hubbard GB, Braak E, Braak H, Expression of stress proteins alpha B-crystallin, ubiquitin, and hsp27 in pallido-nigral spheroids of aged rhesus monkeys. Neurobiol. Aging 22, 677–682 (2001). [PubMed: 11445268]

86. Abreu AP, Dauber A, Macedo DB, Noel SD, Brito VN, Gill JC, et al. , Central precocious puberty caused by mutations in the imprinted gene MKRN3. N. Engl. J. Med. 368, 2467–2475 (2013). [PubMed: 23738509]

87. Zannas AS, Jia M, Hafner K, Baumert J, Wiechmann T, Pape JC, 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). [PubMed: 31113877]

88. Zannas AS, Wiechmann T, Gassen NC, Binder EB, Gene-Stress-Epigenetic Regulation of FKBP5: Clinical and Translational Implications. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol. 41, 261–274 (2016).

89. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA, Religious Orders Study and Rush Memory and Aging Project. J Alzheimers Dis 64, S161–S189 (2018). [PubMed: 29865057]

90. Saura CA, Deprada A, Capilla-Lopez MD, Parra-Damas A, Revealing cell vulnerability in Alzheimer's disease by single-cell transcriptomics. Semin Cell Dev Biol 139, 73–83 (2023). [PubMed: 35623983]

91. Wang J, Roeder K, Devlin B, Bayesian estimation of cell type-specific gene expression with prior derived from single-cell data. Genome Res 31, 1807–1818 (2021). [PubMed: 33837133]

92. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. , A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 47, 1091–8 (2015). [PubMed: 26258848]

93. Wang YH, Hou HA, Lin CC, Kuo YY, Yao CY, Hsu CL, 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). [PubMed: 34614508]

94. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. , Analysis of shared heritability in common disorders of the brain. Science 360 (2018).

95. Cross-Disorder Group of the Psychiatric Genomics Consortium, Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet Lond. Engl. 381, 1371–1379 (2013).

96. Gelernter J, Sun N, Polimanti R, Pietrzak R, Levey DF, Bryois J, et al. , Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. Nat Neurosci 22, 1394–1401 (2019). [PubMed: 31358989]

97. Pisanu C, Williams MJ, Ciuculete DM, Olivo G, Del Zompo M, Squassina A, et al. , Evidence that genes involved in hedgehog signaling are associated with both bipolar disorder and high BMI. Transl. Psychiatry 9, 315 (2019). [PubMed: 31754094]

98. Chopra M, McEntagart M, Clayton-Smith J, Platzer K, Shukla A, Girisha KM, 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). [PubMed: 33909992]

99. Reay WR, Atkins JR, Quidé Y, Carr VJ, Green MJ, Cairns MJ, Polygenic disruption of retinoid signalling in schizophrenia and a severe cognitive deficit subtype. Mol. Psychiatry 25, 719–731 (2020). [PubMed: 30532020]

100. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. , Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362, eaat8127 (2018).

101. Wang M, Zhang L, Gage FH, Microglia, complement and schizophrenia. Nat. Neurosci. 22, 333–334 (2019). [PubMed: 30796420]

102. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. , DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 34, D668–672 (2006). [PubMed: 16381955]

103. Tian R, Abarientos A, Hong J, Hashemi SH, Yan R, Dräger N, et al. , Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. Nat. Neurosci. 24, 1020–1034 (2021). [PubMed: 34031600]

104. Bauernfeind AL, Babbitt CC, The predictive nature of transcript expression levels on protein expression in adult human brain. BMC Genomics 18, 322 (2017). [PubMed: 28438116]

105. Moritz CP, Mühlhaus T, Tenzer S, Schulenborg T, Friauf E, Poor transcript-protein correlation in the brain: negatively correlating gene products reveal neuronal polarity as a potential cause. J. Neurochem. 149, 582–604 (2019). [PubMed: 30664243]

106. Carlyle BC, Kitchen RR, Kanyo JE, Voss EZ, Pletikos M, Sousa AMM, et al. , A multiregional proteomic survey of the postnatal human brain. Nat. Neurosci. 20, 1787–1795 (2017). [PubMed: 29184206]

107. Liharska LE, Park YJ, Ziafat K, Wilkins L, Silk H, Linares LM, et al. , A study of gene expression in the living human brain. medRxiv [Preprint] (2023). 10.1101/2023.04.21.23288916.

108. PsychENCODE Consortium (PEC), PyschENCODE Consortium (PEC) Capstone II Cross-study Harmonized Data, Synapse (2023); 10.7303/SYN51111084.1.

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

110. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE261983.

111. Rodin RE, Dou Y, Kwon M, Sherman MA, D'Gama AM, Doan RN, 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). [PubMed: 33432195]

112. Emani P, at el., gersteinlab/PsychENCODE_SingleCell_Integrative: v1.0.0, version v1.0.0 (2024); 10.5281/ZENODO.10849968.

113. Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, et al. , The PsychENCODE project. Nat. Neurosci. 18, 1707–1712 (2015). [PubMed: 26605881]

114. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, et al. , Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. 19, 224 (2018). [PubMed: 30567574]

115. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. , Single-cell transcriptomic analysis of Alzheimer's disease. Nature 570, 332–337 (2019). [PubMed: 31042697]

116. Greenwood AK, Montgomery KS, Kauer N, Woo KH, Leanza ZJ, Poehlman WL, 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). [PubMed: 33085189]

117. Freund M, Taylor A, Ng C, Little AR, The NIH NeuroBioBank: creating opportunities for human brain research. Handb. Clin. Neurol. 150, 41–48 (2018). [PubMed: 29496155]

118. Li B, Gould J, Yang Y, Sarkizova S, Tabaka M, Ashenberg O, et al. , Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. Nat. Methods 17, 793–798 (2020). [PubMed: 32719530]

119. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. , Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049 (2017). [PubMed: 28091601]

120. McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, et al. , MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nat. Methods 16, 619–626 (2019). [PubMed: 31209384]

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

122. Fleming SJ, Chaffin MD, Arduini A, Akkad A-D, Banks E, Marioni JC, et al. , Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Nat. Methods 20, 1323–1335 (2023). [PubMed: 37550580]

123. Rath S, Sharma R, Gupta R, Ast T, Chan C, Durham TJ, et al. , MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. Nucleic Acids Res. 49, D1541–D1547 (2021). [PubMed: 33174596]

124. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. , Conserved cell types with divergent features in human versus mouse cortex. Nature 573, 61–68 (2019). [PubMed: 31435019]

125. Wolock SL, Lopez R, Klein AM, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 8, 281–291.e9 (2019). [PubMed: 30954476]

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

127. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. , Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296 (2019). [PubMed: 31740819]

128. Jorstad NL, Close J, Johansen N, Yanny AM, Barkan ER, Travaglini KJ, et al. , Transcriptomic cytoarchitecture reveals principles of human neocortex organization. Science 382, eadf6812 (2023).

129. Bakken TE, Jorstad NL, Hu Q, Lake BB, Tian W, Kalmbach BE, et al. , Comparative cellular analysis of motor cortex in human, marmoset and mouse. Nature 598, 111–119 (2021). [PubMed: 34616062]

130. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. , Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29 (2021). [PubMed: 34062119]

131. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. , "Scaling accurate genetic variant discovery to tens of thousands of samples" (preprint, Genomics, 2017); 10.1101/201178.

132. Li H, Durbin R, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–60 (2009). [PubMed: 19451168]

133. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. , STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

134. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ, Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015). [PubMed: 25722852]

135. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. , Next-generation genotype imputation service and methods. Nat Genet 48, 1284–1287 (2016). [PubMed: 27571263]

136. Pedersen BS, Quinlan AR, Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am J Hum Genet 100, 406–413 (2017). [PubMed: 28190455]

137. Gursoy G, Emani P, Brannon CM, Jolanki OA, Harmanci A, Strattan JS, et al. , Data Sanitization to Reduce Private Information Leakage from Functional Genomics. Cell 183, 905–917 e16 (2020). [PubMed: 33186529]

138. Wang K, Li M, Hakonarson H, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164 (2010). [PubMed: 20601685]

139. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. , The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443 (2020). [PubMed: 32461654]

140. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M, CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47, D886–D894 (2019). [PubMed: 30371827]

141. Ebler J, Ebert P, Clarke WE, Rausch T, Audano PA, Houwaart T, et al. , Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes. Nat. Genet. 54, 518–525 (2022). [PubMed: 35410384]

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

143. Li X, Kim Y, Tsang EK, Davis JR, Damani FN, Chiang C, et al. , The impact of rare variation on gene expression across tissues. Nature 550, 239–243 (2017). [PubMed: 29022581]

144. Jew B, Alvarez M, Rahmani E, Miao Z, Ko A, Garske KM, et al. , Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. Nat. Commun. 11, 1971 (2020). [PubMed: 32332754]

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

146. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R, Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420 (2018). [PubMed: 29608179]

147. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. , Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21 (2019). [PubMed: 31178118]

148. Hafemeister C, Satija R, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296 (2019). [PubMed: 31870423]

149. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. , Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29 (2000). [PubMed: 10802651]

150. Gene Ontology Consortium, The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, D325–D334 (2021). [PubMed: 33290552]

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

152. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R, Single-cell chromatin state analysis with Signac. Nat. Methods 18, 1333–1341 (2021). [PubMed: 34725479]

153. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. , Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008). [PubMed: 18798982]

154. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. , JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 48, D87–D92 (2020). [PubMed: 31701148]

155. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. , Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015). [PubMed: 26414678]

156. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, 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). [PubMed: 25642630]

157. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. , The Human Phenotype Ontology in 2021. Nucleic Acids Res. 49, D1207–D1217 (2021). [PubMed: 33264411]

158. Gazal S, Finucane HK, Furlotte NA, Loh P-R, Palamara PF, Liu X, et al. , Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat. Genet. 179, 1421–1427 (2017).

159. Stein JL, de la Torre-Ubieta L, Tian Y, Parikshak NN, Hernández IA, Marchetto MC, et al. , A Quantitative Framework to Evaluate Modeling of Cortical Development by Neural Stem Cells. Neuron 83, 69–86 (2014). [PubMed: 24991955]

160. de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, et al. , The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. Cell 172, 289–304.e18 (2018). [PubMed: 29307494]

161. Trevino AE, Sinnott-Armstrong N, Andersen J, Yoon S-J, Huber N, Pritchard JK, et al. , Chromatin accessibility dynamics in a model of human forebrain development. Science 367, eaay1645 (2020).

162. Walker RL, Ramaswami G, Hartl C, Mancuso N, Gandal MJ, de la Torre-Ubieta L, et al. , Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. Cell 179, 750–771.e22 (2019). [PubMed: 31626773]

163. Hoffman GE, Schadt EE, variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics 17, 483 (2016). [PubMed: 27884101]

164. Gandal MJ, Haney JR, Wamsley B, Yap CX, Parhami S, Emani PS, et al. , Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. Nature 611, 532–539 (2022). [PubMed: 36323788]

165. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. , Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050 (2005). [PubMed: 16024819]

166. Dong X, Li X, Chang TW, Scherzer CR, Weiss ST, Qiu W, 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). [PubMed: 34009297]

167. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET, A complete tool set for molecular QTL discovery and analysis. Nat Commun 8, 15452 (2017). [PubMed: 28516912]

168. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O, Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 32, 1479–1485 (2016). [PubMed: 26708335]

169. Storey JD, Tibshirani R, Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100, 9440–9445 (2003). [PubMed: 12883005]

170. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, 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). [PubMed: 31066453]

171. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. , SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism 4, 36 (2013). [PubMed: 24090431]

172. Jia P, Han G, Zhao J, Lu P, Zhao Z, SZGR 2.0: a one-stop shop of schizophrenia candidate genes. Nucleic Acids Res. 45, D915–D924 (2017). [PubMed: 27733502]

173. Franklin C, Dwyer DS, Candidate risk genes for bipolar disorder are highly conserved during evolution and highly interconnected. Bipolar Disord. 23, 400–408 (2021). [PubMed: 32959503]

174. Hu Y-S, Xin J, Hu Y, Zhang L, Wang J, Analyzing the genes related to Alzheimer's disease via a network and pathway-based approach. Alzheimers Res. Ther. 9, 29 (2017). [PubMed: 28446202]

175. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, et al. , Human Ageing Genomic Resources: new and updated databases. Nucleic Acids Res. 46, D1083–D1090 (2018). [PubMed: 29121237]

176. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. , PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81, 559–575 (2007). [PubMed: 17701901]

177. Casella G, Berger RL, Statistical Inference (Duxbury, Pacific Grove, Calif, 2. ed., 2002).

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

179. Hoff PD, 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.

180. Xiong L, Tian K, Li Y, Ning W, Gao X, Zhang QC, Online single-cell data integration through projecting heterogeneous datasets into a common cell-embedding space. Nat. Commun. 13, 6118 (2022). [PubMed: 36253379]

181. Pan L, Dinh HQ, Pawitan Y, Vu TN, Isoform-level quantification for single-cell RNA sequencing. Bioinformatics 38, 1287–1294 (2022). [PubMed: 34864849]

182. Garrido-Martin D, Borsari B, Calvo M, Reverter F, Guigo R, Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. Nat Commun 12, 727 (2021). [PubMed: 33526779]

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

184. Rozowsky J, Abyzov A, Wang J, Alves P, Raha D, Harmanci A, et al. , AlleleSeq: analysis of allele-specific expression and binding in a network framework. Mol. Syst. Biol. 7, 522 (2011). [PubMed: 21811232]

185. Chen J, Rozowsky J, Galeev TR, Harmanci A, Kitchen R, Bedford J, et al. , A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. Nat. Commun. 7, 11101 (2016). [PubMed: 27089393]

186. Rozowsky J, Gao J, Borsari B, Yang YT, Galeev T, Gürsoy G, et al. , The EN-TEx resource of multi-tissue personal epigenomes & variant-impact models. Cell 186, 1493–1511.e40 (2023). [PubMed: 37001506]

187. Konopka G, Wexler E, Rosen E, Mukamel Z, Osborn GE, Chen L, et al. , Modeling the functional genomics of autism using human neurons. Mol. Psychiatry 17, 202–214 (2012). [PubMed: 21647150]

188. Lee D, Shi M, Moran J, Wall M, Zhang J, Liu J, et al. , STARRPeaker: uniform processing and accurate identification of STARR-seq active regions. Genome Biol. 21, 298 (2020). [PubMed: 33292397]

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

190. Quinlan AR, Hall IM, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010). [PubMed: 20110278]

191. Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, et al. , MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. Genome Biol 20, 206 (2019). [PubMed: 31604482]

192. Moerman T, Santos S. Aibar, Gonzalez-Blas C. Bravo, Simm J, Moreau Y, Aerts J, et al. , GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics 35, 2159–2161 (2019). [PubMed: 30445495]

193. Suo S, Zhu Q, Saadatpour A, Fei L, Guo G, Yuan GC, Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas. Cell Rep 25, 1436–1445 e3 (2018). [PubMed: 30404000]

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

195. Gupta C, Xu J, Jin T, Khullar S, Liu X, Alatkar S, 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).

196. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J, Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 29, 1363–1375 (2019). [PubMed: 31340985]

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

198. Kashtan N, Itzkovitz S, Milo R, Alon U, Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. Bioinformatics 20, 1746–58 (2004). [PubMed: 15001476]

199. Liu G, Wong L, Chua HN, Complex discovery from weighted PPI networks. Bioinformatics 25, 1891–7 (2009). [PubMed: 19435747]

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

201. Brunet J-P, Tamayo P, Golub TR, Mesirov JP, Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. U. S. A. 101, 4164–4169 (2004). [PubMed: 15016911]

202. Gaujoux R, Seoighe C, A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11, 367 (2010). [PubMed: 20598126]

203. Huuki-Myers L, Spangler A, Eagles N, Montgomery KD, Kwon SH, Guo B, et al. , "Integrated single cell and unsupervised spatial transcriptomic analysis defines molecular anatomy of the human dorsolateral prefrontal cortex" (preprint, Neuroscience, 2023); 10.1101/2023.02.15.528722.

204. Anders S, Pyl PT, Huber W, HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015). [PubMed: 25260700]

205. Ernst J, Bar-Joseph Z, STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics 7, 191 (2006). [PubMed: 16597342]

206. Dai M, Zhao L, Li Z, Li X, You B, Zhu S, 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).

207. Wei W, Jiang C, Chai X, Zhang J, Zhang C-C, Miao W, 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). [PubMed: 33467569]

208. Song Q, Wang J, Bar-Joseph Z, scSTEM: clustering pseudotime ordered single-cell data. Genome Biol. 23, 150 (2022). [PubMed: 35799304]

209. Ben-Kiki O, Bercovich A, Lifshitz A, Tanay A, Metacell-2: a divide-and-conquer metacell algorithm for scalable scRNA-seq analysis. Genome Biol. 23, 100 (2022). [PubMed: 35440087]

210. Schulz M-A, Yeo BTT, Vogelstein JT, Mourao-Miranda J, Kather JN, Kording K, et al. , Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. Nat. Commun. 11, 4238 (2020). [PubMed: 32843633]

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

212. Wainwright MJ, Jordan MI, Graphical Models, Exponential Families, and Variational Inference. Found. Trends® Mach. Learn. 1, 1–305 (2007).

213. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. , Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. 37, 773–782 (2019). [PubMed: 31061481]

214. Choi SW, Mak TS-H, O'Reilly PF, Tutorial: a guide to performing polygenic risk score analyses. Nat. Protoc. 15, 2759–2772 (2020). [PubMed: 32709988]

215. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. , Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. 51, 431–444 (2019). [PubMed: 30804558]

216. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, 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). [PubMed: 34002096]

217. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. , Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604, 502–508 (2022). [PubMed: 35396580]

218. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, 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). [PubMed: 34493870]

219. Privé F, Arbel J, Vilhjálmsson BJ, LDpred2: better, faster, stronger. Bioinformatics 36, 5424–5431 (2021). [PubMed: 33326037]

220. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, 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). [PubMed: 31768057]

221. Kaplanis J, Akawi N, Gallone G, McRae JF, Prigmore E, Wright CF, et al. , Exome-wide assessment of the functional impact and pathogenicity of multinucleotide mutations. Genome Res. 29, 1047–1056 (2019). [PubMed: 31227601]

222. Xia Y, Dai R, Wang K, Jiao C, Zhang C, Xu Y, 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). [PubMed: 30976086]

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

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

225. Tsimberidou A-M, Skliris A, Valentine A, Shaw J, Hering U, Vo HH, et al. , AKT inhibition in the central nervous system induces signaling defects resulting in psychiatric symptomatology. Cell Biosci. 12, 56 (2022). [PubMed: 35525984]

226. Farrelly LA, Zheng S, Schrode N, Topol A, Bhanu NV, Bastle RM, et al. , Chromatin profiling in human neurons reveals aberrant roles for histone acetylation and BET family proteins in schizophrenia. Nat. Commun. 13, 2195 (2022). [PubMed: 35459277]

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

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

229. West DB, Introduction to Graph Theory (Prentice Hall, Upper Saddle River, 1996).

230. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B, WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. 45, W130–W137 (2017). [PubMed: 28472511]

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

232. Privé F, Albiñana C, Arbel J, Pasaniuc B, Vilhjálmsson BJ, Inferring disease architecture and predictive ability with LDpred2-auto. Am. J. Hum. Genet. 110, 2042–2055 (2023). [PubMed: 37944514]

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

234. Lee T, Lee H, Prediction of Alzheimer's disease using blood gene expression data. Sci. Rep. 10, 3485 (2020). [PubMed: 32103140]

235. Sims R, Hill M, Williams J, The multiplex model of the genetics of Alzheimer's disease. Nat. Neurosci. 23, 311–322 (2020). [PubMed: 32112059]

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

237. Li J, Cai T, Jiang Y, Chen H, He X, Chen C, et al. , Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. Mol. Psychiatry 21, 298 (2016). [PubMed: 25939403]

238. Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE, et al. , FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. Cell 146, 247– 261 (2011). [PubMed: 21784246]

239. Basu SN, Kollu R, Banerjee-Basu S, AutDB: a gene reference resource for autism research. Nucleic Acids Res. 37, D832–D836 (2009). [PubMed: 19015121]

240. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. , Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science 359, 693– 697 (2018). [PubMed: 29439242]

241. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. , Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature 540, 423–427 (2016). [PubMed: 27919067]

242. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, et al. , Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat. Commun. 5, 5748 (2014). [PubMed: 25494366]

243. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. , Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474, 380–384 (2011). [PubMed: 21614001]

244. International Schizophrenia Consortium, Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455, 237–241 (2008). [PubMed: 18668038]

245. Ayalew M, Le-Niculescu H, Levey DF, Jain N, Changala B, Patel SD, et al. , Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Mol. Psychiatry 17, 887–905 (2012). [PubMed: 22584867]

246. Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, et al. , Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. Am. J. Hum. Genet. 73, 34–48 (2003). [PubMed: 12802786]

247. He X, Fuller CK, Song Y, Meng Q, Zhang B, Yang X, et al. , Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. Am. J. Hum. Genet. 92, 667–680 (2013). [PubMed: 23643380]

248. Ng MYM, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T, et al. , Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Mol. Psychiatry 14, 774–785 (2009). [PubMed: 19349958]

249. Chen C, Cheng L, Grennan K, Pibiri F, Zhang C, Badner JA, et al. , Two gene co-expression modules differentiate psychotics and controls. Mol. Psychiatry 18, 1308–1314 (2013). [PubMed: 23147385]
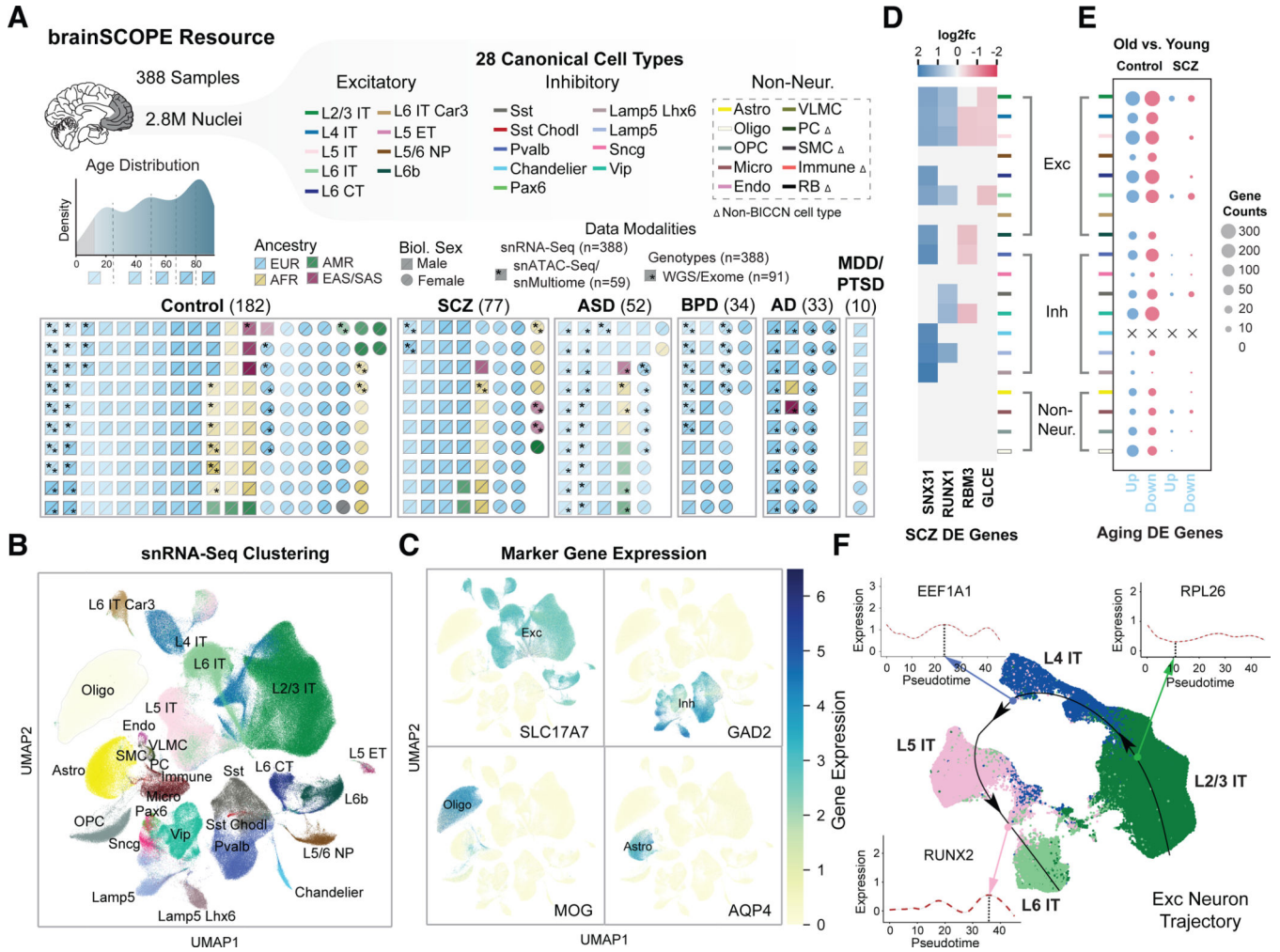
Figure 1. Constructing a single-cell genomic resource for 388 individuals.
(**A**) Overview of the integrative single-cell analysis performed on 388 adult prefrontal cortex samples. (Top) Schematic for 28 cell types grouped by excitatory (Exc), inhibitory (Inh), and non-neuronal cell types (table S3); 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 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.
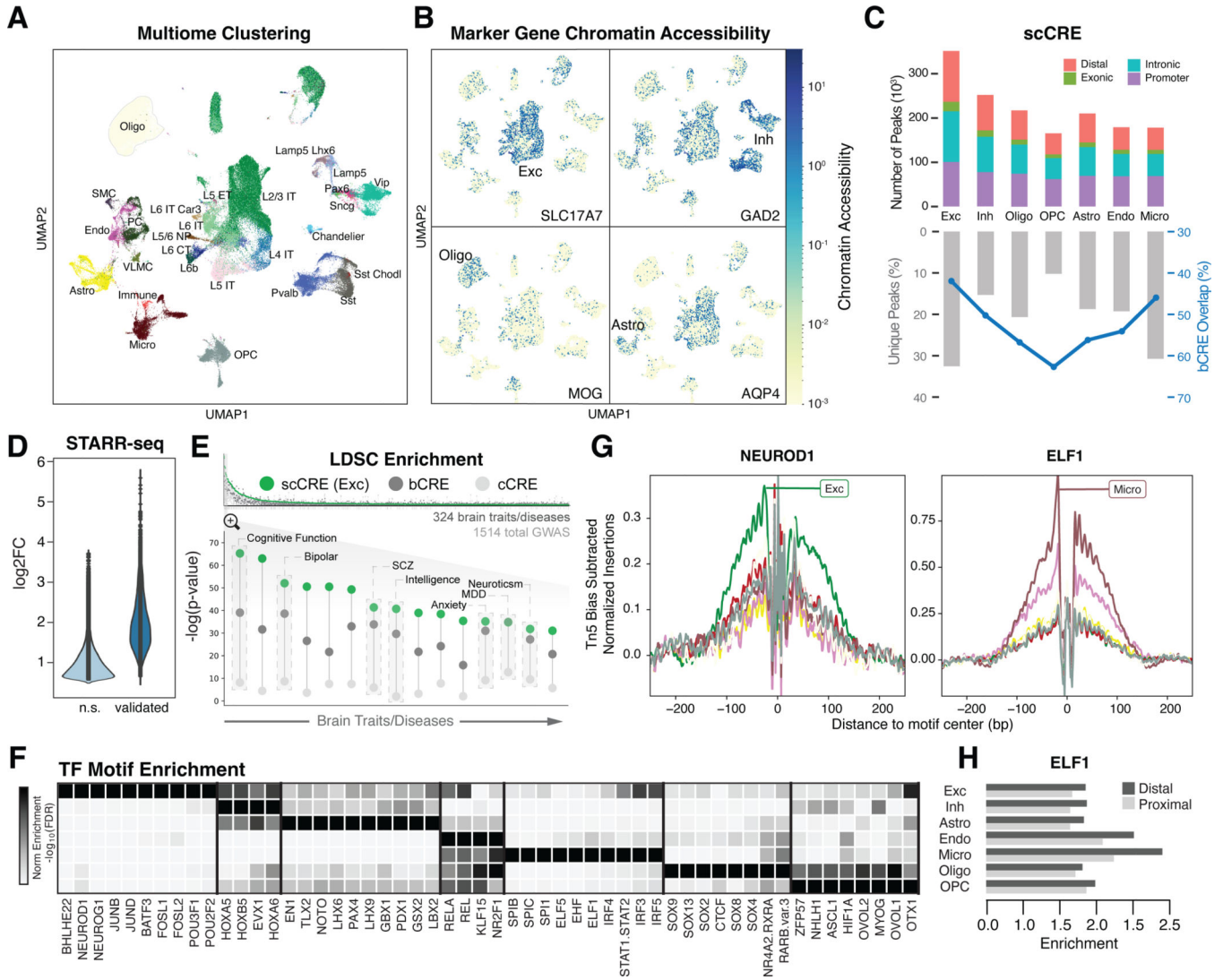
Figure 2. Determining regulatory elements for cell types from snATAC-seq.
(A) UMAP plot for clustering of 28 harmonized cell types from snMultiome data derived from 40 individuals. **(B)** UMAP plots highlighting chromatin accessibility of key marker genes for four broad cell types (see Fig. 1C). **(C)** (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.
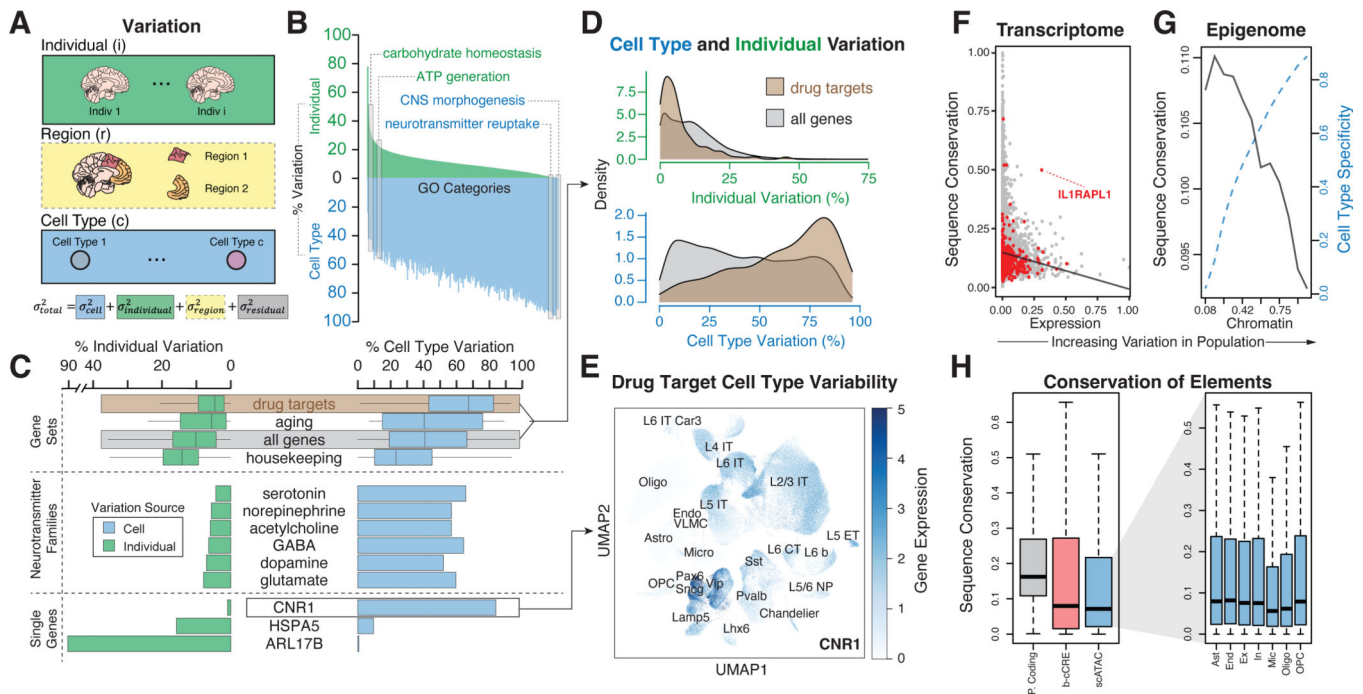
Figure 3. Measuring transcriptome and epigenome variation across the cohort at the single-cell level.
(**A**) Schematic for the calculation of overall gene-level variance partition by integrating individual, brain region, and cell-type-specific variation. Variation analysis using different brain regions (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.
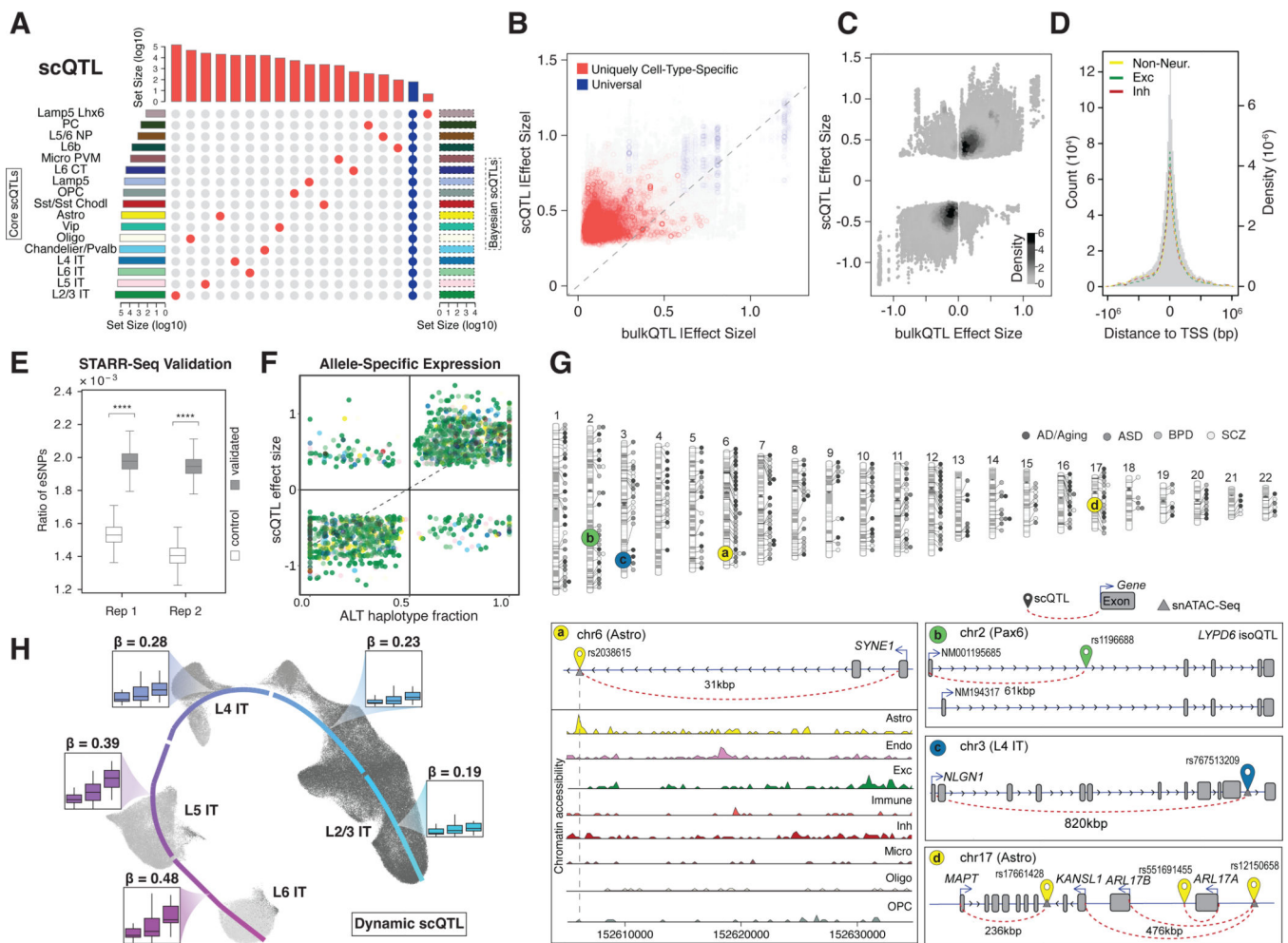
Figure 4. Determining cell-type-specific eQTLs from single-cell data.

(**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 (**β**) values for the eSNP increase over pseudotime. Additional information is shown in fig. S53.
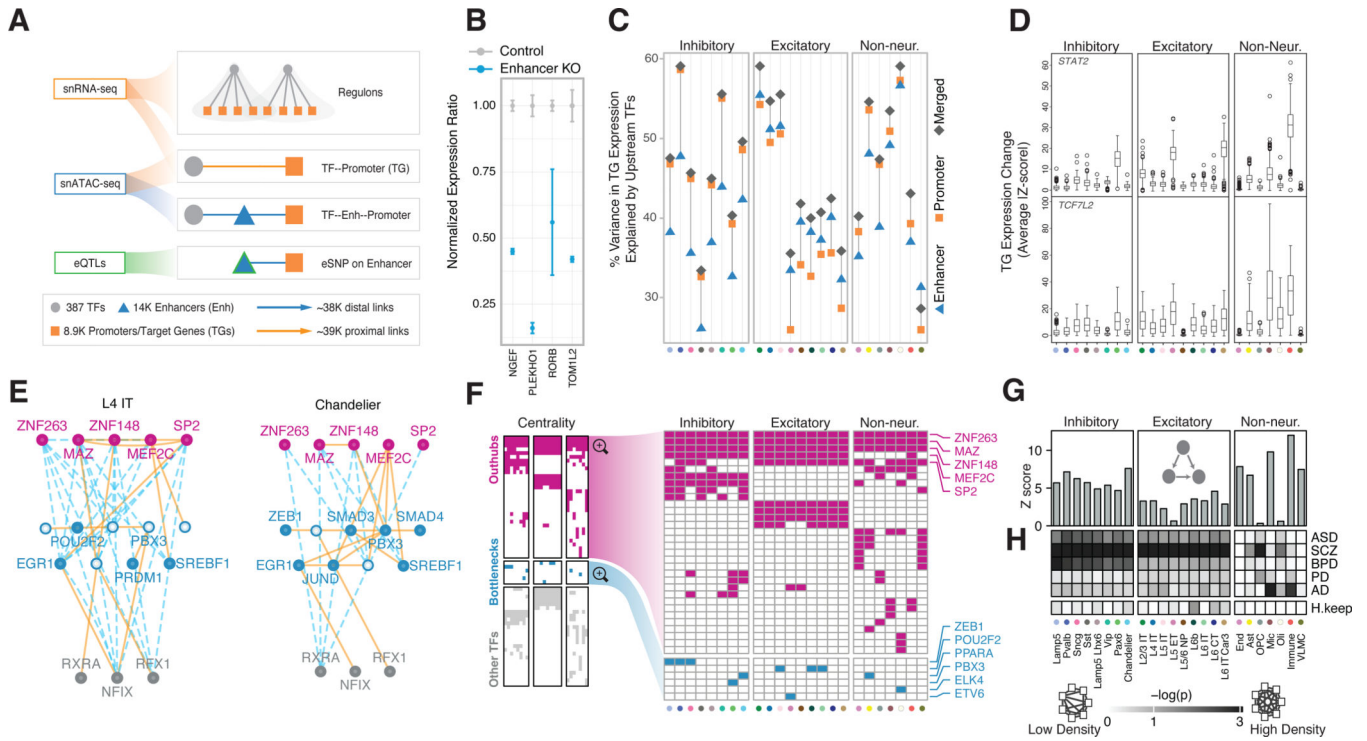
Figure 5. Building a gene regulatory network for each cell type.

(**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.
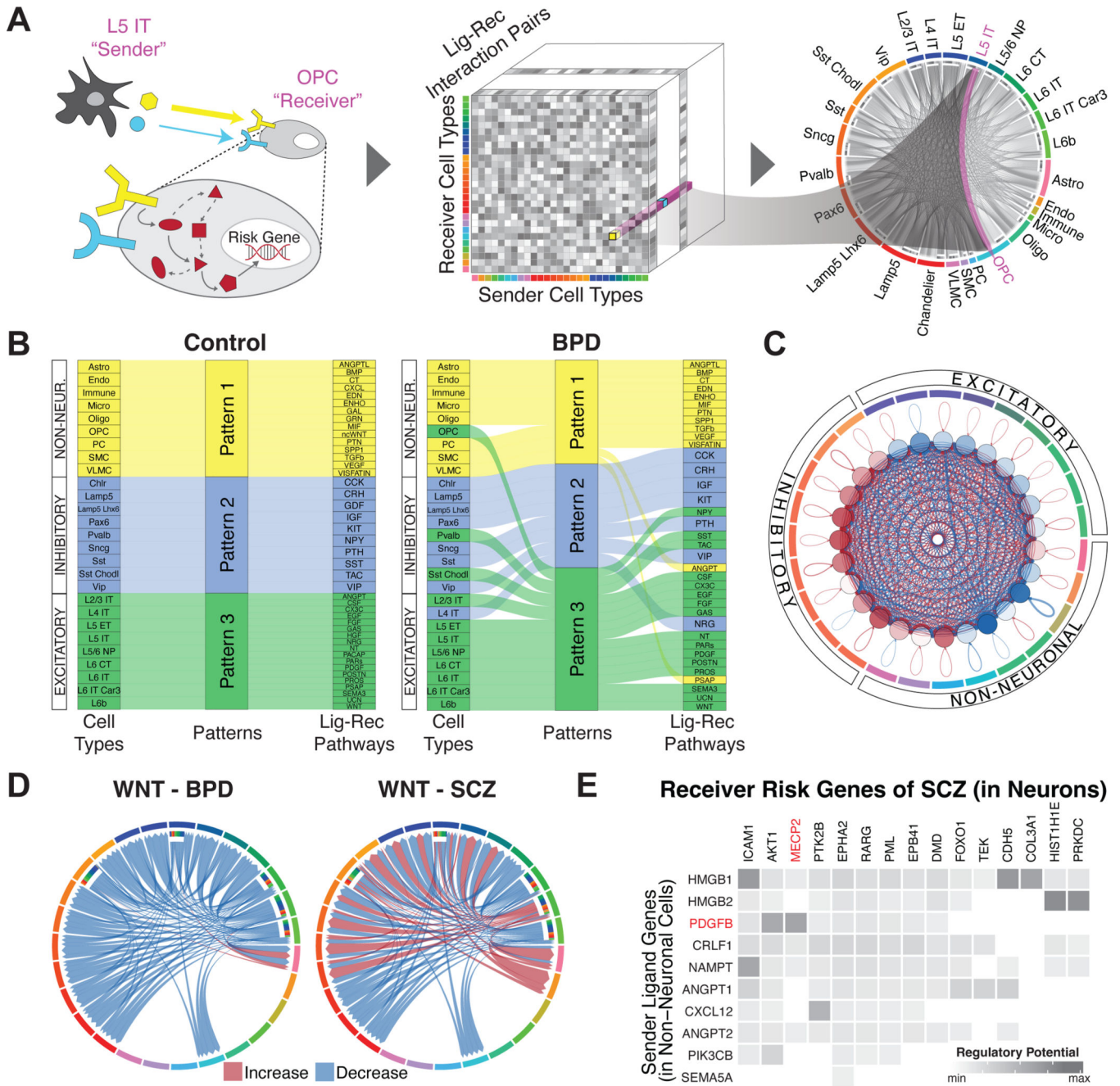
Figure 6. Constructing a cell-to-cell communication network.

(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.
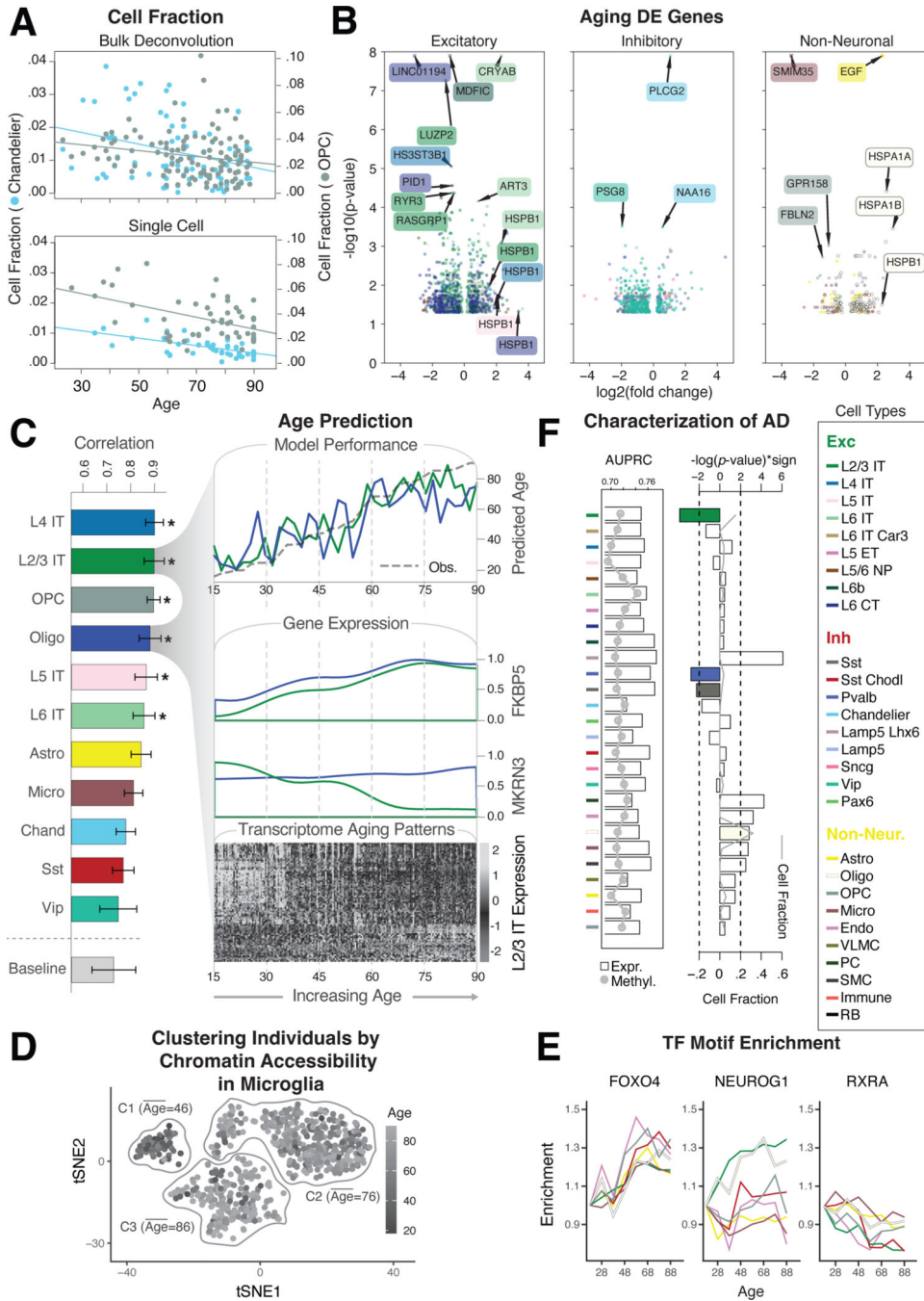
Figure 7. Assessing cell-type-specific transcriptomic and epigenetic changes in aging.
(**A**) Normalized changes in the fraction of OPC (gray) and Chandelier cells (blue) by age, based on bulk RNA-seq deconvolution (top) and single-cell annotation (bottom), with best-fitted lines. (**B**) Log$_2$-fold 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.
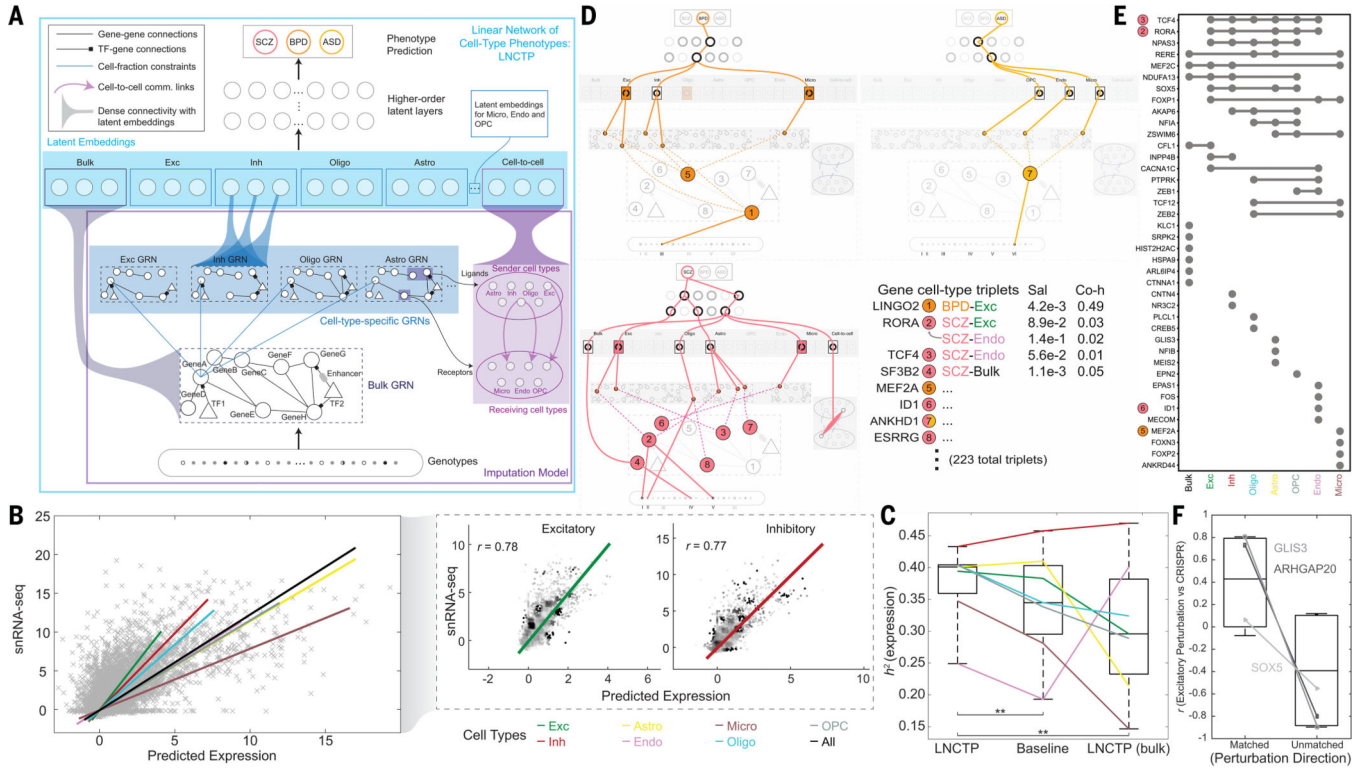
Figure 8. Imputing gene expression and prioritizing disease genes across cell types with an integrative model.

(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).