PsychENCODE Data Analysis and Coordination Center

A collaborative team of investigators from four institutions (Zhiping Weng from UMass Chan Medical School, Mark Gerstein from Yale University, Mette Peters from Sage, and Daifeng Wang from University of Wisconsin-Madison) played essential roles in the data coordination, analysis, and integration of the PsychENCODE Consortium in the past seven years. An NIMH U01 grant has funded our joint effort since 2018 (MH116492, currently on a no-cost extension; MPIs: Weng and Gerstein; co-Is: Peters and Wang). Dr. Anna Greenwood became the Co-I from Sage when Dr. Peters retired several months ago. The results of our joint efforts have formed the foundation of this grant proposal. We (led by the Sage team) coordinated the data submission by all PsychENCODE members to Synapse and built a knowledge portal on the PsychENCODE Project [12]. We (led by the Gerstein and Weng labs) built standardized RNA sequencing (RNA-seq), chromatin immunoprecipitation with sequencing (ChIP-seq), transposase-accessible chromatin with sequencing (ATAC-seq), and genotyping pipelines to uniformly process all PsychENCODE data, and performed integrative data analyses. We (led by the Weng lab) also built PsychSCREEN [13], a web-based platform for data search and visualization at the level of individual genes, cis-regulatory elements (CREs), single-nucleotide polymorphisms (SNPs), and QTLs.

In Phase I of the PsychENCODE Project, we (led by the Yale team) built the PsychENCODE Capstone Collection [5], an effort combining extensive brain sequencing data from individual PsychENCODE studies. The Capstone Collection datasets offer insights into various functional genomic elements, including genes expressed in the human brain, coexpression modules, single-cell expression profiles for major cell types, and brain-active enhancers identified through open chromatin regions enriched in H3K27ac, transcription factors (TFs), and topologically associating domains. These datasets were then used to compute QTLs associated with expression, chromatin activity, splicing-isoform levels, and cell-type proportions. The integrative analyses produced by the Capstone Project underscore the effectiveness of functional genomic data in elucidating the underlying molecular mechanisms of neuropsychiatric diseases. We uniformly processed all datasets, making them publicly available as a comprehensive resource. Our study [5] also presents integrative functional genomics models, including a gene regulatory network (GRN) that connects enhancers, TFs, and target genes. This network aids in linking potential neuropsychiatric disease-related genes with risk variants identified through genome-wide association studies (GWAS) and QTLs. To enhance disease risk prediction beyond conventional approaches, the GRN is embedded within a deep-learning model that predicts psychiatric phenotypes using genotype and expression data. Integration of additional data layers, such as functional genomic elements, QTLs, and GRNs, significantly enhances disease risk prediction accuracy.

In the current PsychENCODE Phase II, we (Yale, Wisconsin, and UMass Chan) integrated single-nucleus multiomic datasets of PsychENCODE to create a uniformly processed resource comprising >3.2M nuclei across 388 individuals, enabling the assessment of expression variation in 28 rigorously defined cell types in the human prefrontal cortex [14]. Overall, we identified >450K cell-type-specific CREs and >1.4 M single-cell expression QTLs (eQTLs), with which we built cell-type-specific GRNs and cell-to-cell communication networks. These networks provide a comprehensive view of cell-type-level gene expression and regulation changes in aging and neuropsychiatric disorders. Finally, we constructed an integrative model that accurately imputes cell-type-specific gene expression and prioritizes disease-risk genes.

Aim 1. Data Coordination

Aim 1.2 Building a PsychENCODE Consortium portal for processed data to facilitate data retrieval and visualization at the level of individual genes and regulatory elements.

Currently, we (led by the Sage team, Co-I Greenwood, see **letter of support** from **Dr. Anna Greenwood**) are actively engaged in the process of migrating all PsychENCODE data from Synapse to the NIMH Data Archive (NDA) [15]. As stipulated by the funding opportunity notice (PAR-23-234), we are committed to submitting all forthcoming PsychENCODE data to the NDA. In alignment with the objectives of the upcoming phase of the PsychENCODE Project, the collaborative efforts of the remaining three institutions (UMass Chan, Yale, and Wisconsin) will meet the DACC's responsibilities as described in this proposal.

The NDA primarily serves as an archive for data storage and retrieval rather than interactive query and visualization. NDA operates on study participant units, emphasizing the necessity for each participant to possess a GUID. Users must download data from NDA and perform subsequent analyses using in-house computational tools. This limitation curtails the potential impact of PsychENCODE data, especially for researchers interested in a particular gene, genetic variant, or regulatory element. To complement NDA, we developed PsychSCREEN, a web-based platform for interactive data search and visualization [13]. The data units for PsychSCREEN are individual genes, variants, and cis-regulatory elements (CREs). PsychSCREEN comprises four portals that offer interactive visualization of PsychENCODE annotations from diverse biological perspectives, supported by raw PsychENCODE and public datasets contributing to these annotations. PsychSCREEN employs the ReactJS framework for interactive visualization, adopting a minimalistic design with clear portal illustrations for intuitive navigation. A novel embedded genome browser supports visualizations in scalable vector graphics, which are exportable for generating figures and sharing insights. The architectural details of PsychSCREEN are expounded in a manuscript in the PsychENCODE Phase II paper package in review at *Science Advances* [18].

PsychSCREEN's Disease and Trait portal allows users to explore GWAS analyses for brain-specific traits. Users input diseases or traits of interest, like schizophrenia or neuroticism. The portal employs trait summary statistics to locate risk loci, plotting them on chromosome cytoband views. Clicking on highlighted loci redirects users to interactive genome browser views, displaying significant SNPs in a Manhattan-style plot alongside annotations. These annotations include a combined set of adult and fetal brain-specific CREs (bCREs), importance scores computed by a deep-learning algorithm ChromBPNet [19,20], and pseudo-bulk tracks from single-cell types (**Fig. 1A**). A related SNP/QTL portal showcases PsychENCODE annotations, encompassing eQTLs, bCREs, TF binding sites, and trait associations for a queried SNP. The portal enables visualization of these annotations in the SNP's vicinity through an embedded genome browser view. PsychSCREEN's Gene portal delivers gene regulatory annotations for individual genes of interest (**Fig. 1B**). Users input a gene's name, leading to a genome browser view that displays the genomic surroundings of the gene. Overlays reveal nearby bCREs, importance scores from ChromBPNet models [20], and TF binding sites. Tracks from single-cell and bulk ATAC-seq datasets from PsychENCODE can also be toggled. The Genes portal presents gene expression tabs, showing gene expression across adult and fetal time points in the brain based on PsychENCODE data. An eQTL and bCRE tab showcases SNPs correlated with gene expression based on PsychENCODE annotations. The single-cell expression tab illustrates gene expression patterns in individual brain cell types using single-cell RNA-seq data from the adult dorsolateral prefrontal cortex by PsychENCODE. Lastly, the Single Cell portal facilitates interactive exploration of PsychENCODE single-cell RNA-seq datasets, encompassing single-cell ATAC-seq peaks, single-cell QTLs (SNPs and associated genes), and cell type-specific GRNs, including enhancer-promoter links (**Fig. 1C**).

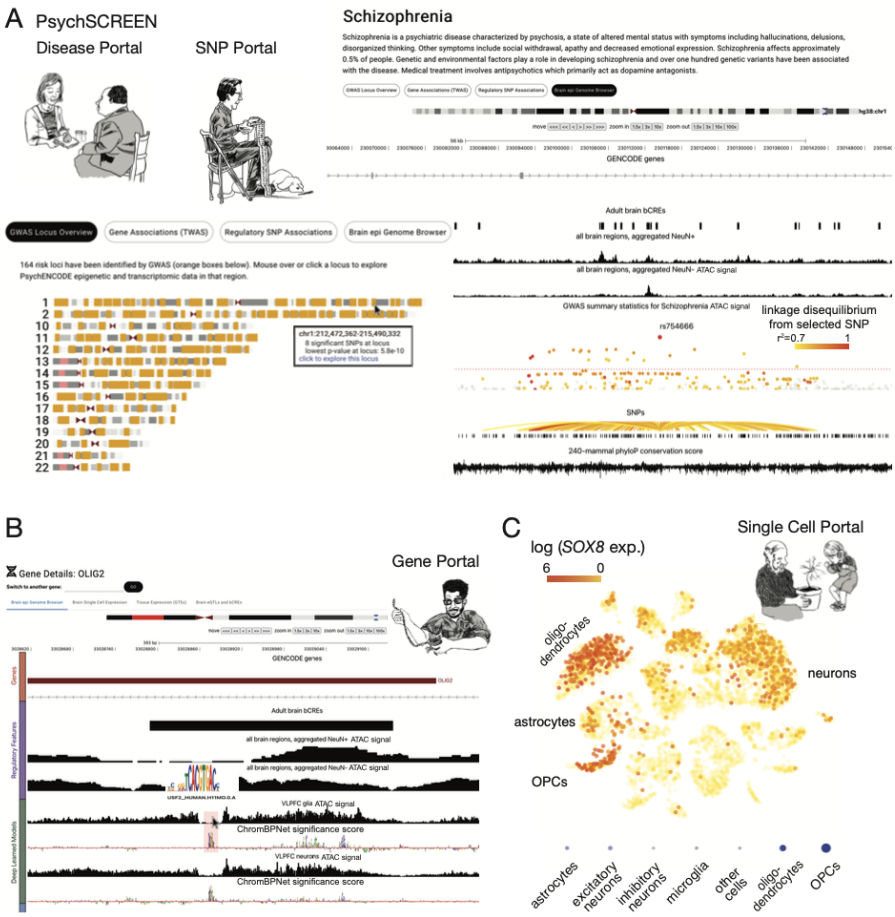

Figure 1. PsychSCREEN is an interactive platform to visualize PsychENCODE data and annotations at individual loci

A. PsychSCREEN's disease/SNP portals show disease queries such as Schizophrenia (top right) and GWAS-identified risk loci in the genome (bottom left). Clicking a locus leads to a genome browser (bottom right) displaying brain-specific CREs (bCREs), neuron and glia ATAC signal, Schizophrenia GWAS summary statistics, SNP linkage disequilibrium, and mammalian conservation scores. **B.** PsychSCREEN's gene portal displays screenshots of a genome browser view for the *OLIG2* gene, a transcription factor (TF) involved in determining neural progenitor cell fate, particularly in specifying cells to become oligodendrocyte precursor cells (OPCs). The view includes bCREs, aggregate neuron and glia ATAC signal, ATAC signal from individual experiments, and predicted importance scores from ChromBPNet deep-learning models. The ChromBPNet track assists in locating TF motifs. **C.** PsychSCREEN's single-cell portal screenshots exhibit a UMAP plot from a single-nucleus RNA-seq experiment, indicating the expression of the queried gene *SOX8* (top). A dot plot (bottom) displays gene expression in clusters, where circle size signifies the percentage of expressing cells and opacity represents average expression in each cluster. *SOX8* encodes a TF most highly expressed in OPCs, and it is known for its role in brain cell differentiation and cell fate maintenance.

Aim 2. Data Analysis

Aim 2.1 Uniformly processing PsychENCODE datasets, performing data integration, and generating integrated joint datasets for distribution to PsychENCODE Consortium members and the broader research community.

Recent advancements in high-throughput technologies have led to the rapid accumulation of genomic data extracted from brain tissue samples. The integration and analyses of these extensive datasets pose multiple challenges, from initial data preprocessing to assurance of data consistency and reproducibility. The task of harmonizing divergent datasets from various sources into a coherent and unified resource becomes particularly intricate when applied to brain tissue and cells, given the inherent data variability. This complexity is further compounded by considerations such as cell classification, dataset clustering, and the alignment of diverse assays within the expansive sample pool of the PsychENCODE Project. A major goal of the DACC is to achieve uniform processing of all PsychENCODE data, culminating in their integration into a comprehensive genomic resource for each distinct neuropsychiatric disorder.

Our efforts have been pivotal in numerous human genome projects, including ENCODE [3], model organism ENCODE (modENCODE) [22], GENCODE [23], IGVF, and Zoonomia [24]. Dr. Gerstein's lab has been at the forefront of integrating diverse genomic datasets to construct intricate GRNs encompassing regulatory elements like TFs, micro-RNAs, and their target genes [22,25–28]. Dr. Gerstein has significant involvement in GENCODE, primarily focusing on genome annotation [23]. As part of the IGVF consortium's Data Analysis Coordination Center, the Gerstein team's contributions have resulted in a comprehensive catalog of human genetic variants for the biomedical community. In a parallel endeavor, Dr. Weng's laboratory played a significant role in the ENCODE Phase III consortium paper published in *Nature*, where they defined a registry of human candidate CREs [3]. Dr. Weng's group further extended this impactful work by publishing a study in *Science*'s Zoonomia package, which characterized the evolutionary dynamics of the registry of CREs in the mammalian lineage and illuminated that elements under evolutionary constraint significantly contribute to human traits [24].

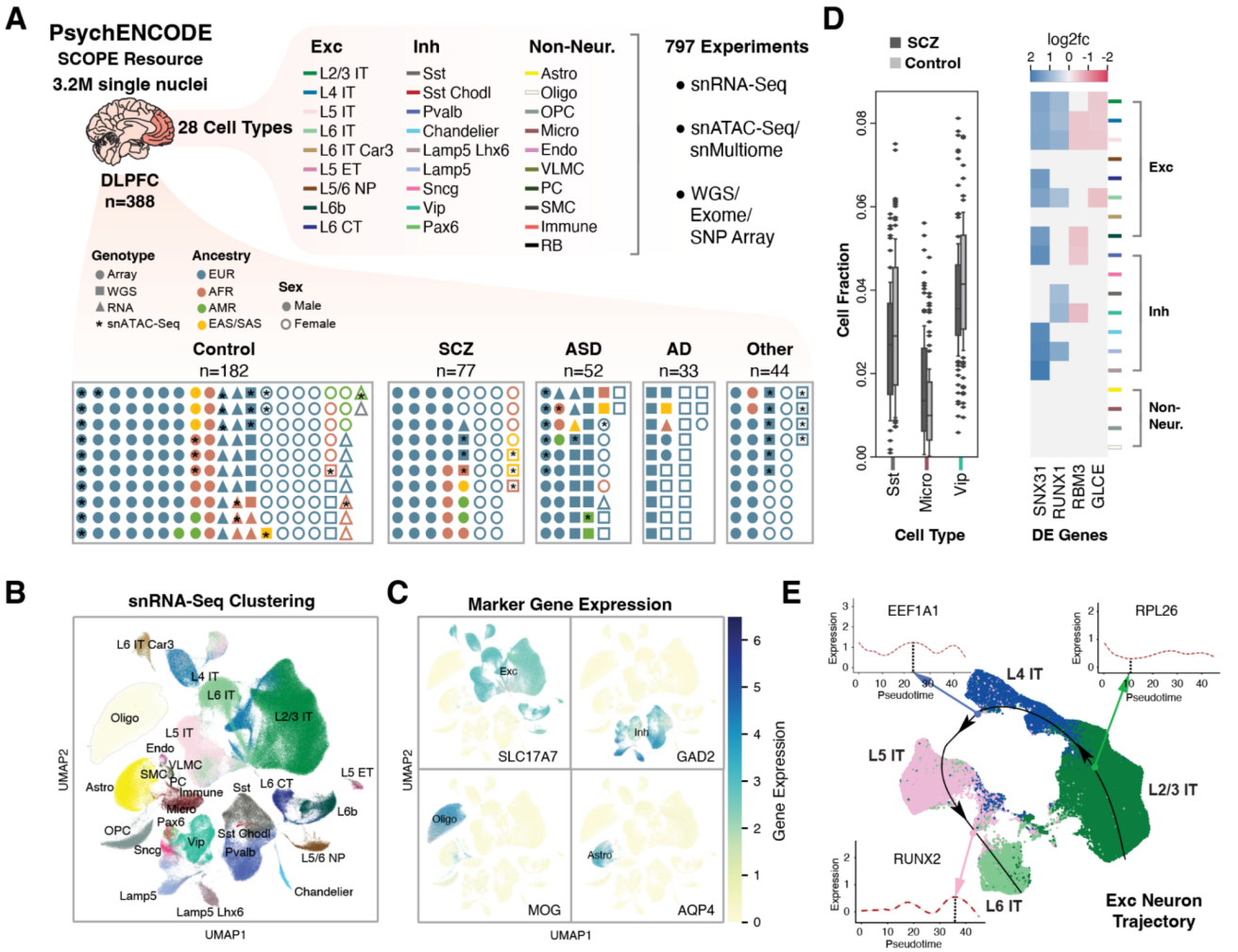

Figure 2. Constructing a single-cell genomic resource for 388 individuals. A. Overview of the integrative single-cell analysis performed on 388 adult prefrontal cortex samples. Top schematic shows 28 cell types grouped by excitatory (Exc), inhibitory (Inh), and non-neuronal cell types. Bottom schematic shows all samples labeled by disease, biological sex, ancestry, and available data modalities. **B.** UMAP plot shows the clustering of 28 harmonized cell types from snRNA-seq data derived from 72 samples in the SZBDMulti-Seq cohort. **C.** UMAP plots highlight the expression of key marker genes in four broad cell types (excitatory: *SLC17A7*; inhibitory: *GAD2*; oligodendrocytes: *MOG*; and astrocytes: *AQP4*). **D.** Box plots (left) highlight three cell types that show significant cell fraction changes in individuals with schizophrenia. Heat map (right) shows differential expression (log_2-fold change) of four schizophrenia-related genes across cell types in samples from individuals with schizophrenia. **E.** UMAP plot shows the 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. Smoothed line plot insets highlight log-normalized gene expression in cells along the pseudo-time axis for three genes.

Notably, Drs. Gerstein and Wang led the data analysis efforts for the PsychENCODE Phase I Capstone Project, employing an innovative approach that utilizes regulatory networks derived from Hi-C, QTLs, and activity relationships. This methodology has facilitated linking non-coding GWAS loci to potential genes associated with psychiatric diseases such as schizophrenia, ASD, bipolar disorder, and Alzheimer's disease [5]. Dr. Gerstein's team has developed integrative tools within the various Consortium frameworks. Notably, they harnessed topic modeling to harmonize various data sources, including gene expression data, to identify groups of genes exhibiting significant co-expression patterns and shared functionalities [29]. Additionally, the team collated data from diverse biological networks such as protein-protein interactions, phosphorylation, signaling, metabolism, genetics, and regulation to engineer a Multinet [30]. This approach facilitated the analysis of gene roles across individual and combined networks and led to the accurate assignment of functional indispensability scores to genes. Their contributions extended to devising methodologies for delineating network hierarchies, exemplified by the HirNet framework [31].

The leadership of Dr. Gerstein's lab (with contributions from Drs. Weng's and Wang's labs) was once again instrumental in the PsychENCODE Phase II Capstone Project in orchestrating the construction of an integrated cohort comprising PsychENCODE data and external data sources such as Accelerating Medicines Partnership for Alzheimer's Disease (AMP-AD) [32]. Our endeavor focused on carefully compiling and comprehensively analyzing single-cell multiomic data from the adult human prefrontal cortex of 388 individuals of diverse composition, encompassing variations in gender, ancestry, and age (**Fig. 2**) [14]. Within this framework, we incorporated 182 healthy control individuals alongside those afflicted with psychiatric or neurodegenerative conditions, including schizophrenia, ASD, bipolar disorder, and Alzheimer's disease.

To construct this resource, we undertook comprehensive and uniform processing of over 447 single-nucleus (sn)RNA-seq, snATAC-seq, and snMultiome (joint ATAC-seq and RNA-seq performed on the same single nuclei) datasets, encompassing more than 3.2M nuclei. These datasets were accompanied by corresponding genotypes and validation experiments from individual studies within the PsychENCODE Consortium and external sources. The intricacies of this processing encompassed the harmonization of datasets stemming from varying technologies and modalities. Notably, we worked towards achieving consistent genotypes across multiple data sources, including whole-genome sequencing, SNP arrays, and RNA-seq data. This harmonization extended to encompass rare single nucleotide variants (SNVs) and structural variants for specific samples. Moreover, we tailored custom datasets to bridge the gaps between various studies, with a particular emphasis on performing snMultiome sequencing for control samples. We devised a comprehensive cell-type annotation framework that effectively harmonizes the BRAIN Initiative Cell Census Network (BICCN) reference atlas [33] with prior analyses uniquely tailored to the prefrontal cortex [34]. This harmonization effort resulted in 28 distinct cell types, most exhibiting robust expression across all studied cohorts. We grouped excitatory and inhibitory subtypes into broader "excitatory" and "inhibitory" categories to facilitate more powerful downstream analyses, leading to seven major cell groupings. Our assessment encompassed 2,728,915 meticulously annotated nuclei of high-quality snRNA-seq data. This annotation scheme was rigorously validated by examining the expression of key marker genes.

In addition to our analyses of snRNA-seq data, our resource included 59 samples with snATAC-seq data, among which 40 were equipped with snMultiome data. Through rigorous QC measures, we extracted 202,510 nuclei with deep sequencing data from the snMultiome datasets. We harnessed transcriptomic and epigenetic information simultaneously to derive robust cell embeddings. This endeavor culminated in the recovery of 28 distinct cell types, aligning closely with the snRNA-seq annotation. These cell types were corroborated by assessing chromatin accessibility associated with marker genes. Furthermore, our uniform processing of snATAC-seq data identified a cumulative total of 449,941 open-chromatin regions across all datasets, a notably larger number than prior brain studies. In line with the established ENCODE convention, we termed these elements single-cell candidate CREs (scCREs). Approximately half of these scCREs displayed cell-type specificity and were distal to genes.

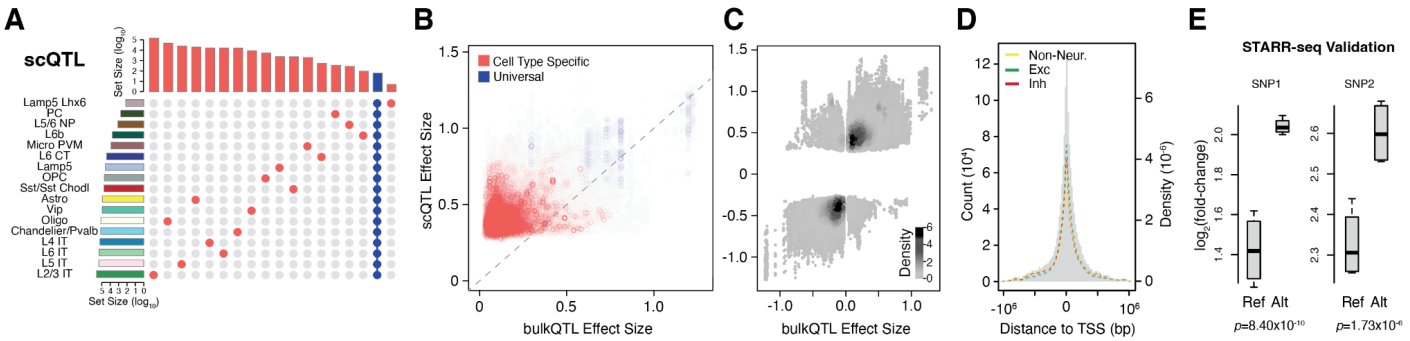

Figure 3. Identification of cell-type-specific eQTLs from single-cell data. A. UpSet plot depicting the count of distinct scQTLs, with unique occurrences in individual cell types (red) or shared across all cell types (blue). **B.** Scatter plots juxtapose absolute eQTL effect sizes between single-cell and bulk RNA-seq datasets, highlighting shared QTLs across >14 cell types (blue) and those confined to a cell type (red). **C.** Density plot facilitates a comparison of eQTL effect sizes between single-cell and bulk RNA-seq datasets. **D.** Histogram profiles the distribution of scQTLs based on distance from the transcription start site (TSS) of the eGene, with normalized distributions emphasized for the collective scQTLs across excitatory, inhibitory, and non-neuronal cell types. **E.** Box plots depict significant expression changes determined through mutSTARR-seq assays for two specific scQTLs in enhancer regions.

We then evaluated expression variation within individual cell types using snRNA-seq data to identify single-cell cis-eQTLs (scQTLs). This process closely mirrored the approach developed by the GTEx consortium [2], which included strict filtering at the cell-type level while generating pseudo-bulk data. Our analysis yielded an average of approximately 85,000 scQTLs within each core cell type, corresponding to around 690 significant eGenes—genes significantly associated with scQTLs (**Fig. 3A**). Aggregating across all cell types culminated in approximately 1.4M scQTLs. Notably, many of these scQTLs demonstrated cell-type specificity, with nearly 47% manifesting across more than one cell type (**Fig. 3B**). Interestingly, approximately 30% of scQTLs demonstrated overlap with bulk cis-eQTLs (**Fig. 3C**). Although the direction of effect remained consistent between these two types, the magnitude of the scQTL effect size tended to be more pronounced than the associated bulk eQTL. These scQTLs exhibited a strong enrichment near transcription start sites (**Fig. 3D**). We validated the functionality of a subset of these elements using targeted STARR-seq (**Fig. 3E**).

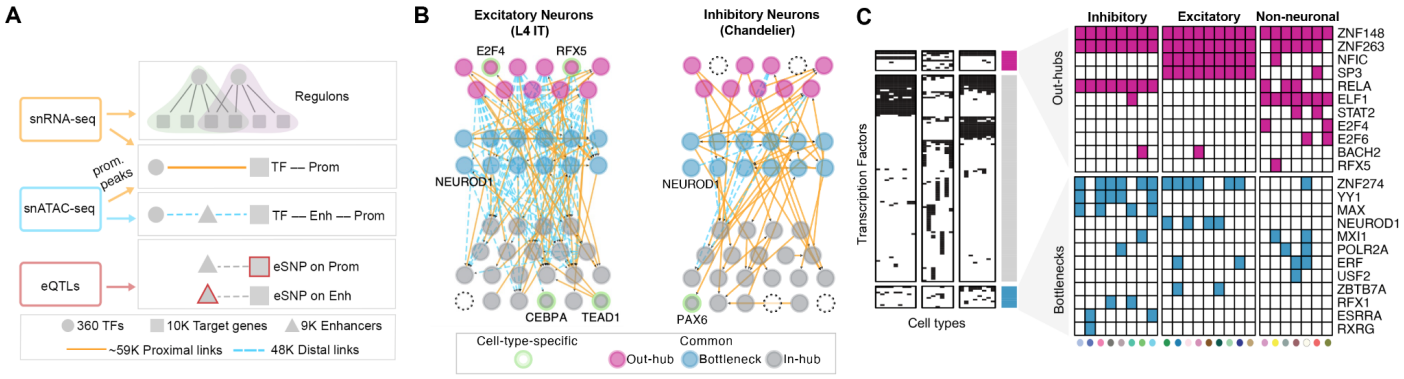

Figure 4. Construction of cell-type-specific GRNs. A. Illustrative diagram outlining the creation of cell-type-specific GRNs using snRNA-seq, snATAC-seq, and scQTL datasets. **B.** Network graphs depicting a selection of excitatory (L4 IT) and inhibitory (Chandelier) GRNs showcasing substantial rewiring. Nodes (transcription factors; TFs) are shaded by type. Solid orange lines denote proximal connections, while dashed blue lines indicate distant links. **C.** The left panel presents the complete array of TFs (y-axis) functioning as hubs or bottlenecks in different cell types (x-axis). Colored cells indicate TFs serving as hubs or bottlenecks within the corresponding cell type. The right panel offers an enlarged view focusing on out-hubs (top) and bottlenecks (bottom).

To gain insights into the regulatory landscape specific to individual cell types, we constructed GRNs tailored to prefrontal cortex cell types (**Fig. 4A**). Our approach incorporated diverse data modalities, including scQTLs, scCREs, TF binding sites, and gene co-expression. The integration of these diverse data types allowed us to link TFs with potential target genes based on their co-expression relationships inferred from snRNA-seq data. Moreover, we mapped eQTLs to promoters and enhancers in the GRNs. Our analyses unveiled intricate patterns of network rewiring across various cell types (**Fig. 4B**). Interestingly, we found that while most highly connected TFs, termed hubs, were largely shared across different cell types, key proteins regulating information flow within the network, termed bottlenecks, demonstrated cell-type specificity (**Fig. 4C**). Enrichment analysis of the targets of these bottleneck proteins revealed clear cell-type-specific functions, including associations with processes such as myelination and axon ensheathment in oligodendrocytes.

Aim 2.2 Coordinating and conducting joint data analyses with PsychENCODE Consortium members to study neuropsychiatric disorders.

The collaborative efforts of the DACC and fellow Consortium members have culminated in the creation of a uniformly processed single-cell functional genomic resource, mostly targeting phenotypically normal human brains at the population level [14]. This comprehensive resource encompasses 3.2M nuclei derived from 388 individuals, delineates 28 well-defined cell types, identifies 85,000 cis-eQTLs, and characterizes 450,000 CREs integrated within GRNs specific to particular cell types (**Aim 2.1**).

Aim 2.3 Leveraging complementary public resources (e.g., the BRAIN Initiative and dGTEx) to expand the PsychENCODE resource in cell type, spatial, and temporal dimensions.

One key outcome of the PsychENCODE Project is integrating its findings within the broader scope of initiatives like the BRAIN Initiative and the dGTEx project. The BRAIN Initiative was launched in 2014 to study brain cell complexities, consisting of the Brain Initiative Cell Census Network (BICCN) from 2017-2022 [4], and the ongoing Brain Initiative Cell Atlas Program (BICAN) as its new phase. We have incorporated BICCN's cell census [33] in our PsychENCODE integrative study [14]. The dGTEx project, linked to GTEx, addresses childhood developmental gene expression gaps. Dr. Gerstein is a member of dGTEx, aiming to understand noncoding genomic variants' impact on gene expression during postnatal brain development. Below we describe our approach to leveraging these initiatives and other public data to enhance the impact of PsychENCODE data.

Aim 2.3.1 Integrating with the BRAIN Initiative

A significant challenge we tackled while integrating single-cell data from various PsychENCODE labs was assigning cell types while overcoming variations in assay timing, the absence of biological replicates, and diverse experimental protocols [14]. Initially, we utilized the cell-type atlas developed by BICCN [33]. However,

we discovered that the atlas had incomplete coverage of glial cell types, despite its comprehensive representation of neuronal cell types. Consequently, we integrated cell type annotations from Dr. Sestan's lab within the PsychENCODE Consortium, focusing on the prefrontal cortex [34]. Our harmonization process encompassed five steps: (1) aggregating samples from both studies using advanced demultiplexing techniques, (2) selecting and jointly analyzing raw count matrices through the robust AnnData framework, (3) linking Leiden clusters from the Sestan study [34] to the BICCN reference cell types via the infer-cell-types function in the Pegasus tool [39], (4) preprocessing raw count matrices from both studies in preparation for anchor transfer, and (5) scrutinizing cell type labels and reconciling discrepancies between the two annotation schemes to create a unified and harmonized cell classification system. Our innovative approach strategically leverages BICCN's extensive neuronal sampling and Sestan's non-neuronal cell data, categorizing cell types under a cohesive annotation scheme. Our systematic methodology ensures data quality by excluding cells with inconsistent classifications, enhancing the reliability of final annotations. Our comprehensive effort results in a collection of 28 distinct cell subclasses within the prefrontal cortex, encompassing excitatory and inhibitory neurons, as well as diverse non-neuronal cells. Through rigorous analysis of over 2.7M annotated nuclei, we validate our annotation scheme by evaluating key marker gene expression patterns and calculating cell type fractions. Moving forward, we plan to incorporate spatial [40], phenotypic (morphological and electrophysiological) [41], and circuit-wiring [42] data from the BICCN phase and human data from the BICAN phase of the BRAIN Initiative with PsychENCODE data. An active PsychENCODE grant led by Dr. Keri Martinowich generates spatial gene expression data in the human brain [43,44], and we are working with her on integrative analysis and data visualization (see **letter of support** from **Dr. Keri Martinowich**).

Aim 2.3.2 Building a time-course multiomic brain resource from prenatal development to maturity

Multiple consortia (PsychENCODE [1], ENCODE [3], Roadmap [45], and dGTEx [2]) have generated multiomic datasets (RNA-seq, ATAC-seq, DNase I hypersensitive sites sequencing [DNase-seq], and ChIP-seq) for the human brain at distinct developmental stages. If effectively integrated, these datasets offer the opportunity to construct a comprehensive resource for investigating molecular regulatory dynamics in brain development. However, several challenges arise. Firstly, the disparate timing of assays during development complicates cross-modality comparisons. Secondly, the scarcity of biological replicates for human samples impedes batch effect assessment. Lastly, variations in experimental protocols among consortia introduce batch effects, including variations in sampling techniques and sequencing platforms. Addressing these challenges, we formulated an integrative resource that collates multiomic time-series brain developmental data from PsychENCODE, ENCODE, and Roadmap [7]. A pivotal aspect of our approach involves grouping samples into three-week time windows to maximize assay availability across the time course and enhance statistical power by treating same-window samples as pseudoreplicates. We are developing a pipeline encompassing a centered log-ratio transformation and batch correction to mitigate technical artifacts while preserving temporal trajectory information. Our methodology extends to integrating different chromatin accessibility assays (e.g., ATAC-seq and DNase-seq) into the same time course. This compiled resource includes DNase-seq, H3K27ac ChIP-seq, and RNA-seq data from brain organoids and fetal brains spanning 7 to 21 post-conception weeks. Utilizing this framework, we intend to merge this resource with time-series data from pediatric brain samples generated by the developmental GTEx project (dGTEx). This integration will encompass RNA-seq, ATAC-seq, DNase-seq, and ChIP-seq data from postnatal samples ranging from infancy to 18 years, thereby expanding the time-series scope from prenatal development to maturity. Concurrently, we will establish harmonized standards and analysis pipelines for PsychENCODE and dGTEx.

Aim 3. Consortium Administration and Coordination

With a strong history of consortium administration and coordination, including experiences with ENCODE, modENCODE, and PsychENCODE, we have a proven ability to foster collaboration both within and outside consortia. We are dedicated to open science practices: our expertise extends to extracting insights from multiomic datasets through coordinating integrative analysis working groups, complemented by our proactive approach to knowledge dissemination through workshops, webinars, and online materials. We have also demonstrated proficiency in developing and managing websites for data sharing and communication, as showcased by our involvement in PsychENCODE and the Extracellular RNA Communication consortium. By forming steering committees, establishing governance documents, and facilitating onboarding processes, we are dedicated to supporting PsychENCODE science on technical and operational aspects, promoting Consortium resources, enhancing inclusivity, and fostering education.

REFERENCES

1. PsychENCODE Consortium, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, et al. The PsychENCODE project. Nat Neurosci. 2015;18:1707–12.

2. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369:1318–30.

3. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583:699–710.

4. Ecker JR, Geschwind DH, Kriegstein AR, Ngai J, Osten P, Polioudakis D, et al. The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. Neuron. 2017;96:542–57.

5. 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 [Internet]. 2018;362. Available from: http://dx.doi.org/10.1126/science.aat8464

7. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science [Internet]. 2018;362. Available from: http://dx.doi.org/10.1126/science.aat7615

12. Welcome to the PsychENCODE knowledge portal [Internet]. [cited 2023 Aug 23]. Available from: http://psychencode.synapse.org

13. PsychSCREEN [Internet]. [cited 2023 Aug 23]. Available from: https://psychscreen.wenglab.org/

14. PsychENCODE consortium et al. Single-cell genomics & regulatory networks for 388 human brains. Science.

15. National Institute of Mental Health Data Archive [Internet]. 2019 [cited 2023 Aug 23]. Available from: https://nda.nih.gov/

18. Pratt HE, Andrews G, Shedd N, Phalke N, Li T, Pampari A, Jensen M, Wen C, PsychENCODE Consortium, Gandal MJ, Geschwind DH, Gerstein M, Moore J, Kundaje A, Colubri A, Weng Z. Using a Comprehensive Atlas and Predictive Models to Reveal the Complexity and Evolution of Brain-Active Regulatory Elements. Science Advances.

19. Avsec Ž, Weilert M, Shrikumar A, Krueger S, Alexandari A, Dalal K, et al. Base-resolution models of transcription-factor binding reveal soft motif syntax. Nat Genet [Internet]. 2021; Available from: https://doi.org/10.1038/s41588-021-00782-6

20. Pampari, Anusri and Shcherbina, Anna and Nair, Surag and Schreiber, Jacob and Patel, Aman and Wang, Austin and Kundu, Soumya and Shrikumar, Avanti and Kundaje, Anshul. Bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. 2023; Available from: https://github.com/kundajelab/chrombpnet

22. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, et al. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. 2010;330:1775–87.

23. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47:D766–73.

24. Andrews G, Fan K, Pratt HE, Phalke N, Zoonomia Consortium§, Karlsson EK, et al. Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. Science. 2023;380:eabn7930.

25. Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, et al. Comparative analysis of the transcriptome across distant species. Nature. 2014;512:445–8.

26. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489:91–100.

27. Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, et al. Comparative analysis of regulatory information and circuits across distant species. Nature. 2014;512:453–6.

28. Cheng C, Yan K-K, Hwang W, Qian J, Bhardwaj N, Rozowsky J, et al. Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. PLoS Comput Biol. 2011;7:e1002190.

29. Lou S, Li T, Kong X, Zhang J, Liu J, Lee D, et al. TopicNet: a framework for measuring transcriptional regulatory network change. Bioinformatics. 2020;36:i474–81.

30. Khurana E, Fu Y, Chen J, Gerstein M. Interpretation of genomic variants using a unified biological network approach. PLoS Comput Biol. 2013;9:e1002886.

31. Cheng C, Andrews E, Yan K-K, Ung M, Wang D, Gerstein M. An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. Genome Biol. 2015;16:63.

32. Accelerating Medicine Partnership in Alzheimer's disease [Internet]. Sage Bionetworks. 2018 [cited 2023 Aug 23]. Available from: https://sagebionetworks.org/research-projects/accelerating-medicine-partnership-in-alzheimers-disease/

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

34. 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. 2022;377:eabo7257.

39. 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. 2020;17:793–8.

40. 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. 2021;598:137–43.

41. Scala F, Kobak D, Bernabucci M, Bernaerts Y, Cadwell CR, Castro JR, et al. Phenotypic variation of transcriptomic cell types in mouse motor cortex. Nature. 2021;598:144–50.

42. Muñoz-Castañeda R, Zingg B, Matho KS, Chen X, Wang Q, Foster NN, et al. Cellular anatomy of the mouse primary motor cortex. Nature. 2021;598:159–66.

43. Maynard KR, Collado-Torres L, Weber LM, Uytingco C, Barry BK, Williams SR, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci. 2021;24:425–36.

44. Tran MN, Maynard KR, Spangler A, Huuki LA, Montgomery KD, Sadashivaiah V, et al. Single-nucleus transcriptome analysis reveals cell-type-specific molecular signatures across reward circuitry in the human brain. Neuron. 2021;109:3088–103.e5.

45. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30.