Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mapping the genetic regulation of long non-coding RNAs across cellular contexts advances understanding of complex brain diseases

Tianyi Zhao

zty2009@hit.edu.cn

Harbin Institute of Technology

Yuran Jia

Harbin Institute of Technology

Li Chen

Westlake University

Liyang Song

Westlake University

Tianyi Zheng

Harbin Institute of Technol

Jian Yang

Westlake University https://orcid.org/0000-0003-2001-2474

Yadong Wang

Harbin Institute of Technol

Article

Keywords:

Posted Date: April 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8945185/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on August 29th, 2026. See the published version at https://doi.org/10.1038/s41467-026-77304-7.

Mapping the genetic regulation of long non-coding RNAs across cellular contexts advances understanding of complex brain diseases

Yuran Jia[1#], Li Chen [4,5,6#], Liyang Song[4,5], Tianyi Zheng[2], Jian Yang[4,5], Yadong Wang[1,2,3*], Tianyi Zhao[1,2,3*]

[1] Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China
[2] School of Medicine and Health, Harbin Institute of Technology, Harbi n 150001, China
[3] Zhengzhou Research Institute, Harbin Institute of Technology, Harbin 450000, China
[4] New Cornerstone Science Laboratory, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China
[5] Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China
[6] School of Engineering, Westlake University, Hangzhou, Zhejiang 310024, China

[*] **Correspondence:**
Tianyi Zhao (zty2009@hit.edu.cn)
Yadong Wang (ydwang@hit.edu.cn)

Abstract

Long non-coding RNAs (lncRNAs) play essential roles in the pathogenesis of neurological disease, but the cellular heterogeneity of their genetic regulation remains poorly understood, with a key barrier being the resolution–power trade-off between bulk and single-nucleus eQTL studies. Here, we profiled 17,541 lncRNAs including both GENCODE-annotated and newly assembled transcripts in the dorsolateral prefrontal cortex (DLPFC) of 2,443 individuals and performed interaction eQTL (ieQTL) mapping based on computational deconvolution. We identified 3,763 interaction-effect lncRNAs (ieLncRNAs) across brain cell types, including 2,783 (74%) previously unannotated in GENCODE and 3,262 (86%) not captured in current single-nucleus eQTL resources. Integration of lncRNA-ieQTLs with GWAS data of brain-related diseases indicated that nearly two-thirds of colocalization signals were missed by bulk eQTL analyses, underscoring the added interpretive resolution afforded by modeling cellular context. We further illustrate this with representative loci, including *AC096667.1* and lncRNAs overlapping *NRGN*, which show cell-context–dependent colocalization with neuroticism and schizophrenia risk, respectively, refining the cellular interpretation of genetic risk. Collectively, this resource advances our understanding of lncRNA contributions to brain disease by uncovering regulatory effects that are obscured in bulk tissues and not captured by single-nucleus eQTL resources.

Main

Expression quantitative trait locus (eQTL) mapping offers an effective route for functionally interpreting noncoding genetic variants identified by genome-wide association studies (GWAS)[1–5]. Bulk transcriptomic analyses of the brain have catalogued maps of genetic variants regulating gene expression at the tissue level[6–11]. The advance of single-cell transcriptomics has revealed pervasive cellular heterogeneity in brain gene expression[12,13], and single-cell eQTL analyses have demonstrated that genetic regulatory effects can vary across cell types and states, thereby extending the functional interpretation of variants beyond tissue-averaged signals to cellular contexts[14–16].

Despite these advances, long noncoding RNAs (lncRNAs) remain a particularly important yet understudied class. These transcripts are increasingly recognized as components of regulatory programs operating within defined neuronal and glial contexts, suggesting that their genetic regulation may be strongly cell-type-dependent[17,18]. Most efforts to date to characterize lncRNA genetic regulation have relied on bulk brain eQTL studies, where tissue-averaged effects can obscure cell-type–specific regulatory signals[7,19]. However, current single-cell eQTL study designs face limited sample sizes and reduced statistical power for detecting effects on lowly expressed transcripts, including lncRNAs[20,21]. Consequently, the cell-type specificity of lncRNA genetic regulation remains poorly understood, limiting the interpretation of noncoding GWAS signals for complex brain disorders.

In this study, we employed interaction eQTL (ieQTL) mapping, a framework that leverages single-cell derived reference profiles to model genotype-by-cell-abundance interactions in bulk RNA-seq data, enabling the identification of regulatory variants whose effects depend on cellular context[22–25]. We integrated and rigorously harmonized genotype and dorsolateral prefrontal cortex (DLPFC) RNA-seq data from 2,443 individuals across seven cohorts, quantified lncRNA expression using a transcriptome model developed in our previous work that captures both annotated and previously unannotated transcripts[19], and applied BayesPrism[26] to infer cellular composition after systematic benchmarking of deconvolution methods. We revealed widespread cell-type–dependent genetic regulation of lncRNAs, a large fraction of which is absent from current GENCODE annotations and prior single-nucleus eQTL studies. We subsequently demonstrated the robustness and cell-type specificity of lncRNA-ieQTLs by leveraging existing single-cell eQTL datasets and cell-type–specific epigenomic features. Integration with brain-related GWAS additionally identified a substantial number of cell-context–dependent colocalization signals that were not detectable in bulk eQTL analyses, highlighting the additional resolution afforded by modeling of cellular context. Collectively, we provide a resource for understanding cell-type–specific genetic regulation of lncRNAs in the human brain transcriptome, filling a major gap in our understanding of lncRNA genetic regulation and offering new insights into how transcriptional regulation shapes brain disease susceptibility (http://brain.hitxqtl.org.cn/Brain-ieQTLAtlas/#/).

Results

Study design and data overview

The overall design of this study is shown in **Fig. 1**, and it comprises three major components: benchmarking of cellular deconvolution, ieQTL mapping, and GWAS integration. Investigating genotype–cellular context interactions requires accurate quantification of cellular composition across individuals[27]. To this end, we evaluated the robustness and reproducibility of computational deconvolution in complex brain tissues based on both simulated and real data (**Fig. 1a**). We then assembled genotype and RNA-seq data from seven cortical cohorts from the PsychENCODE[28] and AMP-AD[29] consortia (**Supplementary Table 1 and Fig. 1b**), restricting analyses to the DLPFC to ensure consistency in tissue origin. After uniform expression quantification and quality control, the final dataset comprised RNA-seq profiles from 2,443 unrelated individuals and 12 million common genetic variants. Across these samples, we identified 17,541 lncRNAs and 14,850 PCGs, encompassing both GENCODE-annotated and unannotated transcripts (Methods). Using a single-nucleus RNA-seq (snRNA-seq) reference profile derived from representative cortical samples (referred to as PFC48), we performed cellular deconvolution of bulk transcriptomic data and mapped genotype–cell-type interaction effects for both lncRNAs and PCGs[13]. Finally, we integrated brain-related GWAS traits to assess the potential role of cell-type–dependent lncRNA regulation in disease susceptibility. (**Fig. 1c**).

Deconvolution benchmarking in the brain cortex

To test the accuracy and stability of cell-type abundance estimation in bulk human brain transcriptomes, we conducted a benchmarking analysis of four widely used deconvolution methods, including MuSiC[30], CIBERSORTx[31], BayesPrism[26], and Bisque[32]. The deconvolution accuracy was measured by the Spearman's rho and mean absolute error (MAE) between the inferred abundance and the ground truth. Our simulation was based on the previously described PFC48 dataset[13], with cell-type annotations obtained from the original publication. We split PFC48 into two portions, randomly assigning one portion to generate synthetic bulk mixtures based on a predefined method, and the other as the reference for deconvolution (Methods). Performance was evaluated against the known cellular composition. Across simulated mixtures, BayesPrism showed the highest accuracy (median correlation = 0.98; median MAE = 0.01), compared with CIBERSORTx (0.96; 0.03), MuSiC (0.95; 0.04), and Bisque (0.94; 0.04) (**Fig. 2a and Fig. 2b**). To better approximate realistic transcriptomic heterogeneity, we further assessed cross-dataset performance using a large-scale snRNA-seq atlas comprising 427 cortical donors[33] (referred to as PFC427). Pseudo-bulk profiles were constructed by aggregating cells at the individual level, thereby explicitly incorporating inter-individual variation and dataset-specific heterogeneity. Under this setting, deconvolution performance was more stable for higher-abundance cell populations, whereas reduced performance was observed for lower-abundance populations (**Supplementary Fig. 1a**). These observations underscore inherent challenges in

accurately resolving rare cell types in realistic biological settings, where reference and target profiles may differ in cellular composition, technical origin, and biological context.

We next assessed the robustness of each method using fluorescence-activated cell sorting (FACS)-purified, lineage-specific expression profiles from the SuperAgerEpiMap[34] and EpiGABA[35] projects (Methods), where cellular identities are unambiguous, thereby testing each method's ability to preserve cell-type specificity and resist spurious contributions from unrelated reference signals during deconvolution. In the $NeuN^+$ and $NeuN^-$ samples (**Supplementary Fig. 1b**), BayesPrism, MuSiC, and CIBERSORTx correctly identified the dominant compartment. In contrast, Bisque produced estimates that deviated from the expected cell-type identity, assigning appreciable abundance to non-neuronal populations in neuron-purified samples. Similar trends were observed in lineage-enriched samples (**Fig. 2c**), where BayesPrism and CIBERSORTx consistently returned estimates aligned with known identities, effectively suppressing noise from non-target cell types. In comparison, MuSiC and Bisque showed greater deviations. Based on the aforementioned evaluations, we selected BayesPrism as the method for estimating cell-type abundances in downstream analyses.

Inferring cellular abundance within seven DLPFC cohorts

We applied BayesPrism to infer cell-type composition from bulk DLPFC transcriptomes across seven cohorts, resolving eight brain cell types including excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, oligodendrocyte precursor cells, microglia, endothelial cells and pericytes. As expected, excitatory neurons represented the most abundant population (**Fig. 2d**), while estimated cell-type proportions varied markedly among cohorts (**Supplementary Fig. 2**). Pairwise correlations between inferred cell types were generally low to moderate (median Spearman rho = **0.26**; **Fig. 2e**), with excitatory neurons showing negative correlations with all other cell types, consistent with their dominant representation in the cortical transcriptome. In samples annotated with Braak stage, we additionally observed a progressive decline in neuronal proportions with increasing pathology, consistent with expected neurodegenerative trajectories[36] (**Supplementary Fig. 3**). Given the limited reliability of deconvolution estimates for rare cell populations, downstream analyses were restricted to cell types with a median estimated proportion above 0.05, namely excitatory neurons, inhibitory neurons, astrocytes, and oligodendrocytes[22]. Although deconvolution provides relative rather than absolute estimates, these high-abundance cell types exhibited strong concordance when the reference profile was replaced with PFC427 (**Supplementary Fig. 4**; average Spearman rho = 0.94), confirming the robustness of proportion inference for these major cell types.

Mapping and validation of ieQTLs in the DLPFC

To increase statistical power, we integrated RNA-seq and genotype data across seven cohorts and performed ieQTL mapping within a ± 1 Mb cis window (**Supplementary Fig. 5**), after which

interaction effects were classified according to how allelic effects varied with cellular context into positive, negative, and uncertain categories (**Fig. 3a**). At a 5% false discovery rate (FDR), we identified 9,306 genes (ieGenes) with at least one significant ieQTL (**Fig. 3b and Supplementary Table 2**). We observed that ieQTLs exhibiting negative interaction effects in one cell type were frequently detected as positive-direction ieQTLs in other cell types (**Supplementary Fig. 6a-c**). This pattern of cross–cell-type sharing corresponded with correlations in the estimated abundances of those cell types, suggesting that some effects may reflect the projection of a genetic signal onto correlated or anticorrelated cellular contexts[37]. To evaluate the robustness of the originally identified genotype × cell-type interaction effects, we implemented a joint model incorporating interaction terms for all four cell types simultaneously and tested whether the originally significant interaction term remained significant (**Supplementary Fig. 6d**). Under this more stringent model, positive-direction ieQTLs consistently exhibited higher rates of significance retention for their originally assigned interaction terms, compared to negative-direction ieQTLs. In contrast, negative-direction ieQTLs more frequently lost statistical support for their initial cell-type assignments, suggesting that these effects may be more susceptible to confounding due to correlations in cellular composition, rather than being anchored to the tested cellular context. Approximately 15–21% of ieQTLs showed uncertain interaction effect directions (**Fig. 3b**), characterized by lower minor allele frequencies (**Supplementary Fig. 7a**) and greater distances from transcription start sites (TSSs) (**Supplementary Fig. 7b**). The proportion of uncertain-direction ieQTLs decreased progressively as FDR thresholds were increasingly stringent (**Supplementary Fig. 7c and 7d**), suggesting that many of these signals likely represent false positives and were therefore excluded from subsequent analyses.

After filtering, we retained 3,445 interaction-effect PCGs (iePCGs) and 3,763 interaction-effect lncRNAs (ieLncRNAs) with ieQTLs showing well-defined effect directions in at least one cell type. Notably, 74% (2783 out of 3763) of the identified ieLncRNAs were previously unannotated in GENCODE (**Supplementary Fig. 8a**). Comparison with cis-eQTLs mapped in bulk tissue showed that the vast majority of ieGenes (93.5%) corresponded to genes harboring at least one cis-eQTL (hereafter referred to as eGenes; **Supplementary Fig. 8b**). A fraction of ieQTLs (19.2%) exhibited weak linkage disequilibrium (coefficient of determination $(R^2) < 0.2$) with any of the corresponding eGene's conditionally independent bulk eQTLs, based on an empirical threshold commonly used to distinguish strongly from weakly correlated regulatory variants (**Supplementary Fig. 8c**)[37]. This pattern is consistent with prior studies showing that cell-type–dependent regulatory signals captured by interaction-based eQTL analyses are often incompletely captured by tissue-level eQTL mapping[22]. Among ieGenes exhibiting positive or negative ieQTL effects in at least one cell type, 6,480 (85.4%) were not previously reported in brain ieQTL studies (**Supplementary Fig. 8d**), likely reflecting advances in lncRNA annotation and genome-wide mapping[6]. We next assessed the concordance of ieQTL using published single-cell eQTL datasets from human cortex, including the Fujita et al.

study[15] (n = 424, partially overlapping with the present cohort) and the Haglund et al. study[16] (n = 391, independent). For positive-direction ieQTLs, consistent replication was observed in matched cell (sub)types across both datasets (**Fig. 3c and Supplementary Fig. 8e**), with median π_1 estimates of 0.82 (from 0.68 to 0.90) in the Fujita et al. dataset and 0.76 (from 0.54 to 0.84) in the Haglund et al. dataset. Cross-cell-type replication patterns reflected the expected mixture of shared and cell-type-enriched effects, as well as (anti)correlations with relative cell type abundances[38]. In contrast, negative-direction ieQTLs showed a more dispersed replication pattern across cell types, with high replication observed in cell types different from the originally tested context (**Supplementary Fig. 8f–h**). This pattern is consistent with observations described earlier, suggesting that negative-direction ieQTLs may not primarily reflect regulatory effects in the tested cell type. Together, the expanded lncRNA annotation and increased statistical power enabled the identification of 3,262 lncRNA–cell type pairs not reported in existing single-cell eQTL studies, many of which correspond to previously unannotated transcripts (**Supplementary Fig. 8i**).

lncRNA-ieQTLs characteristics

Both lncRNA-ieQTLs and PCG-ieQTLs were enriched in transcription start sites (TSSs) (**Supplementary Fig. 9a**). Compared with PCG-ieQTLs, lncRNA-ieQTLs were located closer to TSSs (median lncRNA-ieQTLs distance = 30.1 kb; median PCG-ieQTL distance = 33.4 kb; $P = 1.10 \times 10^{-5}$, Wilcoxon test; **Supplementary Fig. 9b**) and exhibited a more compact cis-regulatory architecture (**Fig. 3d**), consistent with a simpler regulatory dependency[9]. We observed that only 6–48% of intronic and antisense lncRNAs with positive-direction ieQTLs showed significant ieQTL signals in their host PCGs within the same cellular context (**Fig. 3e**), indicating that lncRNA regulatory effects observed in specific cellular contexts are not attributable to corresponding ieQTL effects on their host PCGs. In addition, the distribution of lncRNA-ieQTLs across cell types was highly uneven, with neuronal cell types harboring approximately twice as many lncRNAs with positive-direction ieQTLs as glial cell types (**Fig. 3e**). This pattern may partially reflect higher neuronal abundance and greater inter-individual variability, and may also be influenced by cell-type differences in deconvolution estimation accuracy and power for detecting genotype–cell-type interaction effects[37].

We applied the multivariate adaptive shrinkage (mashr) to lncRNA-ieQTLs to identify patterns of sharing and specificity[39]. Based on mashr posterior inference, we identified 139 and 142 cell-type-specific regulatory effects among iePCGs and ieLncRNAs, respectively (defined as effects with a local false sign rate[40] (LFSR) < 0.05 in only a single cell type), with comparable effect sizes (**Supplementary Fig. 9c**). We next examined how ieQTL effect sizes varied with the degree of cross–cell-type sharing. For both lncRNA- and PCG-ieQTLs, effect sizes decreased significantly as the number of cell types in which the effect was detected increased (LFSR < 0.05; Kruskal–Wallis test;

PCG-ieQTLs, $P = 1.3 \times 10^{-10}$; lncRNA-ieQTLs, $P = 5.6 \times 10^{-15}$; **Supplementary Fig. 9d**), and this trend remained robust when applying a more stringent significance threshold ($P_{ieQTL} < 5 \times 10^{-8}$; **Supplementary Fig. 9e**). To further characterize effect-sharing patterns, we compared pairwise sharing across cell types, defined as consistent effect direction with effect size differences within a twofold range[39]. Within the neuronal lineage, lncRNA-ieQTLs showed lower levels of pairwise sharing compared with PCG-ieQTLs (**Fig. 3f)**. Collectively, lncRNA-ieQTLs and PCG-ieQTLs showed comparable levels of cell-type specificity and regulatory effect magnitude, while exhibiting distinct cross–cell-type effect-sharing patterns.

Epigenomic annotation

To gain insight into the functional properties of ieQTLs, we annotated ieQTLs with cell-lineage-specific ATAC-seq peaks obtained from FACS-sorted $NeuN^+$ and $NeuN^-$ nuclei[41]. Overall, ieQTLs were enriched within cell-lineage–specific open chromatin regions (**Fig. 4a**). When stratified by interaction direction, positive-direction ieQTLs showed stronger enrichment in open chromatin regions of the corresponding cellular lineage than negative-direction ieQTLs. This direction-dependent enrichment pattern is consistent with the differential interpretation of positive- and negative-direction interaction effects described above.

Using cell-type–specific accessibility peaks derived from single-nucleus ATAC-seq (snATAC-seq)[12], we found that positive-direction ieQTLs for glial cell types, including astrocytes (PCG-ieQTLs fold enrichment = 2.20, 95% CI 1.82–2.57; lncRNA-ieQTLs fold enrichment = 1.80, 95% CI 1.40–2.20) and oligodendrocytes (PCG-ieQTLs fold enrichment = 3.25, 95% CI 2.92–3.58; lncRNA-ieQTLs fold enrichment = 2.33, 95% CI 1.99–2.68), were strongly enriched within their respective accessible chromatin regions, showing progressively stronger enrichment with increasing significance stringency. In contrast, positive-direction ieQTLs for excitatory neurons (PCG-ieQTLs fold enrichment = 1.78, 95% CI 1.60–1.96; lncRNA-ieQTLs fold enrichment = 1.51, 95% CI 1.33–1.68) and inhibitory neurons (PCG-ieQTLs fold enrichment = 1.36, 95% CI 1.18–1.54; lncRNA-ieQTLs fold enrichment = 1.22, 95% CI 1.03–1.41), showed comparatively modest enrichment (**Fig. 4b**). This overall pattern was similarly observed when using cell-type–specific chromatin immunoprecipitation followed by sequencing (ChIP-seq) marks[42] (H3K4me3 and H3K27ac), consistent with prior reports that neuronal regulatory variants often display weaker cell-type specificity than those in glial lineages[14] (**Fig. 4c**).

Figure 4d illustrates the chromatin landscape surrounding the lncRNA locus *HAR1A*, which contains the excitatory neuron–specific ieQTL rs3746746. This SNP resides in an exon of *HAR1A* and, despite lacking a strong ATAC-seq signal, shows extensive overlap with neuron-specific ChIP-seq marks,

indicating that it is located within an active regulatory environment. The *HAR1A* locus is located within a human accelerated region (HAR), a class of noncoding regions that often function as enhancers in the nervous system[43], regulating neuronal development and maturation, and have been linked to central nervous system disorders such as Huntington's disease[44–46]. In addition, PLAC-seq data indicate a chromatin interaction between the *HAR1A*-associated regulatory region and the downstream *MIR124-3* locus, which encodes a key neuronal regulator[47]. These features implicate rs3746746 as part of a neuron-specific regulatory architecture, potentially modulating transcriptional activity at the *HAR1A* locus.

Colocalization with disease susceptibility

To integrate lncRNA-ieQTLs with brain-related genetic risk, we compiled GWAS summary statistics for 36 phenotypes spanning four major categories: neuropsychiatric disorders, neurodegenerative diseases, cognitive traits, and brain morphological or developmental features (**Supplementary Table 3 and Methods**). We observed that lncRNA-ieQTLs were enriched for GWAS-associated variants (**Supplementary Fig. 10**). This noteworthy pattern underscores the functional relevance of lncRNA-ieQTLs and offers insight into the cellular-context–dependent regulatory mechanisms underlying complex traits.

We performed colocalization analyses[48] between lncRNA ieQTLs (FDR < 0.25) and the GWAS traits described above, following the relaxed threshold adopted in previous interaction QTL studies to accommodate reduced statistical power in detecting interaction effects[22,37] (**Supplementary Fig. 11 and Supplementary Table 4**). Colocalization was most frequently observed for traits such as schizophrenia, neuroticism, and educational attainment, whereas no colocalization signals were detected for multiple sclerosis or autism spectrum disorder. Notably, 86.6% (123 out of 142) of lncRNA-ieQTLs showing colocalization involved lncRNAs absent from current GENCODE annotations. For comparison, we also performed colocalization analyses using PCG-ieQTLs as well as lncRNA- and PCG- eQTLs (FDR < 0.25; **Fig. 5a**). Across all analyses, 45% (113 out of 248) of colocalization signals were uniquely identified through lncRNA-ieQTLs, underscoring the additional resolution gained by incorporating cellular context into regulatory mapping. An additional 21% (52 out of 248) of colocalizations were supported by both lncRNA-ieQTLs and matched lncRNA eQTLs, reinforcing the cell-type–specific regulatory context of these trait associations and helping to refine their likely cellular origin.

As shown in **Fig. 5b**, the GENCODE-annotated lncRNA *AC096667.1* (chr2: 192,711,265–192,908,909; lead variant rs1019706) is specifically expressed in inhibitory neuron and exhibited strong colocalization with neuroticism risk loci in the inhibitory neuron ieQTLs (PP4 = 0.99; **Fig. 5c**), whereas no colocalization signal was detected using standard bulk eQTL (PP4 = 0.00).

Supplementary Figs. S12 and S13 provide additional examples of lncRNA-ieQTLs colocalizing with brain-related trait associations. These cases illustrate how lncRNA-ieQTLs can refine the cellular interpretation of GWAS associations and reveal regulatory signals that are masked in tissue-averaged analyses.

As colocalization analysis alone cannot distinguish true pleiotropy from confounding due to LD, we additionally applied summary-based Mendelian randomization[49] (SMR) with heterogeneity testing (HEIDI) to refine causal inference. We identified 134 loci (**Supplementary Table 5)**, of which 52 were also supported by COLOC within the same cellular context (**Fig. 5d and Supplementary Fig. 14**). One illustrative locus involves the schizophrenia risk variant rs12293670[50,51], previously reported as a region-dependent eQTL in the DLPFC affecting expression of *Neurogranin*[52] (*NRGN*), a neuron-specific gene and direct thyroid hormone target implicated in synaptic plasticity[53]. lncRNA-ieQTL revealed that the SCZ GWAS signal colocalizes with the newly assembled antisense lncRNA *NRGN-AS1* (chr11:124,610,987–124,616,233; lead variant rs12293670; PP.H4 = 0.95; **Fig. 5e**), rather than with NRGN (PP.H4 = 0.00). The allelic effect of rs12293670 increased with excitatory neuron abundance and decreased with inhibitory neuron abundance. Epigenomic annotations further showed that rs12293670 resides within a neuron-specific enhancer marked by H3K27ac and H3K4me1 (**Supplementary Fig. 15**). Notably, this specificity was observed despite highly correlated expression of NRGN and *NRGN-AS1* in excitatory neurons (r = 0.71; **Supplementary Fig. 15**). SMR evidence further supported an association between schizophrenia risk and *NRGN-AS1* expression ($P_{SMR} = 4.99 \times 10^{-9}$; $P_{HEIDI} = 0.03$). SMR yielded a negative gene–trait effect estimate for NRGN-AS1 ($\beta_{SMR} = -0.21$), indicating that higher expression levels of this lncRNA are associated with lower SCZ risk. Given the observed downregulation of *NRGN* in SCZ patients[54], we hypothesize that *NRGN-AS1* may function to stabilize *NRGN* expression, mitigating synaptic vulnerability linked to disease risk. These analyses demonstrate how lncRNA-ieQTLs can nominate previously uncharacterized non-coding transcripts as candidate regulatory targets and inform hypotheses about the cell-type specificity of disease-associated genetic effects.

Discussion

Despite extensive efforts to characterize cell-type-specific genetic regulation in the human brain, lncRNAs remain a substantially understudied component. Here, leveraging a large cortical dataset (n = 2,443) assembled through cross-cohort integration together with an expanded lncRNA transcript annotation, we systematically evaluated genotype–cell-type interaction effects on lncRNA expression. We identified 3,763 ieLncRNA regulatory associations whose effects vary with cellular context, many of which map to transcripts not represented in current GENCODE annotations. To facilitate downstream exploration and reuse, we developed an online resource (http://brain.hitxqtl.org.cn/Brain-ieQTLAtlas/#/) that enables users to browse and download summary statistics for lncRNA ieQTLs,

providing a foundation for investigating cell-type–specific genetic regulation of lncRNAs and interpreting lncRNAs in relation to brain-related complex traits within defined cellular contexts.

Current brain single-nucleus eQTL studies remain limited in their ability to comprehensively characterize lncRNA genetic regulation, owing to both limited sample sizes and technical constraints of prevailing protocols. In particular, the widespread reliance on oligo(dT)-primed poly(A) capture and intact 3′ transcript ends renders transcript detection highly sensitive to RNA integrity, posing additional challenges for frozen or partially degraded brain tissue and further reducing recovery of noncoding transcripts. In this context, ieQTL analysis provides a complementary framework for interrogating cell-type-specific genetic regulation of lncRNAs by incorporating cellular context. Consistent with this complementarity, approximately 87% of the lncRNA–cell-type pairs identified in this study were not captured by existing single-nucleus eQTL datasets. Enrichment of ieQTLs within lineage-specific epigenomic annotations further supports the biological plausibility of these signals, indicating consistency with cellular-context–dependent cis-regulatory features rather than statistical artifacts driven by compositional confounding. When stratified by interaction direction, positive- and negative-direction ieQTLs exhibited distinct enrichment profiles: positive-direction signals showed stronger and more coherent enrichment within regulatory elements characteristic of their corresponding cellular lineages, whereas negative-direction signals displayed weaker or less consistent correspondence. Within positive-direction ieQTLs, enrichment extended to cell-type–specific epigenomic annotations, reinforcing their interpretability as cell-type–specific regulatory loci.

Integrating lncRNA-ieQTLs with brain-related GWAS datasets deepens our understanding of the molecular mechanisms underlying noncoding disease variants. From 13 brain-related phenotypes, we identified 142 lncRNAs that colocalize with disease risk loci, the majority of which are absent from current GENCODE annotations. Notably, 68% of the lncRNA-ieQTL–supported colocalization signals (113 of 165) were not detected using bulk eQTLs, underscoring regulatory associations that are not captured by tissue-level analyses. Although ieQTLs may not always localize the precise cellular source of regulatory activity, they provide meaningful insight into the cellular contexts most likely to mediate genetic risk, thereby refining the interpretive potential of GWAS findings. Further, colocalization analyses prioritized lncRNAs over nearby protein-coding genes, underscoring that current efforts have yet to fully account for regulatory contributions from noncoding transcripts (**Fig. 5e and Supplementary Fig. 15**). Together, these observations position lncRNA-ieQTLs as a critical complement to existing regulatory maps for interpreting disease-associated genetic variation in the human brain, while simultaneously expanding the catalog of functionally relevant lncRNAs with potential roles in disease biology.

Despite these advances, several limitations should be noted. Bulk-based ieQTL analyses can be influenced by uncertainty in deconvolution for low-abundance cell populations, dilution of rare cell-type–specific signals in heterogeneous tissue, and reduced sensitivity for detecting regulatory effects on very lowly expressed lncRNAs. In addition, correlations among cell-type proportions mean that genotype–cell-abundance interaction effects can be shaped by coordinated variation across cellular populations. Consistent with this, negative-direction ieQTLs tend to show greater sharing across inversely correlated cell types, suggesting that some interaction signals may reflect broader cellular composition patterns rather than regulation confined to a single cell type (Supplementary Fig. 6). Together, these considerations indicate that cell-type–specific genetic regulation of lncRNAs may not always be fully resolved from bulk tissue data alone, underscoring the importance of complementary validation using population-scale single-nucleus eQTL studies.

Despite these limitations, we generated the largest population-scale lncRNA-ieQTL resource for the human brain to date. Our analyses reveal widespread cell-type–specific genetic regulation of lncRNAs, a regulatory layer that has remained difficult to capture in previous studies. This resource links brain-related GWAS loci to candidate regulatory lncRNAs acting in specific cellular contexts, and is expected to facilitate hypothesis generation regarding disease-relevant regulatory pathways by distinguishing regulatory effects across cellular environments, thereby supporting downstream functional and translational studies of neurological and psychiatric disorders.

Methods

Ethical approval

This study received ethical approval from the Ethics Committee of Westlake University, under approval number 20200722YJ001.

Datasets collection and description

Bulk RNA-seq resources. We integrated genotype and bulk RNA-seq data from seven publicly available cohorts contributed by the PsychENCODE (https://nda.nih.gov/pec, BrainGVEx, LIBD, CMC, HBCC)[28] and AMP-AD (https://adknowledgeportal.synapse.org/)[29] (MSBB, Mayo, ROSMAP) consortia (**Supplementary Table 1**). To ensure consistency in tissue origin, analyses were restricted to DLPFC samples. For the MSBB cohort, which includes four cortical regions (BM10, BM22, BM36, and BM44), only Brodmann area 10 (BM10; prefrontal cortex) samples were retained. Across cohorts, RNA-seq FASTQ files were processed using a uniform pipeline, including read-level quality assessment with FastQC, adapter and low-quality base trimming with Cutadapt[55], and alignment to GRCh37 using STAR (v2.7.8a) in two-pass mode[56]. Gene-level expression of lncRNAs and PCGs was quantified using RNA-SeQC (v2.3.5)[57], based on a transcript model generated by integrating our

cohort-derived lncRNA catalog with GENCODE v37[19]. For downstream analyses, we retained lncRNAs that were detected (TPM > 0.5) in all seven cohorts.

snRNA-seq resources.

PFC48 (https://www.synapse.org/#!Synapse:syn18485175)[13]: The PFC48 dataset generated by Mathys et al. comprises prefrontal cortex (PFC) samples from 48 individuals in the ROSMAP cohort, including 24 neuropathologically normal or minimally affected controls and 24 age-matched individuals with varying degrees of Alzheimer's disease–related pathology. In total, 80,660 nuclei were profiled, of which 70,634 passed quality control and were retained. Nuclei were annotated into eight major brain cell types based on transcriptomic profiles: astrocytes (n = 3,392), endothelial cells (n = 121), excitatory neurons (n = 34,976), inhibitory neurons (n = 9,196), microglia (n = 1,920), oligodendrocytes (n = 18,235), oligodendrocyte precursor cells (n = 2,627), and pericytes (n = 167). Cell-type annotations were taken directly from the original publication.

PFC427 (https://www.synapse.org/#!Synapse:syn52293417)[33]: The PFC427 dataset also obtained from Mathys et al., the PFC427 dataset comprises approximately 2.3 million nuclei isolated from PFC samples of 427 participants in the ROSMAP cohort. Participants spanned a spectrum of Alzheimer's disease (AD) pathology, ranging from non-AD and early-stage (Braak 0–II) to intermediate (III–IV) and late-stage AD (V–VI). After quality control and doublet removal, 2,359,994 nuclei were retained for analysis. The original study identified 54 transcriptionally defined cell types grouped into 12 major classes, including excitatory neurons (n = 1,043,230), inhibitory neurons (n = 329,699), oligodendrocytes (n = 645,142), oligodendrocyte precursor cells (n = 90,502), astrocytes (n = 149,558), five immune cell types—including microglia, CNS-associated macrophages (CAMs), and T cells (n = 83,889)—and several vascular cell types, including endothelial cells, smooth muscle cells, fibroblasts, and pericytes (n = 17,974).

FACS resources.

EPIGABA (https://www.synapse.org/Synapse:syn12034263)[35]. Gene expression profiles were generated from human PFC samples by Kozlenkov et al. using a FACS protocol. Nuclei were isolated and enriched for inhibitory GABAergic interneurons (SOX^+, n = 4), excitatory glutamatergic neurons (GLU^+, n = 4), and oligodendrocytes ($OLIG^+$, n = 4).

SuperAgerEpiMap (https://www.synapse.org/#!Synapse:syn25672226)[34]. FACS-based RNA-seq data generated by Dong et al. from human cortex. Nuclei were sorted into neuronal ($NeuN^+$, n = 10) and non-neuronal ($NeuN^-$, n = 10) populations across five cortical regions (Brodmann areas 10, 17, 22, 36, and 44) from ten neurotypical adult donors.

Epigenomic resources.

Cell-type-specific ChIP–seq peaks for H3K27ac and H3K4me3 were obtained from Nott et al.[42]. Lineage-specific ATAC-seq peaks distinguishing $NeuN^+$ and $NeuN^-$ nuclei were obtained from Fullard et al.[41], whereas cell-type-specific chromatin accessibility peaks were derived from Emani et al.[42]

GWAS resources.

We curated publicly available GWAS summary statistics for 36 brain-related phenotypes, classified into four major categories: (1) Neuropsychiatric disorders (e.g., schizophrenia and bipolar disorder); (2) Cognitive traits (e.g., intelligence and educational attainment); (3) Neurodegenerative disorders (e.g., Alzheimer's disease and Parkinson's disease); (4) Morphological and developmental traits (e.g., regional brain volume and age at menarche). Full details on included phenotypes, original publications, and GWAS sample sizes are provided in **Supplementary Table 3**.

Benchmarking of deconvolution methods

Basic settings. We systematically benchmarked four widely used cell-type deconvolution methods: MuSiC[30], CIBERSORTx[31], BayesPrism[26], and Bisque[32]. For CIBERSORTx (https://cibersortx.stanford.edu/), we constructed a reference signature matrix from the PFC48 dataset by computing cell-type–average expression profiles, followed by counts-per-million (CPM) normalization and log transformation; this matrix was then submitted for deconvolution. For MuSiC, BayesPrism and Bisque, deconvolution was performed using each method's R package with default parameters, using the raw UMI count matrix from PFC48 as the reference input. Method performance was evaluated using Spearman's rank correlation coefficient to assess concordance between inferred and true cell-type proportions, and MAE to quantify estimation error.

Simulation evaluation settings. To generate synthetic bulk RNA-seq mixtures with predefined cell-type compositions, we first randomly divided the PFC48 dataset into two subsets: one serving as the source for constructing simulated mixtures and the other used as the reference for deconvolution. Each synthetic sample consisted of 500 nuclei drawn from a single donor in the source subset, thereby preserving inter-individual expression heterogeneity[58]. For each mixture, between five and all available cell types from that donor were included. One cell type was randomly chosen and assigned a free proportion $f_c \in [0,1]$. The remaining cell types were assigned unnormalized proportions r_c', which were then normalized and scaled so that the combined proportion equaled $1 - f_c$, ensuring that total composition summed to 1: $f_c' = \frac{r_c'}{\sum r_c'} \times (1 - f_c)$. The number of nuclei contributed by each cell type was computed by multiplying the final proportions by 500. Expression profiles were generated

by summing counts from nuclei sampled with replacement according to these proportions, yielding a total of 5,000 simulated mixtures.

Real-data evaluation settings. To complement simulation-based assessments, we further evaluated deconvolution accuracy using real datasets with known or high-confidence cell-type compositions. Specifically, we used FACS-purified samples from the SuperAgerEpiMap and EpiGABA projects as ground-truth benchmarks. Each deconvolution method was applied to the bulk RNA-seq profiles of these samples using PFC48 as the single-cell reference. Performance was evaluated based on whether the inferred cell-type abundances were predominantly assigned to the expected cell type. For example, a $NeuN^+$ sample should be almost entirely decomposed into neuronal components, with minimal contributions from glial or other non-target populations.

Mapping ieQTLs

For ieQTL analysis, genotype quality control was performed using a standardized pipeline consistent with our previous work[59]. SNPs were excluded if they had a genotype call rate below 95%, Hardy-Weinberg equilibrium p-value less than 1×10^{-8}, minor allele frequency (MAF) below 0.01, or imputation information score below 0.3. We further excluded individuals of non-European ancestry, as determined by principal component analysis, and one of each pair of related individuals (genetic relatedness > 0.05). After quality control, we retained RNA-seq data from 2,443 unrelated individuals and 12 million genotyped or imputed common variants for subsequent analyses. Gene expression levels for lncRNAs and PCGs were first normalized for library size using the trimmed mean of M-values (TMM) method, followed by rank-based inverse normal transformation (RINT) to mitigate the influence of outliers and better satisfy linear model assumptions.

We employed TensorQTL[60] to perform ieQTL mapping. For each gene, genetic variants located within ± 1 Mb of the TSS were tested. To minimize potential bias caused by extreme cell-type proportions, we applied a stratified filtering approach: only variants with a MAF > 1% in both the lower and upper 50% of samples (based on cell-type abundance) were included. A linear interaction model ($Y \sim G + E + G \times E + C$) was fitted for each gene–SNP pair, where: Y represents the RINT-transformed gene expression; G is the genotype; E is the RINT-transformed cell-type abundance estimated via BayesPrism; $G \times E$ is the genotype–cell abundance interaction term; C denotes a set of covariates, including genotype principal components, sex, age at death, and probabilistic estimation of expression residuals (PEER)[61] factors. The number of PEER factors was dynamically adjusted based on sample size: 15 for $n < 150$; 30 for $150 \leq n < 250$; 45 for $250 \leq n < 350$; and 60 for $n \geq 350$. To increase statistical power, we combined data across cohorts and used a unified linear regression model to remove batch effects prior to downstream analyses. Multiple testing correction was conducted in two stages: first, eigenMT[62] was used to correct for the effective number of tests per gene based on

local LD structure; second, the Benjamini-Hochberg procedure was applied genome-wide to control the false discovery rate (FDR), with significance defined as FDR < 0.05.

Based on the dynamic behavior of the genotype main effect (β_G) across levels of cell-type abundance, we classified significant ieQTLs into three functional subtypes following a framework that has been extensively benchmarked and adopted by the GTEx Consortium[22]. Specifically, Positive direction: $|\beta_G|$ increases significantly in the high-abundance (top 25%) compared to the low-abundance (bottom 25%) group; Negative direction: $|\beta_G|$ decreases significantly in the high-abundance group; Uncertain direction: the sign of β_G reverses between low- and high-abundance subgroups ($\beta_{G\ low} \times \beta_{G\ high} < 0$).

SNP annotation was performed using ANNOVAR[63], using the integrated transcriptome model established in our previous work[19], incorporating both GENCODE (v37) annotations and newly assembled lncRNAs.

Replication of ieQTLs

We assessed the reproducibility of ieQTLs by comparison with previously published single-cell eQTL datasets. Specifically, we evaluated reproducibility using datasets from Fujita et al.[15], which were derived from the ROSMAP cohort and mapped eQTLs across seven major brain cell types and 54 transcriptionally defined cell subtypes, and from Haglund et al.[16], which profiled eQTLs across eight major brain cell types in an independent cohort not included in this study. Replication was quantified using the π_1 statistic, which estimates the proportion of true positive. The π_1 statistic was calculated using the qvalue R package[64], with the tuning parameter λ set to 0.5 when more than 20 tests were available and to 0.85 when more than 100 tests were available in the replication dataset[38].

Cell-Type specificity of ieQTL effects

To quantify cross–cell-type sharing of ieQTLs, we applied multivariate adaptive shrinkage (mashR)[39]. For each gene, we selected a lead ieQTL defined as the variant with the smallest interaction P value across the four cell types within the cis-regulatory window. The corresponding effect sizes and standard errors were compiled to form a strong signal set for learning data-driven covariance structures. In addition, we randomly sampled 250,000 genome-wide SNP–gene pairs to estimate the background correlation structure of effect sizes across cell types. This set was used to compute a data-driven correlation matrix that captures shared and distinct patterns of regulatory effects. Using the learned covariance and correlation structures, MashR was then used to estimate posterior effect sizes and local false sign rates (LFSRs) for each SNP–gene pair.

Epigenomic enrichment

We assessed the functional enrichment of ieQTLs in cell-type-associated regulatory annotations. Let N denote the total number of ieQTLs tested and n_{obs} the number overlapping a given regulatory annotation, yielding an observed overlap proportion $x = \frac{n_{\text{obs}}}{N}$. To mitigate ascertainment bias, we constructed an empirical null distribution by randomly sampling N control SNPs from the set of all SNPs tested in the analysis, matching minor allele frequency and distance to the transcription start site. This procedure was repeated 1,000 times, and for each iteration i, the overlap proportion y_i between the control variants and the same regulatory annotation was calculated. The expected overlap proportion was defined as $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$. Functional enrichment was quantified as the fold enrichment $FE = \frac{x}{\bar{y}}$. The sampling variance of the fold enrichment estimate was approximated using the delta method[24,65]. In general, the variance of the ratio $\frac{x}{\bar{y}}$ can be expressed as $var\left(\frac{x}{\bar{y}}\right) \approx \left(\frac{x}{\bar{y}}\right)^2 [\frac{var(x)}{x^2} + \frac{var(\bar{y})}{\bar{y}^2} - \frac{2cov(x,\bar{y})}{x\bar{y}}]$. Assuming that the covariance term $cov(x, \bar{y})$ is negligible and that $var(x)$ can be approximated by the empirical variance of the control overlap proportions, the variance simplifies to $var\left(\frac{x}{\bar{y}}\right) \approx \left(\frac{x}{\bar{y}}\right)^2 \frac{\widehat{var}(x)}{x^2} + \frac{\widehat{var}(\bar{y})}{m\bar{y}^2}$, where $\widehat{var}(\bar{y})$ denotes the sample variance of the control overlap proportions across resampling iterations.

Colocalization analysis

We integrated GWAS summary statistics with ieQTL signals using two complementary summary-level approaches. Given the inherently reduced statistical power of interaction QTL mapping and the additional constraints imposed by downstream causal inference analyses, ieQTLs were selected for downstream analyses using a relatively permissive significance threshold (FDR < 0.25). We performed Bayesian colocalization analysis using the COLOC R package[48] to evaluate whether an ieQTL signal and a brain disorder–associated GWAS signal were likely driven by the same causal variant. For each locus, COLOC estimates posterior probabilities for five competing hypotheses, and colocalization was considered supported when the posterior probability for a shared causal variant (PP.H4) exceeded 0.8. We applied summary data–based Mendelian randomization (SMR; https://yanglab.westlake.edu.cn/software/smr/) analysis to further assess putative causal relationships between gene expression and disease risk[49]. Statistical significance was determined using a Bonferroni-corrected threshold ($p_{SMR} < 0.05/m$ where m denotes the total number of tested genes). To distinguish true pleiotropic associations from confounding due to LD between independent causal variants, we further applied the heterogeneity in dependent instruments (HEIDI) test, excluding loci showing evidence of heterogeneity ($p_{HEIDI} \leq 0.01$). For SMR and HEIDI analyses, LD estimates were derived from 8,652 unrelated individuals from the Health and Retirement Study (dbGaP: phs000428), which served as the reference panel. To validate cell type–specific gene expression, we

quantified lncRNAs and PCGs in FACS-purified samples from the SuperAgerEpiMap and EpiGABA projects, applying the same inclusion criteria.

Author contributions

T.Z. and Y.W. conceived and supervised the study. J.Y., L.C., and J.Y. designed the experiments. Y.J. and L.C. performed data collection, cleaning, and quality control. Y.J. conducted data analyses with support and guidance from L.S. and L.C. Y.J. drafted the initial manuscript. All authors contributed to manuscript review and revision and approved the final version.

Acknowledgements

We thank the participants involved in this research project. We thank Ting Qi, Ruilei Ma, and Yanyu Xiao for valuable discussions. This research was partly supported by the National Key R&D Program of China (2024YFC3405800, 2024YFC3405802). We thank the Westlake University High-Performance Computing Center for assistance in computing. This study makes use of data from HRS (dbGaP accession: phs000428), GTEx (dbGaP accession: phs000424), PsychENCODE (Synapse accession: syn4921369), and AMP-AD (Synapse accession: syn5550382). A full list of acknowledgements to these data sets can be found in the **Supplementary Note**.

Data Availability

PsychENCODE data: https://www.synapse.org/#!Synapse:syn4921369/files/. AMP-AD data: https://www.synapse.org/#!Synapse:syn5550382. Full summary statistics for PCG-ieQTLs and lncRNA-ieQTLs: https://zenodo.org/records/13690169. Other publicly available datasets used in this study are listed below: Fujita et al. eQTL summary statistics: https://www.synapse.org/Synapse:syn52335807. Haglund et al. eQTL summary statistics: https://zenodo.org/records/13343729. MetaBrain ieQTL summary statistics: https://www.metabrain.nl. EPIGABA: https://www.synapse.org/Synapse:syn4588488. SuperAgerEpiMap: https://www.synapse.org/Synapse:syn25716854.

Code availability

Custom scripts used in this study are available at GitHub: https://github.com/yrjia1015/Brain-ieQTL. Additional software tools used for data analysis are listed below: eQTL mapping software: https://github.com/broadinstitute/tensorqtl. SMR: https://yanglab.westlake.edu.cn/software/smr/. coloc: https://chr1swallace.github.io/coloc/. mashr: https://stephenslab.github.io/mashr/. ANNOVAR: http://annovar.openbioinformatics.org/. qvalue: https://github.com/StoreyLab/qvalue.

References

1. Uffelmann, E. *et al.* Genome-wide association studies. *Nat Rev Methods Primers* **1**, 1–21 (2021).
2. Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *The American Journal of Human Genetics* **110**, 179–194 (2023).
3. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research* **51**, D977–D985 (2023).
4. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491–504 (2018).
5. Qi, T., Song, L., Guo, Y., Chen, C. & Yang, J. From genetic associations to genes: methods, applications, and challenges. *Trends in Genetics* **40**, 642–667 (2024).
6. de Klein, N. *et al.* Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat Genet* **55**, 377–388 (2023).
7. He, Y. *et al.* Genetic regulation of lncRNA expression in whole human brain and their contribution to CNS disorders. *Brief Bioinform* **26**, bbaf291 (2025).
8. Vattathil, S. M. *et al.* Mapping the microRNA landscape in the older adult brain and its genetic contribution to neuropsychiatric conditions. *Nat Aging* **5**, 306–319 (2025).
9. De Goede, O. M. *et al.* Population-scale tissue transcriptomics maps long non-coding RNAs to complex disease. *Cell* **184**, 2633-2648.e19 (2021).
10. Consortium, T. Gte. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* https://doi.org/10.1126/science.aaz1776 (2020) doi:10.1126/science.aaz1776.
11. Liu, Z. *et al.* Detection of circular RNA expression and related quantitative trait loci in the human dorsolateral prefrontal cortex. *Genome Biol* **20**, 99 (2019).
12. Emani, P. S. *et al.* Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 (2024).
13. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
14. Bryois, J. *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).

15. Fujita, M. *et al.* Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat Genet* **56**, 605–614 (2024).
16. Haglund, A. *et al.* Cell state-dependent allelic effects and contextual Mendelian randomization analysis for human brain phenotypes. *Nat Genet* **57**, 358–368 (2025).
17. Srinivas, T., Siqueira, E. & Guil, S. Techniques for investigating lncRNA transcript functions in neurodevelopment. *Mol Psychiatry* **29**, 874–890 (2024).
18. Liau, W.-S. *et al.* Fear extinction is regulated by the activity of long noncoding RNAs at the synapse. *Nat Commun* **14**, 7616 (2023).
19. Chen, L. *et al.* Genetic control of non-coding RNAs in the human brain and their implications for complex traits. *The American Journal of Human Genetics* **0**, (2025).
20. Liau, W.-S., Samaddar ,Sarbani, Banerjee ,Sourav & and Bredy, T. W. On the functional relevance of spatiotemporally-specific patterns of experience-dependent long noncoding RNA expression in the brain. *RNA Biology* **18**, 1025–1036 (2021).
21. Xue, A., Yazar, S., Neavin, D. & Powell, J. E. Pitfalls and opportunities for applying latent variables in single-cell eQTL analyses. *Genome Biology* **24**, 33 (2023).
22. Kim-Hellmuth, S. *et al.* Cell type–specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).
23. Sheng, X. *et al.* Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat Genet* **53**, 1322–1333 (2021).
24. Song, L., Sun, X., Qi, T. & Yang, J. Mixed model-based deconvolution of cell-state abundances (MeDuSA) along a one-dimensional trajectory. *Nat Comput Sci* **3**, 630–643 (2023).
25. Boltz, T. *et al.* Cell-type deconvolution of bulk-blood RNA-seq reveals biological insights into neuropsychiatric disorders. *The American Journal of Human Genetics* **111**, 323–337 (2024).
26. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer* **3**, 505–517 (2022).
27. Herrera-Luis, E., Benke, K., Volk, H., Ladd-Acosta, C. & Wojcik, G. L. Gene–environment interactions in human health. *Nat Rev Genet* **25**, 768–784 (2024).

28. THE PSYCHENCODE CONSORTIUM. Revealing the brain's molecular architecture. *Science* **362**, 1262–1263 (2018).

29. Ertekin-Taner, N. & Petanceska, S. S. Ten year anniversary of AMP AD: Enabling a Precision Medicine Approach to Target and Biomarker Discovery for Alzheimer's Disease. *Alzheimer's & Dementia* **20**, e086464 (2024).

30. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* **10**, 380 (2019).

31. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773–782 (2019).

32. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat Commun* **11**, 1971 (2020).

33. Mathys, H. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365-4385.e27 (2023).

34. Dong, P. *et al.* Population-level variation in enhancer expression identifies disease mechanisms in the human brain. *Nat Genet* **54**, 1493–1503 (2022).

35. Kozlenkov, A. *et al.* A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Science Advances* **4**, eaau6190 (2018).

36. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239–259 (1991).

37. Kasela, S. *et al.* Interaction molecular QTL mapping discovers cellular and environmental modifiers of genetic regulatory effects. *The American Journal of Human Genetics* **111**, 133–149 (2024).

38. Kasela, S. *et al.* Interaction molecular QTL mapping discovers cellular and environmental modifiers of genetic regulatory effects. *The American Journal of Human Genetics* **111**, 133–149 (2024).

39. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet* **51**, 187–195 (2019).

40. Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).

41. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252 (2018).

42. Nott, A. *et al.* Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).

43. Keough, K. C. *et al.* Three-dimensional genome rewiring in loci with human accelerated regions. *Science* **380**, eabm1696 (2023).

44. Policarpo, R., Sierksma, A., De Strooper, B. & d'Ydewalle, C. From Junk to Function: LncRNAs in CNS Health and Disease. *Front. Mol. Neurosci.* **14**, (2021).

45. Yang, S., Lim, K.-H., Kim, S.-H. & Joo, J.-Y. Molecular landscape of long noncoding RNAs in brain disorders. *Mol Psychiatry* **26**, 1060–1074 (2021).

46. Diverse lncRNA mechanisms in brain development and disease. *Current Opinion in Genetics & Development* **65**, 42–46 (2020).

47. An update on the role of miR-124 in the pathogenesis of human disorders. *Biomedicine & Pharmacotherapy* **135**, 111198 (2021).

48. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics* **10**, e1004383 (2014).

49. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481–487 (2016).

50. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).

51. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381–389 (2018).

52. Collado-Torres, L. *et al.* Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia. *Neuron* **103**, 203-216.e8 (2019).

730 53. Hwang, H. *et al.* Neurogranin, Encoded by the Schizophrenia Risk Gene NRGN,
731 Bidirectionally Modulates Synaptic Plasticity via Calmodulin-Dependent Regulation of the
732 Neuronal Phosphoproteome. *Biological Psychiatry* **89**, 256–269 (2021).
733 54. Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders:
734 the emerging picture and its implications. *Nat Rev Genet* **13**, 537–551 (2012).
735 55. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
736 *EMBnet.journal* **17**, 10–12 (2011).
737 56. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
738 57. Graubert, A., Aguet, F., Ravi, A., Ardlie, K. G. & Getz, G. RNA-SeQC 2: efficient RNA-seq
739 quality control and quantification for large cohorts. *Bioinformatics* **37**, 3048–3050 (2021).
740 58. Tran, K. A. *et al.* Performance of tumour microenvironment deconvolution methods in breast
741 cancer using single-cell simulated bulk mixtures. *Nat Commun* **14**, 5758 (2023).
742 59. Qi, T. *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation.
743 *Nat Genet* **54**, 1355–1363 (2022).
744 60. Taylor-Weiner, A. *et al.* Scaling computational genomics to millions of individuals with
745 GPUs. *Genome Biology* **20**, 228 (2019).
746 61. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian Framework to Account for Complex
747 Non-Genetic Factors in Gene Expression Levels Greatly Increases Power in eQTL Studies. *PLOS*
748 *Computational Biology* **6**, e1000770 (2010).
749 62. Davis, J. R. *et al.* An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts
750 for Linkage Disequilibrium between Variants. *The American Journal of Human Genetics* **98**, 216–
751 224 (2016).
752 63. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants
753 from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164 (2010).
754 64. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell
755 transcriptomics. *BMC Genomics* **19**, 477 (2018).
756 65. Leal, S. M. Genetics and Analysis of Quantitative Traits. *The American Journal of Human*
757 *Genetics* **68**, 548–549 (2001).

Figure Legends

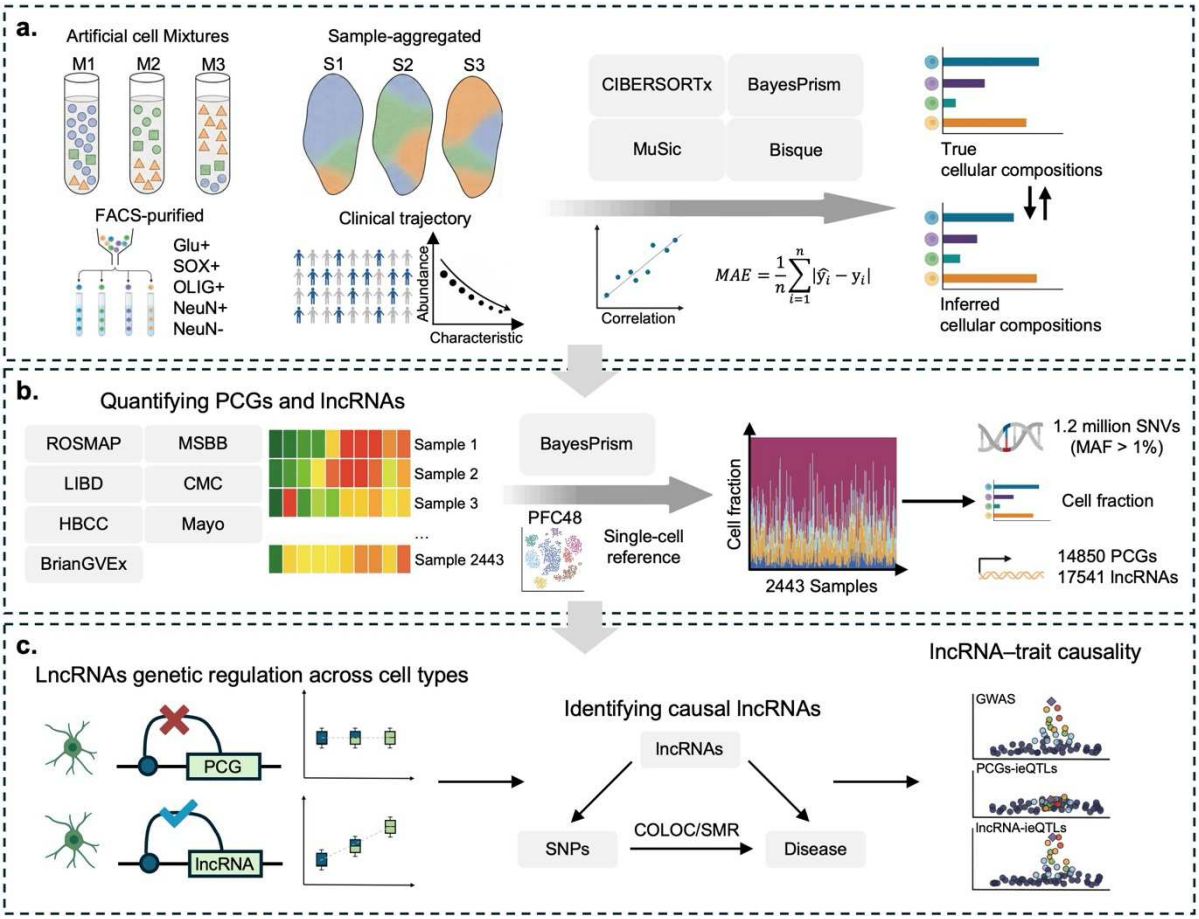


Figure 1. Overview of the study design. a. We established a benchmarking pipeline to evaluate the robustness of four transcriptome-based deconvolution methods (CIBERSORTx, MuSiC, BayesPrism, and Bisque) under diverse conditions. Evaluation was performed using both simulated and real-world data: (1) artificial mixtures generated by recombining single-cell profiles within PFC48[13] for within-dataset benchmarking; (2) pseudo-bulk mixtures constructed by aggregating cells from different donors in the PFC427[33] for cross-dataset evaluation; (3) FACS-purified cortical populations with defined cell identities; and (4) matched bulk RNA-seq and clinical phenotype–linked data from ROSMAP. **b.** We applied BayesPrism to infer cell-type composition across 2,443 DLPFC samples from PsychENCODE (BrainGVEx, CMC, CMC_HBCC and LIBD) and AMP-AD consortia (Mayo, ROSMAP and MSBB BM10). Integrating cell abundance estimates with matched genotypes and transcript quantification of 17,541 lncRNAs, including both GENCODE-annotated and unannotated transcripts, we performed genome-wide mapping of lncRNA-ieQTLs based on an interaction model. **c.** Finally, we integrated lncRNA-ieQTLs with GWAS summary statistics to nominate lncRNAs with putative roles in mediating genetic risk for complex brain traits and disorders.

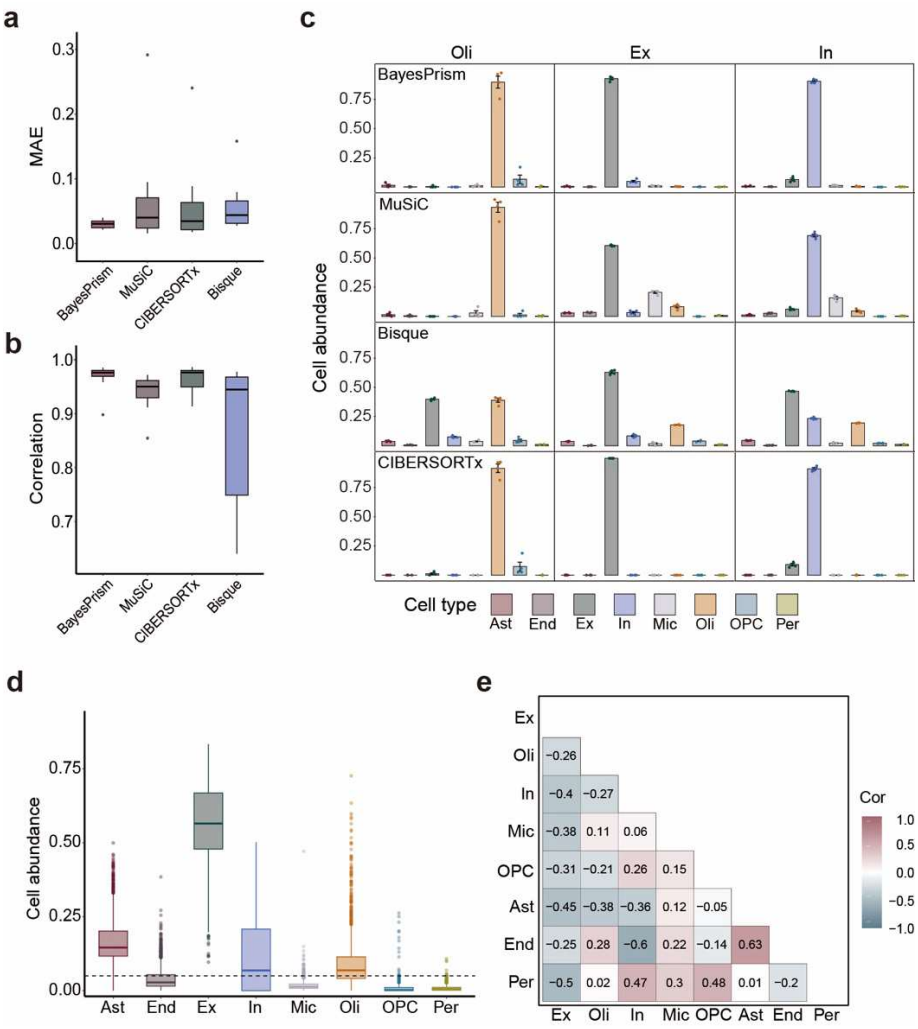


Figure 2. Evaluation of deconvolution performance under diverse benchmarking conditions. a–b. Boxplots showing the Spearman's rho and mean absolute error (MAE) between estimated and true cell type proportions across 5,000 artificial mixtures, comparing four deconvolution methods: BayesPrism, MuSiC, CIBERSORTx, and Bisque. **c.** Deconvolution accuracy in FACS-purified samples with known cell identities (Ex, In, Oli), using PFC48 as reference. Bisque and MuSiC showed limited robustness in filtering reference noise, misattributing fractions to incorrect cell types despite clear ground-truth definitions. **d.** Estimated cell type proportions for 2,443 DLPFC samples from seven cohorts using BayesPrism with PFC48 reference. A threshold of 0.05 (dashed line) was applied to define detectable cell types for downstream ieQTL mapping. **e.** Pairwise Spearman correlation between estimated cell type proportions. Cell type abbreviations: Ast, astrocytes; End, endothelial cells; Ex, excitatory neurons; In, inhibitory neurons; Mic, microglia; Oli, oligodendrocytes; OPC, oligodendrocyte precursor cells; Per, pericytes.

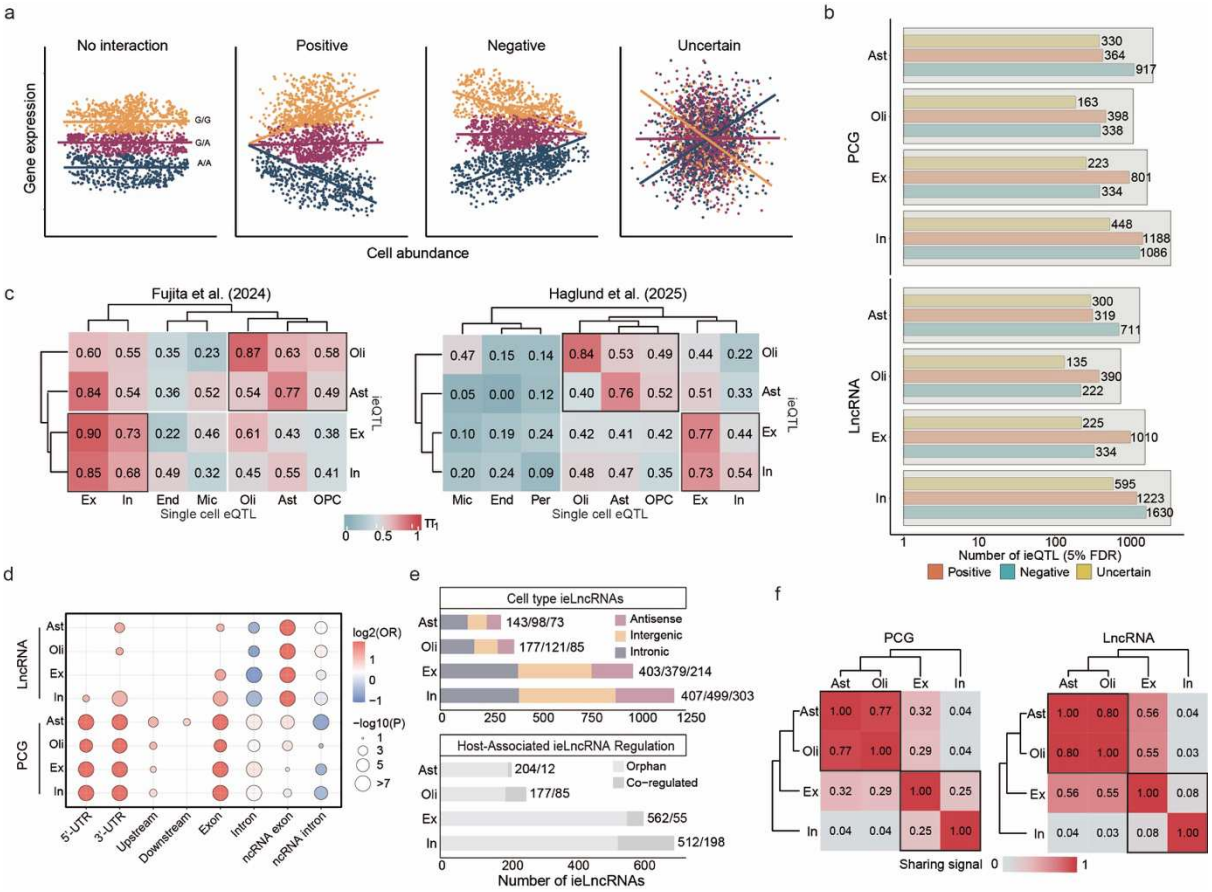


Figure 3. Discovery of lncRNA-ieQTLs. a. Conceptual illustration of ieQTL, including no interaction, positive direction, negative direction, and uncertain direction. **b.** Number of significant ieQTLs (FDR < 0.05) detected across four cell types, stratified by gene class and ieQTL directionality. Bar labels indicate the total count of ieQTLs in each category. **c.** π_1 statistics comparing ieQTL with single-nucleus eQTLs summary statistics from Fujita et al.[15] (left) and Haglund et al.[16] (right). Rows and columns are clustered using Euclidean distance and complete linkage. **d.** Enrichment analyses comparing the distribution of lncRNA-ieQTLs and PCG-ieQTLs across genomic location categories. The circle color represents log_2 odds ratio, and size indicates $-log_{10}\,P$, derived from two-sided Fisher's exact tests. **e.** (Top) Number of ielncRNAs with significant positive-direction ieQTLs across cell types, stratified by lncRNA biotype, including intronic, antisense, and intergenic lncRNAs. (Bottom) Among intronic and antisense lncRNAs with significant positive-direction ieQTLs, the fraction whose host protein-coding genes are also identified as ieGenes in the same cell type. **f.** Pairwise sharing of ieQTL effects between cell types, estimated by mashr[39]. Pairwise sharing is defined as the proportion of significant ieQTLs (LFSR < 0.05) with concordant effect direction and similar magnitude (within a 0.5-fold range) between two cell types.

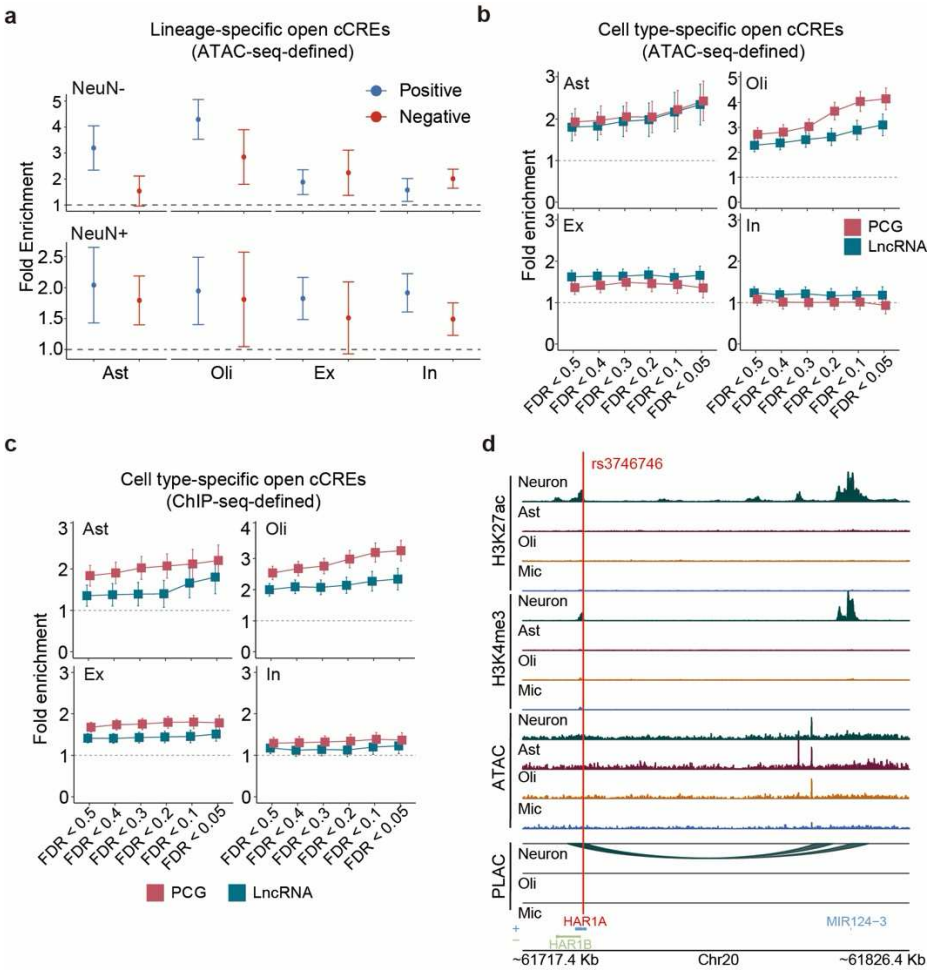


Figure 4. Epigenomic annotation of lncRNA-ieQTLs. a. Enrichment of ieQTLs in non-neuron ($NeuN^-$) and neuron ($NeuN^+$) lineage ATAC-seq peaks, stratified by ieQTL direction. The x-axis indicates the fold enrichment, and the y-axis lists the corresponding cell types. The error bar represents the 95% CI computed using permutations. **b–c.** Enrichment of positive-direction ieQTLs in cell-type–specific regulatory regions. Each box represents the fold enrichment of positive-direction ieQTLs, stratified by FDR thresholds (x-axis), within open chromatin regions (ATAC-seq, b) and active regulatory regions marked by H3K4me3 and H3K27ac (ChIP-seq, c) in the corresponding brain cell type. Colors indicate gene categories; error bars denote 95% confidence intervals derived from permutation testing. **d.** Chromatin landscape surrounding the excitatory neuron-specific ieQTL rs3746746. The x-axis denotes the genomic coordinates on chromosome 20 encompassing HAR1A and nearby loci. From top to bottom: ChIP-seq tracks for H3K27ac (4 cell types), H3K4me3 (4 cell types), and ATAC-seq (4 cell types), followed by PLAC-seq tracks highlighting chromatin contacts in the same cell type.

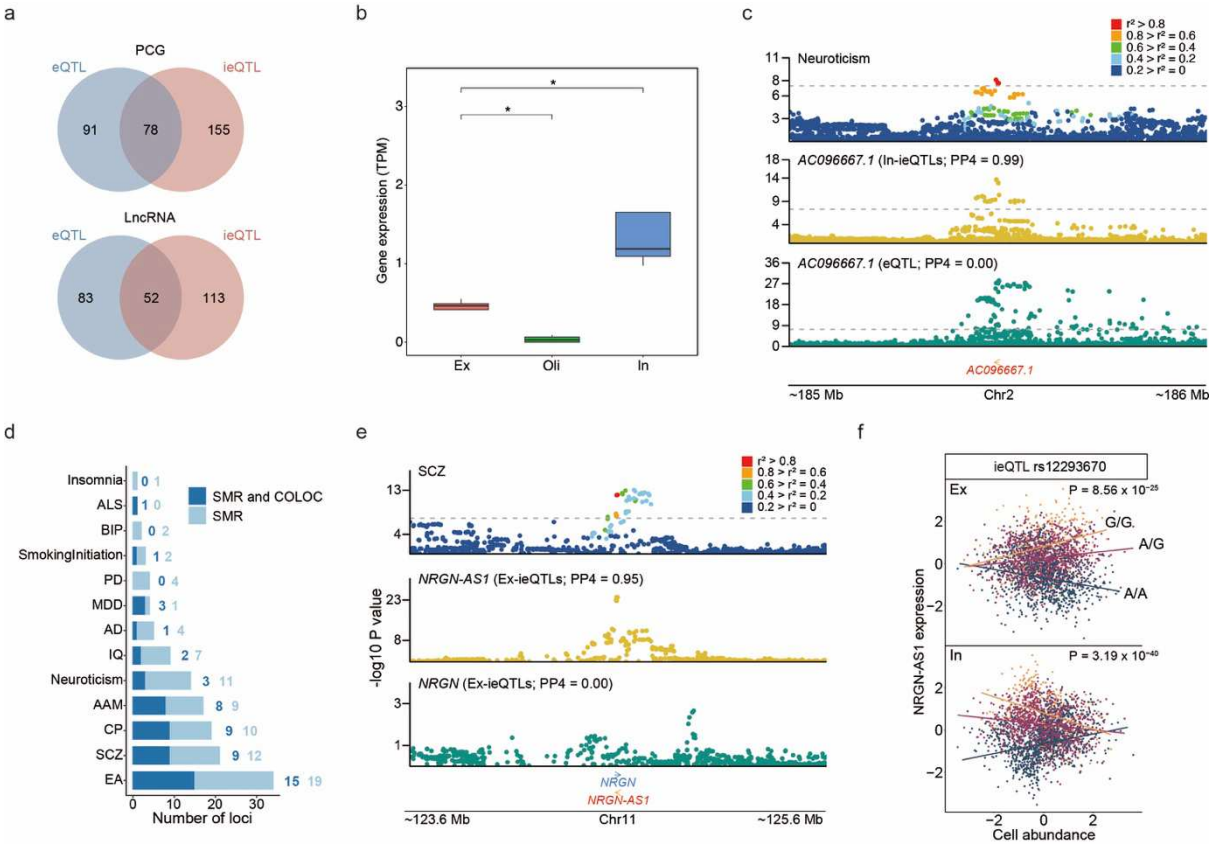


Figure 5. Integrating lncRNA regulation in cellular context refines GWAS functional interpretation. a. Comparison of colocalization supported by ieQTLs versus the eQTLs, showing the additional colocalization signals revealed when incorporating cellular context (PP.H4 > 0.8, ieQTLs FDR < 0.25). **b.** Expression levels of lncRNA *AC096667.1* in specific cell types: excitatory neurons (Ex), oligodendrocytes (Oli), and inhibitory neurons (In) from EPIGABA samples, highlighting its cell type-specific expression. Statistical significance was assessed using a two-sided Wilcoxon rank-sum test. *p < 0.05. **c.** Locus plot showing colocalization of lncRNA *AC096667.1* in inhibitory neurons and at the bulk-tissue level with neuroticism GWAS signal. Association p-values ($-log_{10}$ scale) are shown for inhibitory neuron ieQTLs (middle) and bulk eQTLs (bottom). **d.** Across the 13 traits with significant SMR results, the number of loci prioritized by SMR and the subset with concordant evidence from COLOC in the same cellular context. Bars are stratified by evidence type: light color represents loci uniquely prioritized by SMR, while dark color indicates loci supported by both SMR and COLOC. **e.** Locus plot showing colocalization between the SCZ GWAS locus rs12293670, *NRGN*, and the newly assembled antisense lncRNA *NRGN-AS1* in excitatory neuron ieQTLs. **f.** ieQTLs of *NRGN-AS1* in excitatory (Top) and inhibitory neurons (Bottom). The p value of the interaction effect from the linear model fitted with tensorQTL[60] is shown. Dots are colored on the basis of the genotype of rs12293670.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure.pdf
- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- Supplementarytable3.xlsx
- Supplementarytable4.xlsx
- Supplementarytable5.xlsx