PSYCHENCODE2

Evaluating performance and applications of sample-wise cell deconvolution methods on human brain transcriptomic data

Rujia Dai[1], Tianyao Chu[2], Ming Zhang[2], Xuan Wang[2], Alexandre Jourdon[3], Feinan Wu[3], Jessica Mariani[3], Flora M. Vaccarino[3,4], Donghoon Lee[5], John F. Fullard[5], Gabriel E. Hoffman[5], Panos Roussos[5], Yue Wang[6], Xusheng Wang[7], Dalila Pinto[8], Sidney H. Wang[9], Chunling Zhang[10], PsychENCODE consortium†, Chao Chen[2]*, Chunyu Liu[1,2,10]*

Sample-wise deconvolution methods estimate cell-type proportions and gene expressions in bulk tissue samples, yet their performance and biological applications remain unexplored, particularly in human brain transcriptomic data. Here, nine deconvolution methods were evaluated with sample-matched data from bulk tissue RNA sequencing (RNA-seq), single-cell/nuclei (sc/sn) RNA-seq, and immunohistochemistry. A total of 1,130,767 nuclei per cells from 149 adult postmortem brains and 72 organoid samples were used. The results showed the best performance of dtangle for estimating cell proportions and bMIND for estimating sample-wise cell-type gene expressions. For eight brain cell types, 25,273 cell-type eQTLs were identified with deconvoluted expressions (decon-eQTLs). The results showed that decon-eQTLs explained more schizophrenia GWAS heritability than bulk tissue or single-cell eQTLs did alone. Differential gene expressions associated with Alzheimer's disease, schizophrenia, and brain development were also examined using the deconvoluted data. Our findings, which were replicated in bulk tissue and single-cell data, provided insights into the biological applications of deconvoluted data in multiple brain disorders.

INTRODUCTION

Brain tissue transcriptome is essential for studying brain biology and related disorders, but important cell-type information can be obscured. Several large-scale projects have generated valuable transcriptomic resources from human brains, such as GTEx (*1*), PsychENCODE (*2*), CommonMind (CMC) (*3*), Brainspan (*4*), and ROSMAP (*5*). However, most of the existing transcriptomes are from bulk tissues, which are mixtures of many different cell types. Given that gene expressions and regulatory mechanisms are known to vary across brain cell types, using bulk tissue data to study gene expression changes will obscure the underlying cellular context.

Sorted-cell RNA sequencing (RNA-seq) and single-cell/nuclei (sc/sn) RNA-seq (*6*, *7*) offer solutions for profiling brain transcriptome at the cell-type resolution but with several limitations. Cell sorting relies on marker genes of cell types, which are not always available. The specificity of these marker genes is frequently a concern. Moreover, a combination of several marker genes only can sort limited cell types. Regarding sc/snRNA-seq, considerable gene expressions cannot be detected due to the limited RNA input from each cell (*8*). It is also challenging to discriminate between two possible causes of missing data: biologically true zero expression and technically missing data (*9*). These missing data may result in potential problems in gene expression quantification and the quality of downstream analysis. Another limitation is the high cost of sc/snRNA-seq. Although multiplexing methods have been developed to simultaneously profile cells from numerous samples (*10*), using sc/snRNA-seq in studies requiring hundreds of subjects, such as disease association and expression quantitative trait loci (eQTL) mapping (*11*), can be cost-prohibitive.

Computational algorithms for cell deconvolution have been developed to estimate cell proportions. These algorithms can be classified into two types: Supervised deconvolution uses prior information from cell-type reference data to facilitate the estimation of the cell proportions in bulk tissue samples, while unsupervised deconvolution does not need a reference. This study focused on evaluating supervised deconvolution methods.

The performance of methods for estimating cell proportions has been previously evaluated (*12*–*16*) with data from the brain and other tissues. Methods such as Digital Sorting Algorithm (DSA) (*17*), ordinary least squares regression (OLS), CIBERSORT (*18*), dtangle (*19*), and MuSiC (*20*) showed good performance in these evaluations. In addition, the effects of cell-type marker gene selection, covariates, data transformation and normalization, and cell subtypes on cell deconvolution have been evaluated, which provided guidelines for data processing before cell deconvolution.

Cell-type gene expressions can also be deconvoluted from bulk tissue expression data. Sample-wise deconvolution methods have been developed to estimate cell-type gene expressions for each sample,

[1]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA. [2]MOE Key Laboratory of Rare Pediatric Diseases & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, and Department of Psychiatry, The Second Xiangya Hospital, Central South University, Changsha, China. [3]Child Study Center, Yale University, New Haven, CT, USA. [4]Department of Neuroscience, Yale University, New Haven, CT, USA. [5]Center for Disease Neurogenomics, Departments of Psychiatry and Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [6]Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. [7]Department of Biology, University of North Dakota, Grand Forks, ND, USA. [8]Departments of Psychiatry and Genetics and Genomic Sciences, Mindich Child Health and Development Institute, and Icahn Genomics Institute for Data Science and Genomic Technology, Seaver Autism Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [9]Center for Human Genetics, The Brown foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA. [10]Department of Neuroscience & Physiology, SUNY Upstate Medical University, Syracuse, NY, USA.
*Corresponding author. Email: liuch@upstate.edu (C.L.); chenchaor@gmail.com (C.C.)
†PsychENCODE consortium authors, collaborators, and affiliations are listed in the Supplementary Materials.

such as bMIND (*21*), swCAM (*22*), and Tensor Composition Analysis (TCA) (*23*). All these methods are supervised and use expression references from sc/snRNA-seq or sorted-cell expressions. The cell-type expressions of the individual samples enable eQTL mapping and differential expression analysis in a cell type–specific way. The deconvoluted data can cover most of the genes detected in bulk tissue, which is less sparse than sc/snRNA-seq data. Cell deconvolution makes the large-scale study of cell-type expressions affordable since bulk tissue data are either ready to use or can be generated at a relatively low cost.

The methods for estimating sample-wise cell-type expressions have been partially evaluated, although there are major blind spots. The performance of bMIND, TCA, and swCAM has been evaluated in their original methodology papers; however, these studies used artificially constructed pseudo-bulk data rather than bulk tissue data to benchmark the different methods' performance. Pseudo-bulks were constructed by simulating cell proportions and multiplying simulated proportions with expressions from sc/snRNA-seq or sorted-cell expression data; therefore, pseudo-bulk data are less complex than the data from real bulk tissue (*24*). The cell-type differences in the pseudo-bulk are easier to be captured than those in bulk tissue data. Thus, the benchmarking conclusion based on the pseudo-bulk data may not be applicable to data from brain tissues. On the other hand, head-to-head comparisons of all these methods on brain tissue data have not been conducted to date. The downstream applications based on deconvoluted data, such as eQTL mapping and differential expression, have also not been evaluated to showcase the validity of deconvolution.

The current study aimed at evaluating the performance of algorithms for sample-wise deconvoluting cell proportions and cell-type expressions, as well as research applications based on the deconvoluted data. Specifically, we evaluated six commonly used deconvolution methods for estimating cell proportions and three deconvolution methods for estimating the cell-type expressions of individual samples. Data from bulk tissue RNA-seq, sc/snRNA-seq, and immunohistochemistry (IHC) of matched adult postmortem brains and brain organoids were used for evaluation. Downstream analyses of the deconvoluted results were also conducted, including their use in eQTL mapping, schizophrenia (SCZ) genome-wide association study (GWAS) heritability enrichment, the differential expression for Alzheimer's disease (AD), SCZ, and brain development in cell types. On the basis of the evaluation, we recommended the best practice for cell deconvolution of brain transcriptome.

RESULTS

Benchmarking of sample-wise deconvolution methods with brain transcriptome data

To evaluate commonly used deconvolution methods, we selected six methods [DSA, OLS, CIBERSORT, dtangle, MuSiC, and Bisque; (*25*)] for estimating cell proportions and three methods (bMIND, swCAM, and TCA) for estimating cell-type expressions (Fig. 1). Bulk tissue RNA-seq, snRNA-seq, and IHC data from ROSMAP were used as primary data for evaluation (*26*). Data from adult brains in CMC (*27*) and brain organoids (*28*) were used for confirmation (Table 1). Cell proportions from IHC and sc/snRNA-seq data were used as ground truth for evaluating the accuracy of estimated cell proportions. Gene expressions in sc/snRNA-seq data were used as ground truth for evaluating the accuracy of estimated cell-type expressions. The root mean square error (RMSE) and Spearman's correlation coefficient were used as evaluation metrics. After method evaluation, eQTL mapping, GWAS heritability enrichment, and differential expression analysis were performed on the cell-type expressions estimated by the best-performing method. To further evaluate the quality of deconvolution outputs by actual applications, the cell-type eQTLs, explained GWAS heritability, and phenotype-associated genes (PAGs) derived from deconvoluted expressions were compared to corresponding results based on sc/snRNA-seq and bulk tissue data.

Evaluation of cell proportions estimated by deconvolution methods

The overall performance of six deconvolution methods (DSA, OLS, CIBERSORT, dtangle, MuSiC, and Bisque) for estimating cell proportions was evaluated with ground truth from matched samples. To ensure the deconvolution performance, the intersection of marker genes identified at the single-cell level and the pseudo-bulk level was used to guide deconvolution (see details in Materials and Methods). When using ROSMAP IHC data as the reference (with a sample size of 49), both dtangle and OLS demonstrated lower RMSE values compared to the other methods, as depicted in Fig. 2A. In the case of ROSMAP, CMC (with a sample size of 94), and organoid data (with a sample size of 55) using cell proportions computed from sc/snRNA-seq data as the ground truth, Bisque exhibited superior performance compared to the other methods, followed by dtangle. MuSiC did not perform well in these three datasets despite being specifically designed to utilize sc/snRNA-seq data as a reference. When averaging the RMSE values across all datasets, both dtangle and Bisque achieved an RMSE of 0.11, establishing them as the two best-performing methods. The accuracy of deconvoluted cell proportions in major cell types was better than the accuracy in rare cell types (Fig. 2B and fig. S1). The RMSE increased sharply when the cell proportion was below 5%, such as in oligodendrocyte precursor cells (Opcs), microglia, endothelial cells, and pericytes in adult brains. Similar results were observed using Spearman's correlation as an evaluation metric (fig. S2).

Evaluation of cell-type expressions estimated by sample-wise deconvolution methods

The accuracy of sample-wise cell-type expressions deconvoluted by bMIND, swCAM, and TCA was evaluated using ground truth generated from sc/snRNA-seq expressions of matched samples ($n = 35$ for ROSMAP, $n = 94$ for CMC, and $n = 55$ for brain organoids). Cell proportions estimated by Bisque and dtangle were selected as input for the three methods because they showed the best performance in the above evaluation for estimating cell proportions. bMIND displayed the best performance for estimating cell-type expressions in all datasets, followed by swCAM (Fig. 3A). For bMIND, the averaged correlation coefficient between estimated expressions and sc/snRNA-seq data was 0.62 in ROSMAP data, 0.75 in CMC data, and 0.85 in brain organoid data. We did not observe a substantial difference in performance for estimating expressions of major and rare cell types (Fig. 3B). However, bMIND performed more steadily and better overall in major cell types than in rare cell types. The expression profiles of the deconvoluted cell types exhibited a greater resemblance to their corresponding cell types in sc/snRNA-seq data (e.g., deconvoluted neurons to neurons detected in sc/snRNA-seq data), while they showed less similarity to other cell types [e.g., deconvoluted

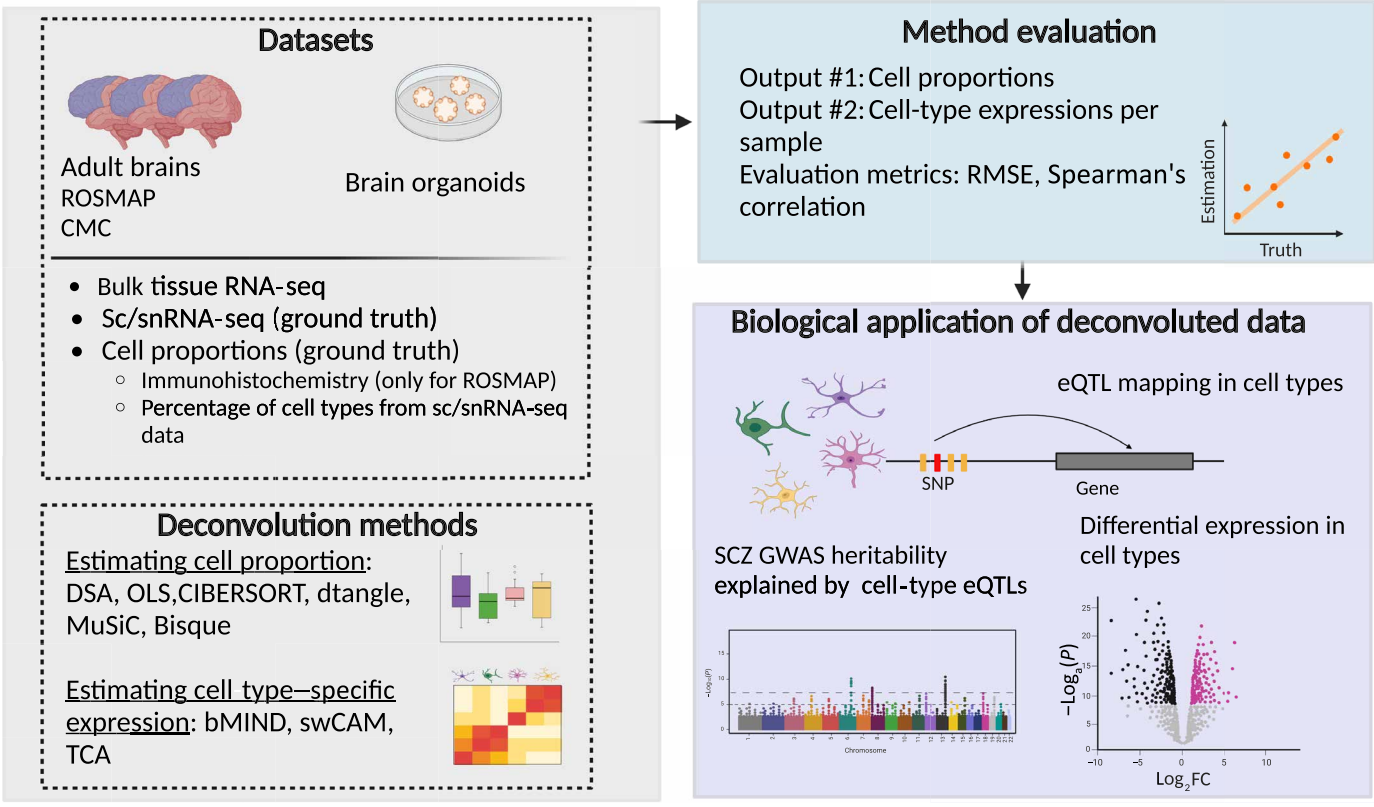

Fig. 1. Study overview. This figure was created with BioRender software.

Table 1. Datasets used for evaluation. PFC, prefrontal cortex.

Study	Brain region	Data type	Sample size	Number of cells	Number of cell types	Number of genes
ROSMAP	PFC	Bulk tissue RNA-seq	1112	–	–	17,128
	PFC	snRNA-seq	48	69,611	8	17,926
	PFC	IHC	49	–	5	–
CMC	PFC	Bulk tissue RNA-seq	572	–	–	25,774
	PFC	snRNA-seq	101	569,289	7	33,822
Brain organoid	–	Bulk tissue RNA-seq	130	–	–	20,125
	–	scRNA-seq	72	490,844		33,538

neurons to glial types in sc/snRNA-seq data (fig. S3)]. A number of well-known marker genes were highly expressed in corresponding cell types, thus indicating that the deconvoluted data have good cell-type specificity (Fig. 3C).

Cell-type eQTL mapping with deconvoluted sample-wise expression data

To identify single-nucleotide polymorphisms (SNPs) that cis-regulate gene expression in specific cell types, cell-type eQTL mapping was performed for the association between genotypes and deconvoluted gene expressions of individual samples. The cell-type eQTLs identified with deconvoluted gene expressions were named deconvolution eQTLs (decon-eQTLs). RNA-seq data of 1112 bulk tissue samples in ROSMAP collection were deconvoluted. Of the 1112 samples, 861 samples had genotype data and were used for decon-eQTL mapping. Cell proportions and cell-type expressions were estimated with dtangle and bMIND, respectively. The effect of SNPs within a 1-megabase window around the transcription start site (TSS) of genes was tested. The number of input genes for decon-eQTL mapping ranged from 8521 to 12,418, across eight cell types. The number of input SNPs was 4,954,561 for all cell types. Effects of known and hidden covariates on deconvoluted gene expressions were corrected. A total of 3,185,333 unique decon-eQTLs were detected across eight cell types at a genome-wide significant level [false discovery rate (FDR) < 0.05].

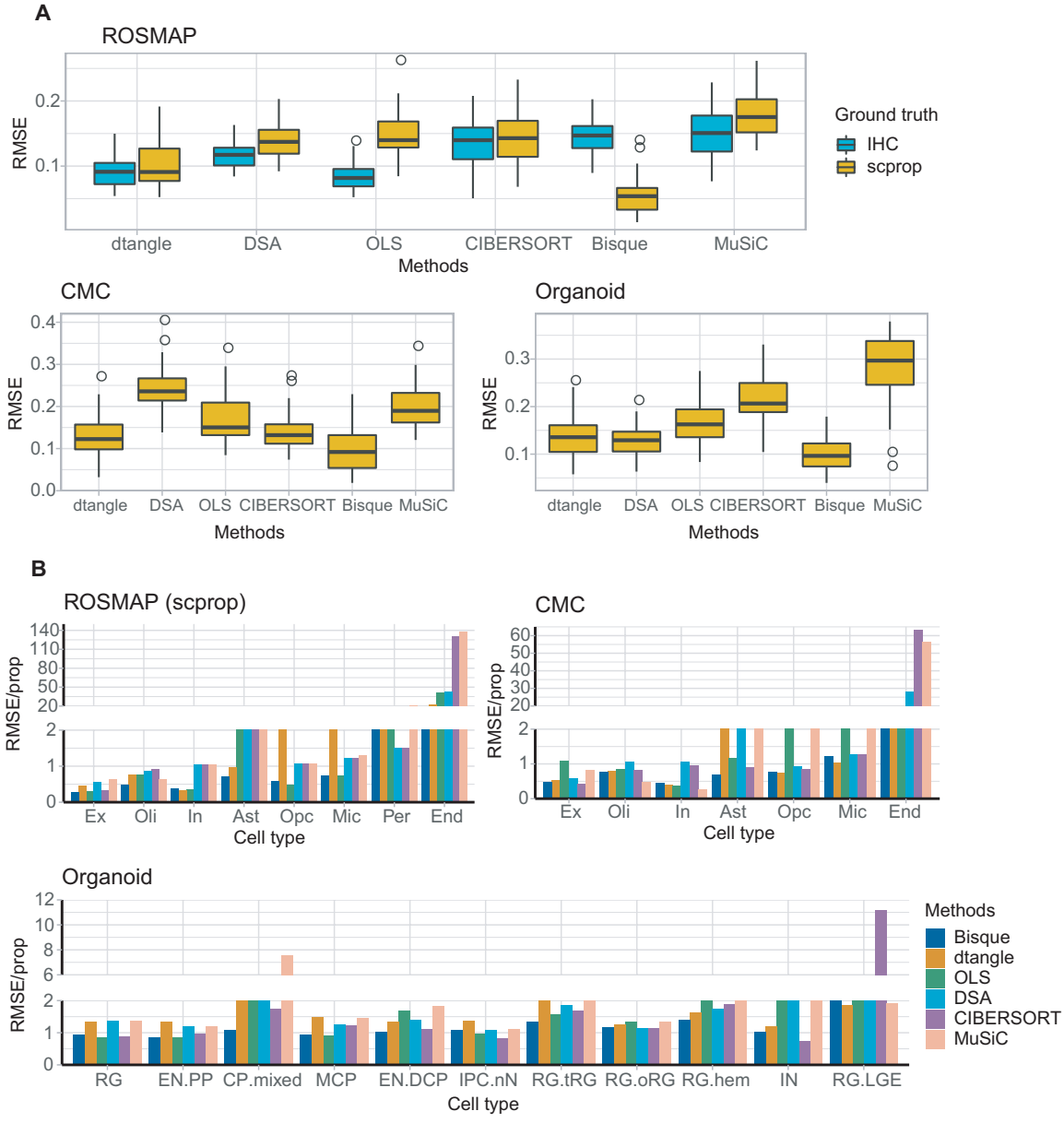

Fig. 2. Assessment of cell proportions estimated by examined deconvolution methods. (**A**) Sample-level RMSE values between estimated cell proportions and ground truth. IHC, immunohistochemistry; scprop, cell proportions calculated from sc/snRNA-seq data; scprop, the number of cells of specific cell type/number of total cells. (**B**) Cell-type level RMSE values between estimated cell proportions and ground truth data. RMSE values were normalized by the value of cell proportions to make them comparable across cell types. Cell types were ordered by cell proportions in a decreasing way. Ex, excitatory neurons; In, inhibitory neurons; Ast, astrocytes; Opc, oligodendrocyte precursor cells; Mic, microglia; Per, pericytes; End, endothelial cells; RG, radial glia; EN.PP, early born excitatory neurons of the preplate/subplate; CP.mixed, cortical plate mixed neurons; MCP, medial cortical plate; EN-DCP, dorsal cortical plate excitatory neurons; IPC-nN, intermediate progenitor cell or newborn neuron; RG.tRG, truncated radial glia; RG.oRG, outer radial glia; RG.hem, radial glia in cortical hem; IN, inhibitory neurons; RG-LGE, progenitors corresponding to a putative ventrolateral ganglionic eminence fate.

The number of decon-eQTLs detected in each cell type ranged between 1,088,634 and 2,245,945. To identify the strongest eQTL in SNPs, a permutation test was performed for each gene. A total of 25,273 (4541 to 8149) independent decon-eQTLs were identified at FDR < 0.05 for eight cell types (Fig. 4B). As expected, eQTL SNPs (eSNPs) were enriched around the TSS region of eQTL genes (eGenes) (fig. S4). The numbers of detected decon-eQTLs were positively correlated with the proportions of cell types in the tissue (Fig. 4B). To test the robustness of identified decon-eQTLs, sample IDs were randomly shuffled before the eQTL mapping. The absence of significant eQTL in the shuffled data supported the notion that the identified decon-eQTLs were not due to random noise (fig. S5).

Identified decon-eQTLs from ROSMAP data were replicated with deconvoluted data from BrainGVEX (*29*). The same deconvolution and eQTL mapping procedures were performed on RNA-seq data of 400 postmortem brain samples from BrainGVEX to obtain

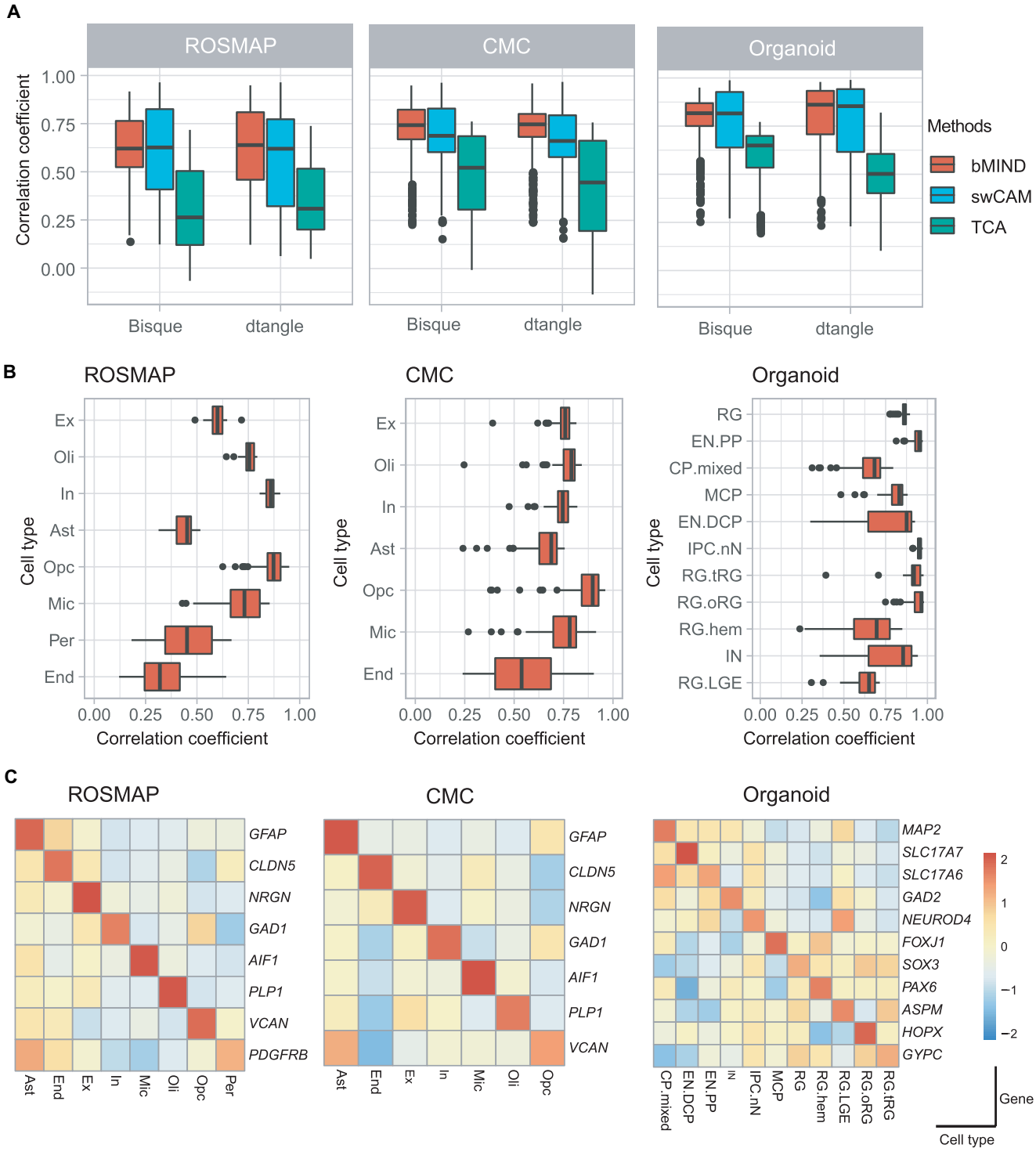

Fig. 3. Assessment of sample-wise cell-type expressions deconvoluted from bulk tissue data. (**A**) Overall assessment of methods for estimating cell-type expressions. Spearman's correlations between deconvoluted data and sc/snRNA-seq data from matched samples. The averaged expression by cell types was used as ground truth. (**B**) Cell-type level assessment of methods for estimating cell-type expressions. Correlations between deconvoluted data by bMIND and sc/snRNA-seq data were calculated for each cell type. Cell proportions estimated by dtangle were used for input. Cell types on the *y* axis were ordered by cell proportions computed from sc/snRNA-seq. (**C**) Assessment of cell-type specificity in estimated expressions. The figure shows the expression of marker genes in deconvoluted data by bMIND.

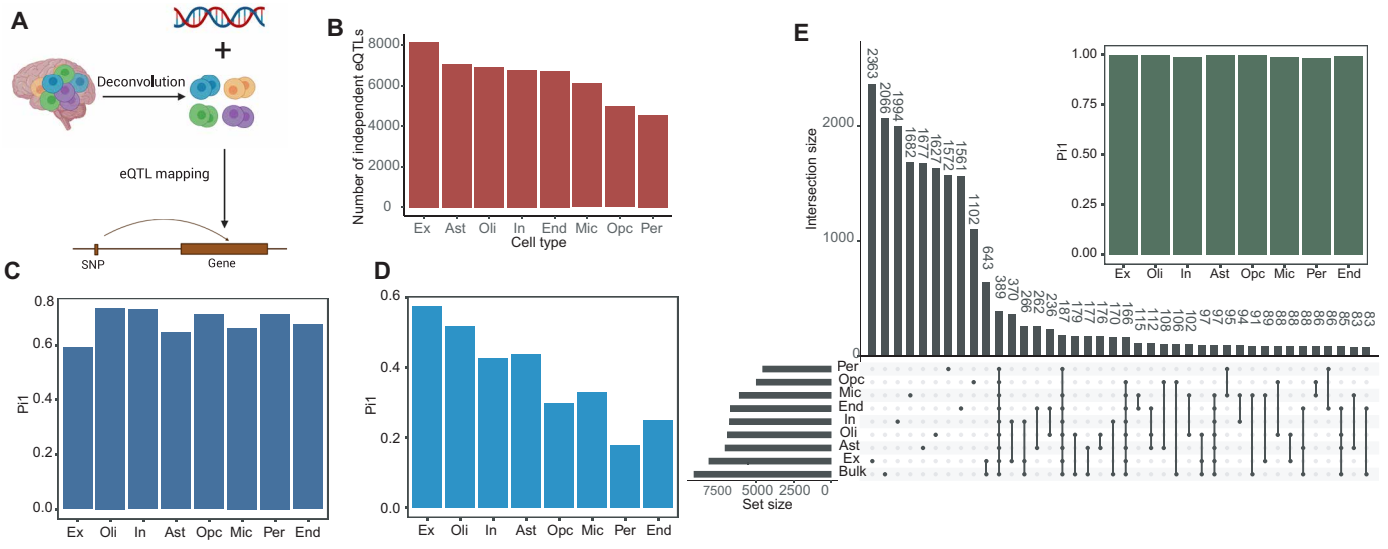

Fig. 4. Cell-type eQTL mapping based on deconvoluted sample-wise expression data. (**A**) Illustration of decon-eQTL mapping. (**B**) Number of decon-eQTLs identified in different cell types at FDR < 0.05 in the permutation test. (**C**) *Pi*1 statistics of decon-eQTLs in BrainGVEX decon-eQTLs and (**D**) eQTLs from snRNA-seq study of Bryois *et al.* (*31*). For calculating *Pi*1 statistics, decon-eQTLs from ROSMAP were used as testing data, and eQTLs from BrainGVEX and Bryois *et al.* (*31*) were used as references. (**E**) Comparison of decon-eQTLs and bulk tissue eQTLs. The top barplot shows the *Pi*1 values of decon-eQTLs (testing data) in bulk tissue eQTLs (reference). The bottom plot shows the intersections between decon-eQTLs and bulk tissue eQTLs, as well as intersections of decon-eQTLs across various cell types.

the decon-eQTLs. Across eight cell types, 3479 to 5718 independent eQTLs were identified in the deconvoluted data from BrainGVEX at FDR < 0.05. To measure the replication rate of ROSMAP decon-eQTLs in BrainGVEX data, the *Pi*1 statistic (*30*), which is the proportion of true eQTL associations in the replication data, was calculated. The *Pi*1 of ROSMAP decon-eQTLs in BrainGVEX data was 0.59 to 0.74 for the matched cell types (Fig. 4C). Decon-eQTLs of oligodendrocyte had relatively better replication than other cell types.

Cell-type eQTLs from snRNA-seq data in Bryois *et al.* (*31*) were also used to replicate our decon-eQTLs. This replication study performed genotyping and snRNA-seq on 192 cortical samples. A total of 7607 independent eQTLs across eight cell types were identified. Although the replication data had less statistical power than our deconvoluted data, 17 to 57% of decon-eQTLs were replicated (Fig. 4D). Excitatory neurons (*Pi*1 = 0.57) had higher *Pi*1 values than other cell types (averaged *Pi*1 = 0.38).

To illustrate the value of decon-eQTLs, we compared decon-eQTLs to bulk tissue eQTLs from ROSMAP (Fig. 4E). Overall, decon-eQTLs had good replication in bulk tissue data, with *Pi*1 > 0.95. The eQTLs that were significant at the cell-type level but insignificant at the bulk tissue level were defined as cell type–specific eQTLs. A total of 1206 to 3006 (24.3 to 36.89%) cell type–specific eQTLs were identified in the deconvoluted data. Cell type–specific eQTLs had *Pi*1 values of 0.17 to 0.52 in single-cell eQTLs, which were similar to the *Pi*1 values of decon-eQTLs that were shared with bulk tissue eQTLs (fig. S6). This demonstrated that a good proportion of eQTLs regulate gene expressions in a cell type–specific way, and they can be detected by decon-eQTLs.

Cell-type eQTLs enriched for the risk heritability in SCZ GWAS data

To test whether cell-type eQTLs are enriched for genetic risk heritability of SCZ, stratified linkage disequilibrium score regression (sLDSC) (*32*) was used to calculate the heritability of SCZ GWAS mediated by decon-eQTLs. Single-cell eQTLs and bulk tissue eQTLs were also included for comparison. Decon-eQTLs explained more SCZ GWAS heritability (averaged $h2 = 37\%$) than single-cell eQTLs (averaged $h2 = 6\%$) for all cell types (Fig. 5A). Bulk tissue eQTLs explained 49% of SCZ GWAS heritability. Integrating decon-eQTLs and bulk tissue eQTLs increased the explained heritability to 63%, whereas the integration of single-cell eQTLs only resulted in an increase of heritability to 53%. The total proportion of explained heritability was correlated with the proportions of cell type in the tissue. To control the effect of SNP numbers, heritability was normalized by the number of decon-eQTLs, which was called GWAS heritability enrichment. Decon-eQTLs of all cell types were enriched for SCZ GWAS heritability (*P* value < 0.05; Fig. 5B). Decon-eQTLs of oligodendrocytes showed the strongest per-SNP enrichment across all cell types. The SCZ GWAS heritability was only significantly enriched in single-cell eQTLs from oligodendrocytes and excitatory neurons. Decon-eQTLs of most of the cell types showed higher enrichment of SCZ GWAS heritability than bulk tissue eQTLs (Fig. 5B), indicating that some of the SCZ risk SNPs may affect gene expression in cell type–specific way. Deconvolution analyses uncovered more of these cell type–specific regulations associated with the genetic risk of SCZ than single-cell or bulk tissue eQTL.

Identification of gene expression changes associated with brain diseases and brain development within cell types

To identify genes associated with various phenotypes in specific cell types, differential gene expression analysis was conducted using the deconvoluted gene expressions. Associations with AD, SCZ, and brain development modeled by organoids were tested in three deconvoluted datasets independently. More samples were included in AD ($n_{AD} = 743$, $n_{control} = 367$) and SCZ ($n_{SCZ} = 246$, $n_{control} = 279$) data. For AD and SCZ, the Wilcoxon signed-rank test was performed on the deconvoluted data. For brain development, the linear regression model was used to test the correlation between

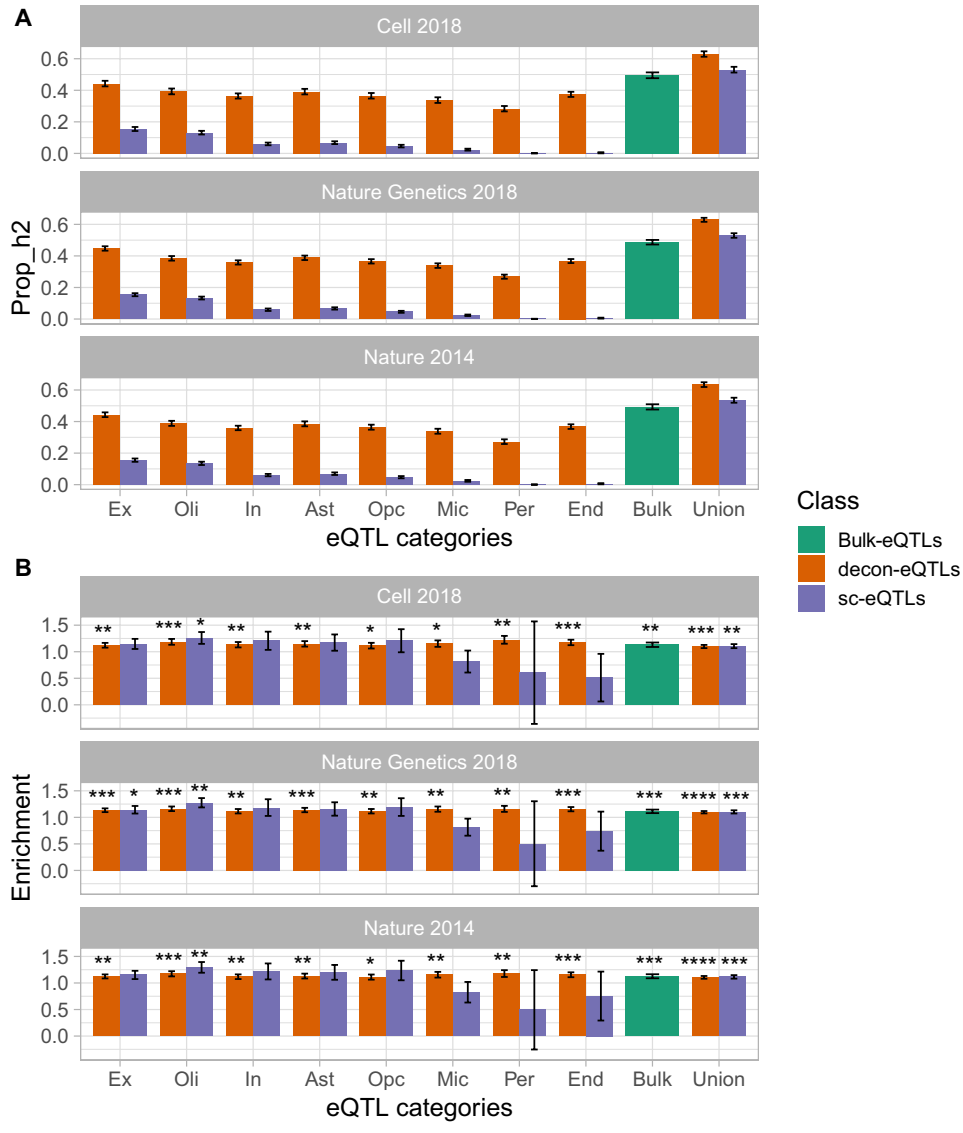

Fig. 5. SCZ GWAS heritability explained by cell-type eQTLs and bulk tissue eQTLs. (**A**) Total SCZ GWAS heritability (*h2*) explained by eQTLs. (**B**) SCZ GWAS heritability enrichment in eQTLs. Enrichment = *h2*/number of SNPs in each eQTL category.

deconvoluted data and culture days of organoids (n_{day0} = 15, n_{day30} = 22, and n_{day60} = 18). With a threshold of FDR < 0.05, 4419, 10,964, and 9562 PAGs were identified for AD, SCZ, and brain development, respectively.

To test the reliability of PAGs identified from deconvoluted data, we conducted a comparison between the PAGs across all deconvoluted cell types (the union set, gene set A) and the PAGs identified in bulk tissue data (gene set B). To determine the replication rate, we calculated the proportion of the overlaps between gene set A and B in gene set A (Fig. 6). In total, 81, 49, and 89% of PAGs for AD, SCZ, and brain development, respectively, were replicated in bulk tissue data. Among these PAGs, most of them (>95%) had the same direction of expression changes in bulk tissue data. For AD and SCZ, less than 15% of PAGs overlapped with PAGs from snRNA-seq data. However, 35% of development-related PAGs could be replicated in scRNA-seq data. The possible explanation for the difference in replication rate in the three datasets is that the expression changes associated with brain development were larger than the changes associated with AD or SCZ (fig. S7). The low replication rate with sc/snRNA-seq data suggested that sc/snRNA-seq data were underpowered to detect PAGs of small effect size.

DISCUSSION

Using bulk tissue RNA-seq, IHC, and sc/snRNA-seq data from matched samples, the performance of six methods for estimating cell proportions and three methods for estimating sample-wise cell-type gene expressions was systematically evaluated. The transcriptome data used for evaluation were from adult brains and cultured brain organoids, providing data representative of different states of cell maturity and developmental processes. In addition, the results of eQTL mapping, SCZ GWAS heritability enrichment, and differential

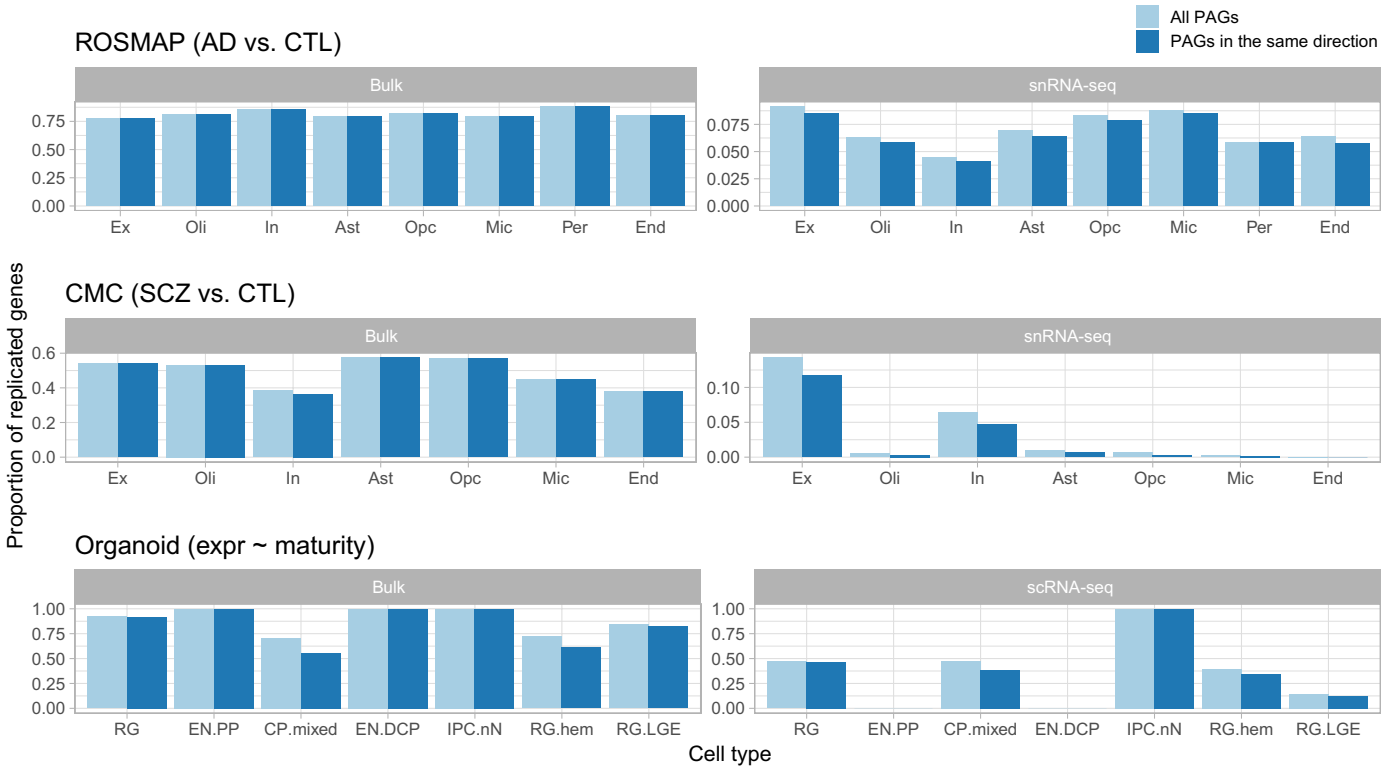

Fig. 6. Replication of PAGs. The proportions of PAGs replicated in bulk tissue data **(left)** and sc/snRNA-seq data **(right)** were displayed. The light blue bar shows the proportions of replicated PAGs with FDR < 0.05. The dark blue bar shows the proportions of replicated PAGs with FDR < 0.05 and having the same direction of changes in replication data. expr denotes expression. The maturity of organoids was measured by the days of cell culture.

expression analysis based on deconvoluted data demonstrate the utility of deconvolution.

dtangle had better accuracy for estimating cell type proportions than other methods. Previous studies have benchmarked the performance of deconvolution methods for estimating cell proportions with ground truth data created from simulated cell proportions. In those studies, dtangle showed good performance in (*15*) but poor performance in (*12*). One possible reason is the difference in pseudo-bulk simulation between studies. Sutton *et al.* (*15*) constructed pseudo-bulks from 500 cells, while Avila Cobos *et al.* (*12*) used only 100 cells for each pseudo-bulk. Given that sc/snRNA-seq data are remarkably sparse, 100 cells may not be representative of cell composition in bulk tissue. This inconsistency suggests the necessity of using real ground truth data in benchmarking studies. By using bulk tissue RNA-seq and IHC data from matched samples, dtangle was found to be the best deconvolution method for estimating cell proportions in this study. The excellent performance of dtangle was preserved in the data using cell proportions from IHC and sc/snRNA-seq data as ground truth.

Deconvolution methods using sc/snRNA-seq data as reference, such as Bisque and MuSiC, did not outperform old methods when using cell proportions from IHC data as ground truth. Given that Bisque learns prior information from the reference of sc/snRNA-seq data, it is expected to see that Bisque showed perfect performance when using cell proportions from sc/snRNA-seq data as ground truth. However, the cell proportions measured by single-cell technologies can be easily biased by the sorting strategy (*33*, *34*). The proportions from single-cell data as prior reference and ground truth should be used with caution.

Head-to-head evaluation of sample-wise deconvolution methods for estimating cell-type expressions was conducted here. bMIND was the best method for estimating cell-type expressions in our evaluation because the correlation coefficients between estimated expressions by bMIND and ground truth were higher than in other methods. Moreover, our evaluation showed that the deconvoluted data by bMIND have good cell-type specificity. The deconvoluted expressions by bMIND had a high correlation with matched cell types but a low correlation with other cell types in the ground truth data (fig. S3). The deconvoluted cell types by bMIND also expressed well-known marker genes. For example, *NRGN* and *GAD1* were highly expressed in excitatory and inhibitory neurons, respectively, but they were poorly expressed in glial cell types. These results indicated that bMIND is the best method for generating cell type–specific gene expressions for each sample directly from bulk tissue data.

The deconvolution performance on rare cell types is in general poor and is thus a challenge for cell deconvolution. The accuracy of deconvoluted cell proportions decreased sharply when cell proportion was less than 5%. Similarly, the accuracy of estimated gene expressions was low for rare cell types in brains, such as endothelial cells (correlation coefficient between deconvoluted expressions and ground truth = 0.33) and pericytes (correlation coefficient = 0.43). Given the fact that the rare cell types may be masked by major cell types in traditional sc/snRNA-seq due to low abundance, rare cell types may need to be studied using RNA-seq of sorted cells, high

coverage sc/snRNA-seq, or with techniques that enrich for rare cell types.

The most important benefit of cell deconvolution was that more cell-type eQTLs were identified using deconvoluted data than using single-cell data and bulk tissue data. Cell-type eQTLs have been generated with sc/snRNA-seq data (*35*, *36*); however, the number of individuals profiled was typically limited (*11*). To date, a total of 7607 eQTLs have been identified in the largest single-cell eQTL study (*31*) of 192 human brain samples. Besides the small sample size, the quality of single-cell data can be potentially affected by poor expression quantification, with serious missing data issues and high technical variability (*9*). Consequently, the eQTLs identified from sc/snRNA-seq data may be affected. In contrast, this study deconvoluted data from 861 human brains and mapped 25,273 decon-eQTLs, far more than eQTLs identified from single-cell studies. The union for decon-eQTLs of all cell types was even more than bulk tissue eQTLs (9148). A total of 24.3 to 36.89% of top decon-eQTLs were not detected by bulk eQTL mapping, indicating that many cell type–specific eQTLs are buried in bulk tissue data since the expression of diverse cell types is mixed. Overall, decon-eQTLs could be replicated in bulk tissue eQTLs (averaged $Pi1 = 0.99$) and single-cell eQTLs (averaged $Pi1 = 0.38$), indicating the reliability of eQTLs identified in deconvoluted data. Sample-wise deconvolution provides a valuable opportunity to study genetic regulations in specific cell types with comparable power to bulk tissue eQTL studies.

Decon-eQTLs explained SCZ GWAS heritability that was missed by single-cell and bulk tissue eQTLs. Nearly six times more SCZ GWAS heritability was explained by decon-eQTLs than that by single-cell eQTLs. Integrating decon-eQTLs and bulk tissue eQTLs explained 63% of SCZ GWAS heritability, which was 14% more than heritability explained only by bulk tissue eQTLs. These results suggest that SCZ GWAS risk may be mediated by genetic regulations in specific cell types, and such an effect can be captured by deconvoluted data.

Decon-eQTL mapping can unveil genes that are associated with the genetic risk of SCZ. A critical application of decon-eQTLs involves identifying genes and associated pathways that link with risk loci, as highlighted by GWAS, within specific cell types. Illustrating this, our study identified a significant association between *rs12466331* and *CALM2* in excitatory neurons—an association not found in bulk tissue data (fig. S8A). *CALM2*, a gene encoding calmodulin, is highly expressed in excitatory neurons (fig. S8B) and has been found downregulated in the postmortem brains of patients with SCZ (*37*). Moreover, *rs12466331* was colocalized with SCZ GWAS risk locus *rs144040771* (fig. S8C). These data suggest that *rs12466331* may regulate the expression of *CALM2* in excitatory neurons and that dysregulation of such a pathway may be associated with SCZ. This integration of decon-eQTLs and GWAS data can be applied to multiple complex diseases. Thus, mapping decon-eQTLs enabled molecular mechanisms of genetic risk of brain disorders in specific cell types.

Cell-type eQTLs mapping with deconvoluted data is an advanced alternative for cell-type interaction eQTLs (ieQTLs) mapping. Cell proportions have been used to map eQTLs associated with cell types, which are called ieQTLs (*27*). The genetic regulators that were associated with gene expressions when the cell proportion varied were mapped. Therefore, ieQTLs are the results of interaction between genetic regulation and cell-type enrichment, while decon-eQTLs are based on deconvoluted gene expressions, which are the direct relationship between genotypes and cell-type expressions for each SNP-gene pair. Both ieQTLs and decon-eQTLs were mapped with ROSMAP data in the current study. Nearly 20 times more decon-eQTLs ($n = 27{,}339$) were identified than ieQTLs (1822) for the same sample size. Moreover, decon-eQTL is more robust than ieQTL. Compared to single-cell eQTLs, the replication rate of decon-eQTLs (averaged $Pi1 = 0.38$) is superior to ieQTLs (averaged $Pi1 = 0.16$; fig. S9).

This study offers a practical guideline for conducting brain cell deconvolution. Using dtangle to estimate cell proportions and bMIND to estimate cell-type gene expressions is recommended. Rare cell types (proportion < 5%) are not recommended to be included in cell deconvolution analysis.

This study has several limitations. The results were based on the analysis of human brain data with specific parameters tested. More tests may be needed to generalize the conclusion to other tissues and situations. In addition, unsupervised deconvolution methods have not been evaluated by this study. This evaluation only focused on the major cell types in brains, and the deconvolution performance of cell subtypes could be further explored to validate our findings.

This study comprehensively evaluated the commonly used methods for sample-wise deconvolution of cell proportions and cell-type gene expressions. The downstream analysis of eQTL mapping, GWAS heritability enrichment, and differential expression was also evaluated. Our analysis is a crucial methodological foundation for other studies where deconvolution can be used. A practical guideline is offered for a broad community interested in cell type–specific studies of brain functions and disorders when only bulk tissue transcriptome is available.

MATERIALS AND METHODS

Data downloading and processing

Three RNA-seq data from brain tissues and brain organoids were used (Table 1). The randomization and blinding of sample collection and ethic approval were described in original studies.

Bulk-tissue RNA-seq data

The trimmed mean of M-values (TMM) normalization (*38*) was applied to the raw count data, and log-transformed counts per million reads mapped ($\log_2$CPM) was used. Genes with $\log_2\text{CPM} > 0.1$ in at least 25% of samples were retained. Connectivity between samples was calculated by weighted correlation network analysis (*39*) and was normalized by the *z*-score method. Samples with *z*-score connectivity < (−3) were labeled as outliers and were removed from downstream analysis. Data were then quantile normalized with the preprocessCore (*40*) package. The batch effect was corrected with the combat function in the sva package (*41*).

sc/snRNA-seq data

The quantified count matrix and metadata were used. The ROSMAP snRNA-seq data were downloaded from www.synapse.org/#!Synapse:syn18681734. For CMC snRNA-seq data, the preprocessing pipeline can be found in the Capstone paper (syn48958066). The preprocessing of scRNA-seq data of brain organoids can be found in (*28*).

IHC data

IHC data were downloaded from https://github.com/ellispatrick/CortexCellDeconv. Cell proportions were normalized according to the sum-to-1 constraint.

Construction of references and pseudo-bulks

Two types of references were created using the sc/snRNA-seq count matrix. DSA, dtangle, OLS, and CIBERSORT require a reference of

pooled cells. To construct the pooled-cell reference, the expression count matrix (genes by cells) from the sc/snRNA-seq data was averaged by cell types for each gene. The averaged count matrix was then normalized into CPM and transformed using a $\log_2$ transformation. On the other hand, MuSiC and Bisque require an input of single-cell reference, which directly used the gene-by-cell count matrix from the sc/snRNA-seq data.

Marker gene identification

Marker genes were identified at both the single-cell level and pseudo-bulk level. At the single-cell level, marker genes were determined by assessing the differential expression between the cell group of target cell type and the cell group of the cell type exhibiting the closest expression profiles. Similarly, at the pseudo-bulk level, marker genes were identified by evaluating the expression differences between pseudo-bulk samples of the target cell type and those of the cell type with the most similar expression patterns to the target cell type. At the single-cell level, marker genes were identified with Seurat (*42*). Genes having a proportion of zero expression > 15% in the target cell type were removed. The Wilcoxon signed-rank test was used to test the expression difference. Genes with $\log_2$FC > 1 and FDR-corrected P value < 0.05 were defined as marker genes at the single-cell level. At the pseudo-bulk level, marker genes were tested in DESeq2 (*43*). The likelihood ratio test was used to test the expression difference between the two cell groups. Marker genes with $\log_2$FC > 2 and FDR-corrected P value < 0.05 were defined as marker genes at the pseudo-bulk level.

Estimation of cell proportions

Three inputs were required for all deconvolution methods: bulk tissue data, reference, and marker genes. Batch-corrected data were used as input for bulk tissue data. The intersection of marker genes identified at the single-cell level and the pseudo-bulk level was used as input of marker genes. For DSA, dtangle, OLS, and CIBERSORT, pooled-cell reference was used. For MuSiC and Bisque, single-cell reference was used. The genes that have no expression variation across cell types were removed from the reference.

Evaluation of estimated cell proportions

Two ground truths were used to evaluate estimated cell proportions: cell proportions from IHC data and cell proportions from sc/snRNA-seq data. For one sample, cell proportions from sc/snRNA-seq were calculated by dividing the number of cells of one specific cell type by the total number of cells. RMSE was used as an evaluation metric. The formula of RMSE is

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n}}$$

y_i is the estimated cell proportion and $\hat{y}$ is the ground truth. n is the number of cell types in the sample-level evaluation, and n is the number of samples in the cell-type level evaluation.

Estimation of cell-type expressions

Estimation of cell-type expression requires three inputs: bulk tissue data, cell proportions, and reference. Batch-corrected data was used as input for bulk tissue data. The proportion from DSA, dtangle, OLS, CIBERSORT, MuSiC, and Bisque was used independently. Pooled-cell reference was used as prior for bMIND and swCAM. For TCA, only bulk tissue data and cell proportions were used to estimate cell type expressions for each sample. Data were transformed into a log scale for bMIND and TCA and a linear scale for swCAM.

Evaluation of estimated cell-type expressions

To construct ground truth for evaluating estimated cell-type expressions, the count matrix from sc/snRNA-seq was averaged by cell types and individuals for each gene. Then, the averaged count matrix was normalized into CPM and was log2-transformed. Sample-to-sample Spearman's correlation was tested between estimated expressions and ground truth for each cell type.

Genotyping quality control

The ROSMAP whole-genome sequencing dataset was downloaded from www.synapse.org/#!Synapse:syn11724057. The data have been imputed. Only individuals with both genotype and deconvolution results were retained for the eQTL analysis. SNPs with minor allele frequency (MAF) < 5% or deviating from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$) were excluded. After quality control, we obtained high-quality genotypes for ~4.9 million SNPs (MAF > 5%) in 861 individuals.

eQTL mapping

decon-eQTLs

To identify decon-eQTLs, we tested the associations between genotypes and deconvoluted expressions (phenotypes). We mapped cis-eQTLs within a 1-Mb window of the TSS of each gene using QTLtools (*44*). For each gene, QTLtools performed permutations of the expression data and recorded the best P value for each SNP in the cis window. We used estimated cell-type expressions by bMIND as phenotype data, and phenotype data of eight cell types were tested independently. Quantile normalization was used for normalizing expression matrixes before eQTL mapping. Probabilistic estimation of expression residuals (PEER) was used to identify hidden covariates in the data (*45*). A total of 8 to 35 PEER factors were included as covariates in eQTL mapping.

Bulk tissue eQTLs

To map bulk tissue eQTLs, the same eQTL mapping procedure was performed on the bulk tissue expression data. A total of 33 PEER factors were included as covariates in bulk tissue eQTL mapping.

Replication of decon-eQTLs in BrainGVEX data

To replicate decon-eQTLs, we deconvoluted RNA-seq data from BrainGVEX (*29*) and mapped eQTLs with the deconvoluted data. A total of 430 brain samples with both genotypes and RNA-seq data were used. dtangle was used to estimate cell proportions, with the marker genes and the reference from ROSMAP snRNA-seq data. Then, bulk tissue data were deconvoluted into cell-type expressions for eight major cell types with bMIND. The same eQTL mapping process was performed on the deconvoluted data to identify decon-eQTLs in BrainGVEX data.

*Pi*1 statistic ($\boldsymbol{\pi}$**1**) was used to measure the replication rate of decon-eQTLs in other datasets. Pi1 is the fraction of deconvolution eSNP-eGene pairs (testing data) that are true associations in single-cell−/bulk tissue−/decon-eQTLs (reference data). To compute the *Pi*1 statistic, we extracted nominal P values of decon-eQTLs (nominal P value < 0.05) that overlapped with eQTLs in reference data. Using the qvalue package (*30*) function, we computed the *Pi*0 value and defined the *Pi*1 as 1 − *Pi*0 for each cell type independently.

Replication of decon-eQTLs in single-cell eQTLs

To measure the replication rate of decon-eQTLs in the sc/snRNA-seq dataset, we downloaded cell-type eQTLs identified from the snRNA-seq data of 192 individuals (*31*). With the single-cell eQTLs as a reference, *Pi*1 statistics were calculated for eight cell types independently.

SCZ GWAS heritability enrichment

sLDSC (*32*) was used to calculate SCZ GWAS heritability enrichment in decon-eQTLs. GWAS summary statistics from three published SCZ studies were downloaded (*46*–*48*). Conditional analysis was performed on decon-eQTLs to select the top SNP for each gene [$r_2 > 0.2$ in 1000 Genomes European individuals (*49*)]. Then, script ldsc.py with the "--l2" parameter was used to generate the gene set–specific annotation and linkage disequilibrium (LD) score files, and ldsc.py with the "--h^2-cts" parameter was used to generate stratified heritability by decon-eQTLs of eight cell types.

Colocalization of decon-eQTLs and GWAS signals

For each gene in decon-eQTLs, the colocalization between eSNP and SCZ GWAS signals (*50*) was tested. The "coloc.abf" function in the Coloc (*51*) package (version 5.1.0) was used for testing. The threshold for significance is SNP.PP.H4 > 0.95.

Differential expression analysis

Differential expression analysis was performed on deconvoluted data and bulk tissue data to identify genes associated with AD, SCZ, and brain development. For AD and SCZ, differential expression analysis was conducted in each cell type with the Wilcoxon rank-sum test. For brain development, the linear regression model was used to identify genes showing significant expression changes across the maturity stage of brain organoids. The *P* values were corrected by FDR. Genes with FDR *q* value < 0.05 were identified as PAGs.

Deconvoluted PAGs were compared to PAGs from sc/snRNA-seq data for replication. Single-cell PAGs for AD (*52*) and SCZ (*53*) were downloaded from the original study. For brain development, PAGs were identified in pseudo-bulk data of brain organoids with the same procedure discussed above in the differential expression analysis section.

Supplementary Materials

This PDF file includes:

PsychENCODE Consortium author and collaborator list
Figs. S1 to S9

REFERENCES AND NOTES

1. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
2. PsychENCODE Consortium, S. Akbarian, C. Liu, J. A. Knowles, F. M. Vaccarino, P. J. Farnham, G. E. Crawford, A. E. Jaffe, D. Pinto, S. Dracheva, D. H. Geschwind, J. Mill, A. C. Nairn, A. Abyzov, S. Pochareddy, S. Prabhakar, S. Weissman, P. F. Sullivan, M. W. State, Z. Weng, M. A. Peters, K. P. White, M. B. Gerstein, A. Amiri, C. Armoskus, A. E. Ashley-Koch, T. Bae, A. Beckel-Mitchener, B. P. Berman, G. A. Coetzee, G. Coppola, N. Francoeur, M. Fromer, R. Gao, K. Grennan, J. Herstein, D. H. Kavanagh, N. A. Ivanov, Y. Jiang, R. R. Kitchen, A. Kozlenkov, M. Kundakovic, M. Li, Z. Li, S. Liu, L. M. Mangravite, E. Mattei, E. Markenscoff-Papadimitriou, F. C. P. Navarro, N. North, L. Omberg, D. Panchision, N. Parikshak, J. Poschmann, A. J. Price, M. Purcaro, T. E. Reddy, P. Roussos, S. Schreiner, S. Scuderi, R. Sebra, M. Shibata, A. W. Shieh, M. Skarica, W. Sun, V. Swarup, A. Thomas, J. Tsuji, H. van Bakel, D. Wang, Y. Wang, K. Wang, D. M. Werling, A. J. Willsey, H. Witt, H. Won, C. C. Y. Wong, G. A. Wray, E. Y. Wu, X. Xu, L. Yao, G. Senthil, T. Lehner, P. Sklar, N. Sestan, The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
3. M. Fromer, P. Roussos, S. K. Sieberts, J. S. Johnson, D. H. Kavanagh, T. M. Perumal, D. M. Ruderfer, E. C. Oh, A. Topol, H. R. Shah, L. L. Klei, R. Kramer, D. Pinto, Z. H. Gumus, A. E. Cicek, K. K. Dang, A. Browne, C. Lu, L. Xie, B. Readhead, E. A. Stahl, J. Xiao, M. Parvizi, T. Hamamsy, J. F. Fullard, Y. C. Wang, M. C. Mahajan, J. M. J. Derry, J. T. Dudley, S. E. Hemby, B. A. Logsdon, K. Talbot, T. Raj, D. A. Bennett, P. L. De Jager, J. Zhu, B. Zhang, P. F. Sullivan, A. Chess, S. M. Purcell, L. A. Shinobu, L. M. Mangravite, H. Toyoshiba, R. E. Gur, C.-G. Hahn, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum, E. E. Schadt, K. Hirai, K. Roeder, K. J. Brennand, N. Katsanis, E. Domenici, B. Devlin, P. Sklar, Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
4. J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley, J. J. Royall, K. Aiona, J. M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S. Caldejon, A. Carey, C. Cuhaciyan, R. A. Dalley, N. Dee, T. A. Dolbeare, B. A. Facer, D. Feng, T. P. Fliss, G. Gee, J. Goldy, L. Gourley, B. W. Gregor, G. Gu, R. E. Howard, J. M. Jochim, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, T. A. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N. K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri, N. V. Shapovalova, E. H. Shen, N. Sjoquist, C. R. Slaughterbeck, M. Smith, A. J. Sodt, D. Williams, L. Zollei, B. Fischl, M. B. Gerstein, D. H. Geschwind, I. A. Glass, M. J. Hawrylycz, R. F. Hevner, H. Huang, A. R. Jones, J. A. Knowles, P. Levitt, J. W. Phillips, N. Sestan, P. Wohnoutka, C. Dang, A. Bernard, J. G. Hohmann, E. S. Lein, Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
5. D. A. Bennett, A. S. Buchman, P. A. Boyle, L. L. Barnes, R. S. Wilson, J. A. Schneider, Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).
6. S. Darmanis, S. A. Sloan, Y. Zhang, M. Enge, C. Caneda, L. M. Shuer, M. G. Hayden Gephart, B. A. Barres, S. R. Quake, A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015).
7. Y. Zhang, S. A. Sloan, L. E. Clarke, C. Caneda, C. A. Plaza, P. D. Blumenthal, H. Vogel, G. K. Steinberg, M. S. B. Edwards, G. Li, J. A. Duncan III, S. H. Cheshier, L. M. Shuer, E. F. Chang, G. A. Grant, M. G. H. Gephart, B. A. Barres, Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).
8. D. Lähnemann, J. Köster, E. Szczurek, D. J. McCarthy, S. C. Hicks, M. D. Robinson, C. A. Vallejos, K. R. Campbell, N. Beerenwinkel, A. Mahfouz, L. Pinello, P. Skums, A. Stamatakis, C. S.-O. Attolini, S. Aparicio, J. Baaijens, M. Balvert, B. de Barbanson, A. Cappuccio, G. Corleone, B. E. Dutilh, M. Florescu, V. Guryev, R. Holmer, K. Jahn, T. J. Lobo, E. M. Keizer, I. Khatri, S. M. Kielbasa, J. O. Korbel, A. M. Kozlov, T.-H. Kuo, B. P. F. Lelieveldt, I. I. Mandoiu, J. C. Marioni, T. Marschall, F. Mölder, A. Niknejad, A. Rączkowska, M. Reinders, J. de Ridder, A.-E. Saliba, A. Somarakis, O. Stegle, F. J. Theis, H. Yang, A. Zelikovsky, A. C. McHardy, B. J. Raphael, S. P. Shah, A. Schönhuth, Eleven grand challenges in single-cell data science. *Genome Biol.* **21**, 31 (2020).
9. P. Brennecke, S. Anders, J. K. Kim, A. A. Kolodziejczyk, X. Zhang, V. Proserpio, B. Baying, V. Benes, S. A. Teichmann, J. C. Marioni, M. G. Heisler, Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).
10. J. Cheng, J. Liao, X. Shao, X. Lu, X. Fan, Multiplexing methods for simultaneous large-scale transcriptomic profiling of samples at single-cell resolution. *Adv. Sci.* **8**, e2101229 (2021).
11. M. Maria, N. Pouyanfar, T. Örd, M. U. Kaikkonen, The power of single-cell RNA sequencing in eQTL discovery. *Genes* **13**, 502 (2022).
12. F. Avila Cobos, J. Alquicira-Hernandez, J. E. Powell, P. Mestdagh, K. De Preter, Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat. Commun.* **11**, 5650 (2020).
13. H. Jin, Z. Liu, A benchmark for RNA-seq deconvolution analysis under dynamic testing environments. *Genome Biol.* **22**, 102 (2021).
14. B. B. Nadel, M. Oliva, B. L. Shou, K. Mitchell, F. Ma, D. J. Montoya, A. Mouton, S. Kim-Hellmuth, B. E. Stranger, M. Pellegrini, S. Mangul, Systematic evaluation of transcriptomics-based deconvolution methods and references using thousands of clinical samples. *Brief. Bioinform.* **22**, bbab265 (2021).
15. G. J. Sutton, D. Poppe, R. K. Simmons, K. Walsh, U. Nawaz, R. Lister, J. A. Gagnon-Bartsch, I. Voineagu, Comprehensive evaluation of deconvolution methods for human brain gene expression. *Nat. Commun.* **13**, 1358 (2022).
16. E. Patrick, M. Taga, A. Ergun, B. Ng, W. Casazza, M. Cimpean, C. Yung, J. A. Schneider, D. A. Bennett, C. Gaiteri, P. L. De Jager, E. M. Bradshaw, S. Mostafavi, Deconvolving the contributions of cell-type heterogeneity on cortical gene expression. *PLoS Comput. Biol.* **16**, e1008120 (2020).
17. Y. Zhong, Y.-W. Wan, K. Pang, L. M. L. Chow, Z. Liu, Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC Bioinformatics* **14**, 89 (2013).
18. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
19. G. J. Hunt, S. Freytag, M. Bahlo, J. A. Gagnon-Bartsch, dtangle: Accurate and robust cell type deconvolution. *Bioinformatics* **35**, 2093–2099 (2019).
20. X. Wang, J. Park, K. Susztak, N. R. Zhang, M. Li, Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).

21. J. Wang, K. Roeder, B. Devlin, Bayesian estimation of cell type-specific gene expression with prior derived from single-cell data. *Genome Res.* **31**, 1807–1818 (2021).
22. L. Chen, C. T. Wu, C. H. Lin, R. Dai, C. Liu, R. Clarke, G. Yu, J. E. Van Eyk, D. M. Herrington, Y. Wang, swCAM: Estimation of subtype-specific expressions in individual samples with unsupervised sample-wise deconvolution. *Bioinformatics* **38**, 1403–1410 (2022).
23. E. Rahmani, R. Schweiger, B. Rhead, L. A. Criswell, L. F. Barcellos, E. Eskin, S. Rosset, S. Sankararaman, E. Halperin, Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. *Nat. Commun.* **10**, 3417 (2019).
24. T. P. Morris, I. R. White, M. J. Crowther, Using simulation studies to evaluate statistical methods. *Stat. Med.* **38**, 2074–2102 (2019).
25. B. Jew, M. Alvarez, E. Rahmani, Z. Miao, A. Ko, K. M. Garske, J. H. Sul, K. H. Pietilainen, P. Pajukanta, E. Halperin, Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
26. P. L. De Jager, Y. Ma, C. McCabe, J. Xu, B. N. Vardarajan, D. Felsky, H.-U. Klein, C. C. White, M. A. Peters, B. Lodgson, P. Nejad, A. Tang, L. M. Mangravite, L. Yu, C. Gaiteri, S. Mostafavi, J. A. Schneider, D. A. Bennett, A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).
27. G. E. Hoffman, J. Bendl, G. Voloudakis, K. S. Montgomery, L. Sloofman, Y. C. Wang, H. R. Shah, M. E. Hauberg, J. S. Johnson, K. Girdhar, L. Song, J. F. Fullard, R. Kramer, C. G. Hahn, R. Gur, S. Marenco, B. K. Lipska, D. A. Lewis, V. Haroutunian, S. Hemby, P. Sullivan, S. Akbarian, A. Chess, J. D. Buxbaum, G. E. Crawford, E. Domenici, B. Devlin, S. K. Sieberts, M. A. Peters, P. Roussos, CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci. Data* **6**, 180 (2019).
28. A. Jourdon, F. Wu, J. Mariani, D. Capauto, S. Norton, L. Tomasini, A. Amiri, M. Suvakov, J. D. Schreiner, Y. Jang, A. Panda, C. K. Nguyen, E. M. Cummings, G. Han, K. Powell, A. Szekely, J. C. McPartland, K. Pelphrey, K. Chawarska, P. Ventola, A. Abyzov, F. M. Vaccarino, ASD modelling in organoids reveals imbalance of excitatory cortical neuron subtypes during early neurogenesis. bioRxiv 484988 [Preprint]. 20 March 2022. https://doi.org/10.1101/2022.03.19.484988.
29. D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, D. Clarke, M. Gu, P. Emani, Y. T. Yang, M. Xu, M. J. Gandal, S. Lou, J. Zhang, J. J. Park, C. Yan, S. K. Rhie, K. Manakongtreecheep, H. Zhou, A. Nathan, M. Peters, E. Mattei, D. Fitzgerald, T. Brunetti, J. Moore, Y. Jiang, K. Girdhar, G. E. Hoffman, S. Kalayci, Z. H. Gumus, G. E. Crawford; PsychENCODE Consortium, P. Roussos, S. Akbarian, A. E. Jaffe, K. P. White, Z. Weng, N. Sestan, D. H. Geschwind, J. A. Knowles, M. B. Gerstein, Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
30. J. D. Storey, A. J. Bass, A. Dabney, D. Robinson, qvalue: Q-value estimation for false discovery rate control. R package version 2.2.2 (2015).
31. J. Bryois, D. Calini, W. Macnair, L. Foo, E. Urich, W. Ortmann, V. A. Iglesias, S. Selvaraj, E. Nutma, M. Marzin, S. Amor, A. Williams, G. Castelo-Branco, V. Menon, P. De Jager, D. Malhotra, 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).
32. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, A. L. Price, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
33. J. Rammohan, S. P. Lund, N. Alperovich, V. Paralanov, E. A. Strychalski, D. Ross, Comparison of bias and resolvability in single-cell and single-transcript methods. *Commun. Biol.* **4**, 659 (2021).
34. E. Denisenko, B. B. Guo, M. Jones, R. Hou, L. de Kock, T. Lassmann, D. Poppe, O. Clement, R. K. Simmons, R. Lister, A. R. R. Forrest, Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 130 (2020).
35. D. Neavin, Q. Nguyen, M. S. Daniszewski, H. H. Liang, H. S. Chiu, Y. K. Wee, A. Senabouth, S. W. Lukowski, D. E. Crombie, G. E. Lidgerwood, D. Hernandez, J. C. Vickers, A. L. Cook, N. J. Palpant, A. Pebay, A. W. Hewitt, J. E. Powell, Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biol.* **22**, 76 (2021).
36. S. Yazar, J. Alquicira-Hernandez, K. Wing, A. Senabouth, M. G. Gordon, S. Andersen, Q. Lu, A. Rowson, T. R. P. Taylor, L. Clarke, K. Maccora, C. Chen, A. L. Cook, C. J. Ye, K. A. Fairfax, A. W. Hewitt, J. E. Powell, Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
37. J. M. Nascimento, D. Martins-de-Souza, The proteome of schizophrenia. *NPJ Schizophr.* **1**, 14003 (2015).
38. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
39. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
40. B. M. Bolstad, A collection of pre-processing functions. R package version 1.60.1 (2022).
41. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
42. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
43. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
44. O. Delaneau, H. Ongen, A. A. Brown, A. Fort, N. I. Panousis, E. T. Dermitzakis, A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* **8**, 15452 (2017).
45. O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
46. A. F. Pardinas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron, M. L. Hamshere, J. Han, L. Hubbard, A. Lynham, K. Mantripragada, E. Rees, J. H. MacCabe, S. A. McCarroll, B. T. Baune, G. Breen, E. M. Byrne, U. Dannlowski, T. C. Eley, C. Hayward, N. G. Martin, A. M. McIntosh, R. Plomin, D. J. Porteous, N. R. Wray, A. Caballero, D. H. Geschwind, L. M. Huckins, D. M. Ruderfer, E. Santiago, P. Sklar, E. A. Stahl, H. Won, E. Agerbo, T. D. Als, O. A. Andreassen, M. Baekvad-Hansen, P. B. Mortensen, C. B. Pedersen, A. D. Borglum, J. Bybjerg-Grauholm, S. Djurovic, N. Durmishi, M. G. Pedersen, V. Golimbet, J. Grove, D. M. Hougaard, M. Mattheisen, E. Molden, O. Mors, M. Nordentoft, M. Pejovic-Milovancevic, E. Sigurdsson, T. Silagadze, C. S. Hansen, K. Stefansson, H. Stefansson, S. Steinberg, S. Tosato, T. Werge; GERARD1 Consortium; CRESTAR Consortium, D. A. Collier, D. Rujescu, G. Kirov, M. J. Owen, M. C. O'Donovan, J. T. R. Walters, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
47. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium, Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715.e16 (2018).
48. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
49. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
50. V. Trubetskoy, A. F. Pardinas, T. Qi, G. Panagiotaropoulou, S. Awasthi, T. B. Bigdeli, J. Bryois, C. Y. Chen, C. A. Dennison, L. S. Hall, M. Lam, K. Watanabe, O. Frei, T. Ge, J. C. Harwood, F. Koopmans, S. Magnusson, A. L. Richards, J. Sidorenko, Y. Wu, J. Zeng, J. Grove, M. Kim, Z. Li, G. Voloudakis, W. Zhang, M. Adams, I. Agartz, E. G. Atkinson, E. Agerbo, M. Al Eissa, M. Albus, M. Alexander, B. Z. Alizadeh, K. Alptekin, T. D. Als, F. Amin, V. Arolt, M. Arrojo, L. Athanasiu, M. H. Azevedo, S. A. Bacanu, N. J. Bass, M. Begemann, R. A. Belliveau, J. Bene, B. Benyamin, S. E. Bergen, G. Blasi, J. Bobes, S. Bonassi, A. Braun, R. A. Bressan, E. J. Bromet, R. Bruggeman, P. F. Buckley, R. L. Buckner, J. Bybjerg-Grauholm, W. Cahn, M. J. Cairns, M. E. Calkins, V. J. Carr, D. Castle, S. V. Catts, K. D. Chambert, R. C. K. Chan, B. Chaumette, W. Cheng, E. F. C. Cheung, S. A. Chong, D. Cohen, A. Consoli, Q. Cordeiro, J. Costas, C. Curtis, M. Davidson, K. L. Davis, L. de Haan, F. Degenhardt, L. E. DeLisi, D. Demontis, F. Dickerson, D. Dikeos, T. Dinan, S. Djurovic, J. Duan, G. Ducci, F. Dudbridge, J. G. Eriksson, L. Fananas, S. V. Faraone, A. Fiorentino, A. Forstner, J. Frank, N. B. Freimer, M. Fromer, A. Frustaci, A. Gadelha, G. Genovese, E. S. Gershon, M. Giannitelli, I. Giegling, P. Giusti-Rodriguez, S. Godard, J. I. Goldstein, J. G. Penas, A. Gonzalez-Pinto, S. Gopal, J. Gratten, M. F. Green, T. A. Greenwood, O. Guillin, S. Guloksuz, R. E. Gur, R. C. Gur, B. Gutierrez, E. Hahn, H. Hakonarson, V. Haroutunian, A. M. Hartmann, C. Harvey, C. Hayward, F. A. Henskens, S. Herms, P. Hoffmann, D. P. Howrigan, M. Ikeda, C. Iyegbe, I. Joa, A. Julia, A. K. Kahler, T. Kam-Thong, Y. Kamatani, S. Karachanak-Yankova, O. Kebir, M. C. Keller, B. J. Kelly, A. Khrunin, S. W. Kim, J. Klovins, N. Kondratiev, B. Konte, J. Kraft, M. Kubo, V. Kucinskas, Z. A. Kucinskiene, A. Kusumawardhani, H. Kuzelova-Ptackova, S. Landi, L. C. Lazzeroni, P. H. Lee, S. E. Legge, D. S. Lehrer, R. Lencer, B. Lerer, M. Li, J. Lieberman, G. A. Light, S. Limborska, C.-M. Liu, J. Lonnqvist, C. M. Loughland, J. Lubinski, J. J. Luykx, A. Lynham, M. Macek Jr., A. Mackinnon, P. K. E. Magnusson, B. S. Maher, W. Maier, D. Malaspina, J. Mallet, S. R. Marder, S. Marsal, A. R. Martin, L. Martorell, M. Mattheisen, R. W. McCarley, C. McDonald, J. J. McGrath, H. Medeiros, S. Meier, B. Melegh, I. Melle, R. I. Mesholam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, M. Mitjans, E. Molden, E. Molina, M. D. Molto, V. Mondelli, C. Moreno, C. P. Morley, G. Muntane, K. C. Murphy, I. Myin-Germeys, I. Nenadic, G. Nestadt, L. Nikitina-Zake, C. Noto, K. H. Nuechterlein, N. L. O'Brien, F. A. O'Neill, S. Y. Oh, A. Olincy, V. K. Ota, C. Pantelis, G. N. Papadimitriou, M. Parellada, T. Paunio, R. Pellegrino, S. Periyasamy, D. O. Perkins, B. Pfuhlmann, O. Pietilainen, J. Pimm, D. Porteous, J. Powell, D. Quattrone, D. Quested, A. D. Radant, A. Rampino, M. H. Rapaport, A. Rautanen, A. Reichenberg, C. Roe, J. L. Roffman, J. Roth, M. Rothermundt, B. P. F. Rutten, S. Saker-Delye, V. Salomaa, J. Sanjuan, M. L. Santoro, A. Savitz, U. Schall, R. J. Scott, L. J. Seidman, S. I. Sharp, J. Shi,

L. J. Siever, E. Sigurdsson, K. Sim, N. Skarabis, P. Slominsky, H. C. So, J. L. Sobell, E. Soderman, H. J. Stain, N. E. Steen, A. A. Steixner-Kumar, E. Stogmann, W. S. Stone, R. E. Straub, F. Streit, E. Strengman, T. S. Stroup, M. Subramaniam, C. A. Sugar, J. Suvisaari, D. M. Svrakic, N. R. Swerdlow, J. P. Szatkiewicz, T. M. T. Ta, A. Takahashi, C. Terao, F. Thibaut, D. Toncheva, P. A. Tooney, S. Torretta, S. Tosato, G. B. Tura, B. I. Turetsky, A. Ucok, A. Vaaler, T. van Amelsvoort, R. van Winkel, J. Veijola, J. Waddington, H. Walter, A. Waterreus, B. T. Webb, M. Weiser, N. M. Williams, S. H. Witt, B. K. Wormley, J. Q. Wu, Z. Xu, R. Yolken, C. C. Zai, W. Zhou, F. Zhu, F. Zimprich, E. C. Atbasoglu, M. Ayub, C. Benner, A. Bertolino, D. W. Black, N. J. Bray, G. Breen, N. G. Buccola, W. F. Byerley, W. J. Chen, C. R. Cloninger, B. Crespo-Facorro, G. Donohoe, R. Freedman, C. Galletly, M. J. Gandal, M. Gennarelli, D. M. Hougaard, H. G. Hwu, A. V. Jablensky, S. A. McCarroll, J. L. Moran, O. Mors, P. B. Mortensen, B. Muller-Myhsok, A. L. Neil, M. Nordentoft, M. T. Pato, T. L. Petryshen, M. Pirinen, A. E. Pulver, T. G. Schulze, J. M. Silverman, J. W. Smoller, E. A. Stahl, D. W. Tsuang, E. Vilella, S.-H. Wang, S. Xu; Indonesia Schizophrenia Consortium; PsychEncode; Psychosis Endophenotypes International Consortium; SynGO Consortium, R. Adolfsson, C. Arango, B. T. Baune, S. I. Belangero, A. D. Borglum, D. Braff, E. Bramon, J. D. Buxbaum, D. Campion, J. A. Cervilla, S. Cichon, D. A. Collier, A. Corvin, D. Curtis, M. D. Forti, E. Domenici, H. Ehrenreich, V. Escott-Price, T. Esko, A. H. Fanous, A. Gareeva, M. Gawlik, P. V. Gejman, M. Gill, S. J. Glatt, V. Golimbet, K. S. Hong, C. M. Hultman, S. E. Hyman, N. Iwata, E. G. Jonsson, R. S. Kahn, J. L. Kennedy, E. Khusnutdinova, G. Kirov, J. A. Knowles, M. O. Krebs, C. Laurent-Levinson, J. Lee, T. Lencz, D. F. Levinson, Q. S. Li, J. Liu, A. K. Malhotra, D. Malhotra, A. McIntosh, A. McQuillin, P. R. Menezes, V. A. Morgan, D. W. Morris, B. J. Mowry, R. M. Murray, V. Nimgaonkar, M. M. Nothen, R. A. Ophoff, S. A. Paciga, A. Palotie, C. N. Pato, S. Qin, M. Rietschel, B. P. Riley, M. Rivera, D. Rujescu, M. C. Saka, A. R. Sanders, S. G. Schwab, A. Serretti, P. C. Sham, Y. Shi, D. St Clair, H. Stefansson, K. Stefansson, M. T. Tsuang, J. van Os, M. P. Vawter, D. R. Weinberger, T. Werge, D. B. Wildenauer, X. Yu, W. Yue, P. A. Holmans, A. J. Pocklington, P. Roussos, E. Vassos, M. Verhage, P. M. Visscher, J. Yang, D. Posthuma, O. A. Andreassen, K. S. Kendler, M. J. Owen, N. R. Wray, M. J. Daly, H. Huang, B. M. Neale, P. F. Sullivan, S. Ripke, J. T. R. Walters, M. C. O'Donovan; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

51. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genet.* **10**, e1004383 (2014).
52. H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang, A. J. Martorell, R. M. Ransohoff, B. P. Hafler, D. A. Bennett, M. Kellis, L. H. Tsai, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
53. W. B. Ruzicka, S. Mohammadi, J. F. Fullard, J. Davila-Velderrain, S. Subburaju, D. R. Tso, M. Hourihan, S. Jiang, H.-C. Lee, J. Bendl; PsychENCODE Consortium, G. Voloudakis, V. Haroutunian, G. E. Hoffman, P. Roussos, M. Kellis, Single-cell multi-cohort dissection of the schizophrenia transcriptome. medRxiv 2022.08.31.22279406 (2022). https://doi.org/10.1101/2022.08.31.22279406.

Acknowledgment: We thank R. Kopp at SUNY Upstate Medical University for the help in polishing words. We thank all the participants involved in the ROSMAP and PsychENCODE study for making the data available. **Funding:** This work was supported by NIH grants U01MH122591, U01MH116489, R01MH110920, U01MH103340, R01MH126459, and R01MH109648; the Simons Foundation grant 632742; the National Natural Science Foundation of China grants 82022024 and 31970572; and the Science and Technology Innovation Program of Hunan Province grants 2021RC4018 and 2021RC5027. Data were generated as part of the PsychENCODE Consortium, supported by the following grants: U01DA048279, U01MH103339, U01MH103340, U01MH103346, U01MH103365, U01MH103392, U01MH116438, U01MH116441, U01MH116442, U01MH116488, U01MH116489, U01MH116492, U01MH122590, U01MH122591, U01MH122592, U01MH122849, U01MH122678, U01MH122681, U01MH116487, U01MH122509, R01MH094714, R01MH105472, R01MH105898, R01MH109677, R01MH109715, R01MH110905, R01MH110920, R01MH110921, R01MH110926, R01MH110927, R01MH110928, R01MH111721, R01MH117291, R01MH117292, R01MH117293, R21MH102791, R21MH103877, R21MH105853, R21MH105881, R21MH109956, R56MH114899, R56MH114901, R56MH114911, R01MH125516, R01MH126459, R01MH129301, R01MH126393, R01MH121521, R01MH116529, R01MH129817, R01MH117406, and P50MH106934 awarded to A. Abyzov, N. Ahituv, S. Akbarian, K. Brennand, A. Chess, G. Cooper, G. Crawford, S. Dracheva, P. Farnham, M. Gandal, M. Gerstein, D. Geschwind, F. Goes, J. F. Hallmayer, V. Haroutunian, T. M. Hyde, A. Jaffe, P. Jin, M. Kellis, J. Kleinman, J. A. Knowles, A. Kriegstein, C. Liu, C. E. Mason, K. Martinowich, E. Mukamel, R. Myers, C. Nemeroff, M. Peters, D. Pinto, K. Pollard, K. Ressler, P. Roussos, S. Sanders, N. Sestan, P. Sklar, M. P. Snyder, M. State, J. Stein, P. Sullivan, A. E. Urban, F. Vaccarino, S. Warren, D. Weinberger, S. Weissman, Z. Weng, K. White, A. Jeremy Willsey, H. Won, and P. Zandi. **Author contributions:** Conceptualization: C.L., S.H.W., X.W., C.C., R.D., and Y.W. Methodology: X.W., C.L., T.C., X.W., C.C., R.D., and Y.W. Formal analysis: X.W., T.C., M.Z., C.Z., X.W., and R.D. Software: T.C., C.Z., G.E.H., X.W., and R.D. Investigation: F.M.V., J.M., T.C., J.F.F., M.Z., D.P., D.L., X.W., A.J., C.C., and R.D. Visualization: T.C., D.P., X.W., and R.D. Supervision: C.L., J.F.F., P.R., S.H.W., C.C., and R.D. Writing—original draft: C.L., X.W., and R.D. Writing—review and editing: F.M.V., C.L., T.C., D.P., C.Z., S.H.W., X.W., and R.D. Resources: F.M.V., C.L., J.M., D.P., P.R., C.Z., D.L., G.E.H., X.W., F.W., C.C., and R.D. Funding acquisition: C.L., D.P., P.R., and C.C.. Data curation: C.L., T.C., C.Z., D.L., X.W., A.J., and R.D. Validation: T.C., M.Z., D.P., X.W., and R.D. Project administration: C.L., P.R., X.W., C.C., and R.D. The PsychENCODE Consortium has made significant contributions to the collection, preprocessing, and management of data. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The source data described in this manuscript are available via the PsychENCODE Knowledge Portal (https://doi.org/10.7303/syn48955488; https://nda.nih.gov/edit_collection.html?id=5063; https://doi.org/10.7303/syn3270015). The PsychENCODE Knowledge Portal is a platform for accessing data, analyses, and tools generated through grants funded by the National Institute of Mental Health (NIMH) PsychENCODE Consortium. Data are available for general research use according to the following requirements for data access and data attribution (https://psychencode.synapse.org/DataAccess). The code and results can be accessed at https://doi.org/10.5281/zenodo.8175819 and https://github.com/RujiaDai/BrainDeconBenchmark.

Submitted 13 March 2023
Accepted 5 January 2024
Published 23 May 2024
10.1126/sciadv.adh2588