Multivariate variance components analysis uncovers genetic architecture of brain isoform expression and novel psychiatric disease mechanisms

Minsoo Kim[1-3*], Daniel D. Vo[1,4], Connor T. Jops[1], Cindy Wen[1,4], Ashok Patowary[1], Arjun Bhattacharya[5], Chloe X. Yap[1,6,7], Hua Zhou[8,9*], Michael J. Gandal[1-4,10,11*]

[1] Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[2] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[3] Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[4] Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA.
[5] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[6] Mater Research Institute, University of Queensland, Brisbane, Queensland, Australia.
[7] Institute for Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia.
[8] Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA.
[9] Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[10] Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
[11] Lifespan Brain Institute, Children's Hospital of Philadelphia, Philadelphia, PA, USA.
* Correspondence to: minsookim@mednet.ucla.edu, huazhou@ucla.edu, michael.gandal@pennmedicine.upenn.edu

Abstract

Multivariate variance components linear mixed models are fundamental statistical models in quantitative genetics, widely used to quantify SNP-based heritability (h^2_{SNP}) and genetic correlation (r_g) across complex traits. However, maximum likelihood estimation of multivariate variance components models remains numerically challenging when the number of traits and variance components are both greater than two. To address this critical gap, here we introduce a novel statistical method for fitting multivariate variance components models. This method improves on existing methods by allowing for arbitrary number of traits and/or variance components. We illustrate the utility of our method by characterizing for the first time the genetic architecture of isoform expression in the human brain, modeling up to 23 isoforms jointly across ~900 individuals within PsychENCODE. We find a significant proportion of isoforms to be under genetic control (17,721 of 93,293 isoforms) with substantial shared genetic effects among local (or *cis*-) relative to distal (or *trans*-) genetic variants (median $r_{g,cis}$ and $r_{g,trans}$ = 0.31 and 0.06). Importantly, we find that 11.6% of brain-expressed genes (2,900 genes) are heritable only at the isoform-level. Integrating these isoform-specific genetic signals with psychiatric GWAS signals uncovers previously hidden psychiatric disease mechanisms. Specifically, we highlight reduced expression of a specific *XRN2* isoform as the underlying driver of the strongest GWAS signal for autism spectrum disorder. Overall, our method for fitting multivariate variance components models is flexible, widely applicable, and is implemented in the Julia programming language and available online.

Introduction

The genetic contribution to phenotypes of interest (i.e. SNP-based heritability or h^2_{SNP}) and the extent to which they are shared (i.e. genetic correlation or r_g) can be quantified through variance components linear mixed models. The simplest univariate variance components model assumes two variance components, one of which captures the aggregate genome-wide genetic effects[1]. However, this model is likely misspecified for either gene or isoform expression, since SNPs in the vicinity of a gene (herein referred to as *cis*-SNPs) tend to exert stronger effects on its expression than distal SNPs (i.e. *trans*-SNPs)[2]. Similarly, the simplest multivariate variance components model looks at phenotypes pairwise and assumes two variance components, one of which captures the aggregate degree of pleiotropy among genome-wide SNPs[3,4]. Again, this model is likely inadequate for isoform-level expression, since the degree of genetic correlation can differ among different sets of SNPs (e.g. *cis*- and *trans*-SNPs).

Unfortunately, maximum likelihood (ML) and restricted maximum likelihood (REML) estimation of the most general form of multivariate variance components models with more than two phenotypes and more than two variance components still poses a significant numerical and computational challenge. Existing methods lack flexibility by modelling at most two phenotypes with multiple variance components[5,6] or multiple phenotypes with two variance components[7,8]. To address this critical gap, we implemented the minorization-maximization (MM) algorithm[9] for ML and REML estimation in multivariate variance components linear mixed models[10], using the Julia programming language[11]. We refer to the Methods for a comprehensive summary of optimization methods for tackling variance components models, but we note here that the major advantages of the MM algorithm include numerical stability, fast convergence, and graceful adaptation to the positive semidefiniteness constraint of variance components parameters[10].

To demonstrate the utility of our method, we modelled constituent isoforms of a given gene jointly in the PsychENCODE dataset[12,13], partitioning genetic variance and covariance among *cis*- and *trans*-SNPs. We find that genetic variants local to the gene or isoform in question have a sparse but large effect on expression, whereas distal genetic variants have individually small but collectively large effects on expression. We further show that there are substantial shared genetic influences among *cis*-SNPs compared to *trans*-SNPs for isoform expression. Importantly, we find isoform-level analyses to lead to discovery of many more genetic signals than gene-level analyses. Some of these isoform-specific genetic signals were significantly associated with increased risk for psychiatric disorders such as autism spectrum disorder[14]. These findings suggest that isoform-resolution analyses have the potential to uncover novel disease genes and aid in interpreting GWAS results. Overall, we present a comprehensive dissection of genetic influences on brain gene and isoform expression and their relation to brain-related disorders.

Results

MM algorithm for multivariate variance components linear mixed models

The multivariate variance components linear mixed model with $n \times d$ response matrix $\mathbf{Y}$, $n \times p$ predictor matrix $\mathbf{X}$, and m known $n \times n$ positive semidefinite matrices $(\mathbf{V}_1, \dots, \mathbf{V}_m)$ (e.g. genetic relationship matrices) assumes $\mathrm{vec}\,\mathbf{Y} \sim N(\mathrm{vec}(\mathbf{XB}), \mathbf{\Omega})$, where

$$\mathbf{\Omega} = \sum_{i=1}^{m} \mathbf{\Gamma}_i \otimes \mathbf{V}_i.$$

The parameters of the model include $p \times d$ mean effects $\mathbf{B}$ and $d \times d$ variance components $(\mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_m)$ that are positive semidefinite. $\mathrm{vec}\,\mathbf{Y}$ creates an $nd \times 1$ vector from $\mathbf{Y}$ by stacking its columns and $\otimes$ denotes the Kronecker product. The univariate model is subsumed under the multivariate model when $d = 1$. Under this model, when there are no mean effects or there is only the intercept term, phenotypic variance and covariance can be decomposed into the sum of variance components parameters (Methods).

As there are no closed-form solutions to ML estimation, variance components are usually estimated using an iterative algorithm such as Fisher scoring and expectation-maximization (EM) algorithms. Here, we pursue ML estimation with the MM algorithm[10] that alternatively updates $\mathbf{B}$ and $\mathbf{\Gamma}_i$'s. In each iteration, the MM updates are

$$\begin{aligned} \mathrm{vec}\,\mathbf{B}^{(t)} &= [(\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-(t)}(\mathbf{I}_d \otimes \mathbf{X})]^{-1}(\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-(t)}\mathrm{vec}\,\mathbf{Y} \\ \mathbf{\Gamma}_i^{(t+1)} &= \mathbf{L}_i^{-(t)T}[\mathbf{L}_i^{(t)T}(\mathbf{\Gamma}_i^{(t)}\mathbf{R}^{(t)T}\mathbf{V}_i\mathbf{R}^{(t)}\mathbf{\Gamma}_i^{(t)})\mathbf{L}_i^{(t)}]^{1/2}\mathbf{L}_i^{-(t)}, \end{aligned}$$

where $\mathbf{L}_i^{(t)}$ is the Cholesky factor of $(\mathbf{I}_d \otimes \mathbf{1}_n)^T[(\mathbf{1}_d\mathbf{1}_d^T \otimes \mathbf{V}_i) \odot \mathbf{\Omega}^{-(t)}](\mathbf{I}_d \otimes \mathbf{1}_n)$ and $\mathbf{R}^{(t)}$ is the $n \times d$ matrix such that $\mathrm{vec}\,\mathbf{R}^{(t)} = \mathbf{\Omega}^{-(t)}\mathrm{vec}(\mathbf{Y} - \mathbf{XB}^{(t)})$. $\odot$ denotes the Hadamard product. For REML estimation, the updates are basically of the same form. Compared to other iterative algorithms, the MM algorithm is numerically stable, converges fast, and gracefully respects the nonnegativity (or positive semidefinite) constraint of variance components. In the simplest model where there are two variance components, repeated matrix inversion can be avoided by the generalized eigenvalue decomposition of the two kernel matrices[10,15,16]. When there are more than two variance components, a matrix inversion is inevitable in each update, so the MM algorithm is not scalable to biobank-level data, and we recommend applying this method for datasets of size up to $n \times d = 50000$.

Standard errors for our parameter estimates were subsequently calculated using the Fisher information matrix (Methods), where

$$\begin{aligned} \mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vec}\,\mathbf{B})^T\partial(\mathrm{vec}\,\mathbf{B})}\mathcal{L}\right] &= (\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-1}(\mathbf{I}_d \otimes \mathbf{X}) \\ \mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vech}\,\mathbf{\Gamma}_i)^T\partial(\mathrm{vec}\,\mathbf{B})}\mathcal{L}\right] &= \mathbf{0} \\ \mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vech}\,\mathbf{\Gamma}_j)^T\partial(\mathrm{vech}\,\mathbf{\Gamma}_i)}\mathcal{L}\right] &= \frac{1}{2}\mathbf{U}_i^T(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1})\mathbf{U}_j \end{aligned}$$

and $\mathbf{U}_i = (\mathbf{I}_d \otimes \mathbf{K}_{nd} \otimes \mathbf{I}_n)(\mathbf{I}_{d^2} \otimes \mathrm{vec}\,\mathbf{V}_i)\mathbf{D}_d$. $\mathbf{K}_{nd}$ is the $nd \times nd$ commutation matrix and $\mathbf{D}_d$ the $d^2 \times d(d+1)/2$ duplication matrix.

Overview of h^2_{SNP} and r_g analyses for brain gene and isoform expression

In the present study, for all variance components models, whether be it univariate, pairwise bivariate, or multivariate, we specified three variance components that include *cis*- and *trans*-SNP effects, and residual effects. This model specification corresponds to an assumption that *cis*- and *trans*-SNP effects are realized from different distributions of effect sizes. Hereafter, we refer to the variance components parameters corresponding to either *cis*- or *trans*-SNP effects as genetic variances and genetic covariances. We defined *cis*-SNPs as those within ±1 Mb window of gene start and gene end sites and *trans*-SNPs as all other SNPs (**Figure 1**). Indeed, other definitions are possible, and we tested varying windows as part of sensitivity analyses.

We present estimates from the MM algorithm and REML estimation, unless otherwise stated. The standard errors for variance components estimates were calculated from the Fisher information matrix. Inference on variance components parameters in the univariate case was done using a variation of the likelihood ratio test (LRT)[5,17] (Methods). In the multivariate setting, inference on the off-diagonal elements of the variance components parameters was done using the Wald test (Methods).

We first estimated h^2_{SNP} and assessed its significance by fitting the univariate variance components model. Since we specified two separate variance components parameters for *cis*- and *trans*-SNPs, h^2_{SNP} could be decomposed into h^2_{cis} and h^2_{trans} (i.e. $h^2_{SNP} = h^2_{cis} + h^2_{trans}$). To avoid estimating too many parameters in multivariate models, for each gene, we focused on isoforms with significant h^2_{SNP} at $P < 0.05$ in the univariate model. Since we partitioned expression covariances among *cis*- and *trans*-SNPs, and residual effects, there were two genetic correlation parameters (r_g's) that correspond to aggregate degree of pleiotropy among *cis*-SNPs ($r_{g,cis}$) and aggregate degree of pleiotropy among *trans*-SNPs ($r_{g,trans}$) (**Figure 1**). There was also residual correlation (r_e) which captures correlation due to biological and technical factors such as shared gene regulation and measurement error (Methods). It is important to note that r_g is not the sum of $r_{g,cis}$ and $r_{g,trans}$, since these parameters are on a normalized scale. Finally, we conducted *cis*-eQTL analyses[18] in parallel with the same set of *cis*-SNPs to validate h^2_{SNP} results as a positive control and to link isoform-specific signals with GWAS signals (**Figure 1**).

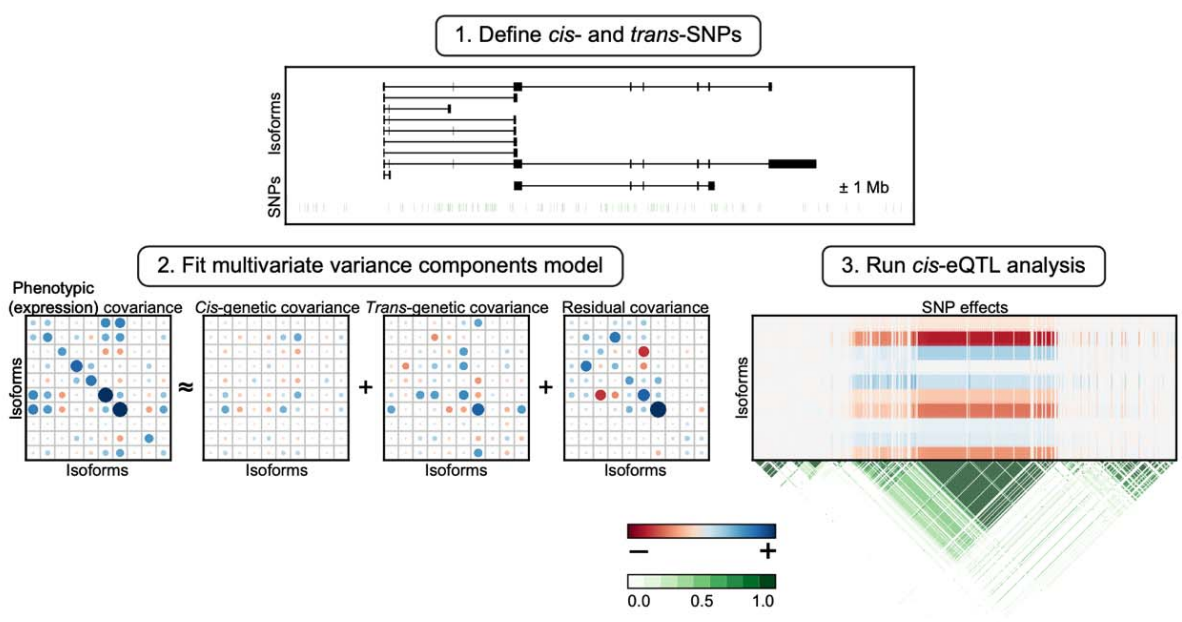


Figure 1: Overview of isoform-centric h^2_{SNP}, r_g, and *cis*-eQTL analyses.
For a given gene and its constituent isoforms, *cis*-SNPs are defined as SNPs within ±1 Mb window of (collapsed) gene start and end sites, while *trans*-SNPs are defined as all other SNPs. Expression variances and covariances are partitioned among *cis*- and *trans*-SNPs for each gene and its constituent isoforms by fitting univariate, pairwise bivariate, and multivariate variance components linear mixed models[10]. The same sets of *cis*-SNPs are used for *cis*-eQTL analyses[18], results of which are compared to heritability and genetic correlation analyses. This schematic figure is based on the real data for ten isoforms belonging to the *KLHL24* gene.

Polygenicity of brain gene and isoform expression

We started with previously harmonized genotype array and frontal cortex RNA-seq data from PsychENCODE[12,13]. We focused on 855 unrelated European individuals with matching genotype and RNA-seq data (Methods). It is worthwhile noting that RNA-seq data was previously quantified using Gencode v19[19] and RSEM[20]. We began with 24,905 genes and 93,293 isoforms in PsychENCODE. On average, each gene had four isoforms, but some genes had as many as 64 brain-expressed isoforms (**Figure S1**). Of these, 22,965 genes and 89,926 isoforms had converged h^2_{SNP} estimates in the univariate model. Median h^2_{SNP} estimates were 0.31 and 0.35 for gene and isoform expression, respectively (**Figure 2a**), and median h^2_{cis} estimates were 0.01 and 0.01 for gene and isoform expression, respectively. 2,822 genes had significant h^2_{SNP} at Bonferroni-adjusted P value < 0.05, while 3,557 genes had at least one isoform with h^2_{SNP} significantly different from zero at Bonferroni-adjusted P value < 0.05. With a more lenient threshold at P < 0.05, 7,239 and 10,139 genes were heritable at the gene- and isoform-level, respectively. This increase in the number of heritable genes reflects added granularity with isoform-level analyses.

It is standard practice when fitting linear mixed models to compare REML and ML estimates, since REML estimates tend to be less biased, while ML estimates can have lower mean squared error (MSE). For the univariate variance components model, we found strong concordance between REML and ML estimates (**Figure S2**). We note that h^2_{trans} and h^2_{SNP} estimates exhibit larger standard errors than h^2_{cis} estimates (**Figure S2**), which is in line with expectation[21], so we caution the readers in interpreting these h^2_{trans} and h^2_{SNP} estimates.

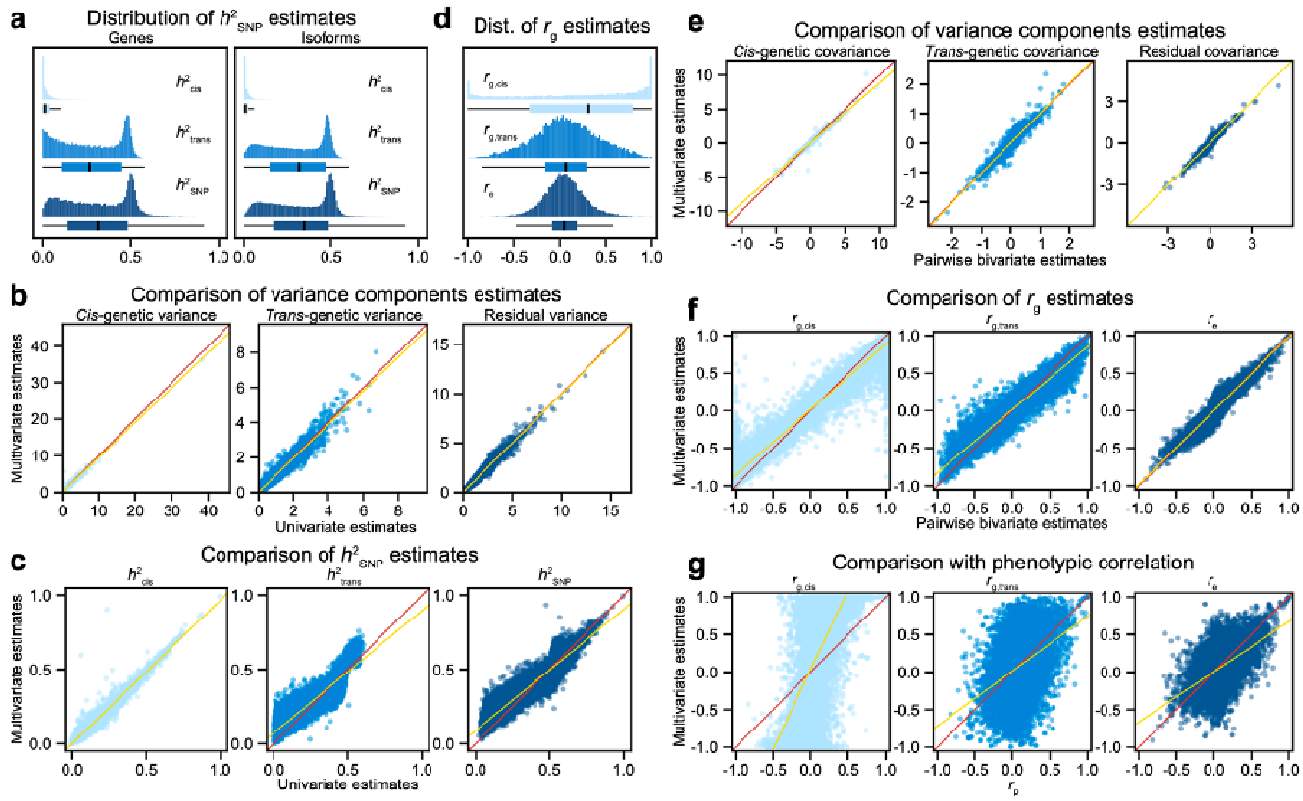


Figure 2: Comparison of h^2_{SNP} and r_g estimates from different variance components models.
a, Shown are the distributions of h^2_{SNP} estimates for gene and isoform expression from fitting a univariate variance components model (22,965 genes + 89,926 isoforms). Note that $h^2_{SNP} = h^2_{cis} + h^2_{trans}$. **b**, Shown are variance components estimates from fitting univariate and multivariate variance components models. **c**, Shown are h^2_{SNP} estimates from univariate and multivariate variance components models. **d**, Shown is the distribution of r_g estimates from fitting a multivariate variance components model. Note that $r_{g,cis}$ captures aggregate degree of pleiotropy among *cis*-SNPs, while $r_{g,trans}$ captures aggregate degree of pleiotropy among *trans*-SNPs. r_e captures residual correlation due to (unknown) biological and technological factors. The bimodal distribution of $r_{g,cis}$ with two modes in the extremes suggests that there are substantial shared genetic influences among *cis*-SNPs. **e**, Shown are (co)variance components estimates from pairwise bivariate and multivariate variance components models. **f**, Shown are r_g estimates from pairwise bivariate and multivariate variance components models. **g**, Shown are r_g estimates from a multivariate variance components model and expression (phenotypic) correlation (r_p). The same r_p values are plotted on the x-axis. All estimates are from REML estimation. The red lines are the diagonal lines, while the yellow lines denote the lines from linear regression.

To ensure that our estimates were robust, we compared gene- and isoform-level h^2_{SNP} estimates for genes with a single isoform (n = 7,246 genes). We found that they were highly concordant (**Figure S3**) despite gene- and isoform-level expression data being processed separately. When we fit a univariate variance components model with two variance components that only specifies *cis*-genetic effects, we observed concordant estimates for h^2_{cis}, although the estimates for the two variance components model were slightly inflated over the three variance components model, indicating the importance of correct model specification. Finally, we ran *cis*-eQTL analyses for gene- and isoform-level expression using QTLtools[18] and found that most (if not all) heritable genes and isoforms harbor a significant eQTL (**Figure S4**).

Using QTLtools, we identified 8,981 genes and 17,174 isoforms that harbor *cis*-eQTL at FDR < 0.05. The top associated SNP or index eQTL explained about 70% of variance from h^2_{cis} (**Figure S5**). Conditional analyses using QTLtools found that most genes and isoforms have a single *cis*-eQTL signal, suggesting sparse and large genetic effects among *cis*-SNPs. We note that the number of independent *cis*-eQTL will likely increase with an increase in sample size, but in the present study, only 25% of genes and 15% of isoforms harbored more than one significant eQTL.

Next, we sought to investigate the polygenicity of *trans*-SNPs by partitioning genetic variances among 22 autosomal chromosomes. Since these estimates were noisy due to the limited sample size of PsychENCODE, we fit a penalized model with lasso penalty[22] (Methods), which had an effect of shrinking variance components estimates and hence prioritizing the most important SNP effects. We found that the chromosome a given gene or isoform belongs to almost always had non-zero estimates, indicating substantial *cis*-SNP effects, while the rest of the chromosomes had non-zero estimates in proportion to their number of SNPs, suggesting polygenic effects (**Figure S6**). Altogether, *cis*-SNP effects were sparse but large, and *trans*-SNP effects were polygenic.

Pleiotropy among brain isoform expression

Given their genomic proximity and co-regulation, we hypothesized that isoforms are under shared genetic influences. To address this question, for genes with multiple heritable isoforms at $P < 0.05$, we jointly modeled isoform expression using the multivariate variance components models, which allowed us to estimate h^2_{SNP} and r_g together. Of note, 3,801 genes had at least two heritable isoforms, and 1,743 genes had at least three heritable isoforms at $P < 0.05$. Thus, our multivariate models included between two to as many as 23 isoforms. We observed concordant variance components (**Figure 2b**) and h^2_{SNP} estimates (**Figure 2c**) between the univariate and multivariate models, although we note that h^2_{trans} and h^2_{SNP} estimates were slightly deflated for multivariate models compared to univariate models. The distribution of $r_{g,cis}$ was bimodal with the two extremes (**Figure 2d**) and a median estimate of 0.31, suggesting that *cis*-SNPs tend to affect or co-regulate nearby isoforms together. The negative $r_{g,cis}$ estimates close to -1 are characteristic of isoform switching events. Meanwhile, $r_{g,trans}$ and r_e estimates followed unimodal distributions with their medians shifted right from zero (**Figure 2d**). This makes biological sense in that the distal genetic regulators and other biological factors tend to influence transcription of nearby isoforms together.

Next, we sought to understand if fitting pairwise bivariate or multivariate models lead to any meaningful differences in r_g estimates. This question is relevant for not only molecular readouts such as isoform expression, but also for complex traits and diseases in general, because most of the genetic correlation estimates are based on fitting pairwise bivariate models[3,5,6,23]. As for isoform expression, covariance terms in variance components estimates were concordant (**Figure 2e**), while r_g estimates were slightly inflated for pairwise bivariate models compared to multivariate models (**Figure 2f**). Further, we observed numerous cases where the signs of r_g estimates flipped between these two models. Based on these findings and previous observations that pairwise bivariate models tend to yield estimates that do not respect the positive semidefinite constraint of variance components parameters[8], we construe that multivariate estimates are generally more accurate than pairwise bivariate estimates.

Previous work demonstrated that genetic correlation mimics phenotypic correlation in the direction of effect[24-30], and thus motivating the use of phenotypic correlation as a proxy for genetic correlation. We sought to test this hypothesis for isoform expression and we observed that genetic correlation (both $r_{g,cis}$ and $r_{g,trans}$) generally recapitulates phenotypic correlation (or co-expression) in the direction of effect with $r_{g,cis}$ being overall larger in magnitude than phenotypic correlation (**Figure 2g**). This is consistent with previous studies and our own observation that the isoforms of a given gene are under substantial shared genetic influences, particularly among *cis*-SNPs (**Figure 2d**).

ATP9B gene as a case study for h^2_{SNP} and r_g analyses

We now focus on *ATP9B* gene to contextualize our h^2_{SNP} and r_g results (**Figure 3**). In Gencode v19, *ATP9B* had 27 annotated isoforms, 17 of which were determined to be brain-expressed in PsychENCODE based on the criteria TPM > 0.1 in at least 25% of samples[12]. Nine of these isoforms were found heritable at $P < 0.05$ after fitting the univariate variance components model (**Figure 3a**). Pairwise bivariate and multivariate models were fit for the nine heritable isoforms, which resulted in concordant h^2_{SNP} estimates with similar magnitudes (**Figure 3b**). Standard errors for h^2_{SNP} estimates were comparable across univariate, pairwise bivariate, and multivariate models, while standard errors for h^2_{trans} were consistently larger than those of h^2_{cis}. Because we modeled nine isoforms, we note that there were eight pairwise bivariate estimates for each h^2_{SNP} parameter (**Figure 3b**). For the r_g parameter, pairwise bivariate and multivariate models yielded estimates that were concordant in direction of effect, but less so in magnitude. That is, pairwise bivariate estimates were relatively inflated compared to multivariate estimates for both $r_{g,cis}$ and $r_{g,trans}$ (**Figure 3c**). Meanwhile, for both pairwise bivariate and multivariate models, $r_{g,cis}$ estimates were substantially larger than $r_{g,trans}$ estimates, suggesting that there were shared genetic influences, particularly among *cis*-SNPs. Not to mention, $r_{g,cis}$ and $r_{g,trans}$ estimates were generally larger than r_e estimates. Further comparison of r_g and r_e estimates with r_p estimates revealed that phenotypic correlation resembled residual correlation most closely (**Figure 3d**), indicating that for this particular gene *ATP9B*, residual (e.g. non-genetic biological and technological) factors overall exert larger effects than genetic factors (i.e. *cis*- and *trans*-SNPs). We note that one might be confused by this observation of high $r_{g,cis}$ estimates but little resemblance between $r_{g,cis}$ and r_p estimates. This is possible because r_g and r_e estimates are on a standardized scale, and hence the magnitude of residual variance-covariance can be larger than genetic variance-covariance while r_e is smaller than r_g[30]. Finally, upon visual inspection of eQTL results, we found that there were distinct isoform-level eQTL signals (**Figure 3e**). For *ATP9B*, gene-level expression was also heritable and harbored strong eQTL signals (**Figure 3e**) unlike in subsequent examples.

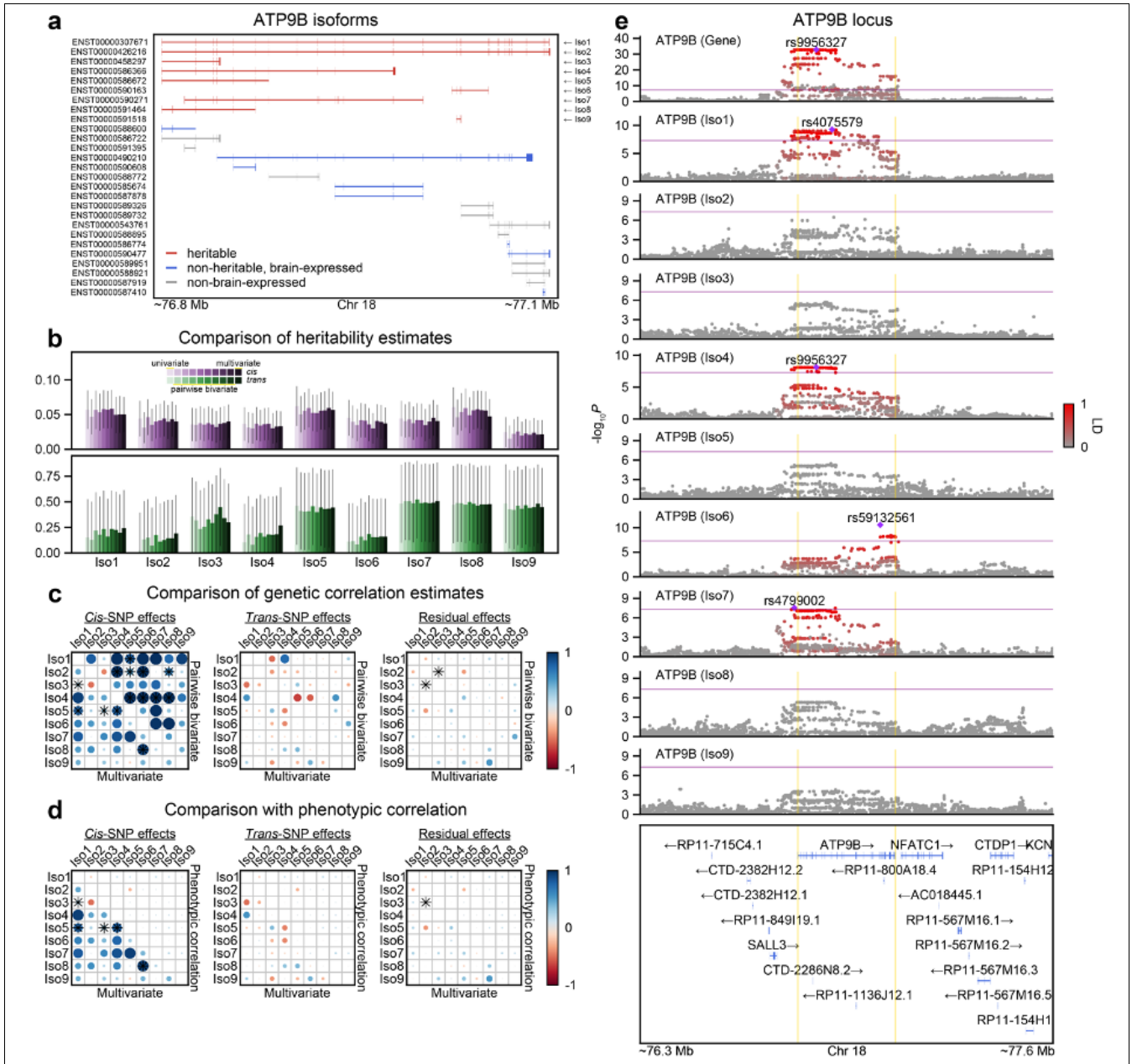


Figure 3: *ATP9B* as an example gene in h^2_{SNP}, r_g, and *cis*-eQTL analyses.

a, Shown are *ATP9B* isoforms that are present in Gencode v19. Nine isoforms that are found heritable in a univariate variance components model with a variation of the likelihood ratio test (LRT)[17] are highlighted in red, and eight isoforms that are determined to be brain-expressed[12] but not heritable are highlighted in blue. **b**, h^2_{cis} and h^2_{trans} estimates are shown in top and bottom rows, respectively. Colors encode estimates from univariate, pairwise bivariate, and multivariate variance components models. Because there are nine isoforms that are modeled jointly, there are eight pairwise bivariate estimates. All error bars denote ± one standard errors calculated from the Fisher information matrix. **c**, Shown are $r_{g,cis}$, $r_{g,trans}$, r_e estimates from pairwise bivariate and multivariate variance components models for *ATP9B*. The lower triangular elements represent multivariate estimates, while the upper triangular elements represent pairwise bivariate estimates. Asterisks denote significance from the Wald test at $P < 0.05$. **d**, Alike panel **c**, except that the upper triangular elements now represent phenotypic correlations (r_p). The same r_p values are plotted. **e**, LocusZoom plot for *ATP9B* gene and isoform expression. Top row shows gene-level results. Index SNPs for features passing $P = 5 \times 10^{-8}$ threshold are shown and corresponding LD between other SNPs are displayed with the intensity of red color. Purple line denotes significance of $P = 5 \times 10^{-8}$, and yellow lines denote gene start and end sites for *ATP9B* gene. LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel.

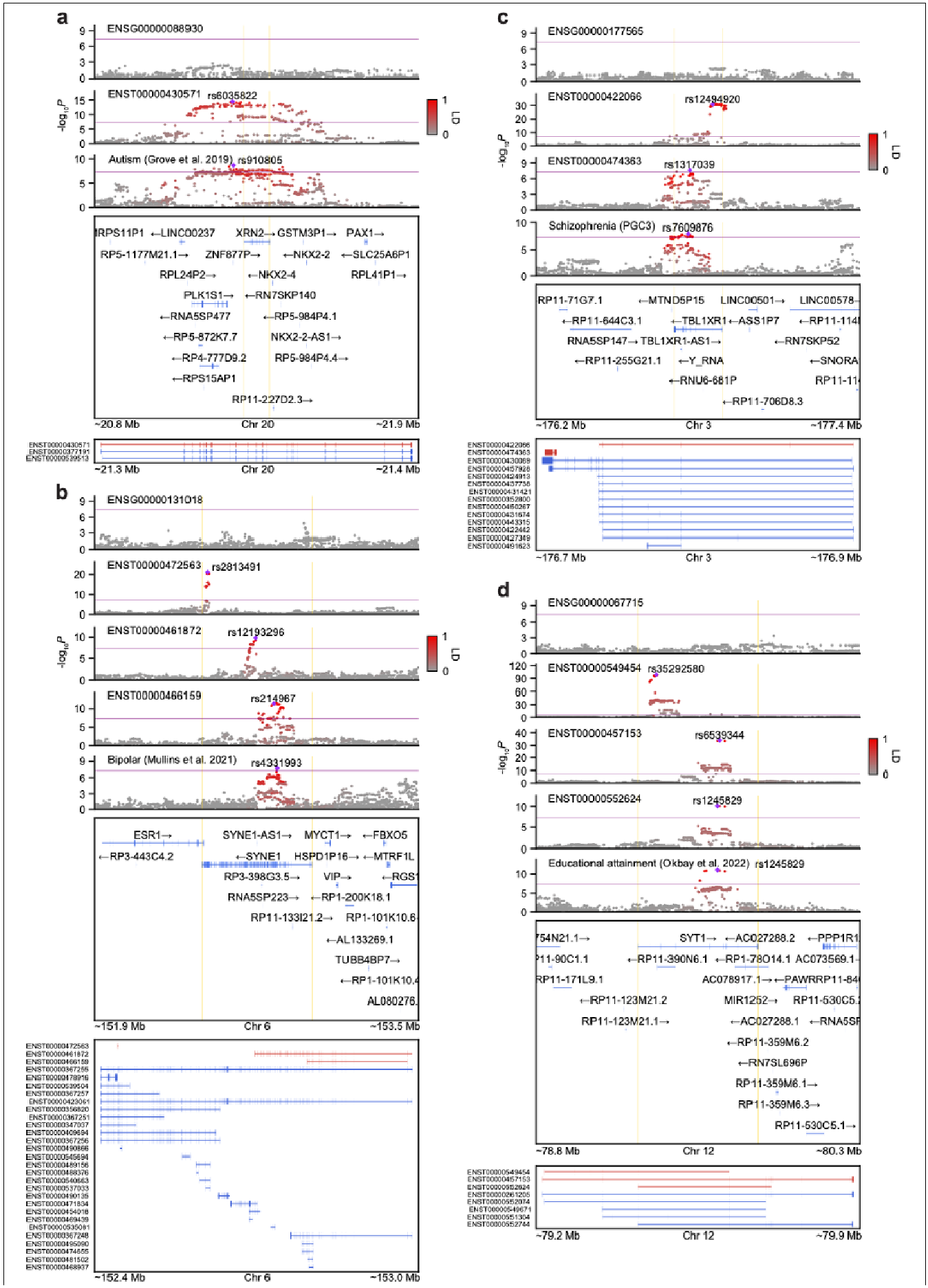

a
ENSG00000088930
ENST00000430571 rs6035822
Autism (Grove et al. 2019) rs910805
-log10P
LD
IRPS11P1 ←LINC00237 XRN2→ GSTM3P1→ PAX1→
RP5-1177M21.1→ ZNF877P→ ←NKX2-2 ←SLC25A6P1
RPL24P2→ ←NKX2-4 RPL41P1→
PLK1S1→ ←RN7SKP140
←RNA5SP477 ←RP5-984P4.1
←RP5-872K7.7 NKX2-2-AS1→
←RP4-777D9.2 RP5-984P4.4→
←RPS15AP1
RP11-227D2.3→
~20.8 Mb Chr 20 ~21.9 Mb
~21.3 Mb Chr 20 ~21.4 Mb
b
ENSG00000131018
ENST00000472563 rs2813491
ENST00000461872 rs12193296
ENST00000466159 rs214967
Bipolar (Mullins et al. 2021) rs4331993
ESR1→ SYNE1-AS1→ MYCT1→ ←FBXO5
←RP3-443C4.2 ←SYNE1 HSPD1P16→ ←MTRF1L
RP3-398G3.5→ VIP→ ←RGS1
RNA5SP223→ ←RP1-200K18.1
RP11-133I21.2→ RP1-101K10.6
←AL133269.1
TUBB4BP7→
←RP1-101K10.4
AL080276.
~151.9 Mb Chr 6 ~153.5 Mb
~152.4 Mb Chr 6 ~153.0 Mb
c
ENSG00000177565
ENST00000422066 rs12494920
ENST00000474363 rs1317039
Schizophrenia (PGC3) rs7609876
RP11-71G7.1 ←MTND5P15 LINC00501→ LINC00578
←RP11-644C3.1 ←TBL1XR1 ←ASS1P7 ←RP11-114
RNA5SP147→ TBL1XR1-AS1→ ←RN7SKP52
←RP11-255G21.1 ←Y_RNA ←SNORA
←RNU6-681P RP11-11
←RP11-706D8.3
~176.2 Mb Chr 3 ~177.4 Mb
~176.7 Mb Chr 3 ~176.9 Mb
d
ENSG00000067715
ENST00000549454 rs35292580
ENST00000457153 rs6539344
ENST00000552624 rs1245829
Educational attainment (Okbay et al. 2022) rs1245829
754N21.1→ SYT1→ ←AC027288.2 ←PPP1R1
11-90C1.1 ←RP11-390N6.1 ←RP1-78O14.1 AC073569.1
RP11-171L9.1 AC078917.1→ ←PAWRRP11-84
←RP11-123M21.2 MIR1252→ RP11-530C5.
RP11-123M21.1→ ←AC027288.1 ←RNA5SP
←RP11-359M6.2
←RN7SL696P
RP11-359M6.1→
RP11-359M6.3→
RP11-530C5.1→
~78.8 Mb Chr 12 ~80.3 Mb
~79.2 Mb Chr 12 ~79.9 Mb

Figure 4: Isoform-level eQTL signals prioritize candidate causal genes in established GWAS loci.
Shown are LocusZoom plots for **a**, *ATP9B*, **b**, *SYNE1*, **c**, *TBL1XR1*, and **d**, *SYT1* genes along with several complex phenotypes. Select isoforms that are included in LocusZoom plots are highlighted in red, while all other isoforms that are brain-expressed are highlighted in blue. Due to space constraints, isoforms that are present in Gencode v19 but not found brain-expressed are omitted. Index SNPs for features passing $P = 5 \times 10^{-8}$ threshold are shown and corresponding LD between other SNPs are displayed with the intensity of red color. Purple line denotes significance of $P = 5 \times 10^{-8}$, and yellow lines denote gene start and end sites. LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel.

Isoform-level eQTL signals prioritize candidate causal genes in GWAS loci

We next sought to understand whether isoform-resolution analyses could help prioritize disease genes in GWAS loci. Here, as a proof of concept, we nominate and share four such cases. The first is *XRN2* gene, which resides in one of two GWAS loci for autism spectrum disorder (ASD)[14]. Of note, the other ASD GWAS locus is a highly pleiotropic genomic region characterized by long-range, complex LD[31] (**Figure S7**), which complicates fine-mapping and gene prioritization efforts. For *XRN2*, there were three brain-expressed isoforms, one of which was significantly heritable (ENST00000430571; $h^2_{cis} = 0.07$, $h^2_{trans} = 0.46$, $P = 5.3 \times 10^{-13}$). eQTL signal of this isoform colocalized[32] strongly with the ASD GWAS signal (PP4 = 0.78; **Figure 4a**; **Figures S8-9**), suggesting that this isoform might be the causal isoform underlying the ASD GWAS signal. The second example is *SYNE1* gene, a SFARI ASD risk gene, of which five isoforms were heritable (**Figure S10**). Three of these isoforms harbored strong and distinct eQTL signals, one of which colocalized with a well-known bipolar disorder (BD) GWAS signal[33] (PP4 = 0.90; **Figure 4b**; **Figures S11-12**). The third example is *TBL1XR1* gene, another SFARI gene, for which loss-of-function is implicated in increased risk for ASD[34] and developmental delay disorder (DDD)[35]. Two *TBL1XR1* isoforms were found heritable, and they harbored distinct isoform-level eQTL signals, one of which colocalized with a SCZ GWAS signal[36] (PP4 = 0.44; **Figure 4c**; **Figure S13-14**). The last example is *SYT1* gene, which is another SFARI gene and a high-confidence DDD risk gene[35]. Three *SYT1* isoforms were found heritable (**Figure S15**), two of which colocalized with an educational attainment (EA) GWAS signal[37] (PP4 = 1.0; **Figure 4d**; **Figures S16-17**).

Given the strong implication of these example genes in neurodevelopmental disorders and their proximity to corresponding GWAS signals, these genes are likely the true causal genes driving the GWAS signals. Relatively straightforward patterns of linkage disequilibrium (LD) and reasonable gene densities also help in reaching this conclusion (**Figure 4**). It is important to note that all these example genes were missing gene-level eQTL signals, while harboring isoform-level eQTL signals, which highlights the potential of isoform-level analyses. Due to incomplete brain isoform annotations, the isoforms we identify may not be the true causal isoforms driving the GWAS signals, but with much more refined and complete transcriptome annotations along with more accurate quantifications of isoform-level expression (e.g. from long-read RNA-seq), we hypothesize that we would gain more fine-grained resolution in both gene- and isoform-level expression and thereby uncover more disease genes in GWAS loci.

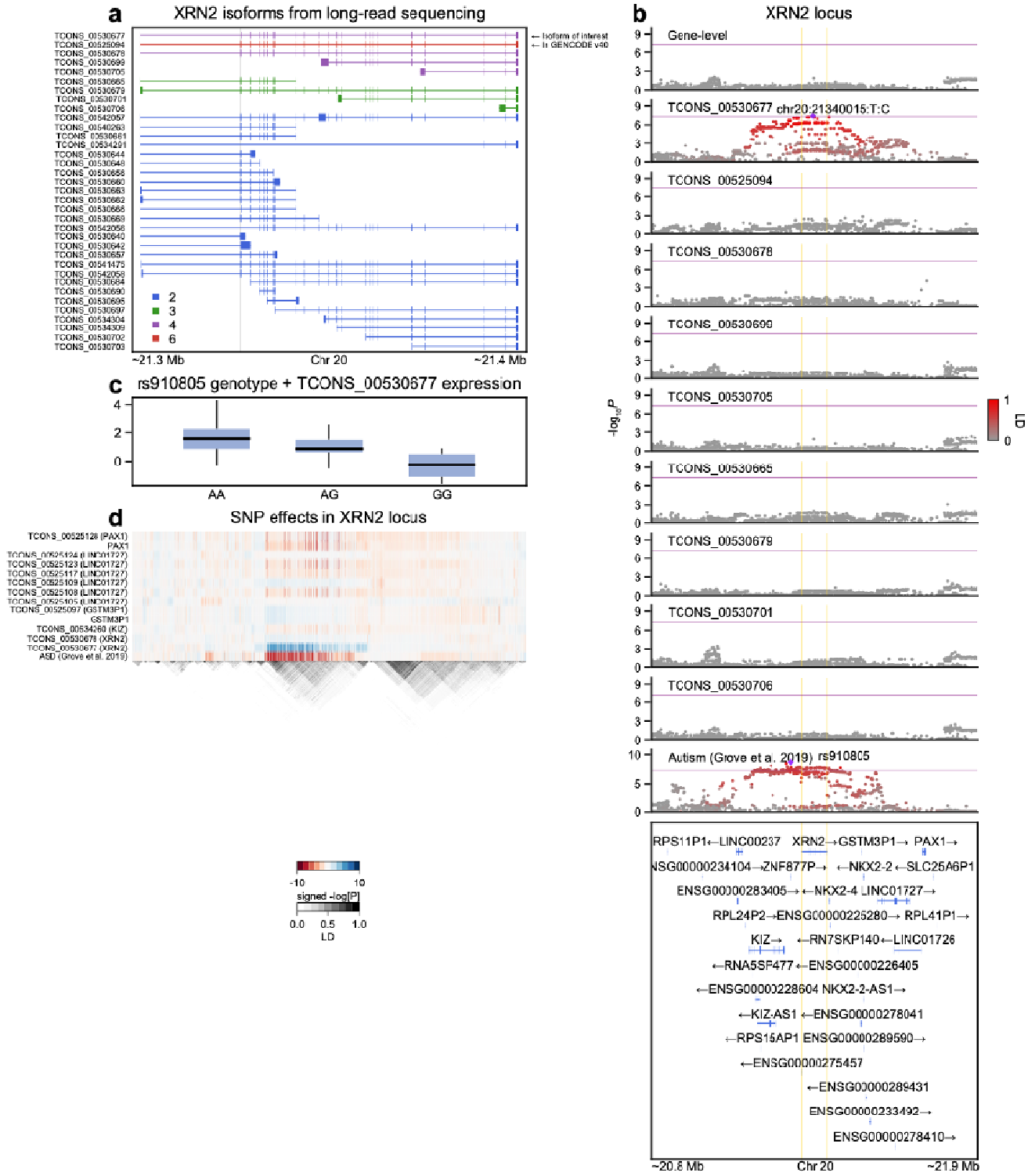


Figure 5: Fine-mapping of top ASD GWAS locus with isoform-level eQTL signals and improved isoform annotations in an independent fetal human brain dataset.

a, Shown are high-confidence, credible *XRN2* isoforms compiled from six existing long-read studies[41-43]. Isoforms are colored with respect to the number of studies each isoform is found in. TCONS_00530677, TCONS_00525094, and TCONS_00541475 correspond to ENST00000430571, ENST00000377191, and ENST00000539513 in Gencode v19, respectively. The skipped exon 2 in TCONS_00530677 is shaded in grey. We subsequently filtered for isoforms found in at least three different studies and updated the transcriptome accordingly. **b**, Shown is LocusZoom plot for *XRN2* gene and its isoform expression based on the updated transcriptome annotation. Isoform expression was quantified in fetal human brain samples[47]. **c**, Index SNP for ASD GWAS (rs910805) and its risk increasing allele G was associated with reduced expression of TCONS_00530677. Note that G is also the minor allele. **d**, Shown is a heatmap of eQTL signals for both gene- and isoform-level expression for features within ±0.5 Mb window of (collapsed) gene start and end sites for *XRN2* gene. Although eQTL analyses were conducted for a total of 30 genes within this region and all their isoforms, we only plot the results of features with minimum *P* values less than 10^{-4} for visual clarity.

which are the same individuals used for eQTL analyses. These analyses were conducted using the GRCh38 human genome build, which leads to slightly different genomic coordinates relative to previous figures that were based on the hg19 reference genome.

Replication of *XRN2* isoform-level eQTL signals in the developing human brain

As another proof of concept, we next sought to definitively fine-map the *XRN2* locus for ASD with improved isoform annotations. This was motivated by relatively simple gene structure of *XRN2*, for example, compared to that of *SYNE1* (**Figure 4b**). Further, accurate quantification of isoform expression in short-read RNA-seq data is challenging, given the presence of multi-mapped reads. The state-of-the-art isoform quantification methods[20,38,39] address multi-mapped reads by probabilistically assigning them via maximizing the log-likelihood of the underlying statistical model with the EM algorithm. Unfortunately, this approach is valid only when all isoforms of a given transcriptome are known[40]. For example, when the transcriptome annotation is incomplete, multi-mapped reads that belong to an "unknown isoform" may be erroneously assigned to different isoforms that are annotated, thereby biasing these estimates. Accordingly, more complete transcriptome annotations are necessary to increase the accuracy of isoform quantification.

In Gencode v19, there were three isoforms for *XRN2*, all of which were determined to be brain-expressed. Upon close inspection of *XRN2* exons, we observed that the only difference between the ASD-associated isoform ENST00000430571 and the canonical isoform ENST00000377191 was skipped exon 2 (**Figure 4a**). On the contrary, the difference between the remaining isoform ENST00000539513 and the canonical isoform was a different exon 1 and the rest of the sequence was the same. Interestingly, we then noticed that all isoforms except the canonical isoform had been removed in subsequent Gencode versions past v19 due to low transcript support level. To ensure that our isoform is expressed and to improve the transcriptome annotation of the human brain, we combined the latest Gencode v40 annotation with existing long-read RNA-seq data from six studies[41-43] (Methods). The data for these studies were mostly generated from fetal and adult human brain samples. To compile a list of high-confidence isoforms, we filtered for isoforms that were found in at least two or three different sources. Remarkably, we recovered all three *XRN2* isoforms in Gencode v19 with ENST00000377191, ENST00000430571, and ENST00000539513 found in six, four, and two studies, respectively (**Figure 5a**).

Given that ASD genetic risk factors are known to converge in cell-types and biological pathways during brain development[44-46], we then sought to replicate our adult brain isoform-level eQTL findings in fetal brain samples[47]. The previous work[47] performed isoform-level eQTL analyses with Gencode v23 and did not detect any significant association for *XRN2*, presumably because ENST00000377191 is the only isoform present in that version of Gencode. In contrast, when we re-quantified fetal brain RNA-seq data, using Salmon[39] with the updated transcriptome annotation, and repeated isoform-level *cis*-eQTL analyses, we observed a strong and specific eQTL signal for ENST00000430571 (or TCONS_00530677 equivalently) (**Figure 5b**). Further, the index SNP for ASD GWAS result (rs910805) and its risk increasing (minor) allele G was associated with reduced expression of ENST00000430571 (**Figure 5c**), which direction of effect is consistent with the adult brain findings (**Figure 4a**). As another sanity check, we tested all genes within ±0.5 Mb window of *XRN2* and their constituent isoforms from the updated transcriptome annotation and observed that no other gene- or isoform-level expression harbored eQTL signals that colocalized with the ASD GWAS signal (**Figure 5d**). Altogether, we replicated our adult brain *XRN2* findings with more complete isoform annotations in an independent fetal brain dataset and confidently narrowed down expression changes in ENST00000430571 as the causal signal for the ASD GWAS signal.

The *XRN2* isoform ENST00000430571 has exon 2 skipped, which in theory should be uncovered by annotation-free methods that detect local splicing patterns[48]. Indeed, the intron that corresponds to the end of exon 1 and the start of exon 3 were shown to harbor a significant splicing QTL (sQTL) in previous studies[45,49,50]. However, several other studies failed to detect a significant sQTL[2,51]. We suspect that this discrepancy is due to a combination of low level of expression for ENST00000430571 and differences in RNA-seq library preparation across studies with polyA selection methods being more susceptible to 3' bias than rRNA depletion methods. Indeed, changes in splicing patterns between rs910805 genotypes were subtle (**Figure S18**), which could be overshadowed by polyA selection methods that induce 3' bias.

Discussion

In this study, we introduce a statistical method for fitting multivariate variance components linear mixed models that allows for arbitrary number of traits and variance components[10]. This method permits more flexible multi-trait analyses. We note that our method is not scalable to biobank-level data, but it is competitive on large data problems, and we recommend applying this method for datasets of size up to $n \times d = 50000$, where n is the number of samples and d is the number of phenotypes. By leveraging this method and the functional genomic dataset from PsychENCODE[12,13], we jointly model isoform expression—for the first time—and dissect the genetic influences (i.e. h^2_{SNP} and r_g) on human brain gene and isoform expression. We find a substantial proportion of the human brain transcriptome to be under genetic control. We find *cis*-SNP effects to be sparse and large individually, while *trans*-SNP effects to be polygenic and large in aggregate. We conclude that the isoforms of a given gene are under shared genetic influences, particularly among *cis*-SNPs. By comparing pairwise bivariate to multivariate models, we empirically show that multivariate models yield more accurate estimates. Finally, several genes are found heritable only at the isoform-level and their genetic signals colocalize strongly with GWAS signals, suggesting that isoform expression changes might be the underlying drivers of these GWAS signals. One notable example is *XRN2*, a specific isoform of which reduced expression appears to increase ASD risk. Overall, the present study comprehensively estimates genetic parameters for brain gene and isoform expression and highlights the utility of conducting isoform-resolution analyses.

We first note a major and important caveat of isoform-level analyses: the underlying generative model for the state-of-the-art short-read RNA-seq quantification tools[20,38-40], start with an assumption that all expressed isoforms of a given transcriptome are known. But the human transcriptome annotation is far from complete, especially for the brain, which can result in inaccurate isoform quantifications and hence inaccurate downstream analyses. However, this also means that with more complete transcriptome annotations, we can obtain more accurate estimates of isoform-level expression even with short-read RNA-seq data[40]. Obviously, it is impossible to account for all possible (known and unknown) sample- and isoform-specific biases when resolving multi-mapped short-read RNA-seq reads and estimating isoform-level expression, but the current methods[20,38,39] offer a reasonable starting point for running isoform-level analyses in conjunction with traditional gene-level analyses. Besides, isoforms are the fundamental biological units expressed in cells and in contrast to usual sQTL analyses[2,45,49-51], isoform-level analyses are more interpretable in that they can pinpoint specific isoforms that are influenced by genetic variation. With extensive efforts to comprehensively catalog all brain-expressed isoforms[41-43,52,53], we foresee that the accuracy of isoform-level quantification will improve for even short-read RNA-seq data.

The current study is limited by its sample size. This is evident by relatively large standard errors in our estimates, particularly for *trans*-SNP effects[21]. Hence with larger sample sizes, we expect to uncover many more genes and isoforms that are genetically regulated. We also note that the PsychENCODE dataset is an output of mega-analysis of six different studies, which can lead to decreased signal-to-noise ratio from heterogeneity and various technical artifacts. We envision increased genetic discovery with more homogenous genotype and RNA-seq data that have sufficient sample sizes. Moreover, the choice of RNA-seq normalization methods can have an impact on h^2_{SNP} and r_g estimates. The PsychENCODE expression data was processed in $\log_2$-CPM-TMM scale, which better accounts for differences in library composition between samples in a large mega-analysis dataset such as PsychENCODE, but complicates the relationship between gene and isoform-level expression, including h^2_{SNP} estimates (Methods). In the main text, we only highlight a handful of genes of which isoform expression changes seem to drive GWAS signals such as *XRN2*, *SYNE1*, *TBL1XR1*, and *SYT1*. Whether isoform-level genetic signals are more broadly enriched for disease risk and how many additional putative disease genes we can prioritize from isoform-resolution analyses remains to be further investigated.

Lastly, the *XRN2* gene encodes an RNA-binding protein that possesses 5'-3' exoribonuclease activity. It is known to promote termination of transcription by degrading RNA to resolve R-loops. *XRN2* is mildly constrained (pLI = 0.35, LOEUF = 0.35), shows peak expression in the fetal brain[54], and is enriched in neuronal cell-types[13,55]. Interestingly, visual inspection of phenome-scale LocusZoom plots reveals that there are two distinct GWAS signals within the *XRN2* locus, one for ASD and the other for blood cell-related phenotypes (**Figure S19**). Note that the same GWAS signal for ASD reached genome-wide significance in the latest GWAS for ADHD. We find that at the gene-level, *XRN2* is not heritable and does not harbor a significant eQTL, but at the isoform-level, one specific *XRN2* isoform is heritable and harbors a significant eQTL that colocalizes with the ASD GWAS signal. As a result of the skipping of the second exon, this isoform is missing 76 amino acids in the 5' end compared to the canonical *XRN2* isoform. Structural and

biological consequences of such changes in protein sequences need to be studied in the future. We note that the prioritized and canonical isoforms belong in different brain WGCNA modules (isoM23 and isoM35)[12], which capture known neurobiological pathways and cell-types, potentially suggesting that the two isoforms might be involved in different biological pathways in the human brain.

Acknowledgments

This work was supported by the National Institute of Mental Health (R01MH121521 to M.J.G.; R01MH123922 to M.J.G.; P50HD103557 to M.J.G.; T32MH073526 to M.K.; F30MH125523 to M.K.) and the UCLA Medical Scientist Training Program (T32GM008042 to M.K.). Data were generated as part of the PsychENCODE Consortium, supported by: U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, P50MH106934, U01MH116488, U01MH116487, U01MH116492, U01MH116489, U01MH116438, U01MH116441, U01MH116442, R01MH114911, R01MH114899, R01MH114901, R01MH117293, R01MH117291, R01MH117292 awarded to: Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind (University of California, Los Angeles), Fernando Goes (Johns Hopkins University), Thomas Hyde (Lieber Institute for Brain Development), Andrew Jaffe (Lieber Institute for Brain Development), James Knowles (University of Southern California), Chunyu Liu (SUNY Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan (University of North Carolina), Flora Vaccarino (Yale University), Daniel Weinberger (Lieber Institute for Brain Development), Sherman Weissman (Yale University), Kevin White (University of Chicago), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns Hopkins University). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Author contributions

M.K. conceived the study. M.K. and H.Z. developed the software. M.K. and D.D.V. performed primary analyses. M.K. conducted data visualization. M.K. interpreted the results and wrote the manuscript with additional input from M.J.G. C.T.J., C.W., A.P. collected and managed long-read and fetal brain data. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569 (2010).
2. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330 (2020).
3. Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* 45, 984–994 (2013).
4. Lukowski, S. W. *et al.* Genetic correlations reveal the shared genetic architecture of transcription in human peripheral blood. *Nat. Commun.* 8, 483 (2017).
5. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82 (2011).
6. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28, 2540–2542 (2012).
7. Lee, S. H. & van der Werf, J. H. J. MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* 32, 1420–1422 (2016).
8. de Vlaming, R. *et al.* Multivariate analysis reveals shared genetic architecture of brain morphology and human behavior. *Commun Biol* 4, 1180 (2021).
9. Hunter, D. R. & Lange, K. A Tutorial on MM Algorithms. *Am. Stat.* 58, 30–37 (2004).
10. Zhou, H., Hu, L., Zhou, J. & Lange, K. MM Algorithms For Variance Components Models. *J. Comput. Graph. Stat.* 28, 350–361 (2019).
11. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.* 59, 65–98 (2017).
12. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 362, eaat8127 (2018).
13. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362, (2018).
14. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* 51, 431–444 (2019).
15. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835 (2011).
16. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824 (2012).
17. Molenberghs, G. & Verbeke, G. Likelihood Ratio, Score, and Wald Tests in a Constrained Parameter Space. *Am. Stat.* 61, 22–27 (2007).
18. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* 8, 15452 (2017).
19. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res.* 49, D916–D923 (2021).
20. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323 (2011).
21. Visscher, P. M. *et al.* Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 10, e1004269 (2014).
22. Kim, J. *et al.* VCSEL: PRIORITIZING SNP-SET BY PENALIZED VARIANCE COMPONENT SELECTION. *Ann. Appl. Stat.* 15, 1652–1672 (2021).
23. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47, 1236–1241 (2015).
24. Searle, S. R. Phenotypic, Genetic and Environmental Correlations. *Biometrics* 17, 474–480 (1961).
25. Cheverud, J. M. A Comparison of Genetic and Phenotypic Correlations. *Evolution* 42, 958–968 (1988).
26. Waitt, D. E. & Levin, D. A. Genetic and phenotypic correlations in plants: a botanical test of Cheverud's conjecture. *Heredity* 80, 310–319 (1998).
27. Bitner-Mathé, B. C. & Klaczko, L. B. Heritability, phenotypic and genetic correlations of size and shape of Drosophila mediopunctata wings. *Heredity* 83 (Pt 6), 688–696 (1999).
28. Vattikuti, S., Guo, J. & Chow, C. C. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet.* 8, e1002637 (2012).
29. Sodini, S. M., Kemper, K. E., Wray, N. R. & Trzaskowski, M. Comparison of Genotypic and Phenotypic Correlations: Cheverud's Conjecture in Humans. *Genetics* 209, 941–948 (2018).
30. van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H. & Wray, N. R. Genetic correlations of polygenic disease traits: from theory to practice. *Nat. Rev. Genet.* 20, 567–581 (2019).
31. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* 5,

1564–1573 (2010).
32. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10, e1004383 (2014).
33. Mullins, N. *et al.* Genome-wide association study of over 40,000 bipolar disorder cases provides novel biological insights. *medRxiv* 2020.09.17.20187054 (2020).
34. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* 180, 568–584.e23 (2020).
35. Kaplanis, J. *et al.* Evidence for 28 genetic disorders dicovered by combining healthcare and research data. *Nature* 586, 757–762 (2020).
36. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502–508 (2022).
37. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* 54, 437–449 (2022).
38. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527 (2016).
39. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419 (2017).
40. Sterne-Weiler, T., Weatheritt, R. J., Best, A. J., Ha, K. C. H. & Blencowe, B. J. Efficient and Accurate Quantitative Profiling of Alternative Splicing Patterns of Any Complexity on a Laptop. *Mol. Cell* 72, 187–200.e6 (2018).
41. Leung, S. K. *et al.* Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Rep.* 37, 110022 (2021).
42. Palmer, C. R., Liu, C. S., Romanow, W. J., Lee, M.-H. & Chun, J. Altered cell and RNA isoform diversity in aging Down syndrome brains. *Proc. Natl. Acad. Sci. U. S. A.* 118, (2021).
43. Glinos, D. A. *et al.* Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* 608, 353–359 (2022).
44. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155, 1008–1021 (2013).
45. Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* 179, 750–771.e22 (2019).
46. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997–1007 (2013).
47. O'Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* 19, 194 (2018).
48. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* 50, 151–158 (2018).
49. Aygün, N. *et al.* Brain-trait-associated variants impact cell-type-specific gene regulation during neurogenesis. *Am. J. Hum. Genet.* 108, 1647–1668 (2021).
50. Li, Y. I., Wong, G., Humphrey, J. & Raj, T. Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *Nat. Commun.* 10, 994 (2019).
51. Raj, T. *et al.* Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* 50, 1584–1592 (2018).
52. Flaherty, E. *et al.* Neuronal impact of patient-specific aberrant NRXN1α splicing. *Nat. Genet.* 51, 1679–1690 (2019).
53. Clark, M. B. *et al.* Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene CACNA1C in human brain. *Mol. Psychiatry* 25, 37–47 (2020).
54. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362, (2018).
55. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* 103, 785–801.e8 (2019).
56. Mostafavi, S. *et al.* Normalizing RNA-sequencing data by modeling hidden covariates with prior knowledge. *PLoS One* 8, e68141 (2013).
57. Dahl, A., Guillemot, V., Mefford, J., Aschard, H. & Zaitlen, N. Adjusting for Principal Components of Molecular Phenotypes Induces Replicating False Positives. *Genetics* 211, 1179–1189 (2019).
58. Danisch, S. & Krumbiegel, J. Makie. jl: Flexible high-performance data visualization for Julia. *Journal of Open Source Software* 6, 3349 (2021).
59. Kim, M., Jops, C. T. & Gandal, M. J. GeneticsMakie.jl: A versatile and scalable toolkit for visualizing locus-level genetic and genomic data. *bioRxiv* 2022.04.18.488573 (2022) doi:10.1101/2022.04.18.488573.

Methods

The PsychENCODE genotype dataset

Genotype array and frontal cortex RNA-seq data from Freeze 1 and 2 of PsychENCODE were obtained from www.doi.org/10.7303/syn12080241. This consisted of uniformly processed data from six studies: BipSeq, LIBD_szControl, CMC-HBCC, CommonMind, BrainGVEX, and UCLA-ASD (see Table S1 and Fig. S33 in ref [13]). Genotype data for these individual studies were previously harmonized through phasing and imputation with the Haplotype Reference Consortium (HRC) reference panel. We focused on 860 unique European individuals with matching genotype and frontal cortex RNA-seq data. We started with 5,312,508 HRC imputed SNPs and filtered for SNPs with minor allele frequency (MAF) > 0.01, genotype and individual missingness rate < 0.05, and Hardy-Weinberg equilibrium *P* values > 1e-6. Five pairs of individuals had classic genetic relationship matrix (GRM) values > 0.05 when using all filtered SNPs, while 647 pairs of individuals had GRM values > 0.025. We kept one individual from each of five pairs and only SNPs belonging to autosomal chromosomes, resulting in a total of 855 unrelated European individuals and 4,685,674 SNPs for downstream analyses.

The PsychENCODE RNA-seq dataset

We used post-QC RNA-seq data that were fully processed, filtered, and normalized (see Materials/Methods and Fig. S3 in ref [12]). Of note, RNA-seq reads were previously aligned to the hg19 reference genome with STAR 2.4.2a and gene and isoform-level quantifications calculated using RSEM v1.2.29. Genes and isoforms were filtered to include those with TPM > 0.1 in at least 25% of samples. Gene and isoform expression were separately normalized using TMM normalization in edgeR and $\log_2$-transformed. The same set of known biological and technical covariates were used for mean or fixed effects, which include age, age^2, study, sex, diagnosis, RNA integrity number (RIN), RIN^2, post-mortem interval (PMI), 24 sequencing principal components (PCs), and 5 genetic PCs. RNA-seq data was also restricted to frontal cortex samples from European individuals as well as genes and isoforms belonging to autosomal chromosomes, resulting in a total of 24,905 genes and 93,293 isoforms based on Gencode v19 annotations. The same expression data was used for all downstream analyses unless otherwise stated. We note that we did not correct for latent factors for the main analyses such as hidden covariates with prior (HCP)[56], given that we are modeling *cis*- and *trans*-SNP effects simultaneously and we risk removing trans effects by adjusting for such latent factors[57].

Multiple variance components linear mixed model for univariate response

Given an $n \times 1$ response vector $\mathbf{y}$ and $n \times p$ predictor matrix $\mathbf{X}$, the variance components linear mixed model assumes $\mathbf{y} \sim N(\mathbf{X\beta}, \mathbf{\Omega})$, where

$$\mathbf{\Omega} = \sum_{i=1}^{m} \sigma_i^2 \, \mathbf{V}_i,$$

and $\mathbf{V}_1, \ldots, \mathbf{V}_m$ are m known positive semidefinite matrices. The parameters of the model include mean effects ($\boldsymbol{\beta}$) and variance components ($\sigma_1^2, \ldots, \sigma_m^2$). $\mathbf{\Omega}$ is assumed to be positive definite. The simplest heritability and GWAS model assumes two variance components, where $\mathbf{V}_1$ is a kinship matrix and $\mathbf{V}_2 = \mathbf{I}$. This model is misspecified for gene expression, since SNPs in the vicinity of a gene (i.e. *cis*-SNPs) are known to exert stronger effects on its expression. Therefore, for gene and isoform expression, we specified three variance components, where

$$\mathbf{\Omega} = \sigma_1^2 \mathbf{V}_1 + \sigma_2^2 \mathbf{V}_2 + \sigma_e^2 \mathbf{I},$$

and $\mathbf{V}_1$ and $\mathbf{V}_2$ are empirical kinship matrices constructed from *cis*- and all other SNPs (i.e. *trans*-SNPs), respectively, such that σ_1^2 and σ_2^2 capture aggregate genetic effects of corresponding sets of SNPs. The classic genetic relationship matrix (GRM) was used for both empirical kinship matrices, under which $\frac{1}{n-1}\mathrm{E}[\mathbf{y}^T(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{y}] \approx \frac{n}{n-1}(\sigma_1^2 + \sigma_2^2) + \sigma_e^2 + \frac{1}{n-1}\boldsymbol{\beta}^T\mathbf{X}^T(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{X}\boldsymbol{\beta}$. This equality is approximate, since the diagonal elements of GRM are not exactly one. When there are no mean effects or there is only the intercept term, the overall variance can be decomposed into the sum of variance components as above, so it makes natural sense to define SNP-based heritability as

$$h^2_{\mathrm{SNP}} := \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \sigma_e^2}.$$

Multiple variance components linear mixed model for multivariate response

The multivariate response variance components model assumes an $n \times d$ response matrix $\mathbf{Y}$ with $\text{vec}\,\mathbf{Y} \sim N(\text{vec}(\mathbf{XB}), \mathbf{\Omega})$, where

$$\mathbf{\Omega} = \sum_{i=1}^{m} \mathbf{\Gamma}_i \otimes \mathbf{V}_i.$$

$\text{vec}\,\mathbf{Y}$ creates an $nd \times 1$ vector from $\mathbf{Y}$ by stacking its columns and $\otimes$ denotes the Kronecker product. The parameters of the model include $p \times d$ mean effects $\mathbf{B}$ and $d \times d$ variance components $(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_m)$ that are positive semidefinite. $\mathbf{\Omega}$ is assumed to be positive definite. The univariate response model is subsumed under the multivariate response model when there is only a single phenotype. The simplest genetic correlation model looks at phenotypes pairwise and assumes two variance components. However, this model is likely inadequate for isoform-level expression, so for constituent isoforms of a given gene, we modeled them jointly and specified three variance components such that

$$\mathbf{\Omega} = \mathbf{\Gamma}_1 \otimes \mathbf{V}_1 + \mathbf{\Gamma}_2 \otimes \mathbf{V}_2 + \mathbf{\Gamma}_e \otimes \mathbf{I}.$$

Corresponding to the univariate response model, $\mathbf{\Gamma}_1$ captures genetic variances and covariances among *cis*-SNPs, while $\mathbf{\Gamma}_2$ captures genetic variances and covariances among *trans*-SNPs. Let $(\mathbf{Y})_{\cdot j}$ denote the jth column of $\mathbf{Y}$ and $(\mathbf{Y})_{jk}$ the (j,k)th element of $\mathbf{Y}$. Then $\frac{1}{n-1}\mathrm{E}[(\mathbf{Y})_{\cdot j}^T(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)(\mathbf{Y})_{\cdot k}] \approx \frac{n}{n-1}[(\mathbf{\Gamma}_1)_{jk} + (\mathbf{\Gamma}_2)_{jk}] + (\mathbf{\Gamma}_e)_{jk} + \frac{1}{n-1}(\mathbf{B})_{\cdot j}^T\mathbf{X}^T(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{X}(\mathbf{B})_{\cdot k}$. In other words, just as the overall variance could be decomposed into the sum of genetic variances and residual variance, the overall covariance can be decomposed into the sum of genetic covariances and residual covariance following this model. Genetic correlation is defined as genetic covariance divided by the product of square root of corresponding genetic variances. For example, genetic correlation for j,kth isoforms among *cis*-SNPs is $\frac{(\mathbf{\Gamma}_1)_{jk}}{\sqrt{(\mathbf{\Gamma}_1)_{jj}(\mathbf{\Gamma}_1)_{kk}}}$. It is worthwhile to note that the magnitude of genetic covariance or genetic correlation can differ between *cis*- and *trans*-SNPs.

Residual variance could be further decomposed into $\mathbf{\Gamma}_e = \mathbf{\Gamma}_\tau + \mathbf{\Gamma}_\epsilon$, where $\mathbf{\Gamma}_\tau$ captures covariance from sampling error and measurement error from shared transcript structures among isoforms, while $\mathbf{\Gamma}_\epsilon$ captures covariance arising from other biological factors. In practice, these two are indistinguishable from one another, but one could make use of bootstrap samples (i.e. also known as technical replicates) from Kallisto[38] or Salmon[39] to estimate $\mathbf{\Gamma}_\tau$. Then we can let $\hat{\mathbf{\Gamma}}_\epsilon = \hat{\mathbf{\Gamma}}_e - \hat{\mathbf{\Gamma}}_\tau$.

Relationship between gene and isoform expression h^2_{SNP}

Gene-level expression in unit of transcripts per million (TPM) is conventionally assumed to be the sum of expression of its constituent isoforms in TPM. Suppose there are d isoforms for a given gene, then

$$\mathbf{Y}\mathbf{1}_d \sim N(\mathbf{XB}\mathbf{1}_d, \sum_{i=1}^{2}(\mathbf{1}_d^T\mathbf{\Gamma}_i\mathbf{1}_d)\mathbf{V}_i + (\mathbf{1}_d^T\mathbf{\Gamma}_e\mathbf{1}_d)\mathbf{I}).$$

As a result, the overall variance in gene expression is approximately equal to $\mathbf{1}_d^T\mathbf{\Gamma}_1\mathbf{1}_d + \mathbf{1}_d^T\mathbf{\Gamma}_2\mathbf{1}_d + \mathbf{1}_d^T\mathbf{\Gamma}_e\mathbf{1}_d$ and gene-level SNP-based heritability becomes

$$h^2_{\text{SNP}} := \frac{\mathbf{1}_d^T\mathbf{\Gamma}_1\mathbf{1}_d + \mathbf{1}_d^T\mathbf{\Gamma}_2\mathbf{1}_d}{\mathbf{1}_d^T\mathbf{\Gamma}_1\mathbf{1}_d + \mathbf{1}_d^T\mathbf{\Gamma}_2\mathbf{1}_d + \mathbf{1}_d^T\mathbf{\Gamma}_e\mathbf{1}_d}.$$

Four scenarios are possible based on the above equation: 1. both gene and isoforms are not heritable, 2. both gene and (a subset of) isoforms are heritable, 3. only gene is heritable, 4. only (a subset of) isoforms are heritable.

In practice, gene and isoform-level expression are $\log_2$-transformed, which leads to the distortion of the above-mentioned relationship, where $\log_2(\sum_{j=1}^{d}\text{TPM}_j) \geq \frac{1}{d}\sum_{j=1}^{d}\log_2(d \cdot \text{TPM}_j) = \log_2 d + \frac{1}{d}\sum_{j=1}^{d}\log_2(\text{TPM}_j)$. The PsychENCODE expression

data was further processed not in TPM but instead in $\log_2$-CPM-TMM, which normalization better accounts for differences in library composition across samples in a large mega-analysis dataset such as PsychENCODE, but complicates the relationship between gene and isoform-level expression.

Estimation in variance components model

We consider both maximum likelihood (ML) and restricted (or residual) maximum likelihood (REML) approaches. REML estimates tend to give less biased estimates of variance components, while ML estimates can still have smaller mean squared error (MSE) and are useful for the likelihood ratio test (LRT). For the univariate response model, the log-likelihood function is

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\det(\boldsymbol{\Omega}) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

where $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_m^2)$. The corresponding score (gradient) vector is

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} \mathcal{L} &= \mathbf{X}^T \boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial}{\partial \sigma_i^2} \mathcal{L} &= -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{V}_i) + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Omega}^{-1}\mathbf{V}_i \boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

The observed information matrix has elements

$$\begin{aligned} -\nabla_{\boldsymbol{\beta}}^2 \mathcal{L} &= \mathbf{X}^T \boldsymbol{\Omega}^{-1}\mathbf{X} \\ -\frac{\partial}{\partial \sigma_i^2 \partial \boldsymbol{\beta}} \mathcal{L} &= \mathbf{X}^T \boldsymbol{\Omega}^{-1}\mathbf{V}_i \boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ -\frac{\partial}{\partial \sigma_i^2 \partial \sigma_j^2} \mathcal{L} &= -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{V}_j \boldsymbol{\Omega}^{-1}\mathbf{V}_i) + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^T)\boldsymbol{\Omega}^{-1}\mathbf{V}_j \boldsymbol{\Omega}^{-1}\mathbf{V}_i \boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

The expected (Fisher) information matrix has elements

$$\begin{aligned} \mathrm{E}\left[-\nabla_{\boldsymbol{\beta}}^2 \mathcal{L}\right] &= \mathbf{X}^T \boldsymbol{\Omega}^{-1}\mathbf{X} \\ \mathrm{E}\left[-\frac{\partial}{\partial \sigma_i^2 \partial \boldsymbol{\beta}} \mathcal{L}\right] &= \mathbf{0} \\ \mathrm{E}\left[-\frac{\partial}{\partial \sigma_i^2 \partial \sigma_j^2} \mathcal{L}\right] &= \frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{V}_j \boldsymbol{\Omega}^{-1}\mathbf{V}_i). \end{aligned}$$

Suppose $\mathrm{rank}(\mathbf{X}) = r$ and columns of $\mathbf{A} \in \mathbb{R}^{n \times (n-r)}$ span the null space of $\mathbf{X}^T$, then the log-likelihood function for REML estimation is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\sigma}^2) &= -\frac{n-r}{2}\ln(2\pi) - \frac{1}{2}\ln\det(\mathbf{A}^T \boldsymbol{\Omega}\mathbf{A}) - \frac{1}{2}\mathbf{y}^T \mathbf{A}(\mathbf{A}^T \boldsymbol{\Omega}\mathbf{A})^{-1}\mathbf{A}^T \mathbf{y} \\ &= -\frac{n-r}{2}\ln(2\pi) - \frac{1}{2}\ln\det(\mathbf{A}^T \boldsymbol{\Omega}\mathbf{A}) - \frac{1}{2}\mathbf{y}^T \mathbf{P}\mathbf{y}, \end{aligned}$$

where $\mathbf{P} = \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{X}(\mathbf{X}^T \boldsymbol{\Omega}^{-1}\mathbf{X})^{-}\mathbf{X}^T \boldsymbol{\Omega}^{-1}$. Due to the uniqueness of $\mathbf{P}$, we can choose the columns of $\mathbf{A}$ to form an orthogonal basis such that $\mathbf{A}^T\mathbf{A} = \mathbf{I}$. The score vector, observed and Fisher information matrices for this residual log-likelihood are similar to the above with $\boldsymbol{\Omega}$ replaced by $\mathbf{A}^T \boldsymbol{\Omega}\mathbf{A}$, $\mathbf{V}_i$ by $\mathbf{A}^T \mathbf{V}_i \mathbf{A}$ for $i = 1, \dots, m$, and $\mathbf{y}$ by $\mathbf{A}^T\mathbf{y}$, and hence

$$\begin{aligned} \frac{\partial}{\partial \sigma_i^2} \mathcal{L} &= -\frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_i) + \frac{1}{2}\mathbf{y}^T \mathbf{P}\mathbf{V}_i \mathbf{P}\mathbf{y} \\ -\frac{\partial}{\partial \sigma_i^2 \partial \sigma_j^2} \mathcal{L} &= -\frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_j \mathbf{P}\mathbf{V}_i) + \mathbf{y}^T \mathbf{P}\mathbf{V}_j \mathbf{P}\mathbf{V}_i \mathbf{P}\mathbf{y} \\ \mathrm{E}\left[-\frac{\partial}{\partial \sigma_i^2 \partial \sigma_j^2} \mathcal{L}\right] &= \frac{1}{2}\mathrm{tr}(\mathbf{P}\mathbf{V}_j \mathbf{P}\mathbf{V}_i). \end{aligned}$$

For the multivariate response model, the log-likelihood function is

$$\mathcal{L}(\mathbf{B},\mathbf{\Gamma}) = \quad -\frac{nd}{2}\ln(2\pi) - \frac{1}{2}\ln\det(\mathbf{\Omega}) - \frac{1}{2}\mathrm{vec}(\mathbf{Y}-\mathbf{XB})^T\mathbf{\Omega}^{-1}\mathrm{vec}(\mathbf{Y}-\mathbf{XB}),$$

where $\mathbf{\Gamma} = (\mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_m)$. To ensure that $\mathbf{\Gamma}_i$ is positive semidefinite, we reparametrize with its Cholesky factor $\mathbf{L}_i$ such that $\mathbf{\Gamma}_i = \mathbf{L}_i\mathbf{L}_i^T$ for $i = 1, \dots, m$. Then the score vector is

$$\begin{aligned}
&\mathrm{D}_{\mathrm{vec}\,\mathbf{B}}\mathcal{L} = \mathrm{vec}(\mathbf{Y}-\mathbf{XB})^T\mathbf{\Omega}^{-1}(\mathbf{I}_d \otimes \mathbf{X})\\
&\mathrm{D}_{\mathrm{vech}\,\mathbf{L}_i}\mathcal{L}\\
&= \frac{1}{2}[\mathrm{vec}(\mathbf{\Omega}^{-1}\mathrm{vec}(\mathbf{Y}-\mathbf{XB})\mathrm{vec}(\mathbf{Y}-\mathbf{XB})^T\mathbf{\Omega}^{-1}) - \mathrm{vec}\,\mathbf{\Omega}^{-1}]^T\\
&\times (\mathbf{I}_d \otimes \mathbf{K}_{nd} \otimes \mathbf{I}_n)([(\mathbf{I}_{d^2} + \mathbf{K}_{dd})(\mathbf{L}_i \otimes \mathbf{I}_d)] \otimes \mathrm{vec}\,\mathbf{V}_i)\mathbf{D}_d,
\end{aligned}$$

where $\mathbf{K}_{nd}$ is the $nd \times nd$ commutation matrix and $\mathbf{D}_d$ the $d^2 \times d(d+1)/2$ duplication matrix.

Let $\mathbf{W}_i = (\mathbf{I}_d \otimes \mathbf{K}_{nd} \otimes \mathbf{I}_n)([(\mathbf{I}_{d^2} + \mathbf{K}_{dd})(\mathbf{L}_i \otimes \mathbf{I}_d)] \otimes \mathrm{vec}\,\mathbf{V}_i)\mathbf{D}_d$ and $\mathbf{R}$ be the $n \times d$ matrix satisfying $\mathrm{vec}\,\mathbf{R} = \mathbf{\Omega}^{-1}\mathrm{vec}(\mathbf{Y}-\mathbf{XB})$. Then the observed information matrix has elements

$$\begin{aligned}
&-\frac{\partial^2}{\partial(\mathrm{vec}\,\mathbf{B})^T\partial(\mathrm{vec}\,\mathbf{B})}\mathcal{L}\\
&= (\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-1}(\mathbf{I}_d \otimes \mathbf{X})\\
&-\frac{\partial^2}{\partial(\mathrm{vech}\,\mathbf{L}_i)^T\partial(\mathrm{vec}\,\mathbf{B})}\mathcal{L}\\
&= (\mathrm{vec}\,\mathbf{R}^T \otimes [(\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-1}])\mathbf{W}_i\\
&-\frac{\partial^2}{\partial(\mathrm{vech}\,\mathbf{L}_j)^T\partial(\mathrm{vech}\,\mathbf{L}_i)}\mathcal{L}\\
&= -\frac{1}{2}\mathbf{W}_i^T[\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1} - (\mathrm{vec}\,\mathbf{R}\mathrm{vec}\,\mathbf{R}^T) \otimes \mathbf{\Omega}^{-1}\\
&-\mathbf{\Omega}^{-1} \otimes (\mathrm{vec}\,\mathbf{R}\mathrm{vec}\,\mathbf{R}^T)]\mathbf{W}_j.
\end{aligned}$$

The expected (Fisher) information matrix has elements

$$\begin{aligned}
\mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vec}\,\mathbf{B})^T\partial(\mathrm{vec}\,\mathbf{B})}\mathcal{L}\right] &= (\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-1}(\mathbf{I}_d \otimes \mathbf{X})\\
\mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vech}\,\mathbf{L}_i)^T\partial(\mathrm{vec}\,\mathbf{B})}\mathcal{L}\right] &= \mathbf{0}\\
\mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vech}\,\mathbf{L}_j)^T\partial(\mathrm{vech}\,\mathbf{L}_i)}\mathcal{L}\right] &= \frac{1}{2}\mathbf{W}_i^T(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1})\mathbf{W}_j.
\end{aligned}$$

As a sanity check, one can observe that the above Fisher information boils down to the Fisher information for the univariate response model when $d = 1$. For REML estimation in the multivariate response model, the log-likelihood function is

$$\mathcal{L}(\mathbf{\Gamma}) = \quad -\frac{(n-r)d}{2}\ln(2\pi) - \frac{1}{2}\ln\det(\widetilde{\mathbf{\Omega}}) - \frac{1}{2}\mathrm{vec}(\mathbf{A}^T\mathbf{Y})^T\widetilde{\mathbf{\Omega}}^{-1}\mathrm{vec}(\mathbf{A}^T\mathbf{Y}),$$

where $\widetilde{\mathbf{\Omega}} = \sum_{i=1}^m \mathbf{\Gamma}_i \otimes (\mathbf{A}^T\mathbf{V}_i\mathbf{A})$. Let $\widetilde{\mathbf{V}}_i = \mathbf{A}^T\mathbf{V}_i\mathbf{A}$ and $\widetilde{\mathbf{W}}_i = (\mathbf{I}_d \otimes \mathbf{K}_{(n-r)d} \otimes \mathbf{I}_{n-r})([(\mathbf{I}_{d^2} + \mathbf{K}_{dd})(\mathbf{L}_i \otimes \mathbf{I}_d)] \otimes \mathrm{vec}\,\widetilde{\mathbf{V}}_i)\mathbf{D}_d$, then the corresponding score vector, observed and Fisher information matrices are

$$
\begin{aligned}
&\mathrm{D}_{\mathrm{vech}\mathbf{L}_i}\mathcal{L} \\
&= \frac{1}{2}\mathrm{vec}(\widetilde{\mathbf{\Omega}}^{-1}[\mathrm{vec}(\mathbf{A}^T\mathbf{Y})\mathrm{vec}(\mathbf{A}^T\mathbf{Y})^T - \widetilde{\mathbf{\Omega}}]\widetilde{\mathbf{\Omega}}^{-1})^T\widetilde{\mathbf{W}}_i \\
&- \frac{\partial^2}{\partial(\mathrm{vech}\mathbf{L}_j)^T\partial(\mathrm{vech}\mathbf{L}_i)}\mathcal{L} \\
&= -\frac{1}{2}\widetilde{\mathbf{W}}_i^T[\widetilde{\mathbf{\Omega}}^{-1} \otimes \widetilde{\mathbf{\Omega}}^{-1} - (\widetilde{\mathbf{\Omega}}^{-1}\mathrm{vec}(\mathbf{A}^T\mathbf{Y})\mathrm{vec}(\mathbf{A}^T\mathbf{Y})^T\widetilde{\mathbf{\Omega}}^{-1}) \otimes \widetilde{\mathbf{\Omega}}^{-1} \\
&\quad -\widetilde{\mathbf{\Omega}}^{-1} \otimes (\widetilde{\mathbf{\Omega}}^{-1}\mathrm{vec}(\mathbf{A}^T\mathbf{Y})\mathrm{vec}(\mathbf{A}^T\mathbf{Y})^T\widetilde{\mathbf{\Omega}}^{-1})]\widetilde{\mathbf{W}}_j \\
&\mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vech}\mathbf{L}_j)^T\partial(\mathrm{vech}\mathbf{L}_i)}\mathcal{L}\right] \\
&= \frac{1}{2}\widetilde{\mathbf{W}}_i^T(\widetilde{\mathbf{\Omega}}^{-1} \otimes \widetilde{\mathbf{\Omega}}^{-1})\widetilde{\mathbf{W}}_j.
\end{aligned}
$$

Variance components parameters are usually estimated using an iterative algorithm such as Fisher scoring and expectation-maximization (EM) algorithms. Fisher scoring uses the expected information matrix derived above instead of the observed information matrix in Newton's method. To derive the EM algorithm, we let $\mathbf{\Omega}_i = \mathbf{\Gamma}_i \otimes \mathbf{V}_i$, $r_i = \mathrm{rank}(\mathbf{V}_i)$, $s_i = \mathrm{rank}(\mathbf{\Gamma}_i)$, and $\det^+(\mathbf{\Gamma}_i)$ denote the pseudo-determinant of $\mathbf{\Gamma}_i$ and $\mathbf{\Gamma}_i^+$ the pseudo-inverse of $\mathbf{\Gamma}_i$. Additionally, if we let $\mathrm{vec}\,\mathbf{R}^{(t)} = \mathbf{\Omega}^{-(t)}\mathrm{vec}(\mathbf{Y} - \mathbf{X}\mathbf{B}^{(t)})$, then Q-function for the EM algorithm in the multivariate response model is

$$
\begin{aligned}
&Q(\mathbf{\Gamma} \mid \mathbf{B}^{(t)}, \mathbf{\Gamma}^{(t)}) \\
&= \sum_{i=1}^{m}\left[-\frac{nd}{2}\ln(2\pi) - \frac{r_i}{2}\ln\det{}^+(\mathbf{\Gamma}_i) - \frac{s_i}{2}\ln\det{}^+(\mathbf{V}_i)\right] \\
&- \sum_{i=1}^{m}\frac{1}{2}\mathrm{vec}\,\mathbf{R}^{(t)^T}(\mathbf{\Gamma}_i^{(t)}\mathbf{\Gamma}_i^+\mathbf{\Gamma}_i^{(t)} \otimes \mathbf{V}_i)\mathrm{vec}\,\mathbf{R}^{(t)} \\
&- \sum_{i=1}^{m}\frac{1}{2}\mathrm{tr}(\mathbf{\Gamma}_i^{(t)}\mathbf{\Gamma}_i^+ \otimes \mathbf{V}_i\mathbf{V}_i^+) \\
&+ \sum_{i=1}^{m}\frac{1}{2}\mathrm{tr}[(\mathbf{\Gamma}_i^{(t)}\mathbf{\Gamma}_i^+\mathbf{\Gamma}_i^{(t)} \otimes \mathbf{V}_i)\mathbf{\Omega}^{-(t)}]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{m}\left[-\frac{nd}{2}\ln(2\pi) - \frac{r_i}{2}\ln\det{}^+(\mathbf{\Gamma}_i) - \frac{s_i}{2}\ln\det{}^+(\mathbf{V}_i)\right] \\
&- \sum_{i=1}^{m}\frac{1}{2}\mathrm{tr}(\mathbf{R}^{(t)^T}\mathbf{V}_i\mathbf{R}^{(t)}\mathbf{\Gamma}_i^{(t)}\mathbf{\Gamma}_i^+\mathbf{\Gamma}_i^{(t)}) - \sum_{i=1}^{m}\frac{r_i}{2}\mathrm{tr}(\mathbf{\Gamma}_i^{(t)}\mathbf{\Gamma}_i^+) \\
&+ \sum_{i=1}^{m}\frac{1}{2}\mathrm{tr}[\left(\mathbf{I}_d \otimes \mathbf{1}_n\right)^T\left[\left(\mathbf{1}_d\mathbf{1}_d^T \otimes \mathbf{V}_i\right) \odot \mathbf{\Omega}^{-(t)}\right](\mathbf{I}_d \otimes \mathbf{1}_n)\mathbf{\Gamma}_i^{(t)}\mathbf{\Gamma}_i^+\mathbf{\Gamma}_i^{(t)}\Big],
\end{aligned}
$$

where $\odot$ denotes the Hadamard (elementwise) product. Hence, the ECM updates are

$$
\begin{aligned}
&\mathrm{vec}\,\mathbf{B}^{(t)} \\
&= [(\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-(t)}(\mathbf{I}_d \otimes \mathbf{X})]^{-1}(\mathbf{I}_d \otimes \mathbf{X}^T)\mathbf{\Omega}^{-(t)}\mathrm{vec}\,\mathbf{Y} \\
&\mathbf{\Gamma}_i^{(t+1)} \\
&= \frac{1}{r_i}\mathbf{\Gamma}_i^{(t)}[\mathbf{R}^{(t)T}\mathbf{V}_i\mathbf{R}^{(t)} - (\mathbf{I}_d \otimes \mathbf{1}_n)^T[(\mathbf{1}_d\mathbf{1}_d^T \otimes \mathbf{V}_i) \odot \mathbf{\Omega}^{-(t)}](\mathbf{I}_d \otimes \mathbf{1}_n)]\mathbf{\Gamma}_i^{(t)} + \mathbf{\Gamma}_i^{(t)}.
\end{aligned}
$$

The EM algorithm is a special case of the minorization-maximization (MM) algorithm[9]. A different formulation of the MM algorithm[10] implements a block ascent strategy by alternatively updating $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$ such that

$$\begin{aligned}\boldsymbol{\beta}^{(t)} &= (\mathbf{X}^T\boldsymbol{\Omega}^{-(t)}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Omega}^{-(t)}\mathbf{y}\\ \sigma_i^{2(t+1)} &= \sigma_i^{2(t)}\sqrt{\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^{(t)})^T\boldsymbol{\Omega}^{-(t)}\mathbf{V}_i\boldsymbol{\Omega}^{-(t)}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}^{(t)})}{\mathrm{tr}(\boldsymbol{\Omega}^{-(t)}\mathbf{V}_i)}}.\end{aligned}$$

For the multivariate response model, the updates are

$$\begin{aligned}&\mathrm{vec}\,\mathbf{B}^{(t)}\\ &= [(\mathbf{I}_d\otimes\mathbf{X}^T)\boldsymbol{\Omega}^{-(t)}(\mathbf{I}_d\otimes\mathbf{X})]^{-1}(\mathbf{I}_d\otimes\mathbf{X}^T)\boldsymbol{\Omega}^{-(t)}\mathrm{vec}\,\mathbf{Y}\\ &\boldsymbol{\Gamma}_i^{(t+1)}\\ &= \mathbf{L}_i^{-(t)T}[\mathbf{L}_i^{(t)T}(\boldsymbol{\Gamma}_i^{(t)}\mathbf{R}^{(t)T}\mathbf{V}_i\mathbf{R}^{(t)}\boldsymbol{\Gamma}_i^{(t)})\mathbf{L}_i^{(t)}]^{1/2}\mathbf{L}_i^{-(t)},\end{aligned}$$

where $\mathbf{L}_i^{(t)}$ is the Cholesky factor of $(\mathbf{I}_d\otimes\mathbf{1}_n)^T[(\mathbf{1}_d\mathbf{1}_d^T\otimes\mathbf{V}_i)\odot\boldsymbol{\Omega}^{-(t)}](\mathbf{I}_d\otimes\mathbf{1}_n)$ and $\mathrm{vec}\,\mathbf{R}^{(t)} = \boldsymbol{\Omega}^{-(t)}\mathrm{vec}(\mathbf{Y}-\mathbf{X}\mathbf{B}^{(t)})$. As another sanity check, one can see that the above updates become the updates for the univariate response model when $d = 1$. For complete derivation, please see ref [10] and for REML estimation, the updates are basically of the same form. Compared to Newton's method or Fisher scoring, the MM algorithm is numerically stable and computationally efficient. Although the number of iterations it takes is usually larger than Fisher scoring, the computational burden is smaller in each iteration[10]. The MM algorithm also gracefully respects the nonnegativity (or positive semidefiniteness) constraint of variance components. Compared to the EM algorithm, the MM updates converge more quickly, since the surrogate function hugs the log-likelihood function more tightly[10]. In the simplest heritability and GWAS model where there are two variance components, repeated matrix inversion can be avoided by the generalized eigenvalue decomposition of the two kernel matrices[10]. When there are more than two variance components, a matrix inversion is inevitable in each update, so the MM algorithm is not scalable to biobank-level data, and we recommend applying this method for datasets of size up to $n \times d = 50000$. In this study, we take advantage of the aforementioned MM algorithm in estimating parameters in variance components models.

Univariate variance components linear mixed models were fit for gene and isoform-level expression with the above-mentioned mean effects covariates. We specified three variance components, two of which capture *cis*- and *trans*-SNP genetic effects. We defined *cis*-SNPs as those within 1 Mb window of gene start and gene end sites, and *trans*-SNPs as all the other SNPs. Based on this definition, 24,754 genes and 93,030 isoforms had non-zero *cis*-SNPs. The same set of *cis*-SNPs was used for a given gene and its constituent isoforms for direct comparison. The mean number of *cis*-SNPs were 3,264 and 3,274 for these genes and isoforms, respectively. Variance components parameters were estimated using both ML and REML. The maximum number of iterations was set to 3,000. In total, 22,965 genes and 89,926 isoforms had converged estimates. In most (if not all) cases where the estimates failed to converge, ML estimation was the issue.

As part of sensitivity analyses, we fit additional univariate models, including three variance components models with different definitions of *cis*-SNP windows (i.e. ±250 kb window and the entire chromosome), two variance components model with a single parameter capturing the entire genetic effects, and two variance components model with only *cis*-SNP effects (1 Mb window). In theory, model selection that best fits the data using a statistical criterion such as AIC metric is possible, but we did not conduct such analyses.

Multivariate variance components linear mixed models were fit for isoform-level expression. We specified three variance components, one of which captures *cis*-SNP genetic effects and the other *trans* effects. We used the same set of *cis*-SNPs that were used for univariate models. To reduce computational burden and the number of variance components parameters that need to be estimated, given limited sample size of the PsychENCODE dataset, we ran the multivariate model for isoforms with significant heritability estimates in a univariate model at $P < 0.05$. For isoforms that are perfectly correlated in expression, we included only one isoform of the two. Variance components parameters were estimated using REML. Note that the MM algorithm is numerically stable to fit even the non-heritable isoforms, but we chose not to. Lastly, pairwise bivariate variance components models were fit with the same scheme as multivariate models. For each gene with at least two heritable isoforms, the model was fit to all pairwise combinations of isoforms.

Inference in variance components model

Inference on variance components parameters was done using a variation of the likelihood ratio test (LRT). Here, we tested the null hypothesis that both variance components corresponding to *cis*- and *trans*-SNP genetic effects are zero by fitting a null model with only a single residual variance components parameter. Then we compared the log-likelihood from ML estimation, which difference times two is assumed to follow a mixture of χ^2 distributions[17] as

$$2(\mathcal{L} - \mathcal{L}_0) \sim \frac{1}{4}\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2.$$

Standard errors for both genetic variances and genetic covariances were estimated using the Fisher information matrix, where

$$\mathrm{E}\left[-\frac{\partial^2}{\partial(\mathrm{vech}\,\mathbf{\Gamma}_j)^T\partial(\mathrm{vech}\,\mathbf{\Gamma}_i)}\mathcal{L}\right] = \frac{1}{2}\mathbf{U}_i^T(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1})\mathbf{U}_j$$

and $\mathbf{U}_i = (\mathbf{I}_d \otimes \mathbf{K}_{nd} \otimes \mathbf{I}_n)(\mathbf{I}_{d^2} \otimes \mathrm{vec}\,\mathbf{V}_i)\mathbf{D}_d$. Note that this form is slightly different from the one shown earlier with $\mathbf{U}_i$ replaced by $\mathbf{W}_i$. This is because we previously imposed the positive semidefiniteness constraint by reparametrizing $\mathbf{\Gamma}_i = \mathbf{L}_i\mathbf{L}_i^T$. Then by using the vec-transpose operator,

$$\begin{aligned}
\mathbf{U}_i &= \mathbf{I}_d \otimes (\mathbf{K}_{nd} \otimes \mathbf{I}_n)(\mathbf{I}_d \otimes \mathrm{vec}\,\mathbf{V}_i)\mathbf{D}_d \\
&= \mathbf{I}_d \otimes (\mathbf{K}_{nd} \otimes \mathbf{I}_n)(\mathbf{I}_d \otimes \mathbf{V}_i^{(n)})\mathbf{D}_d \\
&= \mathbf{I}_d \otimes (\mathbf{K}_{nd} \otimes \mathbf{I}_n)(\mathbf{I}_d \otimes \mathbf{V}_i)^{(n)}\mathbf{D}_d \\
&= \mathbf{I}_d \otimes [(\mathbf{I}_d \otimes \mathbf{V}_i)\mathbf{K}_{dn}]^{(n)}\mathbf{D}_d.
\end{aligned}$$

Naïve computation of the Fisher information matrix is not computationally feasible for even reasonable nd, so we must take advantage of the structure inherent in $\mathbf{U}_i$. Standard errors for heritability and genetic correlation estimates were subsequently calculated using the delta method. Inference on genetic covariance or genetic correlation parameters was done using the Wald test for both bivariate and multivariate models.

Penalization of variance components

To investigate the polygenicity of gene and isoform expression, we also fit a univariate, 23 variance components model with separate SNP effects from each autosome such that

$$\mathbf{y} \sim N(\mathbf{X\beta}, \sum_{i=1}^{22} \sigma_i^2\,\mathbf{V}_i + \sigma_e^2\mathbf{I}).$$

However, due to the limited sample size of PsychENCODE, these estimates were noisy and hence we instead minimized the lasso-penalized negative log-likelihood function[22]. One of the advantages of the MM algorithm is that it separates the parameters of a problem such that penalized estimation is conducive[10,22]. For each chromosome, we focused on heritable genes and isoforms (P < 0.05) that reside in that chromosome, and for these features, we calculated the number of times each chromosome appeared in the solution path.

***cis*-eQTL analyses**

We ran eQTL analyses with the same *cis*-SNPs used in variance components models for each gene and isoform using a linear model as implemented in QTLtools[18]. To account for linkage disequilibrium (LD), sample labels were permuted 1,000 times, and each iteration the most significantly associated *P* value was saved. Such a null distribution of *P* values was then fit to a beta distribution and the observed *P* value was subsequently adjusted to give an empirical *P* value. To account for multiple testing, empirical *P* values transcriptome-wide were FDR corrected. Conditional analyses were further performed through QTLtools to discover independently associated eQTL signals for gene and isoform expression. The top associated eQTL SNP for each gene and isoform was then used to calculate R^2 or variance explained by index SNPs.

Colocalization

We tested for colocalization between isoform-level eQTL signals and complex traits GWAS signals at the *XRN2*, *SYNE1*, *TBL1XR1*, and *SYT1* loci by using association results for SNPs within ±1 Mb window of respective gene bodies and present in both eQTL and GWAS results. We used the coloc.abf function in R package *coloc* v5.1.0.1[32]. This method assumes a single causal variant within each locus and assesses evidence of a shared association signal using a Bayesian approach. Sample size and MAF were used instead of sdY for *cis*-eQTL data and the default priors of $p1 = p2 = 10^{-4}$ and $p12 = 10^{-5}$ were applied.

***XRN2* isoform re-quantification with long-read transcriptome annotation in fetal brain short-read bulk RNA-seq data**

To better complete the human brain transcriptome annotation, we compiled long-read sequencing data of the human brain from six different studies[41-43]. This included in-house fetal brain single-cell Iso-Seq data and two adult human brain data shared by PacBio (https://www.pacb.com/connect/datasets/). Note that ref [43] includes data from non-brain tissues, while other studies are strictly from the adult or developing human brain. We started with finalized (e.g. filtered and merged) GTF files from each study, harmonized genomic coordinates with liftover to the GRCh38 human genome build, ran gffcompare v0.12.6 with Gencode v40 as a reference, and filtered for isoforms found in at least two or three studies. gffcompare searches for identical matches for internal exons, while it implements fuzzy matches for terminal exons, which is appropriate here since these exons can be sequenced to different lengths. We subsequently used the above two filtered/munged GTF files to re-quantify ref [47] fetal brain bulk RNA-seq data using Salmon v1.8.0. Gene-level expression was calculated to be the sum of its constituent isoform TPM. To repeat gene-level and isoform-level eQTL analyses, genotype data for ref [47] was filtered for typed variants with individual call rate (> 0.95), minor allele frequency (> 0.01), Hardy-Weinberg equilibrium (HWE) *P* value ($> 10^{-6}$) and individuals based on genotype call rate (> 0.9) using PLINK 1.9. Genotype data was then phased and imputed with the TOPMed freeze 5 reference panel on the Michigan Imputation Server. After removing SNPs with low imputation quality (R2 < 0.3) and subsetting to 86 individuals of European ancestry, we again filtered SNPs based on individual call rate (> 0.95), MAF (> 0.05), HWE *P* value ($> 10^{-6}$) and individuals based on genotype call rate (> 0.95). For eQTL analyses, we modeled either TPM or $\log_2$TPM with age, sex, 4 genetic PCs, RIN, and 21 sequencing PCs from sequencing metrics generated from PicardTools as covariates.

Data availability

PsychENCODE raw data are available at www.doi.org/10.7303/syn12080241 and processed summary-level data are available at Resource.PsychENCODE.org.

Code availability

The Julia implementation of the MM algorithm for fitting multivariate variance components linear mixed models is available at
https://github.com/Hua-Zhou/MultiResponseVarianceComponentModels.jl.
The code used to perform bioinformatic analyses are available at
https://github.com/mmkim1210/isoforms-genetics.
All plots were generated programmatically using Makie.jl[58] and GeneticsMakie.jl[59].

Supplementary Figures

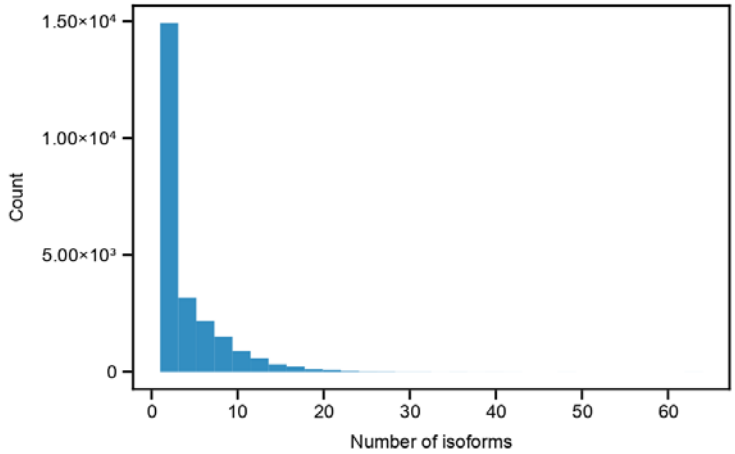
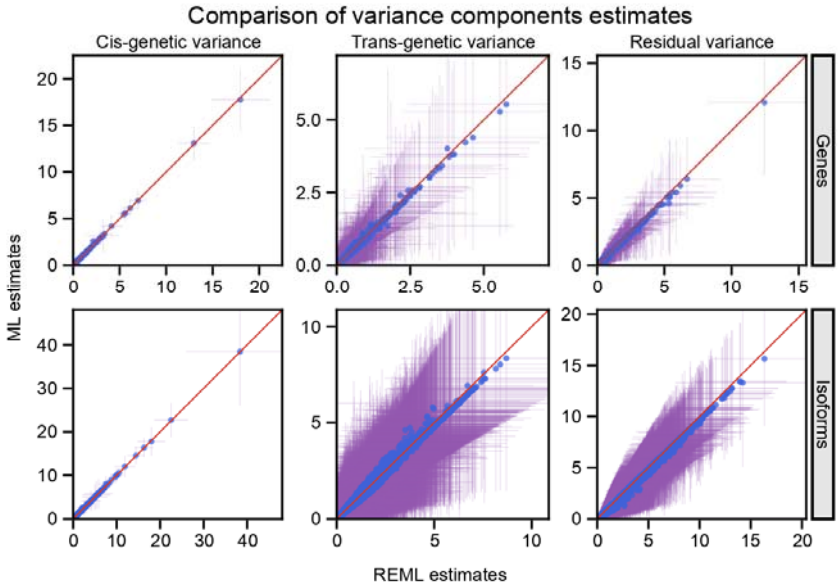


Figure S1: Distribution of the number of isoforms for 24,905 genes in PsychENCODE. The maximum number of isoforms is 64.



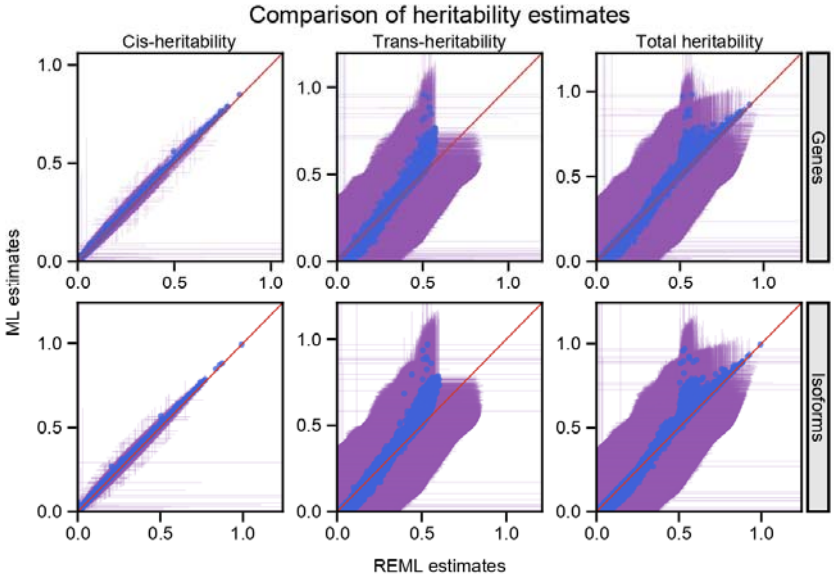


Figure S2: Comparison of REML and ML estimates in the univariate variance components models. REML and ML estimates were highly concordant for h^2_{cis} and to a lesser degree for h^2_{trans} and h^2_{SNP}. Error bars denote ± one standard errors. Note higher standard errors for h^2_{trans} and h^2_{SNP} relative to h^2_{cis}. Standard errors are calculated using the Fisher information matrix and the Delta method.

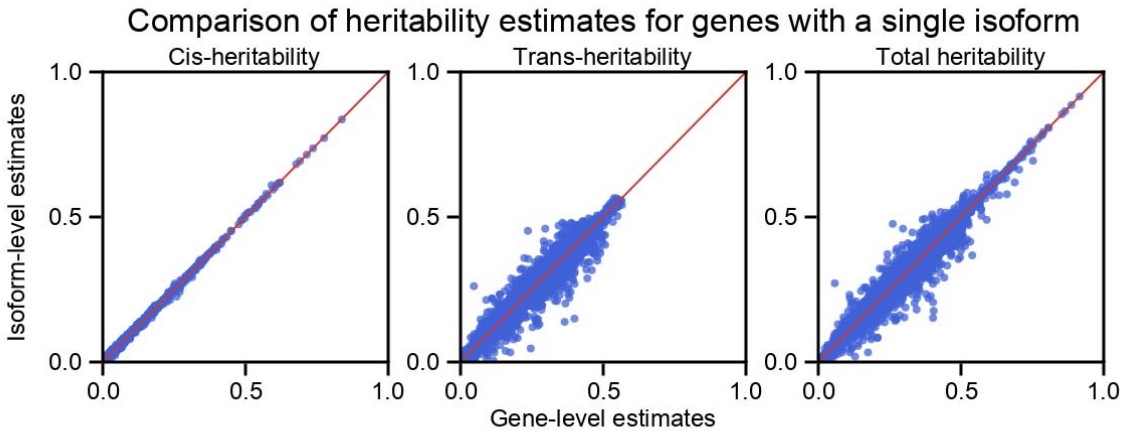


Figure S3: Comparison of heritability estimates for genes with a single isoform. h^2_{cis} estimates were highly concordant, while more variability was observed for h^2_{trans} and h^2_{SNP} estimates.

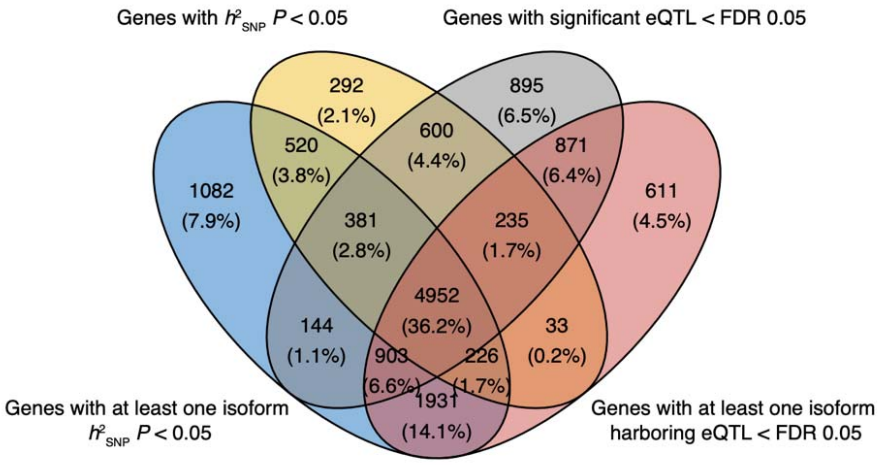


Figure S4: Comparison between heritability and *cis*-eQTL results. Shown is Venn diagram of overlap between heritable genes and genes harboring eQTL at FDR < 0.05.

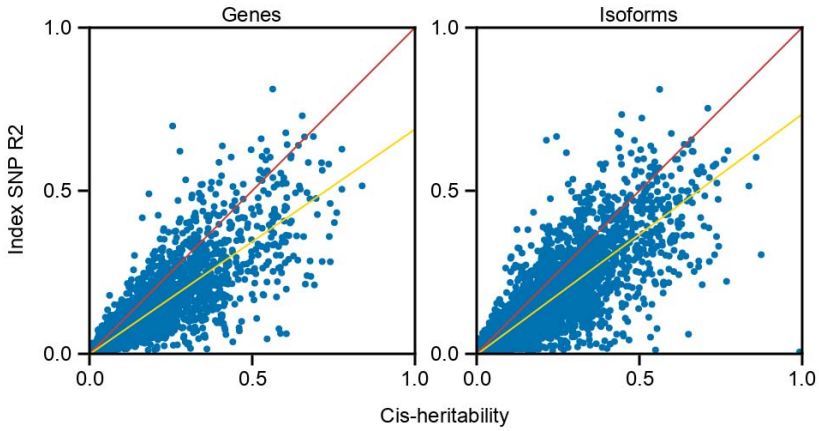


Figure S5: Sparsity of *cis*-SNP effects. Left, comparison between h^2_{cis} estimates and variance explained by top associated eQTL SNPs for heritable genes. Right, the same comparison but for heritable isoforms.

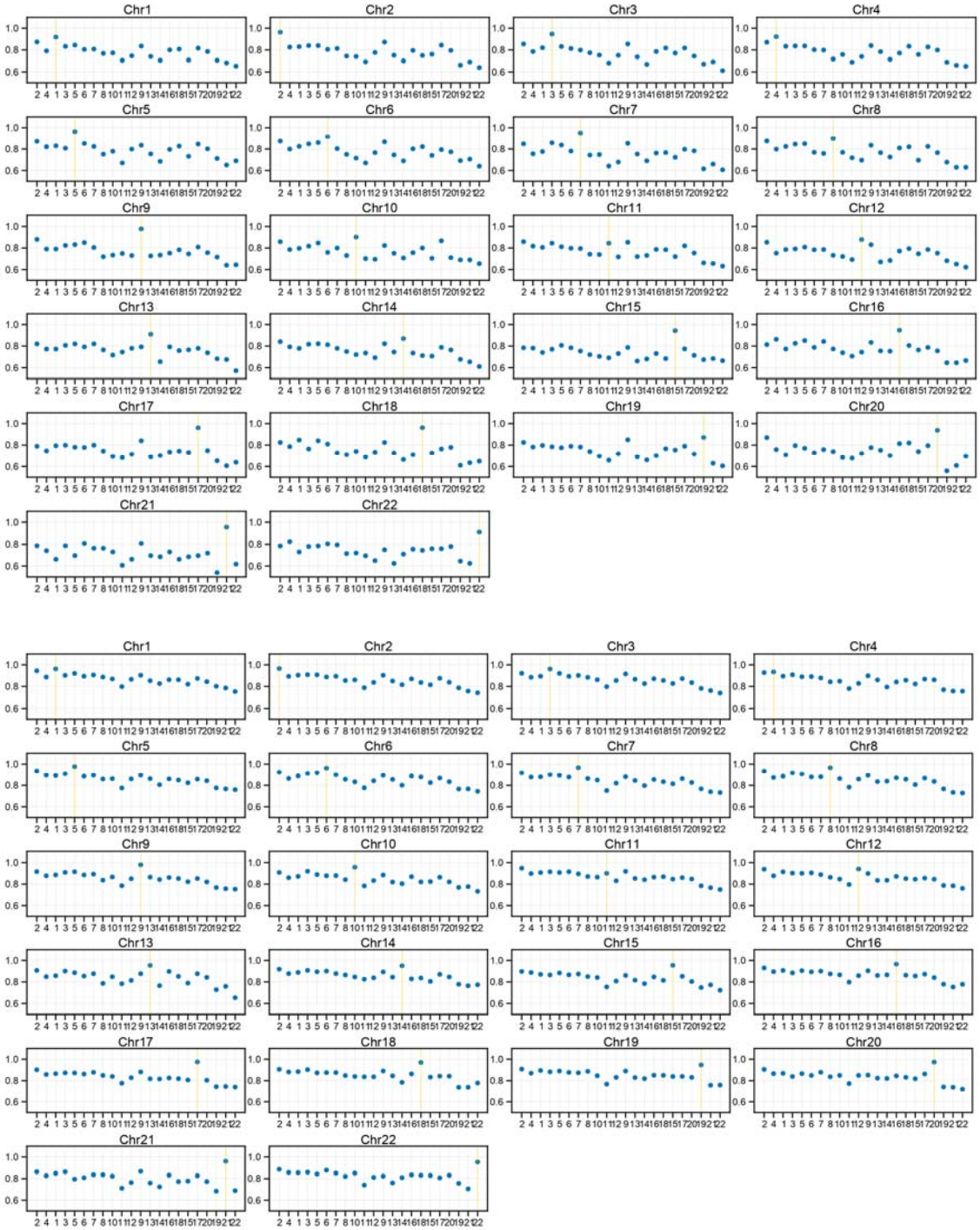


Figure S6: Polygenicity of *trans*-SNP effects. We fit a univariate penalized variance components model with 23 variance components parameters that capture genetic effects from each autosome and lasso penalty. Top, shown are results for genes. Bottom, shown are results for isoforms. For each chromosome, we focus on heritable genes and isoforms ($P < 0.05$) that reside in that chromosome, and for these features, we calculate the number of times each chromosome appeared in the solution path[22]. The chromosome the feature is from almost always appeared in the solution path (i.e. yellow lines have the highest proportions), while the rest of the chromosomes appeared in proportion to their number of SNPs. Chromosomes are listed in descending order of the number of SNPs on the x-axis.

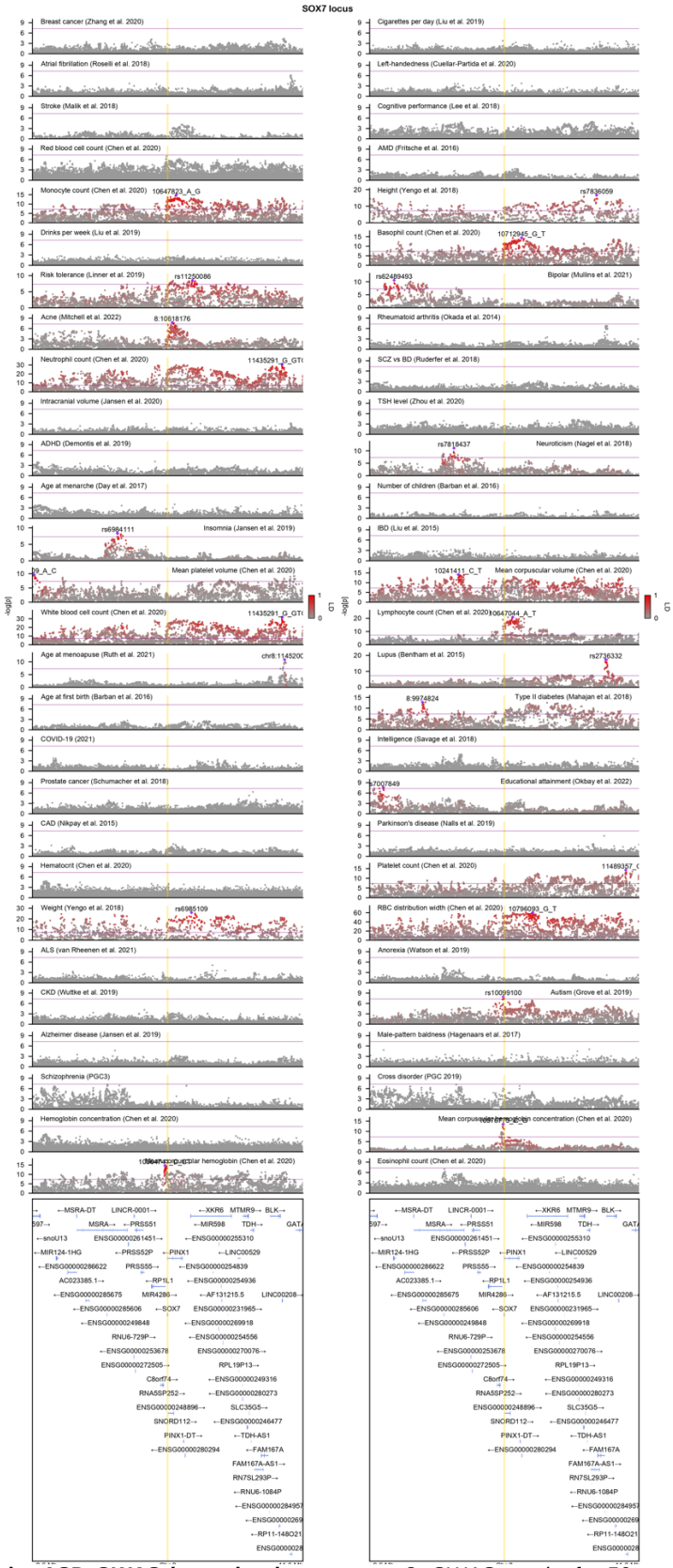


Figure S7: A close look at the ASD GWAS locus in chromosome 8. GWAS results for 56 complex phenotypes are shown, which span autoimmune, endocrine, psychiatric, cardiovascular disorders, and cancer. Index SNPs for phenotypes harboring GWAS hits are labeled and corresponding LD between other SNPs are displayed with the intensity of red color. Purple line denotes genome-wide significance ($P = 5 \times 10^{-8}$), and yellow lines denote gene start and end sites for *SOX7* gene. LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel. ADHD (attention-deficit/hyperactivity disorder), ALS (amyotrophic lateral sclerosis), AMD (age-related macular degeneration), BD (bipolar disorder), CAD (coronary artery disease), CKD (chronic kidney disease), IBD (inflammatory bowel disease), RBC (red blood cell), SCZ (schizophrenia).

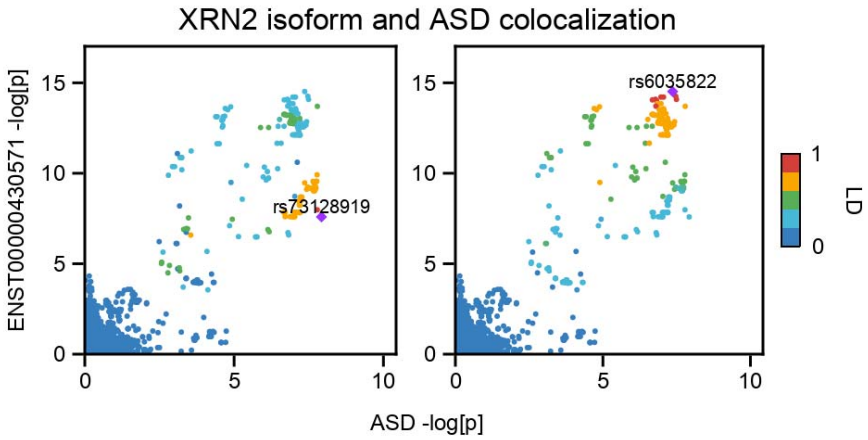


Figure S8: Colocalization between *XRN2* isoform-level eQTL and ASD GWAS results. Left, LD is colored with respect to the index SNP for ASD GWAS. Right, LD is colored with respect to the index SNP for ENST00000430571 eQTL. LD is calculated with individuals of European ancestry in PsychENCODE.

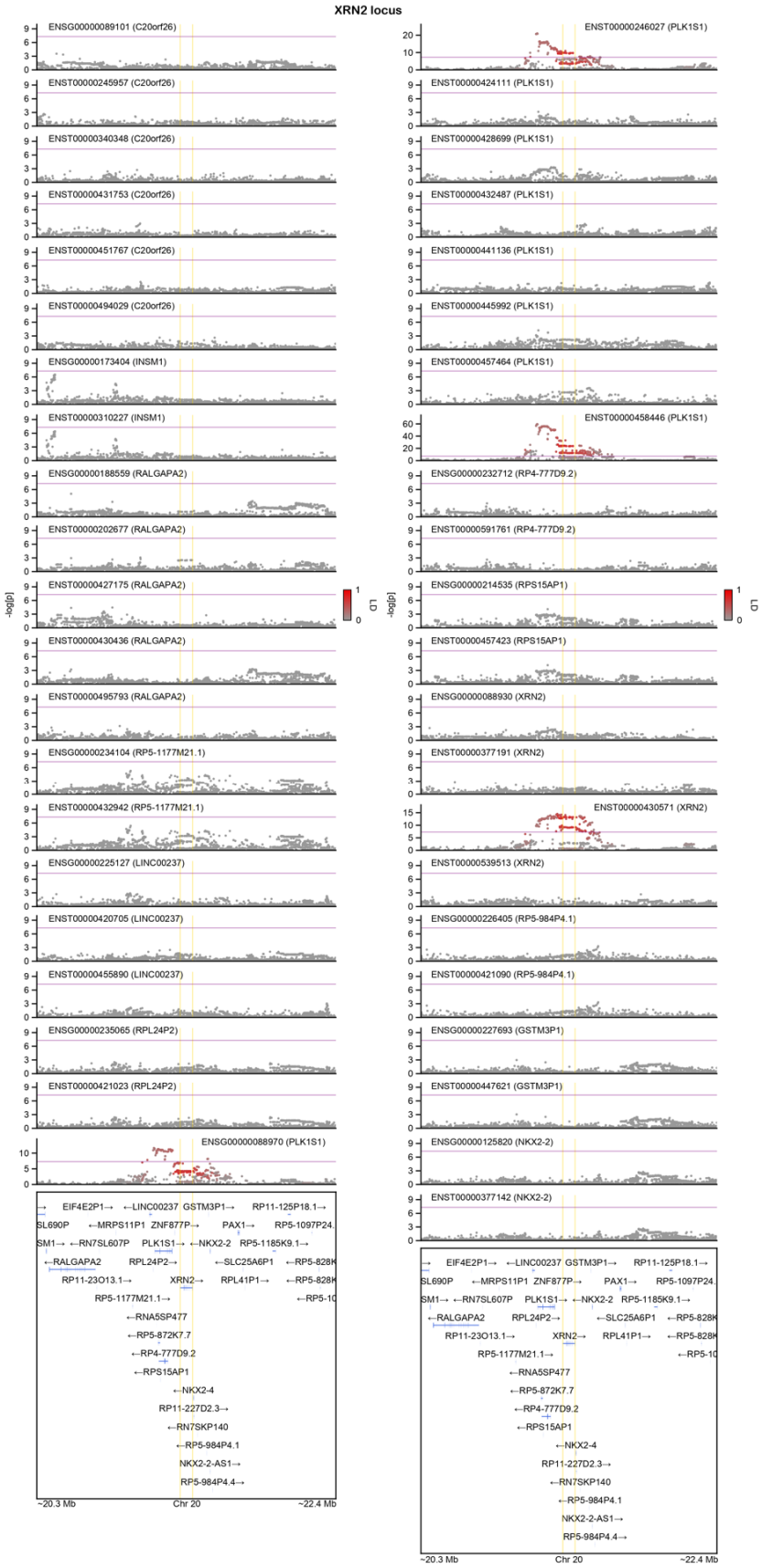


Figure S9: Absence of colocalization for all other features within the *XRN2* locus for the ASD GWAS signal. Shown is LocusZoom of eQTL signals for both gene- and isoform-level expression for all features within ±1 Mb window of (collapsed) gene start and end sites for *XRN2* gene. Gene names are shown in parentheses. LD is colored with respect to the index SNP for ASD GWAS (rs910805). LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel.

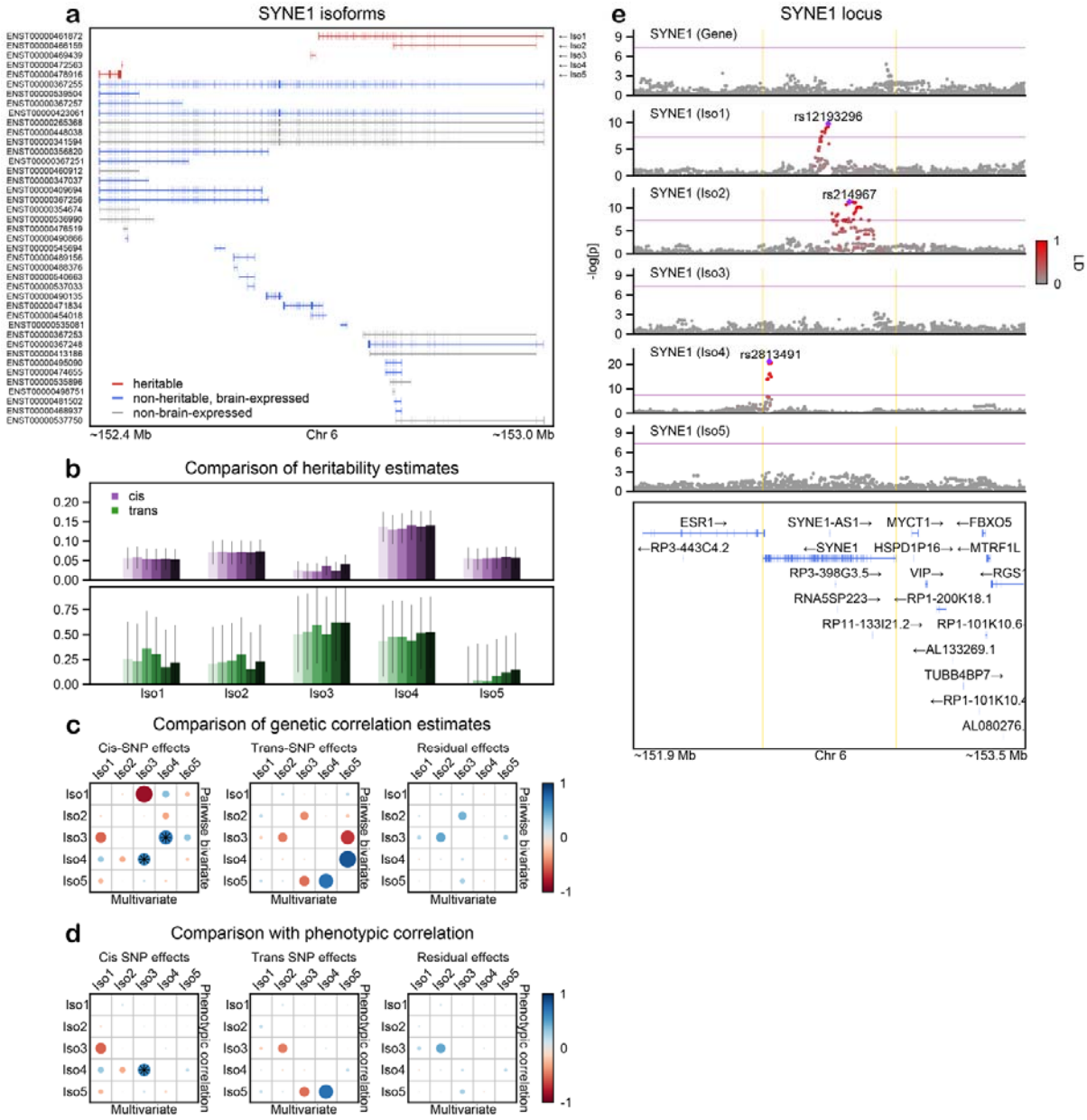


Figure S10: h^2_{SNP}, r_g, and *cis*-eQTL results for *SYNE1* gene. Follows the same outline as **Figure 3** in the main text.

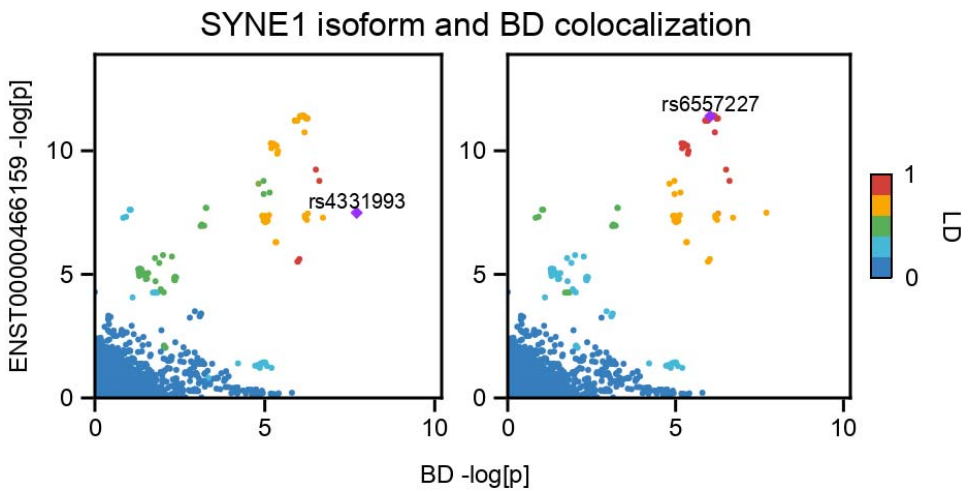
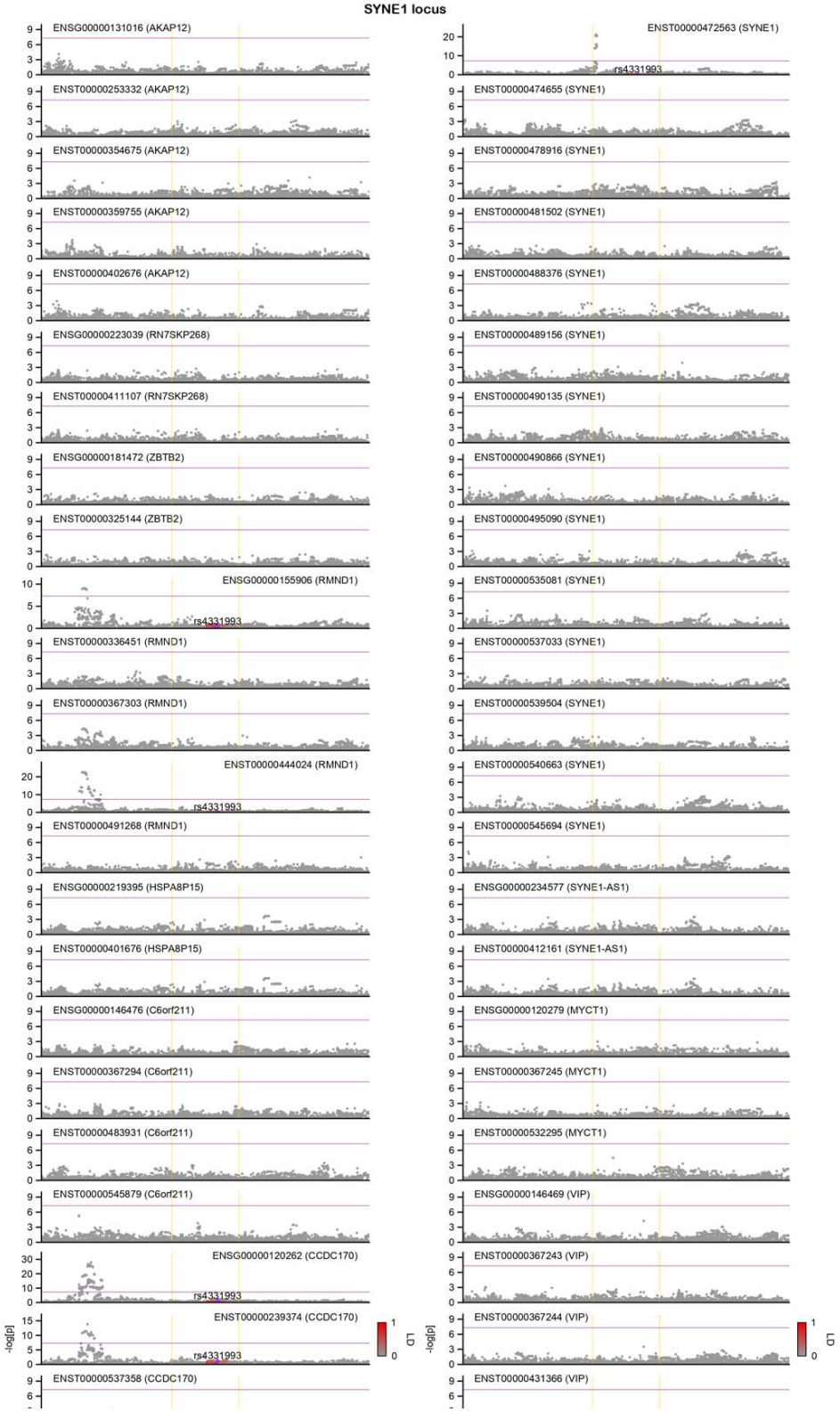


Figure S11: Colocalization between *SYNE1* isoform-level eQTL and BD GWAS results. Left, LD is colored with respect to the index SNP for BD GWAS. Right, LD is colored with respect to the index SNP for ENST00000466159 eQTL. LD is calculated with individuals of European ancestry in PsychENCODE.


SYNE1 locus
ENSG00000131016 (AKAP12)
ENST00000253332 (AKAP12)
ENST00000354675 (AKAP12)
ENST00000359755 (AKAP12)
ENST00000402676 (AKAP12)
ENSG00000223039 (RN7SKP268)
ENST00000411107 (RN7SKP268)
ENSG00000181472 (ZBTB2)
ENST00000325144 (ZBTB2)
ENSG00000155906 (RMND1)
ENST00000336451 (RMND1)
ENST00000367303 (RMND1)
ENST00000444024 (RMND1)
ENST00000491268 (RMND1)
ENSG00000219395 (HSPA8P15)
ENST00000401676 (HSPA8P15)
ENSG00000146476 (C6orf211)
ENST00000367294 (C6orf211)
ENST00000483931 (C6orf211)
ENST00000545879 (C6orf211)
ENSG00000120262 (CCDC170)
ENST00000239374 (CCDC170)
ENST00000537358 (CCDC170)
ENST00000472563 (SYNE1)
ENST00000474655 (SYNE1)
ENST00000478916 (SYNE1)
ENST00000481502 (SYNE1)
ENST00000488376 (SYNE1)
ENST00000489156 (SYNE1)
ENST00000490135 (SYNE1)
ENST00000490866 (SYNE1)
ENST00000495090 (SYNE1)
ENST00000535081 (SYNE1)
ENST00000537033 (SYNE1)
ENST00000539504 (SYNE1)
ENST00000540663 (SYNE1)
ENST00000545694 (SYNE1)
ENSG00000234577 (SYNE1-AS1)
ENST00000412161 (SYNE1-AS1)
ENSG00000120279 (MYCT1)
ENST00000367245 (MYCT1)
ENST00000532295 (MYCT1)
ENSG00000146469 (VIP)
ENST00000367243 (VIP)
ENST00000367244 (VIP)
ENST00000431366 (VIP)
rs4331993
-log[p]
LD

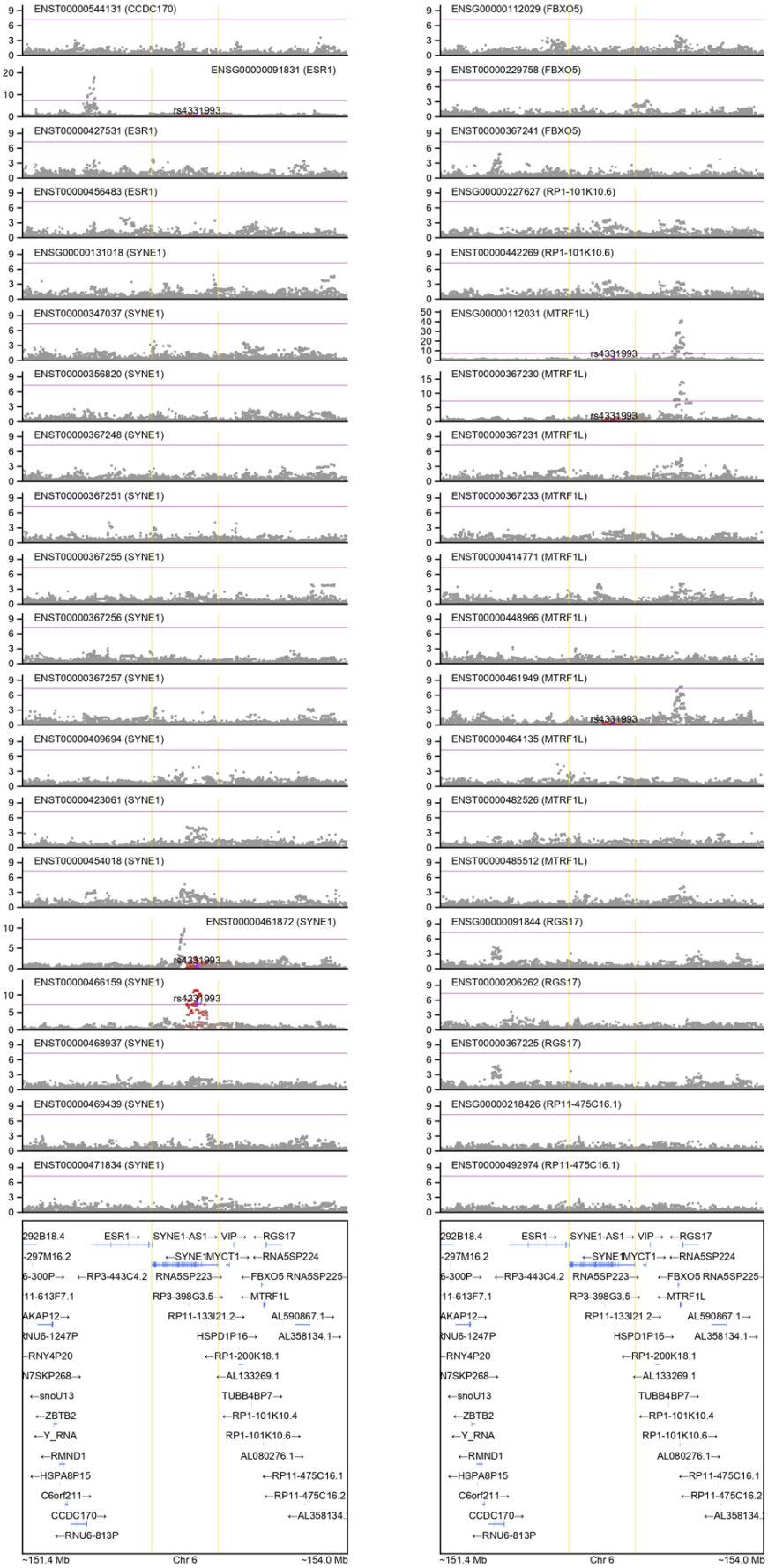


Figure S12: Absence of colocalization for all other features within the *SYNE1* locus for the BD GWAS signal. Shown is LocusZoom of eQTL signals for both gene- and isoform-level expression for all features within ±1 Mb window of (collapsed) gene start and end sites for *SYNE1* gene. Gene names are shown in parentheses. LD is colored with respect to the index SNP for BD GWAS (rs4331993). LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel.

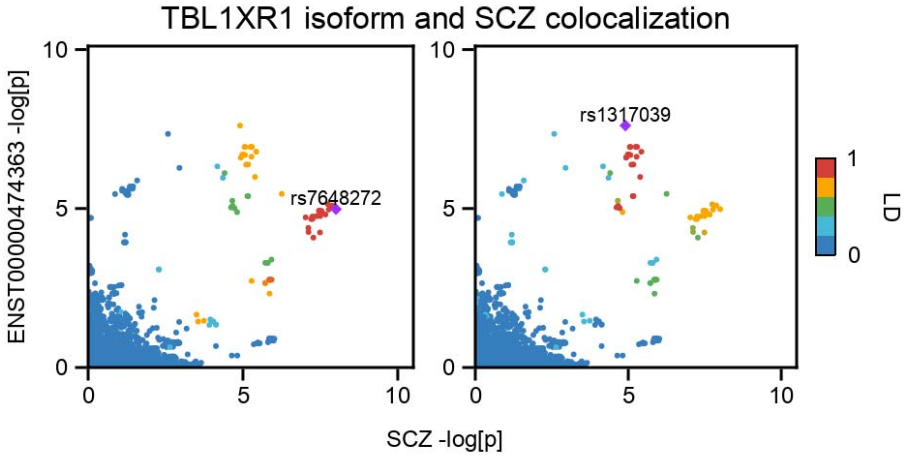


Figure S13: Colocalization between *TBL1XR1* isoform-level eQTL and SCZ GWAS results. Left, LD is colored with respect to the index SNP for SCZ GWAS. Right, LD is colored with respect to the index SNP for ENST00000474363 eQTL. LD is calculated with individuals of European ancestry in PsychENCODE.

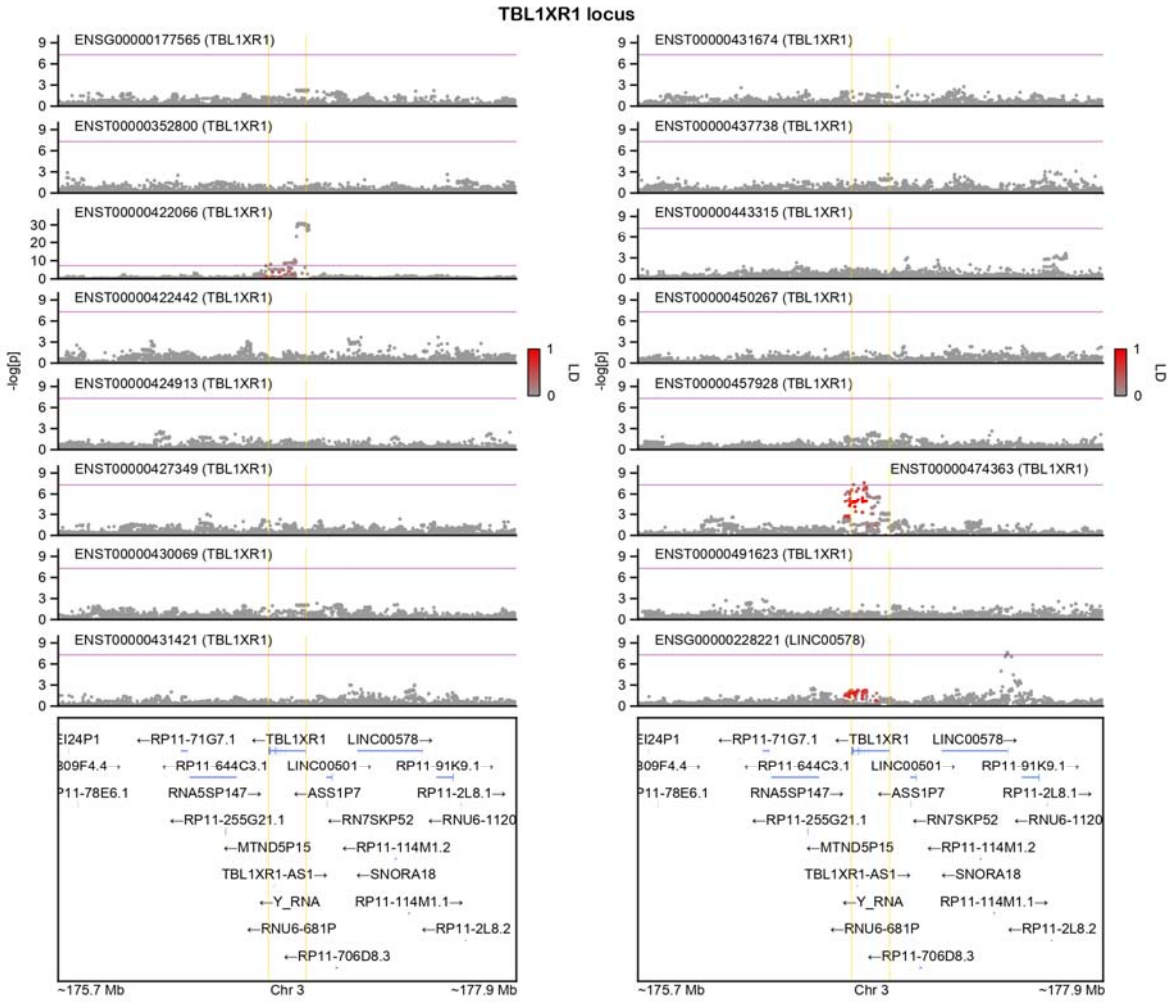


Figure S14: Absence of colocalization for all other features within the *TBL1XR1* locus for the SCZ GWAS signal. Shown is LocusZoom of eQTL signals for both gene- and isoform-level expression for all features within ±1 Mb window of (collapsed) gene start and end sites for *TBL1XR1* gene. Gene names are shown in parentheses. LD is colored with respect to the index SNP for SCZ GWAS (rs7609876). LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel.

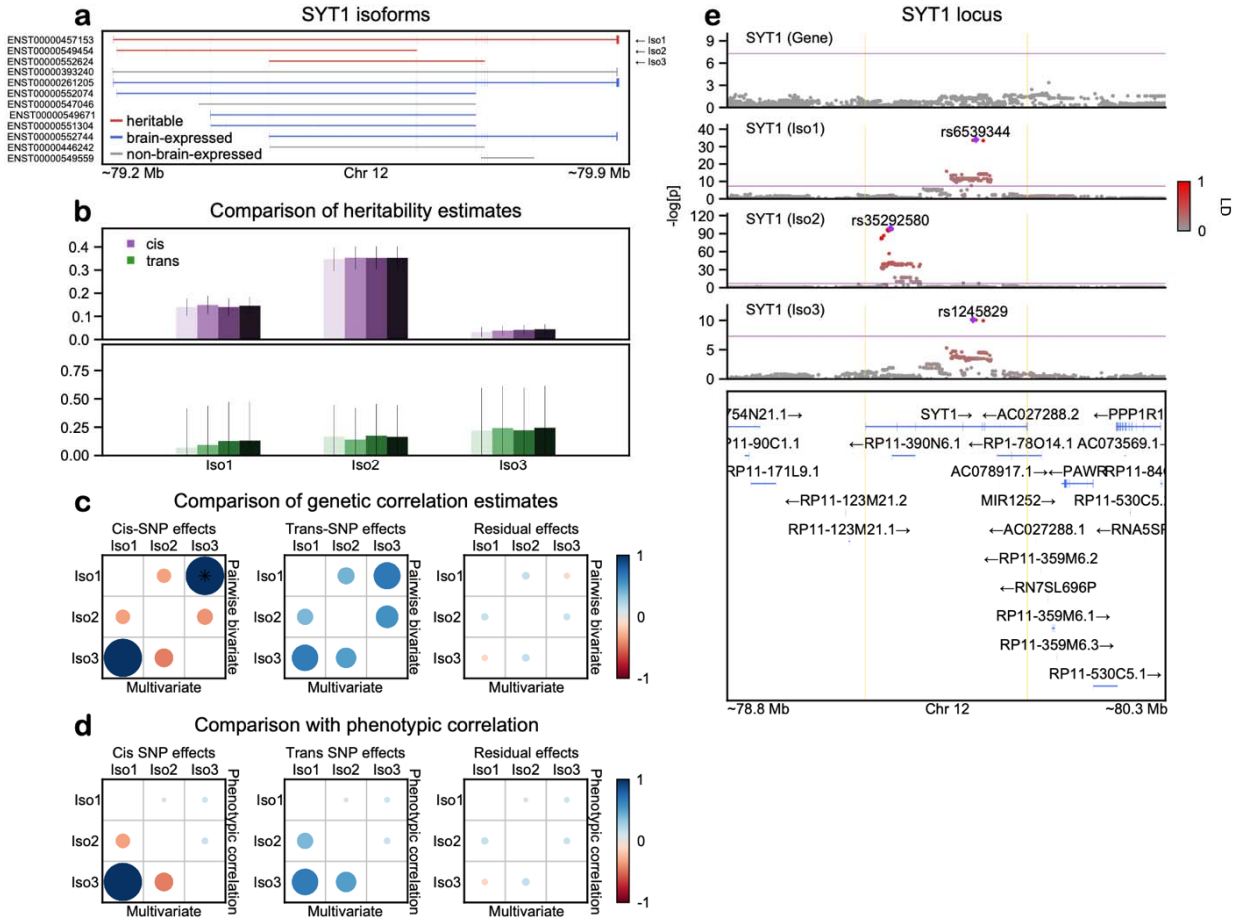


Figure S15: h^2_{SNP}, r_g, and *cis*-eQTL results for *SYT1* gene. Follows the same outline as **Figure 3** in the main text.

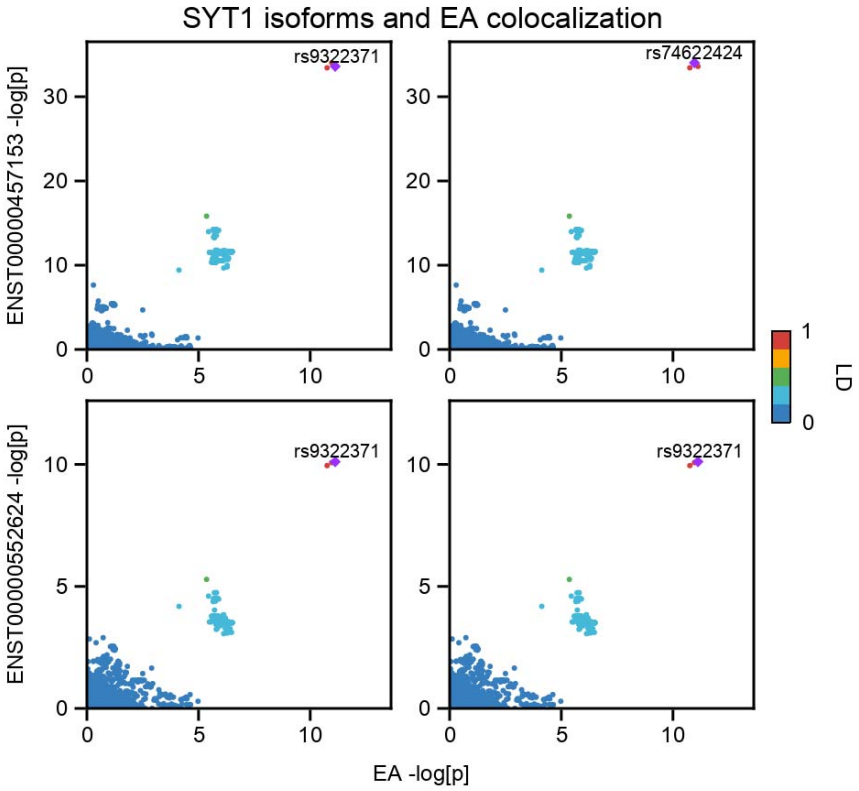


Figure S16: Colocalization between *SYT1* isoform-level eQTL and EA GWAS results. Left, LD is colored with respect to the index SNP for EA GWAS. Right, LD is colored with respect to the index SNP for corresponding isoform-level eQTL. LD is calculated with individuals of European ancestry in PsychENCODE.

Figure S17: Absence of colocalization for all other features within the *SYT1* locus for the EA GWAS signal. Shown is LocusZoom of eQTL signals for both gene- and isoform-level expression for all features within ±1 Mb window of (collapsed) gene start and end sites for *SYT1* gene. Gene names are shown in parentheses. LD is colored with respect to the index SNP for EA GWAS (rs1245829). LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel.

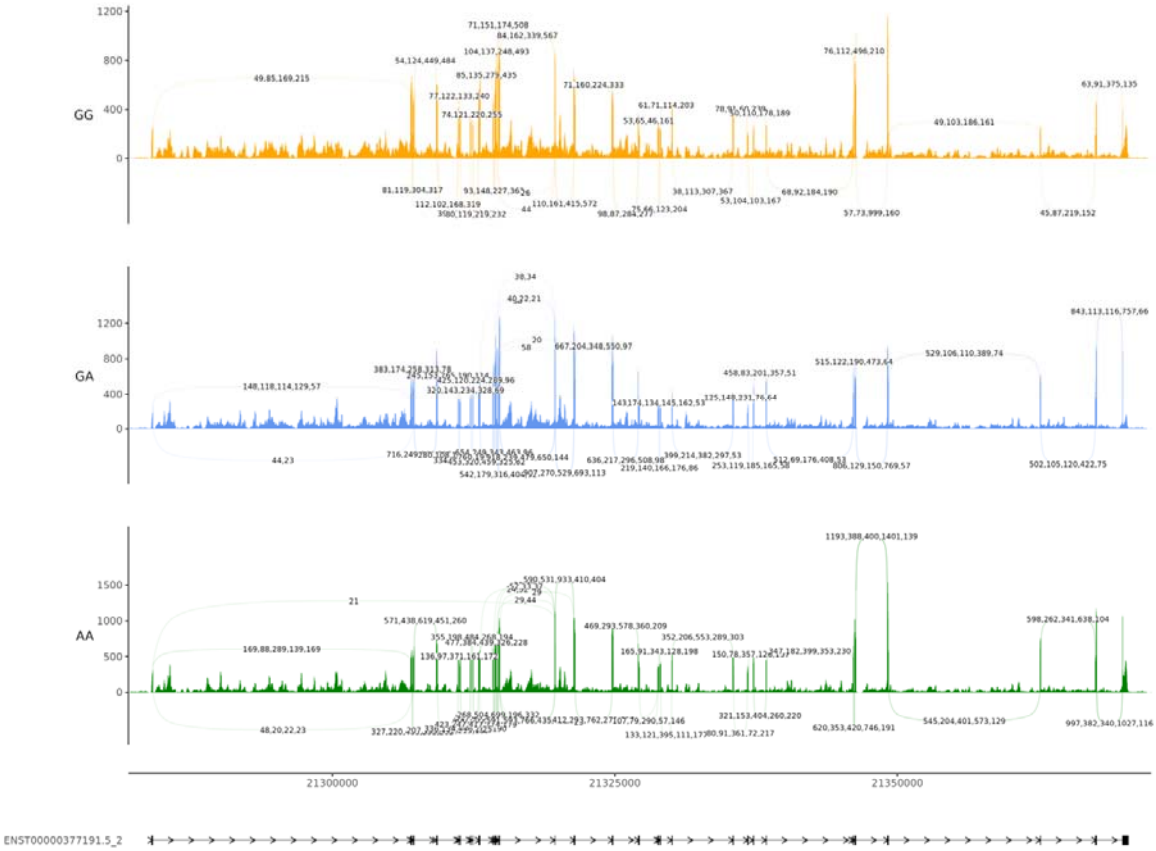


Figure S18: Local splicing events in the *XRN2* gene body across rs910805 genotypes. Five RNA-seq samples[47] with approximately equal library (or read) depths are plotted for each rs910805 genotype. As above, the increase in the number of 1st to 3rd exon junction reads in the 5' end with respect to the A major allele was subtle. This plot was generated with ggsashimi GitHub repository (https://github.com/guigolab/ggsashimi).

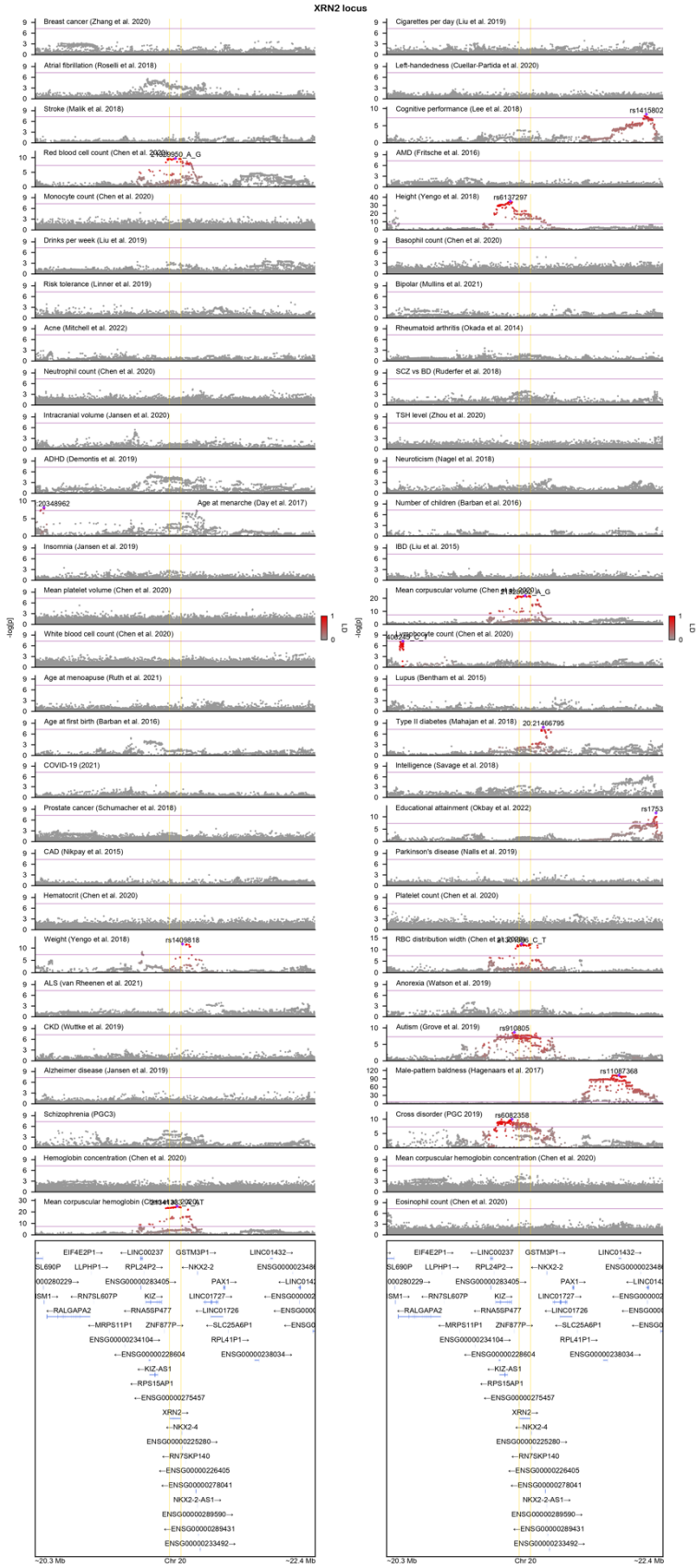


Figure S19: A close look at the pleiotropic *XRN2* locus. GWAS results for 56 complex phenotypes are shown, which span autoimmune, endocrine, psychiatric, cardiovascular disorders, and cancer. Index SNPs for phenotypes harboring GWAS hits are labeled and corresponding LD between other SNPs are displayed with the intensity of red color. Purple line denotes genome-wide significance ($P = 5 \times 10^{-8}$), and yellow lines denote gene start and end sites for *XRN2* gene. LD is calculated with individuals of European ancestry in the 1000 Genomes Project reference panel. ADHD (attention-deficit/hyperactivity disorder), ALS (amyotrophic lateral sclerosis), AMD (age-related macular degeneration), BD (bipolar disorder), CAD (coronary artery disease), CKD (chronic kidney disease), IBD (inflammatory bowel disease), RBC (red blood cell), SCZ (schizophrenia).