Isoform-level transcriptome-wide association uncovers extensive novel genetic risk mechanisms for neuropsychiatric disorders in the human brain

AUTHORS

Arjun Bhattacharya[1,2], Connor Jops[3], Minsoo Kim[3,4], Cindy Wen[3], Daniel D. Vo[3], Jonatan L. Hervoso[4], Bogdan Pasaniuc[1,4,6±], Michael J. Gandal[3,4±]

AFFILIATIONS

1. Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, California, USA
2. Institute for Quantitative and Computational Biosciences, David Geffen School of Medicine, University of California, Los Angeles, California, USA
3. Department of Psychiatry and Biobehavioral Sciences, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, California, USA
4. Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
5. Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA
6. Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

[±]Equal contribution

CORRESPONDENCE

Corresponding author: Arjun Bhattacharya (abtbhatt@ucla.edu)

ABSTRACT

Integrative methods, like colocalization and transcriptome-wide association studies (TWAS), identify transcriptomic mechanisms at only a small fraction of trait-associated genetic loci from genome-wide association studies (GWAS). Here, we show that a reliance on reference functional genomics panels of only total gene expression greatly contributes to this reduced discovery. This is particularly relevant for neuropsychiatric traits, as the brain expresses extensive, complex, and unique alternative splicing patterns giving rise to multiple genetically-regulated transcript-isoforms per gene.

We introduce isoTWAS, a scalable, multivariate framework to integrate genetics, isoform-level expression, and phenotypic associations in a step-wise testing framework. Multivariate predictive models were trained using isoform-level expression across tissues from the Genotype-Tissue Expression Project and in the developing and adult human brain from PsychENCODE. Across five neuropsychiatric traits, isoTWAS dramatically increased discovery of trait associations within GWAS loci, capturing 92 unique loci compared with 27 using gene-level TWAS. These power gains reflected a ~2-fold increase in the number of testable genes, an ~15-35% increase in total gene expression prediction accuracy, and the ability to jointly capture expression and splicing mechanisms. Results from extensive simulations showed no increase in false discovery rate and reinforce isoTWAS's advantages in prediction and trait mapping power over TWAS, especially when genetic effects on expression vary across isoforms of the same gene. We illustrate multiple biologically-relevant isoTWAS-identified trait associations undetectable by gene-level methods, including isoforms of *AKT3*, *GIGYF2*, and *KMT5A* with schizophrenia risk.

The isoTWAS framework addresses an unmet need to consider the transcriptome on the transcript-isoform level to maximize discovery of trait associations, especially for brain-relevant traits.

INTRODUCTION

Over the past decade, the number of genetic associations with complex traits identified by genome-wide association studies (GWAS) has increased considerably[1,2]. However, the slow translation of these genetic associations into concrete molecular mechanisms remains a major obstacle. As GWAS associations predominately localize within non-coding regions of the genome and are often tagged within large blocks of linkage disequilibrium (LD), the first major challenge is prioritizing the underlying causal variant(s) within a given locus and identifying their functional impact on nearby target genes. To address this, numerous methods have been developed to integrate transcriptomic reference panels with GWAS to prioritize genes at trait-associated loci[3–15]. TWAS and related approaches (e.g., PrediXcan) impute the component of gene expression predicted by germline genetics into an association cohort, thereby reducing multiple comparisons and increasing interpretability by identifying a set of genes at a locus that may underlie the genetic association[3,4]. Despite these methodological advances, a majority of GWAS loci still lack robust functional interpretation[16].

Previous integrative analyses have largely focused on aggregated measurements of total gene expression but have not systematically explored the potentially large number of distinct transcript-isoforms that can be generated from a given genetic locus through alternative splicing. Alternative splicing is a fundamental form of tissue-specific gene regulation present in more than 90% of human genes, that vastly expands the coding and regulatory potential of the genome[17–20]. Genes uniquely expressed in the brain, which are longer and contain far more exons, undergo the greatest degree of splicing compared with other tissues and species—a mechanism contributing to the vast proteomic, phenotypic, and evolutionary complexity of the human brain[21–24]. Independent of gene expression, splicing dysregulation has been increasingly implicated as a putative disease mechanism, [23,25–29]. including for neuropsychiatric disorders[10,23,26]. Yet, local splicing events can be computationally intensive to measure, are difficult to systematically integrate across multiple distinct large-scale datasets, and often coordinated across a gene, yielding many non-independent features that increases multiple testing burden. In contrast, transcript-isoform abundance can now be rapidly estimated across large-scale RNA-seq datasets using pseudoalignment methods[30,31]. Furthermore, in the brain, isoform-level expression changes have shown greater enrichment for schizophrenia heritability than gene or local splicing changes[21,32].

Here, we present isoform-level TWAS (isoTWAS), a flexible and scalable approach for complex trait mapping by integrating genetic effects on isoform-level expression with GWAS. In extensive simulations and real data

applications in the Genotype-Tissue Expression (GTEx) Project[33] and the PsychENCODE Consortium[21,23], we show that isoTWAS provides several important advantages to trait mapping. First, in the transcriptomic prediction step, the correlation between isoforms provides additional information that is unavailable when only gene-level expression is measured, which can be leveraged to improve prediction[34] accuracy of total gene expression by 15-35%. In parallel, our isoform-centric framework uncovers heritable predictive models for nearly 2-fold more genes, doubling the number of testable features in the trait mapping step. Third, divergent patterns of genetic effects across isoforms can be leveraged to provide a more granular hypothesis for a mechanism underlying the SNP-trait relationship. Finally, the isoTWAS framework jointly captures expression and splicing disease mechanisms, while maintaining a well-controlled false discovery rate. Altogether, using data from 5 neuropsychiatric traits, isoTWAS substantially increases the power to detect gene-level trait associations uncovering associations at >3 fold more GWAS loci compared to traditional gene-level TWAS. These results stress the need to shift focus to transcript-isoforms to maximize discovery of transcriptomic mechanisms underlying genetic associations with complex traits.

RESULTS

The isoTWAS pipeline

The isoTWAS framework prioritizes transcript-isoforms of genes in which the *cis*-genetic component of expression is significantly associated with a complex trait. Expression of different isoform transcripts of a gene can be modelled as a multivariate data object, with, on average, four GENCODE v40-annotated isoforms per gene (mean 4.05, standard deviation 7.28, median 1), with strong pair-wise correlations between these isoforms[33,35] Here, we assume that local genetic variants directly modulate expression of an isoform. In addition, we assume that the abundance of a gene (total gene expression) is the sum of the abundance of its isoforms (isoform expression) (**Supplemental Figure S1**), on the raw count scale. Flexibly integrating isoform-level expression into trait mapping may lead to novel discoveries in disease mapping and prioritize specific isoforms that explain genetic associations. Accordingly, gene-level trait mapping using traditional TWAS methods may not necessarily detect a trait association on the gene-level if a gene has multiple isoforms but only one is associated with the trait (**Figure 1a**). This scenario may be particularly relevant in the human brain, where some genes may express up to hundreds to thousands of unique isoforms[24]. By modeling the genetic architectures of isoforms of a gene simultaneously, isoTWAS provides a deeper understanding of potential transcriptomic mechanisms that underlie genetic associations.

The isoTWAS pipeline contains three general steps. First, we build multivariate predictive models of isoform-level expression using well-powered functional genomics training datasets, such as GTEx[33] and PsychENCODE[23,26]. Here, we trained and systematically compared 6 multivariate predictive frameworks: (1) multivariate elastic net penalized regression[36], (2) multivariate LASSO penalized regression with simultaneous covariance estimation[37], (3) multivariate Sum of Single Effects (SuSiE)[38], (4) Breiman and Friedman's curds-and-whey procedure[34], (5) ordinary least squares with clustered standard errors[39], and (6) univariate regularized regressions of each individual isoform, as implemented in Gusev et al's FUSION software[4,36,38,40] (see **Methods** and **Supplemental Methods**). Models were trained to predict the expression of heritable isoforms using the set of *cis*-SNPs within 1 Megabase (Mb) of the gene body (**Methods**, **Figure 1b**). Model performance was assessed via 5-fold cross-validation, using McNemar's adjusted R^2 between observed and predicted expression.

Next, we use these models to impute isoform expression into an external GWAS cohort and quantify the association between imputed isoform expression and the target GWAS phenotype (**Figure 1b**). If individual-level genotypes are available, isoform expression can be directly imputed as a linear combination of the SNPs in the models, and these associations can be estimated through appropriate regression analyses. If only GWAS summary statistics are available, imputation and association testing is conducted simultaneously through a weighted burden test[4]. isoTWAS uses a two-step hypothesis testing procedure to account for multiple comparisons: isoform associations are first aggregated to the gene-level to prioritize a gene using the Aggregated Cauchy Association Test (ACAT)[41], where false discovery rates are controlled, and then individual isoforms of prioritized genes are subjected to post-hoc family-wise error control[42] (**Supplemental Figure S2**, **Methods**). After this step, a set of isoforms are identified whose *cis*-genetic component of expression are associated with the trait of interest[4]. For these isoforms, we apply a rigorous permutation test whereby the SNP-to-isoform effects are permuted 10,000 times to generate a null distribution; this permutation test assesses how much signal is added by isoform expression, given the GWAS architecture of the locus, and controls for large LD blocks[4]. Lastly, to identify a set of isoforms that explains the genetic association with the trait at the locus, probabilistic isoform-level fine-mapping can be applied (**Figure 1b**, **Methods**).

By considering isoforms of a gene as a multivariate data structure and aggregating isoform-level tests to the gene-level, isoTWAS boosts both prediction of gene expression and power to detect trait associations, especially

when genetic effects on isoforms are different and isoforms have different effects on a phenotype. Not only does isoTWAS provide substantial power gains, as we show through simulations and empiric data, this last step also provides a testable hypothesis for the mechanism underlying a trait association (**Figure 1b**). isoTWAS is available as an R package through Github (https://github.com/bhattacharya-a-bt/isotwas).

isoTWAS improves prediction of total gene expression in real data

Previous work has demonstrated that isoform-level quantifications, when propagated to the gene-level, can lead to more accurate gene expression estimates and differential expression results[43,44]. Using GTEx data from 13 brain tissues and 35 other tissues with sufficient sample sizes (N > 100, **Supplemental Table S1**; **Methods**), we demonstrate that, by accounting for isoform-level genetic architecture, isoTWAS increases prediction of total gene expression over TWAS by a median of 15-35%, both in- and out-sample (**Figure 2**). As the trait mapping step in TWAS and isoTWAS only tests models with a baseline prediction cross-validation $R^2 > 0.01$, this increase in accuracy leads to nearly 2-fold more testable gene models under the isoTWAS framework. Multivariate SuSiE provided the models with best prediction of isoform expression in 40-55% of models (**Supplemental Figure S3**), with multivariate regression with simultaneous covariate estimation showing the best prediction in 20-30% of models, underscoring the utility of modeling the covariance structure between isoforms to improve prediction.

The advantage for isoTWAS starts at the feature selection step, in which imputation models are filtered based on nominally significant positive *cis*-heritability ($P < 0.05$, **Methods**). Across the 13 brain tissues, there are nearly 3 to 4-fold more the number of genes with at least 1 *cis*-heritable isoform, compared to the number of genes that are *cis*-heritable themselves (mean of 6331 genes with at least 1 *cis*-heritable isoform vs. 1831 genes that are *cis*-heritable across 13 tissues, **Figure 2b**). When we consider genes that pass the *cis*-heritability feature selection cutoff, we trained twice as many isoTWAS models as TWAS models with CV $R^2 \geq 0.01$ (**Figure 2b**). These trends extend for nearly all other tissues as well, with greater than 2.5-fold more cis-heritable and predictable genes on the isoform-level than total gene expression level across GTEx tissues (**Supplemental Figure S4-5, Supplemental Table S2**). The median ratio of numbers of heritable isoforms to genes was significantly larger for brain tissues compared to others (3.40 vs 3.08, respectively; $P = 0.004$), when accounting for sample size (**Figure 2e; Supplemental Figure S8**). Notably, the least complex isoform architecture was exhibited by whole blood samples, with only a 2:1 ratio of *cis*-heritable isoforms to genes. These results may reflect the more complex

splicing patterns in the brain compared to other tissues and suggest that blood-based gene imputation panels may generalize poorly to the complex gene regulatory mechanisms present in the brain.

Even when restricting to the subset of genes with corresponding models at both gene and isoform levels, 60-75% of isoTWAS models showed better prediction of total gene expression than the corresponding TWAS model (**Figure 2c, Supplemental Figure S6-7, Supplemental Table S3-4**). Similarly, across all GTEx tissues, isoTWAS shows considerable improvements in prediction, even when considering tissues not derived from the brain (**Supplemental Figure S6-7. Supplemental Table S3-4**). In a completely independent eQTL reference panel derived from adult frontal cortex brain samples from the PsychENCODE Consortium (N = 850), multivariate isoTWAS models for adult frontal cortex trained in GTEx also outperformed univariate TWAS models in predicting total gene expression (median percent increase in adjusted R^2 of 21.9%; **Figure 2d, Supplemental Table S5-6**).

In sum, as predictive performance is positively related to power to detect trait associations[45], both the increased number and accuracy of trainable imputation models using isoTWAS's multivariate predictive framework will yield substantially improved discovery in trait mapping. Altogether, for 48 tissues in GTEx, we built significant predictive models for 26,000 to 33,000 isoforms across 5,000 to 6,000 unique genes per tissue (**Supplemental Figure S9**). Our results underscore that multivariate methods that incorporate the correlation between isoforms provide utility in prediction of both total gene and individual isoform expression, with strong implications on increased discovery in trait mapping[45].

isoTWAS increases discovery of trait associations across 5 neuropsychiatric disorders

To directly test our central hypothesis that isoform-centric multivariate predictive modeling will yield substantial power gains in complex trait mapping, particularly in brain tissues for relevant neuropsychiatric traits, we next directly compared isoTWAS and TWAS discovery across 5 neuropsychiatric GWAS results using models trained in larger functional reference panels derived from adult[26] and fetal[23] cortex from the PsychENCODE Consortium (**Methods, Supplemental Figure S10**). Our results demonstrate that isoTWAS increases discovery of trait associations without clear indication of elevated test statistic inflation, especially at genetic loci that house trait-associated genetic variants detected by GWAS (**Figure 3**).

Using PsychENCODE data, we trained 7,530/3,402 models using isoTWAS/TWAS in adult cortex and 8,331/2,112 isoTWAS/TWAS models for fetal cortex, passing the heritability and cross-validation R^2 cutoffs mentioned previously (**Methods; Supplemental Tables S5-6**). We next integrated these models with 5 neuropsychiatric GWAS (**Methods, Figure 3a**): SCZ[46], BD[47], ALZ[48], ASD[49], and ADHD[50]. At a conservative permutation-based significance thresholds, we detected far more trait associations with isoform-level mapping than gene-level (**Supplemental Figure S11, Supplemental Data 1-4**). Specifically, across all 5 neuropsychiatric GWAS and both adult and fetal tissue, isoTWAS identified significant associations within 92 unique GWAS loci compared with 27 detected by TWAS - nearly a 3.5-fold increase in discovery (**Figure 3b**; **Methods**). For SCZ, out of the 284 GWAS loci that have been identified to date[46], isoTWAS prioritizes genes within 81 unique loci across fetal and adult reference panels, compared with 22 from gene-level TWAS of which 19 were jointly identified. Over 89% (24/27) of gene-level TWAS associations within these GWAS loci were concordantly identified by isoTWAS. The standardized effect sizes across gene- and isoform-level mapping (for associations at nominal $P < 0.05$) are also positively correlated ($r = 0.90$, $P < 2.2 \times 10^{-16}$), with only 7.8% (103 out of 1220) of these effect sizes in opposite directions (**Figure 3c**). These results emphasize that isoTWAS recovers a large majority TWAS associations and significantly increases the GWAS signal that can be potentially attributed to underlying transcriptomic mechanisms. For example, out of the 284 GWAS loci that have been identified for SCZ[46], we find overlapping isoTWAS-identified genes for 46 and 52 loci in adult and fetal brain, respectively. These isoTWAS associations also recover 16 of 17 and 7 of 9 GWAS-overlapping TWAS associations from adult and fetal brain, respectively.

We next sought to ensure that this increase in trait mapping discovery reflected true biological signal rather than test statistic inflation due to the increased number of tests (approximately 7 to 8-fold increase in number of tests). Quantile-quantile plots of trait association gene-level P-values across isoTWAS and TWAS do not show any apparent differences in test statistic inflation (**Supplemental Figure S12).** Through empirical Bayesian methods[51], we estimated inflation in gene-level test statistics across TWAS and isoTWAS with no marked differences in the 95% credible intervals for test statistic inflation across the 5 traits. In fact, the posterior median for estimated test statistic inflation for isoTWAS were generally smaller than for TWAS (**Supplemental Figure S12, Supplemental Table S7**). Using a heuristic to estimate increases in effective sample size (**Methods**), we observed an approximate increase in effective sample size of 15-35% when using isoTWAS compared to TWAS (**Figure 2d,**

Supplemental Table S7). These analyses indicate that isoTWAS discovery is well calibrated in real data compared with TWAS

Lastly, in our comparison of methods in real data, we compared discovery using isoTWAS to discovery using splicing-event level trait mapping. Briefly, for the fetal brain cohort, we first calculated intron usage using Leafcutter[52] and transformed these usage percentages to M-values[53]. Then, for all introns mapped to a given gene, we used all SNPs within 1 Mb of a splicing-event to predict its usage and mapped trait associations for these splicing events using isoTWAS's multivariate framework (**Methods**). Overall, when aggregated to the gene-level, we found that isoTWAS discovers prioritizes features at far more independent GWAS loci (61 loci) than this splicing-event level trait mapping (31 loci), with 16 loci jointly identified (**Figure 3e**). This minimal overlap may suggest that these analyses could be used complementarily with careful interpretation of results, bridging annotation-based and -free quantifications of splicing patterns in the transcriptome. Taken together, isoTWAS's specific focus on modeling isoforms of a gene provides important gains over considering only total gene expression or intron usage in identifying trait associations for genes and their transcripts.

Simulations support improved power and calibrated null of isoTWAS across genetic architectures

To systematically explore the effects of varied genetic architecture on the performance of isoTWAS and TWAS, we conducted extensive simulations using European-ancestry SNP data from the 1000 Genomes Project[54] (**Methods, Figure 4a, Supplemental Figure S14**). In general, simulations reinforce results from real data: isoTWAS increases prediction of total gene expression and power to detect trait associations, especially when genetic effects on expression are different across isoforms of a gene. First, we assessed scenarios in which isoTWAS improves prediction of gene expression over TWAS. Similar to results from real data, the multivariate prediction models (particularly multivariate SuSiE[38]) provided the largest average adjusted R^2 for isoform expression prediction across most simulation settings (**Supplemental Figure S15, Supplemental Data 5**). At sparser isoQTL architectures, the optimal isoTWAS model greatly outperforms the optimal TWAS model, with the 25% and 75% quantiles of adjusted R^2 increasing by 1.3- and 14-fold (**Figure 4b-c**, **Supplemental Figure S16-17, Supplemental Data 6**). Performance gains decreased somewhat with denser isoQTL architectures, although real data reflects this 0.1-1% sparsity (i.e., 1-10 causal e- and isoQTLs per gene or isoform)[33]. Similar performance gains were observed from simulations in a second simulated test locus around *XRN2* (**Methods**, **Supplemental Figure S18, Supplemental Data 7**).

*W*e next introduced GWAS traits into our simulation framework to benchmark the power and false positive rate (FPR) of isoTWAS (**Methods**). First, we found that the FPR is controlled at 0.05 for isoform-level mapping using ACAT, similar to that of gene-level mapping using TWAS (**Supplemental Figure S19, Supplemental Data 8**). For a simulated trait, we modeled causal effect architectures for a genomic locus with two isoforms under three main scenarios (**Methods**; **Figure 5, Supplemental Figure S20**): (1) where the true trait effect is from total gene expression, (2) where there is only one isoform with a non-zero effect on the trait, called the 'effect isoform'; and (3) where both isoforms are 'effect' isoforms, with varying magnitudes and directions of association. This first scenario shows clear increases in power for TWAS over isoTWAS, but this advantage of TWAS over isoTWAS decreases as the causal proportion of QTLs increases (**Figure 5a, Supplemental Data 9**). As effects on the trait varied for isoforms for the same genes (**Figure 5b-c**, **Supplemental Data 10-11**), we observed stark increases in power for isoTWAS over TWAS. Here, across most scenarios and causal effect architectures, isoTWAS demonstrated improved power compared with gene-level TWAS, particularly when only one isoform of a gene carries a trait effect or when two effect isoforms of the same gene have different directions of effects. Our rigorous simulations reinforce the prediction and power advantages of isoTWAS over TWAS, suggesting varied isoQTL architecture and isoform effects on traits for isoforms of the same gene as key features that may drive advantages we observed in real data.

isoTWAS identifies biologically-meaningful trait associations undetectable by TWAS

Lastly, we applied probabilistic isoform-level fine-mapping to genetic loci where isoTWAS prioritized multiple isoform-trait associations to estimate sets of isoforms that are most likely to explain the trait association at the genetic locus (**Methods**). The full number of gene-level and isoform-level trait associations at each step of the testing pipeline are summarized in **Figure 6a**. In total, forming 90% credible sets explain the association with the trait at the genetic locus, using adult frontal cortex isoform expression, we detected trait associations at 25 genes (26 isoforms) for ADHD, 19 genes (21 isoforms) for ALZ, 17 genes (18 isoforms) for ASD, 28 genes (28 isoforms) for BD, and 136 genes (136 isoforms) for SCZ. Using fetal cortex isoform expression, we detected trait associations at 98 genes (105 isoforms) for ADHD, 20 genes (21 isoforms) for ALZ, 4 genes (4 isoforms) for ASD, 22 genes (22 isoforms) for BD, and 65 genes (66 isoforms) for SCZ (**Supplemental Figure S21-S25**, **Supplemental Tables S8-9**). Overall, we find, for isoforms associated with one of the 5 traits at nominal $P < 0.05$ using both adult and fetal frontal cortex expression, there is only a moderate correlation in isoTWAS effect sizes (r

= 0.63, P = 4.4 x 10^{-11}, **Figure 6b**). This level of correlation may suggest that there is a difference in isoform-specific effects in the developing brain, compared to the adult brain, on these neuropsychiatric traits.

Across all 5 traits, isoTWAS detected 15 trait-associated genes using fetal frontal cortex models and 42 genes using adult cortex models with gnomAD-calculated probability of loss-of-function intolerance (pLI) score of ≥ 0.8 (**Supplemental Tables S8-9**), suggesting that these genes are particularly intolerant to protein-truncating variation[55] and that isoTWAS detects biological-relevant signal due to its isoform-specific focus. Transcripts of many of these genes also house GWAS-significant variants within 10 kilobases, with effect sizes and 95% confidence intervals shown in **Figure 6c**. For example, we identified an association with SCZ with imputed fetal expression of ENST00000586144, an isoform of *CXXC1* (P = 2.56 x 10^{-11} in isoTWAS, pLI = 0.90, chromosomal location of 18q21.1), with a GWAS-significant variant within 10 kilobases upstream of the gene's transcription start site. *CXXC1* belongs to the SET1-COMPASS complex, implicated in transcriptomic pathways underlying schizophrenia risk[56]. Another transcript whose imputed fetal brain expression was associated with SCZ risk is ENST00000511382, an isoform of the gene *PDE4D* (P = 2.4 x 10^{-8}, pLI = 1, chromosomal location of 5q11.2-q12.1), with a GWAS SNP within the gene body (rs7701440). A previous GWAS has implicated *PDE4D* in associations with SCZ risk[57], but our analysis directly ties the genetically-regulated component of its isoform ENST00000511382 to SCZ risk through probabilistic fine-mapping. Imputed adult brain expression of ENST00000492146 (P = 6.9 x 10^{-12}, pLI = 0.94, chromosomal location of 3p21.1), an isoform of *SFMBT1*, showed a strong association with SCZ risk and contains a GWAS SNP within the gene body (rs2710339). In a recent gene-based analysis of GWAS data for SCZ and BD, decreased expression of *SFMBT1* was associated with increased risk of both disorders[58], consistent with the effects for the isoform we identify in this isoTWAS analysis. Lastly, overrepresentation analysis of isoTWAS-priortized genes for these 5 traits showed multiple relevant biological processes, cellular components, and pathways, at Bonferroni-adjusted P < 0.10: cation transportation, immune response ontologies, and plasma membrane and organelle organizations (**Figure 6d**). These results further emphasize that isoTWAS identifies sets of trait-associated genes with biological significance.

As evidenced by our simulation analyses, a main advantage of isoTWAS over TWAS is the identification of trait associations for isoforms of genes, where the gene itself is not associated with the trait. We illustrate a few examples of isoforms prioritized by isoTWAS, all using adult frontal cortex tissues, which show limited evidence of QTLs on the total gene expression level. First, we detected a strong SCZ association with ENST00000492957, an

isoform of *AKT3* (pLI = 1) at 1q43-q44, which encodes a serine/threonine-protein kinase that regulated many processes like metabolism and cell growth, proliferation, and survival. *AKT3* has shown effects on anxiety, spatial and contextual memory, and fear extinction in mice, such that loss-of-function of *AKT3* causes learning and memory deficits[59,60]. At this locus, we found a strong isoQTL signal ($P < 10^{-15}$) that intersects GWAS associations at $P < 5 \times 10^{-8}$; however, the eQTL signal failed to even reach $P < 10^{-4}$ (**Figure 7a**). The lead isoQTL rs2125232 shows a strong positive effect on ENST00000492957 but an insignificant positive association with *AKT3*, as the number of alternative alleles at the SNP increased. This isoQTL effect was in the opposite direction as the negative associations with SCZ from GWAS.

Similarly, we find a strong isoQTL signal for ENST00000373566 but a weak eQTL signal of *GIGYF2* in the 2q37.1 locus (pLI = 1), in another association with SCZ (**Figure 7b**). *GIGYF2* is a translation inhibitor, which, with the protein 4EHP, decreases translation of defective messenger RNAs to assist ribosome-associated quality control; decreased ribosomal quality is a hypothesized mechanism explaining neuropsychiatric disorders, as observed in both mice and humans with decreased *GIGYF2* function[61]. Interestingly, the lead SNP rs2675972 has a positive effect on both isoform and a negative effect on SCZ. Lastly, our analysis implicated ENST00000537270, an isoform of *KMT5A* (pLI = 0.99, found in 12q24.31), in an association with SCZ risk **(Figure 7c)**. There is a strong isoQTL signal at the locus with no significant eQTL signal; the lead isoQTL rs28577594 has a strong negative effect of the transcript-isoform expression and a protective effect on SCZ risk. *KMT5A*, a H4K20 methyltransferase[62], has been previously implicated in GWAS for SCZ but *cis*-eQTLs of the gene have not be colocalized with the GWAS signal previously[46,63].

Lastly, our isoTWAS analysis detected an association between ALZ and ENST00000252491, an isoform of *APOC1* in the 19q13.32 locus, a gene implicated in ALZ risk[64,65], with *APOE* and *ACE*, associated genes in the locus (**Supplemental Figure S26**). Interestingly, when we condition on *APOE* risk variants strongly associated with ALZ risk (rs7412, rs429358)[48,66] using analyses highlighted in Gusev et al[9,67], we find that the association between ENST00000252491 and ALZ persists (without adjustment for *APOE* variants: $P = 1.9 \times 10^{-7}$; with adjustment for *APOE* variants: $P = 2.1 \times 10^{-6}$). Despite limitations in these conditional analyses using GWAS summary statistics[68], our results may suggest an independent signal in this locus with an underlying transcriptomic mechanism through ENST00000252491.

These SNP-by-SNP views into the GWAS, eQTL, and isoQTL signal at a locus demonstrate why isoTWAS gives a clearer picture of the mechanism at the locus over TWAS. Integration with total gene expression cannot detect an overlapping genetic, transcriptomic, and phenotypic signal, especially at these sample sizes. When we specifically model the isoform transcripts belonging to the same gene family, we detect the more nuanced isoform-specific signals that may underlie a genetic association.

DISCUSSION

In this work, we introduce isoTWAS, a scalable framework that integrates genetic and isoform-level transcriptomic variation with GWAS to not only identify genes whose expression are associated with a trait but also prioritize a set of isoforms of the gene that best explains the association. We provide all isoform-level predictive models (https://zenodo.org/record/6795947[69]) and software to conduct isoform-level trait mapping with GWAS summary statistics (https://github.com/bhattacharya-a-bt/isoTWAS).

isoTWAS presents a number of advantages over TWAS. First, by modeling transcript expression on the isoform-level, it can detect isoQTL architecture that varies across isoforms and thus are not reflected in the eQTLs of a gene. This isoform-level approach to prediction showed improved predictive power both in simulated and real data. Second, by aggregating isoform-level associations to the gene-level through ACAT, isoTWAS showed substantially increased power to detect trait associations over traditional TWAS. We attribute this increase in power under several key features: first, isoform-level modeling in isoTWAS increases the number of imputable genes by >2-fold; second, isoTWAS models improve accuracy of gene-level prediction by as much as 35%; third, isoTWAS jointly models both expression and splicing regulation, thereby capturing multiple potential underlying complex trait mechanisms; finally , as genetic control of isoform expression and usage are often much more tissue-specific than eQTLs[25,33], tissue-specific genetically-controlled transcriptomic effects on traits can be more accurately mapped.

Recent work has highlighted the promise of alternative splicing as a mechanism underlying complex trait biology not fully captured through eQTLs[23,25,26,70]. Splicing-QTL integration is a promising vehicle to prioritize alternative exons or splice sites that may explain the genetic association with a trait, but a single exon or splice site can correspond to multiple isoforms of the same gene. Mapping genetic regulation of transcriptomics at the exon- rather than gene-level often leads to more detected signal[71]. However, the overwhelming majority of these

analyses have focused on splicing events or exon-level quantification of expression, rather than the transcriptomic consequences of these events: different isoforms of the same gene. Local splicing events can be computationally intensive to measure and are difficult to systematically integrate across multiple distinct large-scale datasets, which is necessary for achieving sample sizes sufficient for interrogation of population-level allelic effects[26,27]. Furthermore, multiple splicing changes are often coordinated across a gene, yielding a large number of non-independent features that increases multiple testing burden. Our results show that isoform-centric study using isoTWAS increases discovery compared to splice-site-based analysis, but these methods may recover some independent signal. Future work should consider integrating annotation-based and -free quantifications of the transcriptomic and splicing patterns to develop more nuanced mechanistic hypotheses for GWAS loci.

We conclude with some limitations of and future considerations for isoTWAS. First, we note that isoform-level expression quantifications are maximum-likelihood estimates, due to the inherent limitations of short-read RNA-seq. These estimates are generally guided by existing transcriptome annotations. As such, these estimates will be contingent on the completeness and accuracy of the annotations (e.g,. GENCODE). Further, many dataset specific sequencing factors may will affect the accuracy of these estimates, especially sequencing depth, read length and library preparation (mRNA vs total RNA sequencing). The rapid emergence of long-read sequencing technologies, including PacBio IsoSeq and Oxford Nanopore, will be instrumental for improving transcriptome annotations, particularly with tissue specificity. Further, as these methods become cost effective to be performed at population-scale datasets, they will replace the need for short-read sequencing and isoform estimation. Recent analyses have shown that larger sample sizes may outweigh sequencing coverage for eQTL mapping, but this same relationship has not been investigated for isoform-level expression[72]. Thus, for optimal discovery with isoTWAS, larger sample sizes or higher sequencing depth in the reference expression dataset may be imperative. Additionally, it is important to note that transcript quantification is heavily dependent on accurate genomic and transcriptomic annotations. As these annotations improve with more transcript discovery, the ability of isoTWAS to detect gene- and isoform-trait associations should improve.

Second, while inferential replicates from RNA-seq quantification can provide measures of technical variation, they are not incorporated into the predictive models. These inferential replicates are unlikely to affect the point estimates of SNP effects on isoforms but may aid in estimation of the precision of these SNP effects. A more flexible predictive model that estimates standard errors by model averaging across the replicate datasets may

help with trait mapping by providing a prediction interval for isoform- and gene-level GReX. Third, just as TWAS can be cast as a differential gene expression analysis conducted with GReX, isoform-level trait mapping is akin to differential transcript expression analysis. An analogous goal of isoTWAS can be to detect trait associations with genetically-regulated transcript usage. However, it is unclear if the compositional nature of transcript usage data needs to accounted for at the prediction step or the trait mapping steps[73]. Lastly, this framework can suffer from reduced power and inflated false positives in the presence of SNP horizontal pleiotropy, where the genetic variants in the isoform expression model affect the trait, independent of isoform expression[68,74]. Probabilistic fine-mapping may reconcile any pleiotropy for SNPs shared across models for multiple isoforms at the same genetic locus. However, for pathways that are not observed or accounted for in the reference expression panel and GWAS, accounting for horizontal pleiotropy may improve trait mapping. However, as shown in previous analyses, summary-statistic based methods that control for horizontal pleiotropy are not yet effective[75].

In sum, isoTWAS provides a novel framework to fully and scalably interrogate the transcriptomic mechanisms underlying genetic associations with complex traits and generate biologically-meaningful and testable hypotheses about disease risk mechanisms. Based on our results, we emphasize a shift in focus from quantifications of the transcriptome on the gene-level to the transcript-isoform level to maximize discovery of transcriptome-centric genetic associations with complex traits.

METHODS

isoTWAS consists of three steps: (1) training predictive models of isoform expression, (2) imputing Genetically-Regulated eXpression (GReX) into a separate GWAS panel, and (3) estimating the association between GReX and a phenotype (**Figure 1b**). isoTWAS contrasts with TWAS as it models correlations between the expression of isoforms of the same gene. We outline each step below, with further mathematical details in **Supplemental Methods**.

Training predictive models of isoform expression

Model and assumptions

Assume a given gene G has M isoforms with expression levels across N samples, with each sample having R inferential replicates. Let Y_G^* be the $N \times M$ matrix of mean isoform expression for the N samples and M isoforms, using the point estimates from the Expectation-Maximization algorithm of a pseudo-mapping quantification

algorithm, like salmon or kallisto[30,31]. We can jointly model isoform expression with a system of $N \times M \times R$ equations. For sample $n \in \{1, \dots, N\}$, isoform $m \in \{1, \dots, M\}$ of gene G, and replicate $r \in \{1, \dots, R\}$, we have:

$$y_{nmr} = x_n \beta_m + \epsilon_{nmr} \quad (1),$$

where y_{nmr} is the expression of isoform m for the rth inferential replicate of sample n, x_n is the P-vector (vector of length P) of *cis*-genotypes in a 1 Mb window around gene G, β_m is the P-vector of genetic effects of the P genotypes on isoform expression, and ϵ_{nmr} is normally distributed random noise with mean 0 and variance σ^2_{nmr}. We standardize both the genotypes and the isoform expression to mean 0 and variance 1. As the SNP vector x_n does not differ across inferential replicates, we impose the following assumptions on the variance-covariance matrix of $\epsilon = \left(\epsilon_{1,1,1}, \epsilon_{1,1,2}, \dots, \epsilon_{nmr}\right)^T$, the vector of random errors: we assume that ϵ_{nmr} are independent and identically distributed across samples $n \in \{1, \dots, N\}$ and identically distributed across replicates $r \in \{1, \dots, R\}$. Accordingly, the point estimates of the SNP effects on isoform expression are not influenced by differences in expression across replications. Therefore, in matrix form, we consider the following predictive model:

$$Y_G^* = X_G B_G + E_G \quad (2)$$

Here, X_G is the $N \times P$ matrix of genotype dosages, B_G is the $P \times M$ matrix of SNP effects on isoform expression and E_G is a matrix of random errors, such that $vec(E_G) \sim N_{NM}(0, \Sigma = \Omega^{-1} \otimes I_N)$. Σ represents the variance-covariance matrix in the errors (with precision matrix $\Omega = \Sigma^{-1}$), following the aforementioned independence assumptions.

Estimating SNP effects on isoform expression

We apply 6 different methods to estimate $\hat{B}_G$, the matrix of SNP effects on isoform expression. The $\hat{B}_G$ matrix that gives the largest adjusted R^2 in 5-fold cross-validation across the 6 methods is selected as the final model to predict isoform expression for a given gene. We provide an overview of these methods here, with mathematical details in **Supplemental Methods**:

1. *Univariate FUSION*: the simplest implemented method is the univariate predictive modelling used in FUSION[2]. We disregard the correlation structure between isoforms and train a univariate elastic net[36], estimation of the best linear unbiased predictor (BLUP) in a linear mixed model[76], and SuSiE[38] predictive model for each isoform separately. The model with the largest adjusted R^2 out of these three models is outputted.

2. *Curds-and-whey (CW) procedure*: Breiman and Friedman's CW approach takes in univariate estimates of SNP effects and modifies the predicted values from those regressions by shrinking them using the canonical correlations between the isoform expression variables and the SNP dosages[34]. This process significantly reduces prediction error when there exist large correlations between responses.
3. *Multivariate elastic net (MVEnet) regression*: MVEnet is an extension of elastic net, where the response is a matrix of correlated responses[36]. Here, the absolute penalty is imposed on each single coefficient by a group-lasso penalty on each vector of SNP effects across isoforms (rows of B_G); accordingly, a SNP can only have a non-zero effect on an isoform if it has a non-zero effect on all isoforms.
4. *Multivariate LASSO Regression with covariance estimation (MRCE)*: We adapt Rothman et al's proposed procedure to simultaneously and iteratively estimate both $\hat{B}_G$, the SNP effects matrix, and $\hat{\Omega}$, the precision matrix[37]. This procedure accounts for the correlation between isoforms but does not impose the group-lasso penalty as in MVEnet. As such, a single SNP need not have a non-zero effect on all isoforms.
5. *Multivariate SuSiE (mvSuSiE)*: mvSuSiE extends the SuSiE framework to a multivariate response[38], providing SNP effect sizes on expression of each isoform and posterior inclusion probabilities for each SNP. These effect sizes can be incorporated in a predictive model of the isoform expression matrix. Again, mvSuSiE does not impose a group penalty, as in MVEnet.
6. *SuSiE fine-mapping and ordinary least squares with clustered standard errors*: Here, we use SuSiE to generate a 90% credible set of causal SNPs affect isoform expression[38]. We then fit a seemingly unrelated regression[77] with a multivariate response using a design matrix that only includes SNPs found in these credible sets. We compute Liang-Zeger clustered standard errors for the SNP effects to account for inferential uncertainty across replicates[39]. These standard errors can be incorporated to calculate prediction uncertainty when imputing GReX in individual-level genotypes.

Step-wise association testing

We note that the tests of association in isoTWAS are similar to tests in differential transcript expression analyses, as TWAS tests of association are analogous to tests in differential gene expression analyses. isoTWAS and TWAS are distinct, as these methods consider imputed isoform and gene expression, respectively, as predicted by the trained expression models. If individual-level genotypes are available in the external GWAS panel, GReX can be directly imputed by multiplying the SNP weights from the predictive model with the genotype dosages in the GWAS panel. If only summary statistics are available, we adopt the weighted burden test from Gusev et al

with an ancestry-matched linkage disequilibrium panel[4,68]. Compared to TWAS, a main distinction for isoTWAS association testing is the increased number of tests (approximately 4-fold the number of isoforms than genes)[35] and the potentially correlation in test statistics for isoforms of the same gene.

Accordingly, we adopt methods from the stageR framework (**Supplemental Figure S2**)[78]. In the first step, for every isoform with a significant model, we generate a TWAS test statistic using either linear regression with direct GReX imputation for GWAS with individual-level genotypes or the weighted burden test for GWAS with only summary statistics[4]. Given the t test statistics $T_1, \ldots, T_t$ for a single gene, we used an omnibus test to aggregate the t test statistics into a single P-value for a gene. We benchmark different omnibus tests in simulations, but the default omnibus test in isoTWAS is aggregated Cauchy association test (ACAT)[41]. We control for false discovery across all genes via the Bonferroni procedure, but the Benjamini-Hochberg procedure can also be applied for less conservative false discovery control. In the second step, for isoforms for genes with FDR-adjusted omnibus $P < 0.05$, we employ Shaffer's modified sequentially rejective Bonferroni (MSRB) procedure to control the within-gene family-wide error rate. At the end of these two steps, we identify a set of genes and their isoforms that are associated with the trait.

Control for false positives within GWAS loci

In TWAS and related methods, association statistics have been shown to be well calibrated under the null of no GWAS association. However, within loci harboring significant GWAS signal, false positive associations can result when eQTLs and GWAS coincide within overlapping LD blocks. To address this, we have adopted two conservative approaches to control for type 1 error within GWAS loci, namely (1) permutation testing and (2) probabilistic fine mapping. The permutation testing approach, adopted from Gusev et al[4], is a highly conservative test of the signal added by the SNP-transcript effects from the predictive models, conditional on the GWAS architecture of the locus. Briefly, we shuffle the SNP-transcript effects in the predictive models 10,000 times and generate a null distribution for the TWAS test statistic for each isoform. We can use the null distribution to generate a permutation P-value for the original test statistic for each isoform. Additionally, we can aggregate these null distributions using an omnibus test and compare the omnibus P-value to this distribution to generate permutation P-value for each gene. In addition, we employ gene-level probabilistic fine-mapping using methods from FOCUS[79] to generate credible set of isoforms that explain the trait association at a locus. We only run isoform-level fine-mapping for significantly-associated isoforms in overlapping 1 Megabase windows.

GTEx processing and model training

We quantified GTEx v8[33] RNA-Seq samples for all 49 tissues using Salmon v1.5.2[30] in mapping-based mode. We first built a Salmon index for a decoy-aware transcriptome consisting of GENCODE v38 transcript sequences and the full GRCh38 reference genome as decoy sequences[35]. We then extracted reads from the available genome-aligned GTEx BAM files into FASTQ files and used the FASTQ files as input to Salmon with mapping validation and corrections for sequencing and GC bias. We estimated 50 inferential bootstraps for isoform expression using Salmon's Expectation-Maximization algorithm. We then imported Salmon isoform-level quantifications and aggregate to the gene-level using tximeta[44]. Using edgeR, gene and isoform-level quantifications underwent TMM-normalization, followed by transformation into a log-space using the variance-stabilizing transformation using DESeq2[80,81]. We then residualized isoform-level and gene-level expression (as log-transformed CPM) by all tissue-specific covariates (clinical, demographic, genotype PCs, and expression PEER factors) used in the original QTL analyses in GTEx.

SNP genotype calls were derived from Whole Genome Sequencing data for samples from individuals of European ancestry, filtering out SNPs that have less than 5% minor allele frequency and deviate from Hardy-Weinberg equilibrium at $P < 10^{-5}$. We further filtered out SNPs with less than 1% minor allele frequency among the European ancestry samples in 1000 Genomes Project[54].

Details of the model training pipeline for GTEx are similar to those summarized in **Supplemental Figure S10**. Before training genetics models of gene-level and isoform-level expression, we estimated the heritability of gene and isoform expression using all SNPs within 1 Mb with GCTA-GREML[82,83]. We only considered genes and isoforms with positive heritability estimates with nominal $P < 0.05$. We borrowed techniques from FUSION[4] to train gene-level models using the most predictive, based on McNemar's adjusted 5-fold cross-validation (CV) R^2, of models trained using elastic net regression[36], linear mixed modeling[40], and SuSiE[38]. We selected only genes with CV $R^2 \geq 0.01$. We applied multivariate modeling outlined in isoTWAS to train isoform-level predictive models, selecting only those isoform models with CV $R^2 \geq 0.01$; we only used multivariate elastic net, multivariate SuSiE, univariate modeling across isoforms, and multivariate LASSO regression with covariance estimation, as these four methods showed far better predictive performance across all 6 methods (**Supplemental Figure S2**). For gene expression prediction in TWAS, we employed univariate elastic net, linear mixed modeling, and SuSiE, as in

FUSION[4,36,38,76], using all SNPs within 1 Mb of the gene body. All isoTWAS models generated are available at https://zenodo.org/record/6795947[69].

Adult and fetal PsychENCODE processing and model training

We quantified both adult (N = 850)[26,27] and fetal frontal cortex[23] (N = 205) RNA-Seq samples using Salmon v1.5.2[30] in mapping-based mode. We used the same Salmon indexed transcriptome used in the GTEx analysis. We then extracted reads from the available genome-aligned BAM files into FASTQ files and used the FASTQ files as input to Salmon with mapping validation and corrections for sequencing and GC bias. We estimated 50 inferential bootstraps for isoform expression using Salmon's Expectation-Maximization algorithm. We then imported Salmon isoform-level quantifications and aggregate to the gene-level using tximeta[44]. Using edgeR, gene and isoform-level quantifications underwent TMM-normalization, followed by transformation into a log-space using the variance-stabilizing transformation using DESeq2[80,81]. We then residualized isoform-level and gene-level expression (as log-transformed CPM) by covariates (age, sex, 10 genotype PCs, 90 and 70 hidden covariates with prior for fetal and adult brain, respectively) used in the original QTL analyses in Gandal et al and Walker et al.

Typed SNPs with non-zero alternative alleles, MAF > 1%, genotyping rate > 95%, HWE $P < 10^{-6}$ were first imputed to TOPMed Freeze 5 using minimac4 and eagle v2.4[84,85]. genotype calls were derived from Whole Genome We then filtered to SNPs that had imputation accuracy $R^2 > 0.8$, are biallelic, and have rsIDs. Furthermore, we filtered out SNPS with MAF < 0.05 and deviate from Hardy-Weinberg equilibrium at $P < 10^{-6}$.

Details of the model training pipeline are summarized in **Supplemental Figure S10** and are equivalent to those used to train models in GTEx data. All isoTWAS models generated are available at https://zenodo.org/record/6795947[69].

Gene- and isoform-level trait mapping

We conducted gene- and isoform-level trait mapping for five neuropsychiatric traits: (1) schizophrenia (SCZ; N_{cases} = 69,369; $N_{controls}$ = 236,642)[46], (2) bipolar disorder (BD; N_{cases} = 41,917; $N_{controls}$ = 371,549)[47], (3) attention deficit hyperactivity disorder (ADHD; N_{cases} = 20,183; $N_{controls}$ = 35,191)[50], (4) autism spectrum disorder (ASD; N_{cases} = 18,381; $N_{controls}$ = 27,969)[49], and (5) Alzheimer's disease (ALZ; N_{cases} = 90,338; $N_{controls}$ = 1,036,225)[48]. For gene-

level trait mapping, we used the weighted burden test, followed by the permutation test, as outlined by Gusev et al[4]. For isoform-level trait mapping, we used the stage-wise testing procedure outlined in the isoTWAS method. For isoforms, irrespective of their corresponding genes, passing both stage-wise tests and the permutation test, we employed isoform-level probabilistic fine-mapping using FOCUS with default parameters[79].

We estimate the percent increase in effective sample size by employing the following heuristic. We convert gene-level association P-values into χ^2 test statistics with 1 degree-of-freedom. For $\chi^2 > 1$, we then calculate the percent increase for isoTWAS-based associations versus TWAS-based associations. As the mean of the χ^2 distribution is linearly related to power and sample size[86], we can use this percent increase in test statistic as a measure of power or effective sample size.

Simulation framework

We adopt techniques from Mancuso et al's *twas_sim* protocol[79,87] to simulate multivariate isoform expression based on randomly simulated genotypes and environmental random noise. First, for n samples, we generate a matrix of genotype dosages for the SNPs within 1 Megabase of *CACNA1E* (1,107 total SNPs) using an LD reference panel from European samples from 1000 Genomes Project[54]. We use this 'gene for these simulations as it displays interesting patterns of alternative isoform expression[88]; however, we note that the selection of this particular gene is ultimately arbitrary. We extended our simulations to the locus around *XRN2*, as well, with similar results.

Next, we generate a matrix of SNP-isoform effects across different causal SNP proportions p_c, numbers of isoforms t, and "shared" or "different" SNP-isoform architectures. We define "shared SNP-isoform architecture as a setting where a SNP either has non-zero effects on all isoform or none at all, such as a case where a variant lies within a gene's promoter and acts as a broad gene-level *cis*-eQTL. We define "different" SNP-isoform architecture as one in which variant(s) exhibit distinct (zero or non-zero) effects across different isoforms, such as the case where a variant is proximal to an alternative transcription start site shared only by a subset of isoforms. Lastly, we compare "sparse" (all off-diagonal elements are near 0) or "dense" (most or all off-diagonal elements are far from 0 and generally greater than 0.3-0.4) correlation structures for the isoform expression levels within a gene locus, using Joe's C-vine method[89,90]. Using matched moments followed by scaling to ensure a set isoform expression heritability of h_g^2, we can generate the random error matrix that ensures

the desired correlation between isoforms. We generate 10,000 simulations for each configuration of the simulation parameters, varying $n \in \{200, 500\}$, $p_c \in \{0.001, 0.01, 0.10\}$, $h_g^2 \in \{0.10, 0.25\}$, "shared" and different SNP-isoform effects, and "sparse" and "dense" correlation between isoforms. Full mathematical details are provided in **Supplemental Methods** and summarized in **Figure 4a and Supplemental Figure S14**.

We also generate traits under three distinct scenarios, each with 2 isoforms for a given gene:

1. **Only gene-level expression has a non-zero effect on trait.** Here, we sum the isoform expression to generate a simulated gene expression. We randomly simulate the effect size and scale the error to ensure trait heritability $h_t^2 = 0.10$.
2. **Only 1 isoform has a non-zero effect on the trait.** Here, we generate a multivariate isoform expression matrix with 2 isoforms and scale the total gene expression value such that one isoform (called the effect isoform) makes up $p_g \in \{0.10, 0.30, 0.50, 0.70, 0.90\}$ proportion of total gene expression. We then generate effect size for one of the isoforms and scale the error to ensure trait heritability $h_t^2 = 0.10$.
3. **Two isoforms with different effects on traits**. Here, we generate a multivariate isoform expression matrix with 2 isoforms that make up equal portions of the total gene expression. We then generate an effect size of α for one isoform and $p_e\alpha$ for the other isoform, such that $p_e \in \{-1, -0.5, -0.2, 0.2, 0.5, 1\}$. We then scale the error to ensure trait heritability $h_t^2 = 0.10$.

Lastly, to estimate the approximate false positive rate (FPR), we followed the same simulation framework to generate eQTL data and GWAS data. In the GWAS data, however, we set the effect of gene- and isoform-level GReX to 0 to generate simulated trait under the null, where the gene and isoforms are not associated with the trait. We then estimate the FPR by calculating the proportion of gene-trait associations at $P < 0.05$ under this null across 20 sets of 1,000 simulated GWAS panels.

DATA AVAILABILITY

GTEx genetic, transcriptomic, and covariate data were obtained through dbGAP approval at accession number phs000424.v8.p2. Linkage disequilibrium reference data from the 1000 Genomes Project were obtained at this link: https://www.internationalgenome.org/data-portal/sample. GWAS summary statistics were obtained at the following links: ADHD (https://www.med.unc.edu/pgc/download-results/), ALZ

(https://ctg.cncr.nl/software/summary_statistics/), ASD (https://www.med.unc.edu/pgc/download-results/), BD (https://www.med.unc.edu/pgc/download-results/), and SCZ (https://www.med.unc.edu/pgc/download-results/). The RNA-seq and genotype dataset from Walker et al is available at dbGAP with accession number phs001900. The subset of data from the CommonMind Consortium from the PsychENCODE Consortium is available at doi.org/10.7303/syn12080241. isoTWAS models for GTEx and PsychENCODE are available at https://zenodo.org/record/6795947[69].

CODE AVAILABILITY

isoTWAS is available as an R package at https://github.com/bhattacharya-a-bt/isotwas. Sample scripts for analyses are available at https://github.com/bhattacharya-a-bt/isotwas_manu_scripts.

REFERENCES

1. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *The American journal of psychiatry* **175**, 15–15 (2018).
2. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **20**, 467–484 (2019).
3. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* **47**, 1091–1098 (2015).
4. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245–252 (2016).
5. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications* **9**, 1–20 (2018).
6. Hu, Y. *et al.* A statistical framework for cross-tissue transcriptome-wide association analysis. *Nature Genetics* **51**, 568–576 (2019).
7. Zhou, D. *et al.* A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nature Genetics* **52**, 1239–1246 (2020).
8. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genetics* **15**, e1007889 (2019).
9. Bhattacharya, A., Li, Y. & Love, M. I. MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide Association Studies. *PLOS Genetics* **17**, e1009398 (2021).

10. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nature Genetics* **50**, 538–548 (2018).
11. Wu, L. *et al.* A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nature genetics* **50**, 968–978 (2018).
12. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genetics* **10**, e1004383–e1004383 (2014).
13. Gleason, K. J., Yang, F., Pierce, B. L., He, X. & Chen, L. S. Primo: Integration of multiple GWAS and omics QTL summary statistics for elucidation of molecular mechanisms of trait-associated SNPs and detection of pleiotropy in complex traits. *Genome Biology* **21**, 236–236 (2020).
14. He, X. *et al.* Sherlock: Detecting gene-disease associations by matching patterns of expression QTL and GWAS. *American Journal of Human Genetics* **92**, 667–680 (2013).
15. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *American Journal of Human Genetics* **99**, 1245–1260 (2016).
16. Umans, B. D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends in Genetics* **37**, 109–124 (2021).
17. Barrera, L. O. *et al.* Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res* **18**, 46–59 (2008).
18. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
19. Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).
20. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
21. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).
22. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).
23. RL, W. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).
24. 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).

25. Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).
26. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).
27. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).
28. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biology 2021 22:1* **22**, 1–24 (2021).
29. MM, S. & MS, S. RNA mis-splicing in disease. *Nature reviews. Genetics* **17**, 19–32 (2016).
30. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).
31. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).
32. Akula, N. *et al.* Deep transcriptome sequencing of subgenual anterior cingulate cortex reveals cross-diagnostic and diagnosis-specific RNA expression changes in major psychiatric disorders. *Neuropsychopharmacol.* **46**, 1364–1372 (2021).
33. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
34. Breiman, L. & Friedman, J. H. Predicting Multivariate Responses in Multiple Linear Regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**, 3–54 (1997).
35. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766–D773 (2019).
36. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 1–22 (2010).
37. Rothman, A. J., Levina, E. & Zhu, J. Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics* **19**, 947–962 (2010).
38. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B* **82**, 1273–1300 (2020).

39. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).

40. Endelman, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* **4**, 250–255 (2011).

41. Liu, Y. *et al.* ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet* **104**, 410–421 (2019).

42. Van den Berge, K., Soneson, C., Robinson, M. D. & Clement, L. stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biology 2017 18:1* **18**, 1–14 (2017).

43. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research 2015 4:1521* **4**, 1521 (2016).

44. Love, M. I. *et al.* Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLOS Computational Biology* **16**, e1007664 (2020).

45. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLoS Genetics* **17**, e1009405 (2021).

46. Consortium, T. S. W. G. of the P. G., Ripke, S., Walters, J. T. & O'Donovan, M. C. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. 2020.09.12.20192922 Preprint at https://doi.org/10.1101/2020.09.12.20192922 (2020).

47. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet* **53**, 817–829 (2021).

48. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276–1282 (2021).

49. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431–444 (2019).

50. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics* **51**, 63–75 (2019).

51. van Iterson, M., van Zwet, E. W., Heijmans, B. T. & Heijmans, B. T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biology* **18**, 19–19 (2017).

52. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151–158 (2018).

53. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587–587 (2010).
54. Auton, A. *et al.* A global reference for human genetic variation. *Nature* vol. 526 68–74 Preprint at https://doi.org/10.1038/nature15393 (2015).
55. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
56. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci* **19**, 571–577 (2016).
57. Sinha, V. *et al.* Variants in regulatory elements of PDE4D associate with major mental illness in the Finnish population. *Mol Psychiatry* **26**, 816–824 (2021).
58. Reay, W. R. & Cairns, M. J. Pairwise common variant meta-analyses of schizophrenia with other psychiatric disorders reveals shared and distinct gene and gene-set associations. *Transl Psychiatry* **10**, 1–11 (2020).
59. Wong, H. *et al.* Isoform-specific roles for AKT in affective behavior, spatial memory, and extinction related to psychiatric disorders. *eLife* **9**, e56630 (2020).
60. Howell, K. R., Floyd, K. & Law, A. J. PKBγ/AKT3 loss-of-function causes learning and memory deficits and deregulation of AKT/mTORC2 signaling: Relevance for schizophrenia. *PLoS One* **12**, e0175993 (2017).
61. Hickey, K. L. *et al.* GIGYF2 and 4EHP Inhibit Translation Initiation of Defective Messenger RNAs to Assist Ribosome-Associated Quality Control. *Molecular Cell* **79**, 950-962.e6 (2020).
62. Nishioka, K. *et al.* PR-Set7 Is a Nucleosome-Specific Methyltransferase that Modifies Lysine 20 of Histone H4 and Is Associated with Silent Chromatin. *Molecular Cell* **9**, 1201–1213 (2002).
63. Schmidt-Kastner, R., Guloksuz, S., Kietzmann, T., van Os, J. & Rutten, B. P. F. Analysis of GWAS-Derived Schizophrenia Genes for Links to Ischemia-Hypoxia Response of the Brain. *Front Psychiatry* **11**, 393 (2020).
64. Lucatelli, J. F. *et al.* Genetic influences on Alzheimer's disease: evidence of interactions between the genes APOE, APOC1 and ACE in a sample population from the South of Brazil. *Neurochem Res* **36**, 1533–1539 (2011).
65. Zhou, Q. *et al.* Association between APOC1 Polymorphism and Alzheimer's Disease: A Case-Control Study and Meta-Analysis. *PLoS One* **9**, e87017 (2014).
66. Ware, E. B., Faul, J. D., Mitchell, C. M. & Bakulski, K. M. Considering the APOE locus in Alzheimer's disease polygenic scores in the Health and Retirement Study: a longitudinal panel study. *BMC Medical Genomics* **13**, 164 (2020).

67. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245–252 (2016).
68. Bhattacharya, A. *et al.* Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: lessons from the Global Biobank Meta-analysis Initiative. *medRxiv* **3**, 2021.11.24.21266825-2021.11.24.21266825 (2021).
69. Bhattacharya, A. isoTWAS models using 48 GTEx tissues and PsychENCODE data (07.04.22). (2022) doi:10.5281/zenodo.6795947.
70. Qi, T. *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat Genet* (2022) doi:10.1038/s41588-022-01154-4.
71. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nature Communications* **8**, (2017).
72. Schwarz, T. *et al.* Powerful eQTL mapping through low coverage RNA sequencing. 2021.08.08.455466 Preprint at https://doi.org/10.1101/2021.08.08.455466 (2021).
73. Doose, G., Bernhart, S. H., Wagener, R. & Hoffmann, S. DIEGO: detection of differential alternative splicing using Aitchison's geometry. *Bioinformatics* **34**, 1066–1068 (2018).
74. Veturi, Y. & Ritchie, M. D. How powerful are summary-based methods for identifying expression-trait associations under different genetic architectures? in *Pacific Symposium on Biocomputing* vol. 0 228–239 (World Scientific Publishing Co. Pte Ltd, 2018).
75. Zhu, A. *et al.* MRLocus: Identifying causal genes mediating a trait through Bayesian estimation of allelic heterogeneity. *PLOS Genetics* **17**, e1009455–e1009455 (2021).
76. Endelman, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* **4**, 250–255 (2011).
77. Zellner, A. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* **57**, 348–368 (1962).
78. Van den Berge, K., Soneson, C., Robinson, M. D. & Clement, L. stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biology 2017 18:1* **18**, 1–14 (2017).
79. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nature Genetics* **51**, 675–682 (2019).

80. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47–e47 (2015).
81. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550–550 (2014).
82. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics* **47**, 1114–1120 (2015).
83. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76–82 (2011).
84. Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genetics* **15**, e1008500–e1008500 (2019).
85. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* **48**, 1443–1448 (2016).
86. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nature Communications* **10**, 3834–3834 (2019).
87. Mancuso, N. twas_sim. (2021).
88. Lipscombe, D. & Andrade, A. Calcium Channel CaVα1 Splice Isoforms - Tissue Specificity and Drug Action. *Curr Mol Pharmacol* **8**, 22–31 (2015).
89. Qiu, W. & Joe, H. clusterGeneration: Random Cluster Generation (with Specified Degree of Separation). (2020).
90. Joe, H. Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis* **97**, 2177–2189 (2006).

FIGURE LEGENDS

Figure 1: *Isoform-centric approach for complex trait mapping and prioritization of disease mechanisms at a genetic locus.* **(a)** Motivation for isoTWAS. Gene G has three isoforms associated with a trait but only one has an effect on the trait. Gene G itself does not show an association with the trait. Studying genetic associations with an isoform-centric perspective will prioritize Gene G, but not with a gene-centric perspective. **(b)** Schematic comparison of isoTWAS and TWAS. First, isoTWAS differs from TWAS by training a multivariate model of isoform expression, while TWAS models total gene expression, the sum of isoform expression. Second, isoTWAS maps

isoform-trait associations through a step-wise hypothesis testing framework that provides gene-level false discovery control and isoform-level family-wide error rate control. TWAS only maps gene-trait associations. The advantages of isoTWAS over TWAS are summarized.

Figure 2: *Multivariate isoform-level predictive models improve upon gene-level predictive models in predictive gene-level expression.* **(a)** Schematic diagram for predictive modeling and comparison to gene-level expression using gene-level TWAS and isoTWAS models. **(b)** Barplot showing the number of genes passing heritability (error bar) and cross-validation (CV) cutoffs (bar). TWAS heritability (or CV) cutoff is nominal $P < 0.05$ (CV $R^2 > 0.01$). isoTWAS heritability (or CV) cutoff is a gene with at least 1 isoform that is heritabie at $P < 0.05$ (predictable at CV $R^2 > 0.01$) **(c)** For 13 GTEx brain tissues, the percent difference in CV R^2 of predicting total gene expression using isoTWAS versus TWAS models. The median percent increase is labelled and the red line marks 0% difference. **(d)** Using completely independent GTEx frontal cortex and adult PsychENCODE frontal cortex models, the percent difference in external predictive R^2 using either GTEx or PsychENCODE as the training sample. **(e)** Using 48 GTEx tissues, ratio of the number of cis-heritable isoforms over the number of cis-heritable genes in brain tissues (purple) and other tissues (brown). F-test shows the difference in means between the two groups, adjusting for sample size for the tissue.

Figure 3: ***Isoform-level trait mapping increases discovery of genetic associations over gene-level trait mapping.*** **(a)** Schematic diagram for trait mapping using gene-level TWAS and isoTWAS using PsychENCODE data. **(b)** Number of gene-trait associations overlapping GWAS risk SNPs within 0.5 Mb using gene-level TWAS (red), isoTWAS (blue), or either (green) at conservative permutation-based significance thresholds. **(c)** Scatterplot of standardized effect sizes (Z-scores) using isoTWAS and gene-level TWAS, considered associations with nominal $P < 0.05$ using both TWAS and isoTWAS. Gray line shows the 45-degree line and the green line shows an ordinary least squares regression. **(d)** Mean percent increase in approximate χ^2-test statistic (squared Z-score), which is proportional to increase in effective sample size, for isoTWAS trait associations over TWAS trait associations. **(e)** Number of gene-trait associations overlapping GWAS risk SNPs within 0.5 Mb using splicing-event-based TWAS (purple), isoTWAS (blue), or either (orange) at conservative permutation-based significance thresholds.

Figure 4: *IsoTWAS models predict gene expression with more accuracy than TWAS models*. **(a)** Simulation setup to generate isoform expression with specified isoQTL architecture, controlled expression heritability, number of isoforms, and inter-isoform correlation structure. See Supplemental Figure S3 for more details. Boxplots of difference in adjusted R^2 in predicting gene expression between isoTWAS and TWAS models from simulations with sample size 500 where isoform and gene expression heritability are set to **(b)** 0.10 and **(c)** 0.25. Here, we sum predicted isoform expression from isoTWAS models to form predicted gene expression and compare to predicted gene expression from a TWAS model.

Figure 5: *IsoTWAS improves power to detect gene-trait associations, especially when genetic effects differ across isoforms.* **(a)** Simulation settings for gene- and isoform-level associations with complex trait (**Methods**). **(b)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across causal proportion of isoQTLs (X-axis), facetted by sparse and dense correlations between isoforms (right margin) and shared or different isoQTLs across isoforms (top margin). **(c)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across proportion of gene expression explained by effect isoform (X-axis), facetted by sparse and dense correlations between isoforms and shared or different isoQTLs across isoforms. **(d)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across ratio of effect sizes for 2 effect isoforms, facetted by sparse and dense correlations between isoforms and shared or different isoQTLs across isoforms. Across all plots, isoform and total gene expression heritability is set to 0.10 with 2 simulated isoforms and causal proportion of 0.01, unless otherwise specified.

Figure 6: *isoTWAS detects biologically-meaningful trait associations with ADHD, ALZ, ASD, BD, and SCZ using adult and fetal brain tissue.* **(a)** Number of gene and isoform (vertical margin) associations across various significance thresholds (color) in both adult and fetal brain (horizontal margin). **(b)** Scatter plot of isoTWAS effect sizes across adult (Y-axis) and fetal (X-axis) brain for isoform-trait associations. Black diagonal line shows the 45-degree line and the blue line shows the regression line. **(c)** Forest plot of isoTWAS effect sizes and 95% confidence intervals (X-axis) for isoforms that are within 10 kilobases of a significant GWAS SNP at $P < 5 \times 10^{-8}$. **(d)** Lollipop plot of enrichment ratio of various gene ontologies overrepresented by isoTWAS-identified genes at Bonferroni-adjusted $P < 0.10$.

Figure 7: *isoTWAS implicates isoforms of AKT3, GIGYF2, and KMT5A in genetic associations with schizophrenia risk.* In each plot, (top) Manhattan plots of GWAS risk, total gene eQTLs, and isoQTLs, colored by LD to the lead isoQTL and lead isoQTL shown in a triangle and labelled. (bottom) Boxplots of gene and isoform expression by genotype of the lead isoQTL SNP and forest plot of the lead isoQTL's effect and adjusted 95% confidence interval on the trait, gene, and isoform, with P-values labelled. **(a)** SCZ risk, *AKT3* gene expression, and ENST00000492957 isoform expression. **(b)** SCZ risk, *GIGYF2* gene expression, ENST00000373566 isoform expression. **(c)** SCZ risk, *KMT5A* gene expression, ENST00000537270 isoform expression.

SUPPLEMENTAL INFORMATION

Supplemental Methods

Supplemental Methods are provided in the attached SupplementalMethods.pdf.

Supplemental Figure Legends

Figure S1: *Directed acyclic graph (DAG) illustrating causal assumptions in isoTWAS.* We assume that the local genetic variants within 1 Megabase of a gene have direct effects on the expression of a gene G and its isoforms; these genetic effects need not be shared across isoforms and the gene. Furthermore, we assume that the abundance of a gene is the sum of abundances of its isoforms. Lastly, we assume that the isoform and gene need not affect the complex trait through the same pathway. We also acknowledge that genetic variants may have effects on the trait through pathways independent of gene and isoform expression.

Figure S2: *Step-wise hypothesis testing in isoTWAS.* First, isoform-trait associations are estimated either using appropriate regressions of imputed expression in individual-level GWAS or the weighted burden test in Gusev et al in GWAS summary statistics. Then, associations for isoforms of the same gene are aggregated to the gene-level using the Aggregated Cauchy Association Test (ACAT). These aggregated gene-level associations are adjusted for multiple testing burden to control the false discovery rate (FDR) using either a Bonferroni or Benjamini-Hochberg procedure. Lastly, for isoforms of genes that pass gene-level testing, we control the family-wide error rate (FWER) using Shaffer's modified sequentially rejective procedure.

Figure S3: Across GTEx tissues (X-axis), proportion (Y-axis) of isoform expression models that show highest R^2 using one of four models: MRCE, multivariate elastic net, multivariate SuSiE, univariate modeling.

Figure S4: *Comparison of number of cis-heritable genes on at gene and isoform expression level in GTEx*. Across all GTEx tissues (Y-axis), number (X-axis) of genes that are cis-heritable on the total gene expression level (red), with at least 1 isoform that is cis-heritable (blue).

Figure S5: *Comparison of number of predictable genes on isoform expression and total gene expression level*. Across all GTEx tissues (Y-axis), number (X-axis) of genes that are predicted at $R^2 > 0.01$ on the total gene expression level (red) or with at least 1 isoform that is predicted at $R^2 > 0.01$ (blue).

Figure S6: *Comparison of distribution of total gene expression predictive R^2 using TWAS and isoTWAS across all GTEx tissues.* Across all GTEx tissues (Y-axis), boxplots of adjusted R^2 of predicted total gene expression using gene-level TWAS models (red) and isoTWAS models (blue).

Figure S7: *Difference in total gene expression predictive R^2 using TWAS and isoTWAS across all GTEx tissues.* Across all GTEx tissues (Y-axis), boxplots of difference in adjusted R^2 of predicted total gene expression using gene-level TWAS models vs. isoTWAS models. The label shows the proportion of genes that is better predicted using isoTWAS models.

Figure S8: Boxplot of ratio of number of isoforms predicted at $R^2 > 0.01$ using isoTWAS models over number of genes predicted at $R^2 > 0.01$ using TWAS models across 13 brain tissues (purple) and 35 other tissues (brown) in GTEx.

Figure S9: Number (X-axis) of unique genes with at least one isoform (left) and unique isoforms (right) predicted with McNemar's adjusted $R^2 \geq 0.01$ for each tissue (Y-axis) represented in GTEx.

Figure S10: *Schematic diagram for analysis using adult and fetal frontal cortex data from PsychENCODE*. Data sources for eQTL reference data, GWAS cohorts, and reference LD data are provided on the left (black). The full gene-level TWAS (red) and isoTWAS (blue) are summarized on the right.

Figure S11: *Numbers of trait associations detected using TWAS (red), isoTWAS (blue), and both (green) across two testing thresholds.* Using adult and fetal frontal cortex data (right margin), the number of trait associations detected at Bonferroni-adjusted P < 0.05 and permutation P < 0.05 (top margin) for 5 neuropsychiatric traits.

Figure S12: QQ-plots of gene-level P-values for gene-level TWAS (red) and isoTWAS (blue), with expected $-\log_{10}$ P-values on X-axis and observed $-\log_{10}$ P-values on Y-axis, across 5 neuropsychiatric traits. Estimates of test statistic inflation and 95% credible intervals using empirical Bayes methods from bacon.

Figure S13: *Isoform-level trait mapping increases discovery of genetic association over splicing-event level trait mapping.* **(a)** Number of gene-trait associations using splicing-event level TWAS (purple), isoTWAS (blue), or either (orange) at various testing thresholds, comparing genes with both splicing-event TWAS and isoTWAS models. **(b)** Number of gene-trait associations using splicing-event level TWAS (purple) and isoTWAS (blue) with a GWAS locus within 0.5 Mb.

Figure S14: *Detailed schematic of comparison of prediction between TWAS and isoTWAS in GTEx data.*

Figure S15: *Comparison of predictive performance of 6 models implemented in isoTWAS.* Across 1,000 simulations of 5 isoforms with isoform heritability (h_i^2) set to 0.10 **(a)** and 0.25 **(b)**. On the left, boxplots of adjusted R^2 of prediction of isoform expression (Y-axis) across causal isoQTL proportion (X-axis). On the right, plot of proportion of simulations where the model has the maximum adjusted R^2 (Y-axis) across causal isoQTLs proportion (X-axis).

Figure S16: *IsoTWAS models predict gene expression with more accuracy than TWAS models*. Boxplots of difference in adjusted R^2 in predicting gene expression between isoTWAS and TWAS models from simulations with sample size 200 (compare with sample size 500 in **Figure 2**), where isoform and gene expression heritability are set to **(a)** 0.10 and **(b)** 0.25. Here, we sum predicted isoform expression from isoTWAS models to form predicted gene expression and compare to predicted gene expression from a TWAS model.

Figure S17: *IsoTWAS models predict gene expression with more accuracy than TWAS models at lower expression heritability settings*. Boxplots of difference in adjusted R^2 in predicting gene expression between

isoTWAS and TWAS models from simulations with sample size 200 where isoform and gene expression heritability are set to 0.05. Here, we sum predicted isoform expression from isoTWAS models to form predicted gene expression and compare to predicted gene expression from a TWAS model.

Figure S18: *IsoTWAS models predict gene expression with more accuracy than TWAS models at simulated genetic locus around XRN2*. Boxplots of difference in adjusted R^2 in predicting gene expression between isoTWAS and TWAS models from simulations with sample size 200 where isoform and gene expression heritability are set to 0.05 **(a)**, 0.10 **(b)**, and 0.25 **(c)**, using the 1 Megabase window around *XRN2* as a reference. Here, we sum predicted isoform expression from isoTWAS models to form predicted gene expression and compare to predicted gene expression from a TWAS model.

Figure S19: *IsoTWAS and gene-level TWAS show relatively similar false positive rates*. Across 20 iterations of 1,000 simulations, boxplots of false positive rate to detect a gene-trait association using Cauchy-aggregated P-values of isoform-trait associations (red) and gene-level TWAS (blue). In these simulations, we simulate 200 and 5,000 samples in the eQTL and GWAS panels, 5 isoforms, and isoform and gene expression heritability of 0.10. We vary the causal isoQTLs proportion (p_c, shown on X-axis), the QTL architecture (right margin), and correlation between isoforms (top margin). For each iteration, we simulate one eQTL panel and 1,000 GWAS panels where genetically-regulated gene and isoform expression have no effect on the trait. We calculate the false positive rate as the proportion of the 1,000 tests that give $P > 0.05$.

Figure S20: *At low expression heritability settings, isoTWAS improves power to detect gene-trait associations, especially when genetic effects differ across isoforms.* **(a)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across causal proportion of isoQTLs (X-axis), facetted by sparse and dense correlations between isoforms (right margin) and shared or different isoQTLs across isoforms (top margin). **(b)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across proportion of gene expression explained by effect isoform (X-axis), facetted by sparse and dense correlations between isoforms and shared or different isoQTLs across isoforms. **(c)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across ratio of effect sizes for 2 effect isoforms, facetted by sparse and dense correlations between isoforms and shared or different isoQTLs across isoforms. Across all plots, isoform and total gene expression heritability is set to 0.05 with 2 simulated isoforms.

Figure S21: *Miami plot of isoTWAS isoform-trait associations for SCZ.* Z-score of isoTWAS association between trait and isoform (Y-axis) across chromosomal position (X-axis). Points are colored by significance thresholds and labelled with the transcript is within 10 kilobases of a GWAS-significant SNP.

Figure S22: *Miami plot of isoTWAS isoform-trait associations for BD.* Z-score of isoTWAS association between trait and isoform (Y-axis) across chromosomal position (X-axis). Points are colored by significance thresholds and labelled with the transcript is within 10 kilobases of a GWAS-significant SNP.

Figure S23: *Miami plot of isoTWAS isoform-trait associations for ASD.* Z-score of isoTWAS association between trait and isoform (Y-axis) across chromosomal position (X-axis). Points are colored by significance thresholds and labelled with the transcript is within 10 kilobases of a GWAS-significant SNP.

Figure S24: *Miami plot of isoTWAS isoform-trait associations for ALZ.* Z-score of isoTWAS association between trait and isoform (Y-axis) across chromosomal position (X-axis). Points are colored by significance thresholds and labelled with the transcript is within 10 kilobases of a GWAS-significant SNP.

Figure S25: *Miami plot of isoTWAS isoform-trait associations for ADHD.* Z-score of isoTWAS association between trait and isoform (Y-axis) across chromosomal position (X-axis). Points are colored by significance thresholds and labelled with the transcript is within 10 kilobases of a GWAS-significant SNP.

Figure S26: *Colocalization plots for ALZ, APOC1 total gene eQTLs, and ENST00000252491 isoQTLs.* (Top) Manhattan plots of ALZ risk GWAS, *APOC1* total gene eQTLs, and ENST00000252491 isoQTLs, colored by LD to the lead isoQTL and with the lead isoQTL shown in a triangle and labelled. (Bottom) Boxplots of gene and isoform expression by genotype of the lead isoQTL SNP on the left and forest plot of the lead isoQTL's effect and adjusted 95% confidence interval on the trait (black), gene (red), and isoform (blue), with P-values.

Supplemental Tables

Table S1: Sample size, source, and tissues for functional genomics reference panels

Table S2: Number of genes/models that pass heritability and cross-validation prediction cutoffs using TWAS and isoTWAS feature selection criteria and prediction methods

Table S3: Distribution of difference in predictive cross-validation R2 of observed total gene expression vs. predicted total gene expression (isoTWAS -TWAS)

Table S4: Distribution of percent difference in predictive cross-validation R2 of observed total gene expression vs. predicted total gene expression (isoTWAS -TWAS)

Table S5: Distribution of difference in external predictive R2 of observed total gene expression vs. predicted total gene expression using completely independent training and test sets (isoTWAS -TWAS)

Table S6: Distribution of percent difference in external predictive R2 of observed total gene expression vs. predicted total gene expression using completely independent training and test sets (isoTWAS -TWAS)

Table S7: Measures of test statistic inflation across isoTWAS and TWAS and increases in effective sample size for isoTWAS over TWAS using PsychENCODE expression data and across 5 neuropsychiatric traits

Table S8: isoTWAS results using fetal frontal cortex expression across 5 neuropsychiatric traits. Only isoform transcripts that are found in 90% credible intervals using FOCUS are shown

Table S9: isoTWAS results using adult frontal cortex expression across 5 neuropsychiatric traits. Only isoform transcripts that are found in 90% credible intervals using FOCUS are shown

Supplemental Data

Data 1: Predictive performance across all 6 predictive methods in isoTWAS using *CACNA1E* locus and across multiple simulation parameters

Data 2: Difference in predictive performance (isoTWAS - TWAS) using *CACNA1E* locus and across multiple simulation parameters

Data 3: Difference in predictive performance (isoTWAS - TWAS) using *XRN2* locus and across multiple simulation parameters

Data 4: False positive rates across 20 simulations of 1,000 iterations across gene-level TWAS, ACAT, and minimum P aggregation of isoform associations using CACNA1E locus and across multiple simulation parameters

Data 5: Power to detect trait association at $P < 2.5 \times 10^{-6}$ across 1,000 simulations using TWAS and isoTWAS (ACAT) across various simulations. These simulations are under Scenario 1 in Figure 3a (gene has a true effect on the trait, but none of the isoforms have a true effect on the trait)

Data 6: Power to detect trait association at $P < 2.5 \times 10^{-6}$ across 1,000 simulations using TWAS and isoTWAS (ACAT) across various simulations. These simulations are under Scenario 2 in Figure 3b (a gene has multiple isoforms, only one has an effect on the trait, and we vary the usage of this effect isoform)

Data 7: Power to detect trait association at $P < 2.5 \times 10^{-6}$ across 1,000 simulations using TWAS and isoTWAS (ACAT) across various simulations. These simulations are under Scenario 3 in Figure 3c (a gene has two isoforms with differing effects on the trait, and we vary the effect size of one of the isoforms)

Data 8: Raw isoTWAS results across 5 neuropsychiatric traits using adult prefrontal cortex expression models

Data 9: Raw TWAS results across 5 neuropsychiatric traits using adult prefrontal cortex expression models

Data 10: Raw isoTWAS results across 5 neuropsychiatric traits using fetal prefrontal cortex expression models

Data 11: Raw TWAS results across 5 neuropsychiatric traits using fetal prefrontal cortex expression models

ACKNOWLEDGEMENTS

We thank Kangcheng Hou, Tommer Schwarz, Pan Zhang, Leanna Hernandez, Harold Pimentel, Mike Love, and Achal Patel for engaging discussion during the research process. We thank Kanishka Patel for her aesthetic

advice for figures. We thank the Psychiatric Genomics Consortium and Complex Trait Genomics Lab for their publicly available GWAS summary statistics.

BP was partially supported by NIH awards R01 HG009120, R01 MH115676, R01 CA251555, R01 AI153827, R01 HG006399, R01 CA244670, and U01 HG011715. MJG was supported by SFARI Bridge to Independence Award, NIMH R01-MH121521, NIMH R01-MH123922, and NICHD-P50-HD103557.

AUTHOR INFORMATION

Contributions

Conceptualization: AB, BP, MJG; Methodology: AB, BP, MJG; Software: AB; Validation: AB; Formal analysis: AB, MK, CW, CJ; Investigation: AB, MJG; Resources: BP, MJG; Data curation: AB, MK, CW, CJ, MJG; Writing – original draft: AB, MJG; Writing – reviewing and editing: AB, MK, CW, CJ, BP, MJG; Visualization: AB; Supervision: BP, MJG; Project administration: BP, MJG; Funding acquisition: BP, MJG

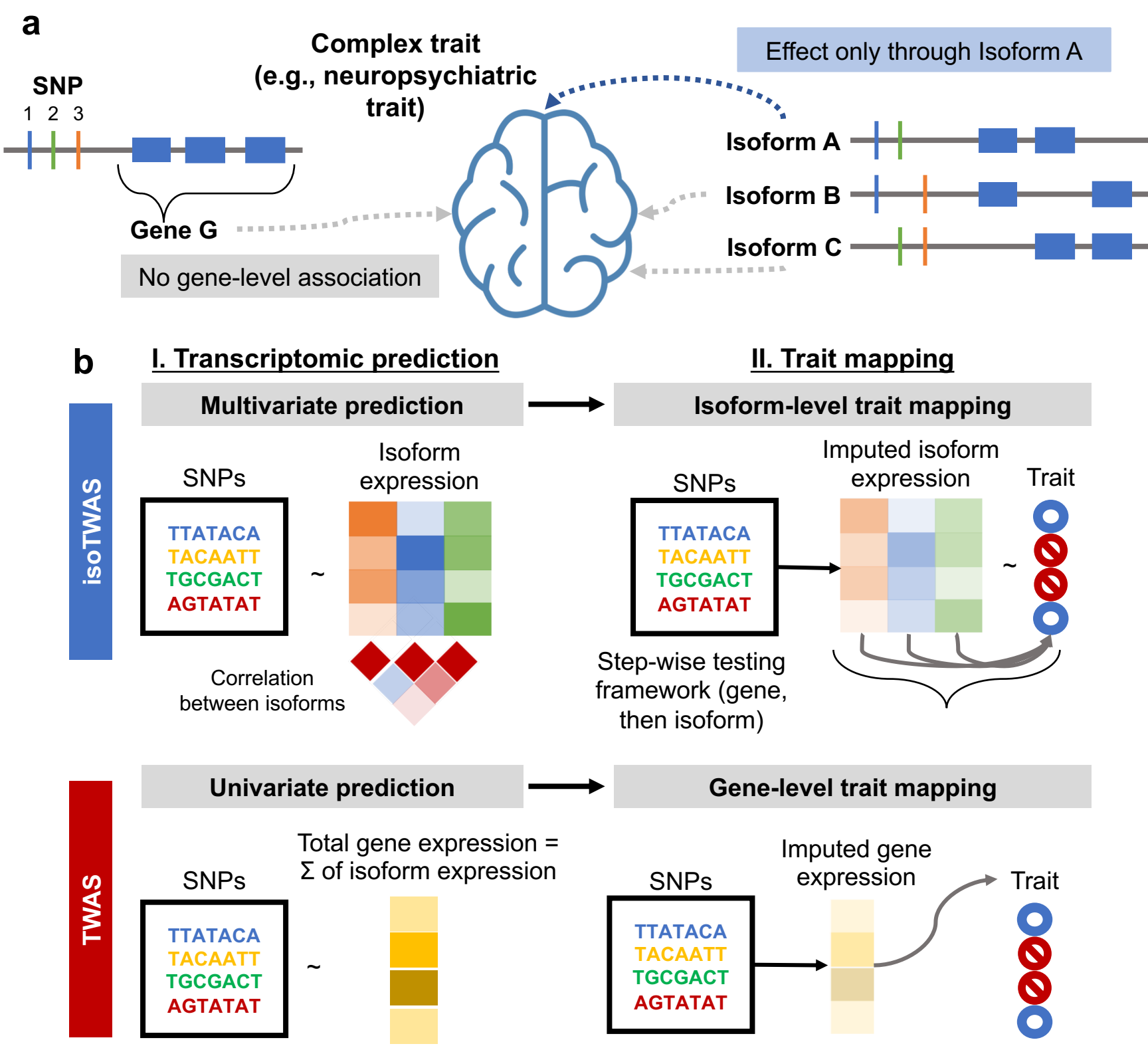
Correspondence

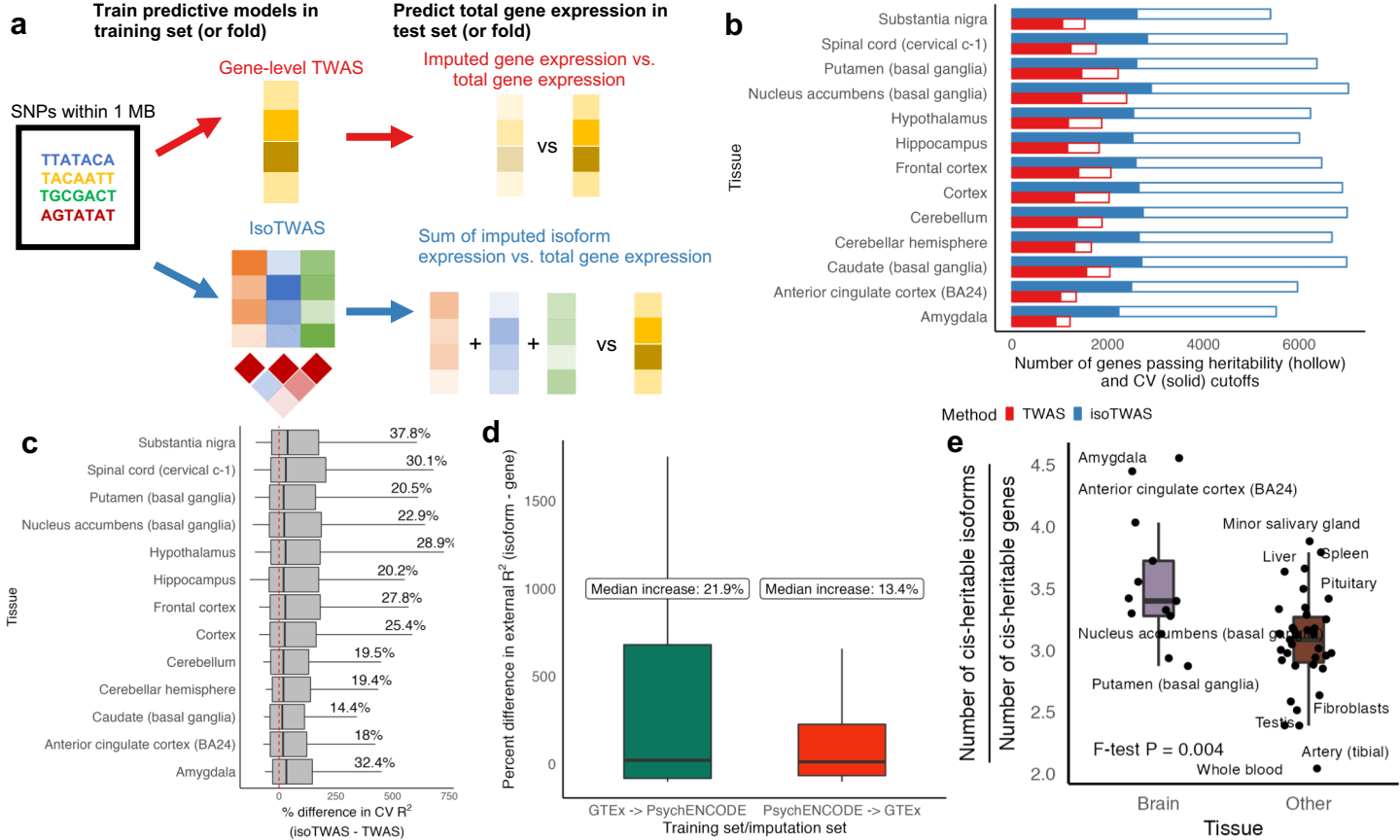
Please direct all correspondence to Arjun Bhattacharya (abtbhatt@ucla.edu).

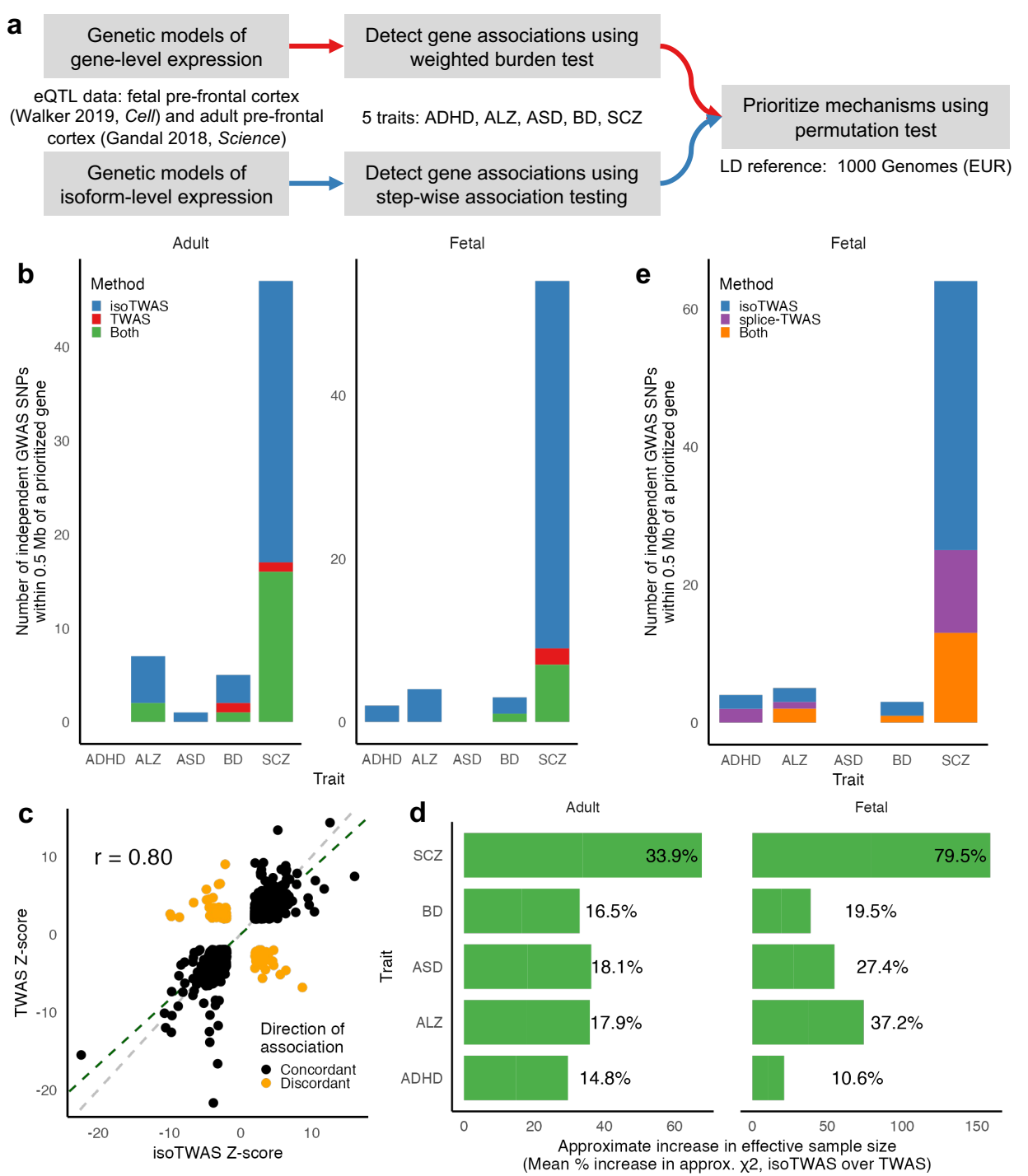
ETHICS DECLARATIONS

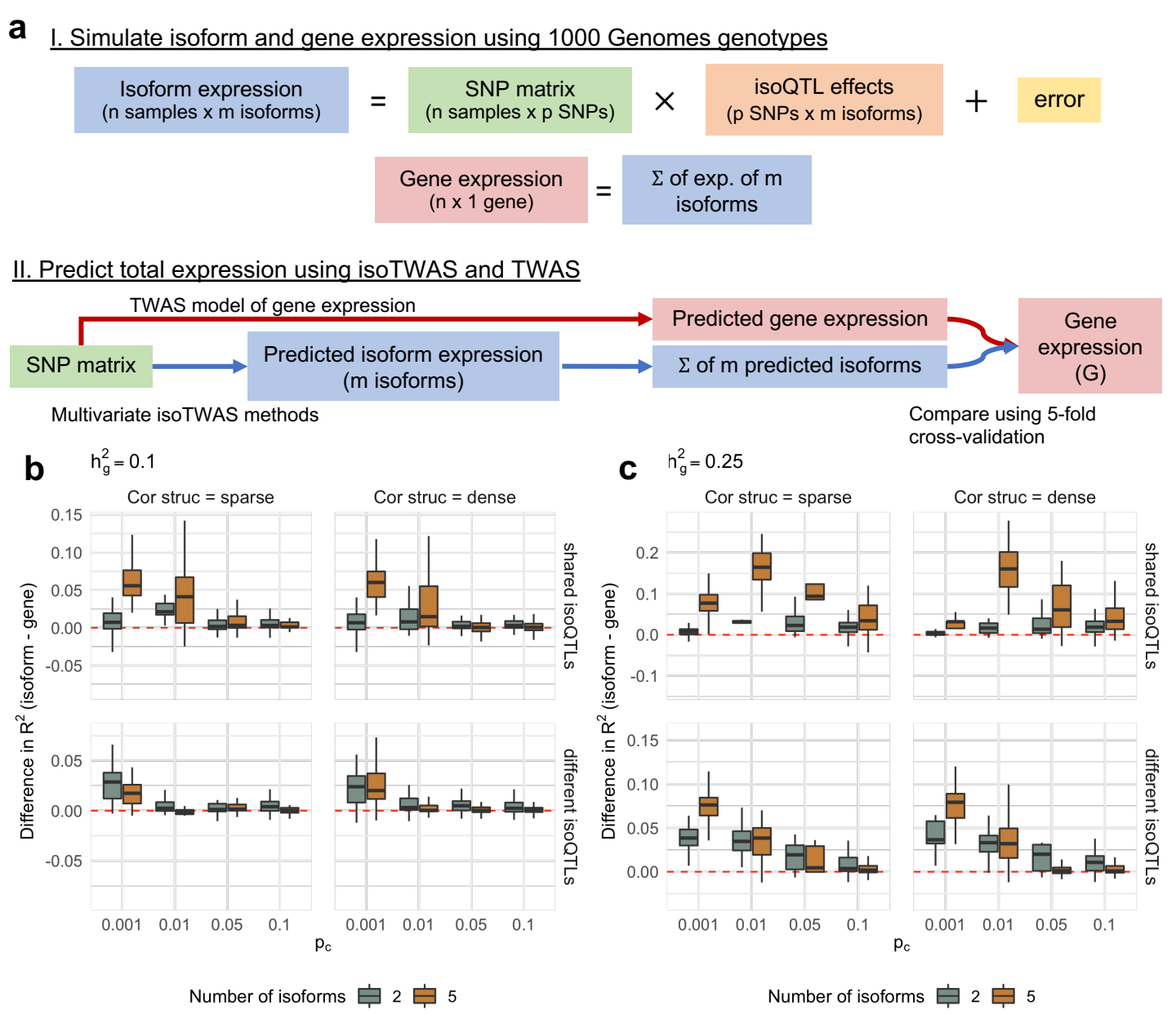
Competing Interests

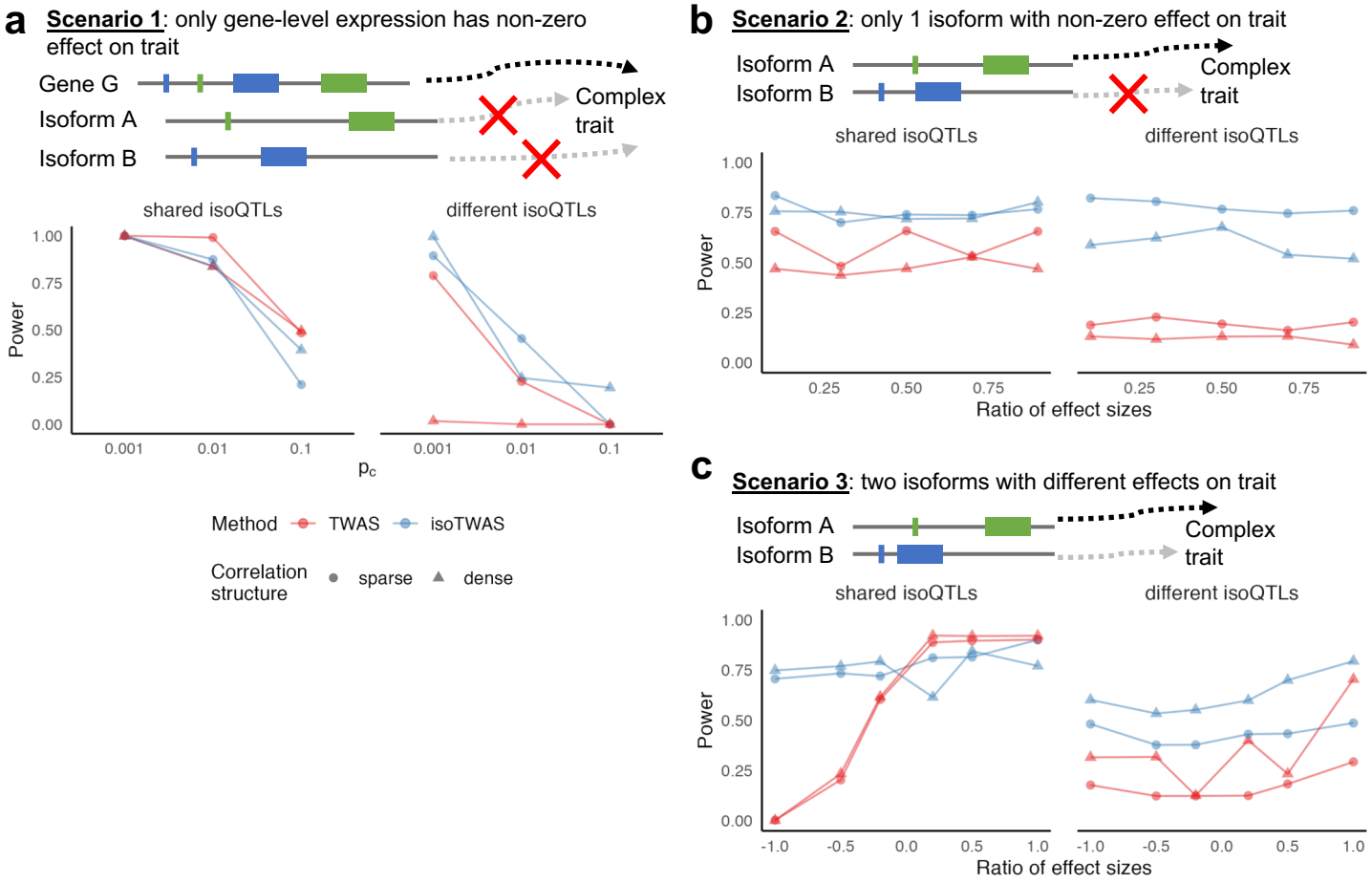
The authors declare no competing interests.

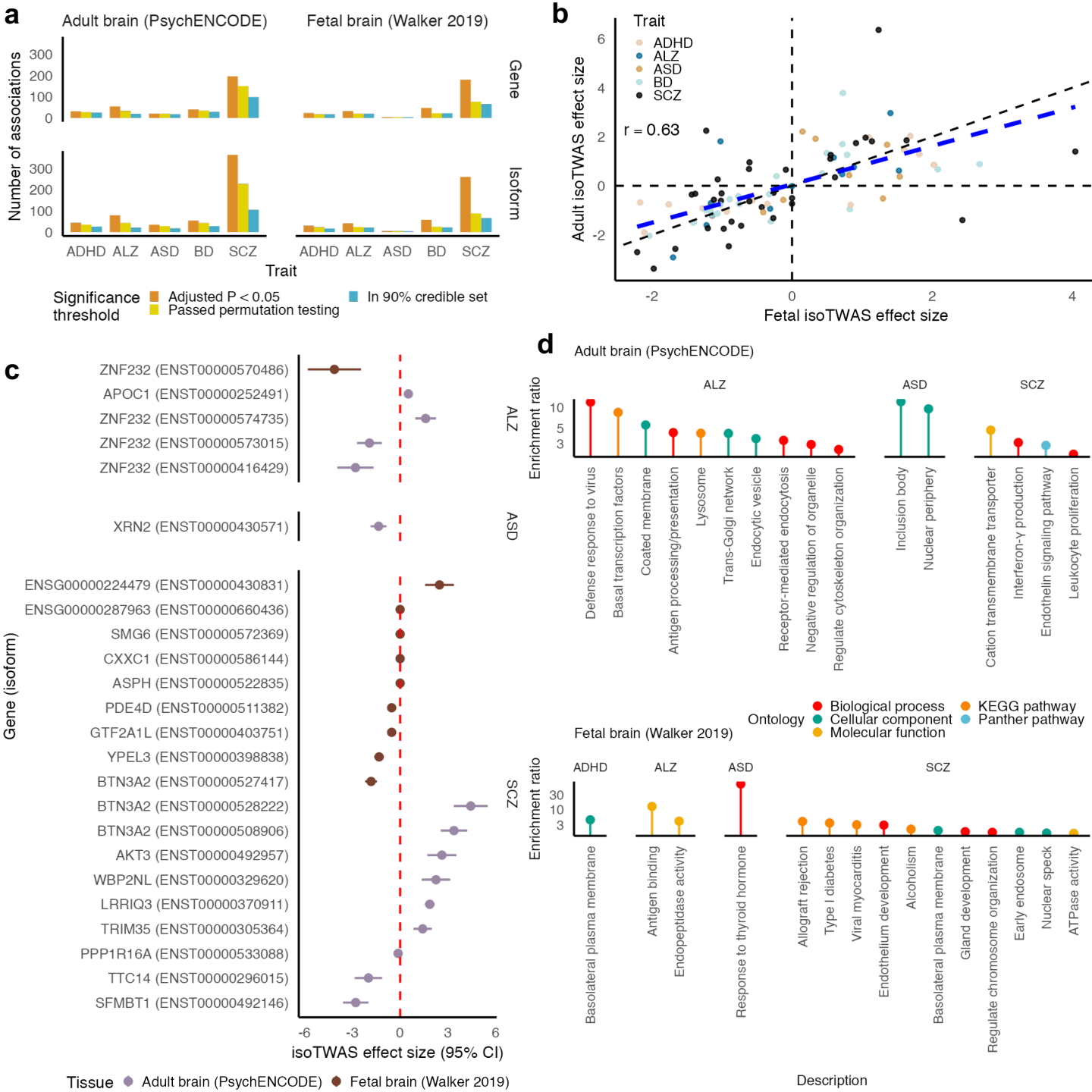

a
SNP
1 2 3
Complex trait
(e.g., neuropsychiatric
trait)
Effect only through Isoform A
Isoform A
Isoform B
Isoform C
Gene G
No gene-level association
b
I. Transcriptomic prediction
II. Trait mapping
isoTWAS
Multivariate prediction
Isoform-level trait mapping
SNPs
Isoform
expression
TTATACA
TACAATT
TGCGACT
AGTATAT
~
Correlation
between isoforms
SNPs
Imputed isoform
expression
Trait
TTATACA
TACAATT
TGCGACT
AGTATAT
~
Step-wise testing
framework (gene,
then isoform)
TWAS
Univariate prediction
Gene-level trait mapping
Total gene expression =
Σ of isoform expression
SNPs
TTATACA
TACAATT
TGCGACT
AGTATAT
~
SNPs
Imputed gene
expression
Trait
TTATACA
TACAATT
TGCGACT
AGTATAT

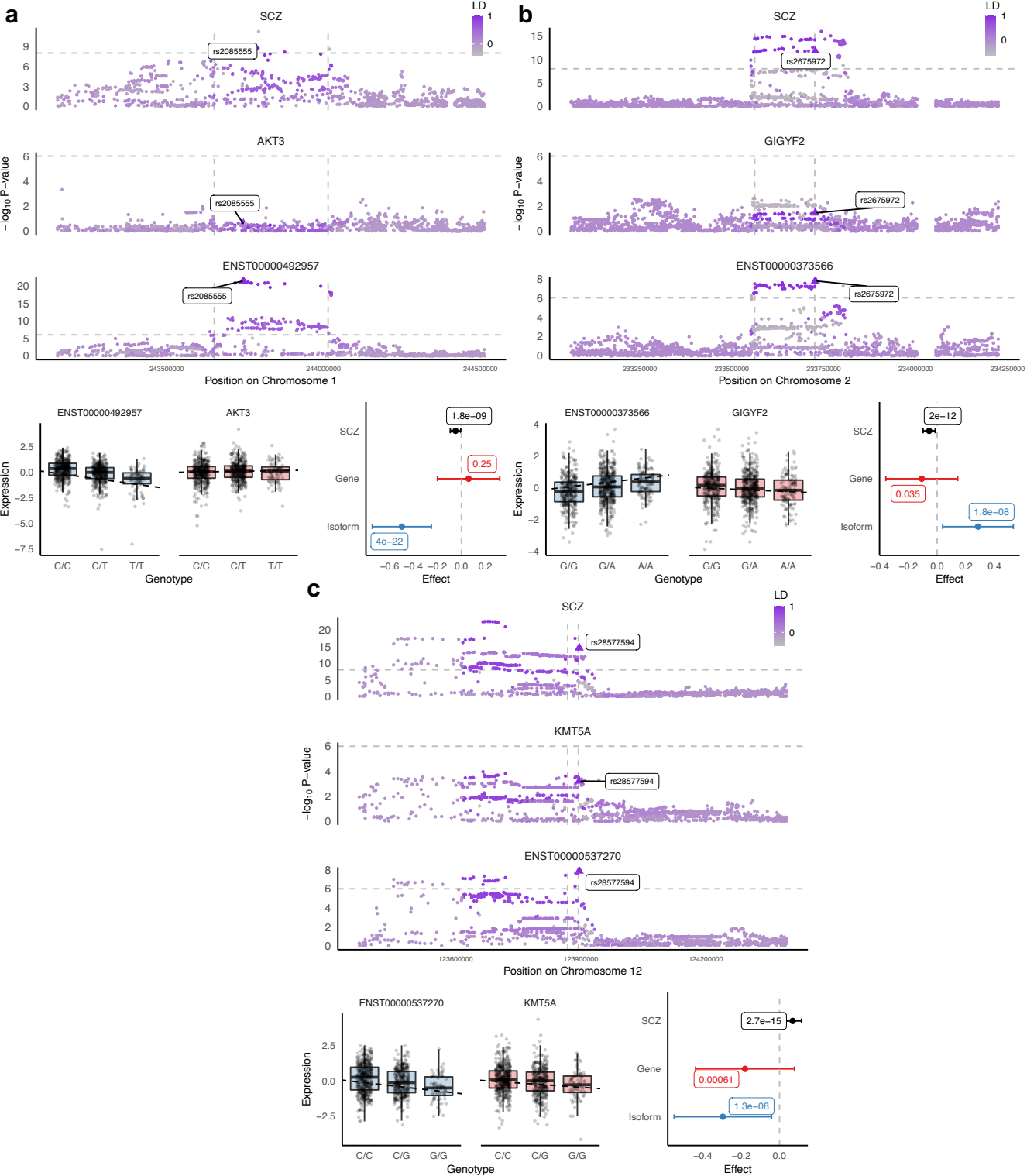

a
Train predictive models in training set (or fold)
Predict total gene expression in test set (or fold)
Gene-level TWAS
Imputed gene expression vs. total gene expression
SNPs within 1 MB
TTATACA
TACAATT
TGCGACT
AGTATAT
vs
IsoTWAS
Sum of imputed isoform expression vs. total gene expression
+
+
vs
b
Tissue
Substantia nigra
Spinal cord (cervical c-1)
Putamen (basal ganglia)
Nucleus accumbens (basal ganglia)
Hypothalamus
Hippocampus
Frontal cortex
Cortex
Cerebellum
Cerebellar hemisphere
Caudate (basal ganglia)
Anterior cingulate cortex (BA24)
Amygdala
0
2000
4000
6000
Number of genes passing heritability (hollow) and CV (solid) cutoffs
Method TWAS isoTWAS
c
Tissue
Substantia nigra 37.8%
Spinal cord (cervical c-1) 30.1%
Putamen (basal ganglia) 20.5%
Nucleus accumbens (basal ganglia) 22.9%
Hypothalamus 28.9%
Hippocampus 20.2%
Frontal cortex 27.8%
Cortex 25.4%
Cerebellum 19.5%
Cerebellar hemisphere 19.4%
Caudate (basal ganglia) 14.4%
Anterior cingulate cortex (BA24) 18%
Amygdala 32.4%
0
250
500
750
% difference in CV R2 (isoTWAS - TWAS)
d
Percent difference in external R2 (isoform - gene)
1500
1000
500
0
Median increase: 21.9%
Median increase: 13.4%
GTEx -> PsychENCODE
PsychENCODE -> GTEx
Training set/imputation set
e
Number of cis-heritable isoforms
Number of cis-heritable genes
4.5
4.0
3.5
3.0
2.5
2.0
Amygdala
Anterior cingulate cortex (BA24)
Minor salivary gland
Liver
Spleen
Pituitary
Nucleus accumbens (basal ganglia)
Putamen (basal ganglia)
Fibroblasts
Testis
F-test P = 0.004
Artery (tibial)
Whole blood
Brain
Other
Tissue


a
Genetic models of gene-level expression
Detect gene associations using weighted burden test
eQTL data: fetal pre-frontal cortex (Walker 2019, Cell) and adult pre-frontal cortex (Gandal 2018, Science)
5 traits: ADHD, ALZ, ASD, BD, SCZ
Prioritize mechanisms using permutation test
LD reference: 1000 Genomes (EUR)
Genetic models of isoform-level expression
Detect gene associations using step-wise association testing
b
Adult
Fetal
Method
isoTWAS
TWAS
Both
Number of independent GWAS SNPs within 0.5 Mb of a prioritized gene
ADHD
ALZ
ASD
BD
SCZ
Trait
e
Fetal
Method
isoTWAS
splice-TWAS
Both
Number of independent GWAS SNPs within 0.5 Mb of a prioritized gene
ADHD
ALZ
ASD
BD
SCZ
Trait
c
r = 0.80
TWAS Z-score
isoTWAS Z-score
Direction of association
Concordant
Discordant
d
Adult
Fetal
Trait
SCZ
BD
ASD
ALZ
ADHD
33.9%
16.5%
18.1%
17.9%
14.8%
79.5%
19.5%
27.4%
37.2%
10.6%
Approximate increase in effective sample size
(Mean % increase in approx. χ2, isoTWAS over TWAS)


a
I. Simulate isoform and gene expression using 1000 Genomes genotypes
Isoform expression (n samples x m isoforms)
=
SNP matrix (n samples x p SNPs)
×
isoQTL effects (p SNPs x m isoforms)
+
error
Gene expression (n x 1 gene)
=
Σ of exp. of m isoforms
II. Predict total expression using isoTWAS and TWAS
TWAS model of gene expression
SNP matrix
Predicted isoform expression (m isoforms)
Predicted gene expression
Σ of m predicted isoforms
Gene expression (G)
Multivariate isoTWAS methods
Compare using 5-fold cross-validation
b
$h_g^2 = 0.1$
Cor struc = sparse
Cor struc = dense
shared isoQTLs
different isoQTLs
Difference in R^2 (isoform - gene)
p_c
Number of isoforms 2 5
c
$h_g^2 = 0.25$
Cor struc = sparse
Cor struc = dense
shared isoQTLs
different isoQTLs
Difference in R^2 (isoform - gene)
p_c
Number of isoforms 2 5


a
Scenario 1: only gene-level expression has non-zero effect on trait
Gene G
Isoform A
Isoform B
Complex trait
shared isoQTLs
different isoQTLs
Power
1.00
0.75
0.50
0.25
0.00
0.001
0.01
0.1
p_c
Method
TWAS
isoTWAS
Correlation structure
sparse
dense
b
Scenario 2: only 1 isoform with non-zero effect on trait
Isoform A
Isoform B
Complex trait
shared isoQTLs
different isoQTLs
Power
0.25
0.50
0.75
Ratio of effect sizes
c
Scenario 3: two isoforms with different effects on trait
Isoform A
Isoform B
Complex trait
shared isoQTLs
different isoQTLs
Power
-1.0
-0.5
0.0
0.5
1.0
Ratio of effect sizes


a
Adult brain (PsychENCODE)
Fetal brain (Walker 2019)
Number of associations
Gene
Isoform
ADHD ALZ ASD BD SCZ
Trait
Significance threshold
Adjusted P < 0.05
Passed permutation testing
In 90% credible set
b
Trait
ADHD
ALZ
ASD
BD
SCZ
r = 0.63
Adult isoTWAS effect size
Fetal isoTWAS effect size
c
ZNF232 (ENST00000570486)
APOC1 (ENST00000252491)
ZNF232 (ENST00000574735)
ZNF232 (ENST00000573015)
ZNF232 (ENST00000416429)
XRN2 (ENST00000430571)
ENSG00000224479 (ENST00000430831)
ENSG00000287963 (ENST00000660436)
SMG6 (ENST00000572369)
CXXC1 (ENST00000586144)
ASPH (ENST00000522835)
PDE4D (ENST00000511382)
GTF2A1L (ENST00000403751)
YPEL3 (ENST00000398838)
BTN3A2 (ENST00000527417)
BTN3A2 (ENST00000528222)
BTN3A2 (ENST00000508906)
AKT3 (ENST00000492957)
WBP2NL (ENST00000329620)
LRRIQ3 (ENST00000370911)
TRIM35 (ENST00000305364)
PPP1R16A (ENST00000533088)
TTC14 (ENST00000296015)
SFMBT1 (ENST00000492146)
Gene (isoform)
isoTWAS effect size (95% CI)
Tissue
Adult brain (PsychENCODE)
Fetal brain (Walker 2019)
d
Adult brain (PsychENCODE)
ALZ
ASD
SCZ
Enrichment ratio
Defense response to virus
Basal transcription factors
Coated membrane
Antigen processing/presentation
Lysosome
Trans-Golgi network
Endocytic vesicle
Receptor-mediated endocytosis
Negative regulation of organelle
Regulate cytoskeleton organization
Inclusion body
Nuclear periphery
Cation transmembrane transporter
Interferon-γ production
Endothelin signaling pathway
Leukocyte proliferation
Ontology
Biological process
Cellular component
Molecular function
KEGG pathway
Panther pathway
Fetal brain (Walker 2019)
ADHD
ALZ
ASD
SCZ
Basolateral plasma membrane
Antigen binding
Endopeptidase activity
Response to thyroid hormone
Allograft rejection
Type I diabetes
Viral myocarditis
Endothelium development
Alcoholism
Basolateral plasma membrane
Gland development
Regulate chromosome organization
Early endosome
Nuclear speck
ATPase activity
Description


a
SCZ
LD
1
0
rs2085555
AKT3
rs2085555
ENST00000492957
rs2085555
−log10 P−value
243500000
244000000
244500000
Position on Chromosome 1
ENST00000492957
AKT3
Expression
C/C
C/T
T/T
Genotype
SCZ
1.8e−09
Gene
0.25
Isoform
4e−22
Effect
b
SCZ
rs2675972
GIGYF2
rs2675972
ENST00000373566
rs2675972
233250000
233500000
233750000
234000000
234250000
Position on Chromosome 2
ENST00000373566
GIGYF2
G/G
G/A
A/A
2e−12
0.035
1.8e−08
c
SCZ
rs28577594
KMT5A
rs28577594
ENST00000537270
rs28577594
123600000
123900000
124200000
Position on Chromosome 12
ENST00000537270
KMT5A
C/C
C/G
G/G
2.7e−15
0.00061
1.3e−08