

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

3

4 **AUTHORS**

5 Arjun Bhattacharya^{1,2}, Daniel D. Vo^{3,4}, Connor Jops^{3,4}, Minsoo Kim^{5,6}, Cindy Wen⁵, Jonatan L. Hervoso⁷,
6 Bogdan Pasaniuc^{1,6,8±}, Michael J. Gandal^{3,4,5,6±}

7

8 **AFFILIATIONS**

- 9 1. Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of
10 California, Los Angeles, California, USA
- 11 2. Institute for Quantitative and Computational Biosciences, David Geffen School of Medicine, University
12 of California, Los Angeles, California, USA
- 13 3. Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia,
14 Pennsylvania, USA
- 15 4. Lifespan Brain Institute at Penn Med and the Children's Hospital of Philadelphia, Philadelphia,
16 Pennsylvania, USA
- 17 5. Department of Psychiatry and Biobehavioral Sciences, Semel Institute, David Geffen School of
18 Medicine, University of California, Los Angeles, California, USA
- 19 6. Department of Human Genetics, David Geffen School of Medicine, University of California, Los
20 Angeles, CA, USA
- 21 7. Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA
- 22 8. Department of Computational Medicine, David Geffen School of Medicine, University of California,
23 Los Angeles, CA, USA

24 ±Equal contribution

25

26 **CORRESPONDENCE**

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

28

29 **ABSTRACT**

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

37

38 We introduce isoTWAS, a multivariate framework to integrate genetics, isoform-level expression, and
39 phenotypic associations in a step-wise testing framework, and evaluate it using data from the Genotype-
40 Tissue Expression (GTEx) Project, PsychENCODE Consortium, and other sources. isoTWAS shows
41 three main advantages. First, joint, multivariate modeling of isoform expression from *cis*-window SNPs
42 improves prediction by ~1.8-2.4 fold, compared to univariate modeling. Second, compared to gene-level
43 TWAS, these improvements in prediction lead to ~1.9-2.5-fold increase in the number of testable genes
44 and a median of 25-70% increase in cross-validated prediction of total gene expression, with the added
45 ability to jointly capture expression and splicing mechanisms. In external validation, isoform-centric
46 models predicted gene expression at percent variance explained >1% for 50% more genes than gene-
47 centric models. Third, across 15 neuropsychiatric traits, isoTWAS increased discovery of trait
48 associations within GWAS loci over TWAS, capturing ~60% more unique loci and 95% of loci detected by
49 TWAS. Results from extensive simulations showed no increase in false discovery rate and reinforce
50 isoTWAS's advantages in prediction and trait mapping power over TWAS, especially when genetic effects
51 on expression vary across isoforms of the same gene. We illustrate multiple biologically-relevant
52 isoTWAS-identified trait associations undetectable by gene-level methods, including isoforms of *AKT3*,
53 *CUL3*, and *HSPD1* with schizophrenia risk, and *PCLO* with multiple disorders.

54

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

57 **INTRODUCTION**

58 Over the past decade, the number of genetic associations with complex traits identified by genome-wide
59 association studies (GWAS) has increased considerably^{1,2}. However, the slow translation of these genetic
60 associations into concrete molecular mechanisms remains a major obstacle. As GWAS associations
61 predominately localize within non-coding regions of the genome and are often tagged within large blocks
62 of linkage disequilibrium (LD), the first major challenge is prioritizing the underlying causal variant(s)
63 within a given locus and identifying their functional impact on nearby target genes. To address this,
64 numerous methods have been developed to integrate transcriptomic reference panels with GWAS to
65 prioritize genes at trait-associated loci^{3–15}. TWAS and related approaches (e.g., PrediXcan) impute the
66 *cis*-component of gene expression predicted by germline genetics into an association cohort, thereby
67 reducing multiple comparisons and increasing interpretability by identifying a set of genes at a locus that
68 may underlie the genetic association^{3,4}. Despite these methodological advances, a majority of GWAS loci
69 still lack robust mechanistic interpretation¹⁶.

70

71 Previous integrative analyses have largely focused on aggregated measurements of total gene
72 expression but have not systematically explored the potentially large number of distinct transcript-
73 isoforms that can be generated from a given genetic locus through alternative splicing. Alternative splicing
74 is a fundamental form of tissue-specific gene regulation present in more than 90% of human genes, that
75 vastly expands the coding and regulatory potential of the genome^{17–20}; GENCODE v40 annotates an
76 average of 4 isoforms per gene (mean 4.05, standard deviation 7.28, median 1)²¹. Genes uniquely
77 expressed in the brain, which are longer and contain far more exons, undergo the greatest degree of
78 splicing compared with other tissues and species—a mechanism contributing to the vast proteomic,
79 phenotypic, and evolutionary complexity of the human brain^{22–25}. Some genes are known to express up to
80 hundreds to thousands of unique isoforms in the human brain²⁵. Independent of gene expression, splicing
81 dysregulation has been increasingly implicated as a putative disease mechanism^{24,26–30} including for
82 neuropsychiatric disorders^{10,24,27,31}. Yet, local splicing events can be computationally intensive to measure
83 and are difficult to systematically integrate across multiple distinct large-scale datasets. Splicing is often
84 coordinated across a gene, yielding many non-independent features that increases multiple testing

85 burden. In contrast, transcript-isoform abundance can now be rapidly estimated across large-scale RNA-
86 seq datasets using pseudoalignment methods^{32,33}. Furthermore, in the brain, isoform-level expression
87 changes have shown greater enrichment for schizophrenia heritability than gene or local splicing
88 changes^{22,31,34–36}. However, to fully integrate transcript-isoform quantifications with GWAS, innovative
89 computational methods are needed that jointly model the highly correlated transcripts of the same gene.
90
91 Here, we present isoform-level TWAS (isoTWAS), a flexible and scalable approach for complex trait
92 mapping by integrating genetic effects on isoform-level expression with GWAS. In extensive simulations
93 and real data applications in the Genotype-Tissue Expression (GTEx) Project³⁷ and the PsychENCODE
94 Consortium^{22,24}, we show that isoTWAS provides several important advantages to trait mapping. First, in
95 the transcriptomic prediction step, the correlation between isoforms provides additional information that is
96 unavailable when only gene-level expression is measured, which can be leveraged to improve
97 prediction³⁸ accuracy of individual isoforms in >80% of cases by a median of ~1.8-2.4-fold improvement
98 and of total gene expression by 25-70%. In parallel, our isoform-centric framework uncovers cross-
99 validated predictive models for ~2-fold more genes, doubling the number of testable features in the trait
100 mapping step. Third, divergent patterns of genetic effects across isoforms can be leveraged to provide a
101 more granular hypothesis for a mechanism underlying the SNP-trait relationship. Finally, the isoTWAS
102 framework jointly captures expression and splicing disease mechanisms, while maintaining a well-
103 controlled false discovery rate. Altogether, using GWAS data for 15 neuropsychiatric traits, isoTWAS
104 greatly increases the power to detect gene-level trait associations uncovering associations at ~60% more
105 GWAS loci compared to traditional gene-level TWAS. These results stress the need to shift focus to
106 transcript-isoforms to increase discovery of transcriptomic mechanisms underlying genetic associations
107 with complex traits.

108

109 **RESULTS**

110 ***The isoTWAS framework***

111 isoTWAS seeks to prioritize genes with transcript-isoforms of genes whose *cis*-genetic component of
112 expression is significantly associated with a complex trait. We extend the traditional gene-level TWAS

113 approach by jointly modeling the expression of distinct transcript-isoforms of a given gene as a matrix
114 while accounting for the pair-wise correlations between these isoforms^{21,37,39,40}. Here, we assume that
115 local genetic variants directly modulate expression of an isoform. In addition, we assume that the
116 abundance of a gene is measured as the is the sum of the abundance of its isoforms, computed as
117 transcripts per million, or TPM (**Supplemental Figure S1**)^{32,33,41,42}. Flexibly integrating isoform-level
118 expression into trait mapping may lead to novel discoveries in disease mapping and prioritize specific
119 isoforms that explain genetic associations. Accordingly, gene-level trait mapping using traditional TWAS
120 methods may not necessarily detect a trait association on the gene-level if a gene has multiple isoforms
121 but only one is associated with the trait (**Figure 1a**). This scenario may be particularly relevant in the
122 human brain, where some genes may express up to hundreds to thousands of unique isoforms²⁵. By
123 modeling the genetic architectures of isoforms of a gene simultaneously, isoTWAS provides a deeper
124 understanding of potential transcriptomic mechanisms that underlie genetic associations.

125

126 The isoTWAS framework contains three general steps (**Figure 1**). First, we build multivariate predictive
127 models of isoform-level expression using well-powered functional genomics training datasets, including
128 GTEx³⁷ and PsychENCODE^{24,27}. Here, we trained and systematically compared 4 multivariate predictive
129 frameworks: (1) multivariate elastic net penalized regression⁴³, (2) multivariate LASSO penalized
130 regression with simultaneous covariance estimation (MRCE)⁴⁴, (3) multivariate elastic net regression with
131 stacked generalization (jointnet)⁴⁵, and (4) sparse partial least squares (SPLS)⁴⁶. As a baseline for
132 comparison, we also modeled each individual isoform independently with univariate regularized
133 regressions, as implemented in Gusev et al's FUSION software^{4,43,47,48} (see **Methods** and **Supplemental**
134 **Methods**). Models were trained to predict isoform expression using the set of *cis*-SNPs within 1
135 Megabase (Mb) of the gene body (**Methods, Figure 1b**). Model performance was assessed via 5-fold
136 cross-validation, using McNemar's adjusted R^2 between observed and predicted expression.

137

138 Next, we use these models to impute isoform expression into an external GWAS cohort and quantify the
139 association with the target GWAS phenotype (**Figure 1b**). If individual-level genotypes are available,
140 isoform expression can be directly imputed as a linear combination of the SNPs in the models, and these

141 associations can be estimated through appropriate regression analyses. If only GWAS summary statistics
142 are available, imputation and association testing is conducted simultaneously through a weighted burden
143 test⁴. Finally, isoTWAS performs step-wise hypothesis testing procedure to account for multiple
144 comparisons and control for local LD structure. Isoform-level P-value are first aggregated to the gene-
145 level to prioritize a gene using the Aggregated Cauchy Association Test (ACAT)⁴⁹, where false discovery
146 rates are controlled, and then individual isoforms of prioritized genes are subjected to post-hoc family-
147 wise error control⁵⁰ (**Supplemental Figure S2, Methods**). After this step, a set of isoforms are identified
148 whose *cis*-genetic component of expression are associated with the trait of interest⁴. For these isoforms,
149 we apply a rigorous permutation test whereby the SNP-to-isoform effects are permuted 10,000 times to
150 generate a null distribution; this permutation test assesses how much signal is added by isoform
151 expression, given the GWAS architecture of the locus, and controls for large LD blocks⁴. Lastly, we can
152 isoform-level Bayesian fine-mapping at each locus with a significant trait association to identify the
153 minimal credible set of isoforms that contains the ‘casual’ isoform at a 90% confidence level and to assign
154 individual posterior inclusion probabilities (**Figure 1b, Methods**). isoTWAS is available as an R package
155 through Github (<https://github.com/bhattacharya-a-bt/isotwas>).

156

157 ***isoTWAS improves prediction of isoform and gene expression in simulations and real data***

158 Previous work has demonstrated that isoform-level quantifications, when propagated to the gene-level,
159 can lead to more accurate gene expression estimates and differential expression results with short read
160 RNA-seq data^{41,42}. We therefore hypothesized that our multivariate SNP-based imputation of isoform
161 expression, when summed to the gene-level, would outperform traditional gene-level TWAS models. To
162 systematically evaluate the performance of TWAS versus isoTWAS models on prediction of total gene
163 expression across a variety of genetic architectures, we conducted an extensive set of simulations across
164 22 different gene loci using European-ancestry data from the 1000 Genomes Project⁵¹ (**Methods, Figure**
165 **2a**). At each gene locus, we controlled gene expression heritability and simulated 2-10 distinct isoforms,
166 varying the proportion of causal isoQTLs (p_{causal}) and their sharing between isoforms (p_{shared}). We then
167 trained cross-validated, multivariate predictive models of isoform expression (isoTWAS) or univariate
168 models of gene expression (TWAS). For isoTWAS, of the specific multivariate prediction models tested,

169 multivariate elastic net⁴³ demonstrated the greatest CV prediction of isoform expression across most
170 simulation settings (**Figure 2b, Supplemental Figure S3, Supplemental Data 1**). For gene expression
171 prediction, the optimal isoTWAS models (in sum) outperformed the optimal TWAS model, particularly at
172 sparser isoQTL architectures, with median absolute increase in adjusted R^2 of 0.6-3.5% (**Figure 2c,**
173 **Supplemental Figure S4, Supplemental Data 2**). Performance gains decreased somewhat with denser
174 isoQTL architectures, although real data is consistent with 0.1-1% sparsity (i.e., 1-10 causal e- and
175 isoQTLs per gene or isoform)³⁷. In simulations, we found that isoTWAS prediction of gene expression
176 also increases as the proportion of shared non-zero effect SNPs across isoforms decreases (**Figure 2b-**
177 **c, Supplemental Figure S4, Supplemental Data 2**).

178
179 Next, we assessed predictive performance in real data from 48 tissues (13 brain) with sufficient sample
180 sizes ($N > 100$) in GTEx for all genes with multiple expressed isoforms (**Supplemental Table S1;**
181 **Methods**). Altogether, for 48 tissues in GTEx, we built significant predictive models for 50,000 to 80,000
182 isoforms across 8,000 to 12,000 unique genes per tissue (**Supplemental Figure S11, Supplemental**
183 **Table S2**). We considered 3 main criteria to evaluate the performance of both the multivariate and
184 isoform-centric approaches of isoTWAS: (1) the number of isoforms whose expression can be imputed
185 using multivariate/univariate models with cross-validation (CV) $R^2 > 0.01$; (2) the number of unique genes
186 with at least one isoform that can be imputed at CV $R^2 > 0.01$; and (3) the number of unique genes in
187 which total gene expression can be imputed at CV $R^2 > 0.01$ using isoTWAS (summed) or TWAS models.
188

189 At the isoform level (criterion 1), through joint multivariate modeling of isoform expression, we trained 2.3-
190 2.5-fold more models at cross-validation (CV) $R^2 > 0.01$ across the 48 tissues, compared to traditional
191 univariate approaches (**Figure 3a, Supplemental Figure S5**). Using these multivariate models, we
192 improve prediction for 79-82% of isoforms with a median increase of ~1.8-2.4 fold increase in adjusted R^2
193 (**Supplemental Figure S6, Supplemental Table S2**). Concordant with simulations, we found that the
194 multivariate elastic net overwhelmingly outperformed other multivariate (and univariate) methods,
195 indicating that leveraging the shared genetic architecture between isoforms of the same gene greatly aids
196 in prediction of each individual isoform (**Supplemental Figure S7, Supplemental Table S2**). Notably, we

197 observed that multivariate models were particularly important for brain tissues compared with non-brain
198 tissues in GTEx, which showed significantly improved performance compared with univariate models
199 **(Figure 3b; P = 0.011 from OLS regression of median percent increase in CV R² for**
200 **multivariate/univariate models against tissue type, adjusted for sample size),** suggesting more shared
201 isoQTL architecture in brain tissues than others which can be leveraged by isoTWAS for improved
202 prediction. These gains in prediction accuracy directly translate into increased power in the trait
203 association step⁵².

204
205 At the gene level (criteria 2 and 3), isoTWAS also increased the number of genes with testable models in
206 the trait mapping step and improved prediction of total gene expression. The number of unique genes
207 with at least 1 isoTWAS model at CV R² > 0.01 (inclusion criterion for isoTWAS trait mapping) was 1.9-
208 2.5 times larger than the number of unique genes with TWAS models achieving CV R² > 0.01 for gene
209 expression prediction **(Figure 3c, Supplemental Figure S8, Supplemental Table S2)**. For a given gene,
210 isoTWAS models (summed) outperformed TWAS models in prediction of total gene expression by a
211 median of 25-70% in cross validation **(Supplemental Figure S9)** with a 50-80% increase in the number
212 of genes that are predicted at CV R² > 0.01 **(Figure 3d, Supplemental Figure S10)**. We replicated these
213 gains in total gene expression prediction using an independent, out of sample QTL reference panel of
214 adult cortex from PsychENCODE/AMP-AD (Methods). Multivariate isoTWAS models outperformed
215 univariate TWAS models in predicting total gene expression with a 15.2% median percent increase in
216 adjusted R² when training in GTEx and testing in PsychENCODE/AMP-AD, and 23.9% vice versa; **Figure**
217 **3e, Supplemental Table S3)**. As predictive performance is positively related to power to detect trait
218 associations⁵¹, both the increased number and accuracy of trainable imputation models using isoTWAS's
219 multivariate predictive framework have strong implications on increased discovery in trait mapping⁵¹.

220

221 ***Performance of isoTWAS across distinct genomic contexts***

222 As genes can differ widely with respect to the number and expression patterns of their constituent
223 isoforms, as well as by other potentially relevant features such as gene length and SNP density, we next
224 sought to characterize the impact of these factors of isoTWAS performance using GTEx models as

225 evaluated by the 3 criteria outlined above (**Methods, Supplemental Note, Supplemental Figures S12-**
226 **19, Supplemental Data 3-4**). Overall, we observed an increase in performance of isoTWAS multivariate
227 modeling of both isoforms and genes with increasing number of isoforms per gene, although there was
228 less conclusive of a pattern with increasing dominant isoform fraction⁵² (**Supplemental Figures S12-13,**
229 **Supplemental Note, Methods**). We also noticed trends in the performance gain using isoTWAS
230 multivariate modeling with respect to both isoform and gene expression prediction across gene length,
231 SNP density at the gene locus, and sample size (**Supplemental Figures S14-16, Supplemental Note**).
232 Finally, as the proportion of non-zero effect SNPs in the isoTWAS model that are shared across isoforms
233 increased (**Supplemental Note**), we found an increasing trend in the gain in prediction of gene
234 expression using isoTWAS compared to TWAS models (**Supplemental Figure S17**), reflecting a similar
235 observation from simulation. Interestingly, as this proportion increased, we found an increase in the gain
236 in prediction of isoform expression, reinforcing the utility of multivariate modelling in marginal prediction of
237 isoform expression.

238

239 Lastly, we investigated how the robustness of isoform abundance estimation from short-read RNA-seq
240 impacted performance gains of isoTWAS, compared with TWAS. Isoform abundance was initially
241 quantified from probabilistic point estimates using Salmon, guided by Gencode annotations. We then
242 assessed the performance of isoTWAS across loci binned by quantification variance measured across 50
243 inferential replicates from Salmon³². In general, we found that for isoform-level prediction, multivariate
244 modeling in isoTWAS substantially outperformed univariate approaches as quantification variance
245 increased. However, comparing isoTWAS isoform-centric and TWAS gene-centric models, there were no
246 discernable trends in prediction of gene expression as the mean count and/or quantification variance of
247 genes increased (**Supplemental Figure S18-19, Supplemental Note**). Finally, we evaluated the impact
248 of reference transcriptome annotation fidelity by generating a synthetic dataset quantified using a
249 reference annotation masking the dominant isoforms for a set of genes. As expected, performance of
250 both isoTWAS and TWAS models declined when isoforms failed to be detected in expression
251 quantification (**Supplemental Note, Supplemental Figure S20**).

252

253 ***Simulations support improved power and calibrated null across genetic architectures***

254 We next introduced GWAS data for complex traits into our simulation framework to benchmark the false
255 positive rate (FPR) and power of isoTWAS (**Methods**). First, we found that the FPR is controlled at 0.05
256 for isoform-level mapping using ACAT, consistent with gene-level mapping using TWAS (**Supplemental**
257 **Figure S21, Supplemental Data 4**). For a simulated trait, we modeled causal effect architectures for a
258 genomic locus with 2-10 isoforms under three main scenarios (**Methods; Figure 4, Supplemental Figure**
259 **S22**): (1) where the true trait effect is from only total gene expression, (2) where there is only one isoform
260 with a non-zero effect on the trait, called the “effect isoform”; and (3) where 2 isoforms are effect isoforms,
261 with varying magnitudes and directions of association. This first scenario showed clear increases in
262 power for TWAS over isoTWAS, but this advantage declined as the causal proportion of isoQTLs
263 increased and as the proportion of shared isoQTLs increased (**Figure 4a, Supplemental Data 5**). For
264 scenarios (2) and (3), as effects on the trait varied across isoforms of the same gene (**Figure 4b-c,**
265 **Supplemental Data 6-7**), we observed large increases in power for isoTWAS over TWAS. Here, across
266 most scenarios and causal effect architectures, isoTWAS demonstrated improved power compared with
267 gene-level TWAS, particularly when only one isoform of a gene carried a trait effect or when two effect
268 isoforms of the same gene had different directions of effects. However, when the effect sizes of two effect
269 isoforms of the same gene converged, TWAS and isoTWAS demonstrated similar power to detect a
270 gene-trait association (**Figure 4c**).

271
272 Finally, we assessed the performance of probabilistic fine-mapping in identifying the true effect isoform in
273 our simulation framework of genes with 5 or 10 isoforms (**Methods, Supplemental Figure S23,**
274 **Supplemental Data 9**). In general, the sensitivity of 90% credible sets (proportion of credible sets that
275 contained the true effect isoform) was under-calibrated, likely due to difficulties in fine-mapping when QTL
276 horizontal pleiotropy is extremely high⁵³. We found that the sensitivity of 90% credible sets decreased and
277 the mean set size increased with increasing proportion of shared isoQTLs. Our simulation results suggest
278 that varied isoQTL architectures and isoform-trait effects for isoforms of the same gene are key features
279 that influence power gains in isoform-centric modeling.

280

281 ***isoTWAS increases discovery of trait associations across 15 neuropsychiatric disorders***

282 To explore our central hypothesis that isoform-centric multivariate predictive modeling will improve
283 discovery for complex trait mapping, particularly for brain relevant traits, we next sought to apply and
284 directly compare isoTWAS and TWAS results across 15 neuropsychiatric disorders and traits. To
285 maximize discovery, we trained both isoTWAS and TWAS models using a greatly expanded adult brain
286 functional genomics reference panel (N = 2,115), comprised of frontal cortex samples from
287 PsychENCODE and AMP-AD Consortia^{27,54}, as well as using a developmental⁵⁵ pre-frontal cortex (N =
288 205) dataset (**Methods; Figure 5**). In the adult cortex, we trained models for 15,127 genes using
289 isoTWAS passing the CV $R^2 > 0.01$ cutoff, compared with 14,283 genes using gene-level TWAS. In the
290 developing cortex, despite a smaller sample size, isoTWAS models were successfully trained for 16,504
291 genes, compared with 10,535 genes using TWAS (**Methods; Supplemental Table S1**).

292
293 We next applied these models to perform trait associations using summary statistics from 15
294 neuropsychiatric and brain-related GWAS⁵⁶⁻⁷⁰ (**Methods, Figure 5a, Supplemental Figure S24**), using
295 the conservative stepwise hypothesis testing procedure (**Figure 1a**; FDR-adjusted $P < 0.05$ and within-
296 locus permutation $P_{ACAT} < 0.05$). We detected far more trait-associated genes with isoTWAS compared
297 with TWAS, across adult (2,595 vs 1,589 genes) and developmental (4,062 vs 890 genes) reference
298 panels, respectively (**Supplemental Figure S25, Supplemental Data 10-13**). In total across both
299 reference panels and all 15 traits, isoTWAS detected 3,436 unique gene and 5,377 unique isoform-trait
300 associations (**Supplemental Figure S26**). In addition, of the 1,335 genes with multiple isoform-trait
301 associations, 661 gene exhibited distinct isoform-level associations in different directions.

302
303 We next sought to compare the performance of isoTWAS/TWAS in prioritizing candidate mechanisms
304 within independent, high-confidence GWAS-significant loci⁷³. Across a combined 1,149 GWAS loci,
305 isoTWAS identified significant associations within 323, compared with 201 detected by TWAS, a ~60%
306 increase in discovery (**Figure 5b; Methods, Supplemental Table S4**). For example, of the 287 GWAS
307 loci identified for SCZ⁷⁴, isoTWAS prioritized genes within 70 and 86 unique loci across adult and
308 developmental reference panels, respectively, compared with 56 and 29 loci for TWAS (**Figure 5b**).

309 Further, across the 15 traits, 96% of gene-level TWAS associations (193/201) were concordantly
310 identified by isoTWAS. Likewise, the standardized effect sizes for significant gene- and isoform-level
311 associations were highly correlated ($r = 0.84$, $P < 2.2 \times 10^{-16}$; **Figure 5c**). Finally, to explore whether these
312 isoTWAS-specific associations were capturing true disease signal, we compared the rate at which each
313 method prioritized constrained genes (probability of loss-of-function intolerance, $pLI \geq 0.9$; **Supplemental**
314 **Tables S5-S8**), which are known to be substantially enriched for disease associations⁷⁵. Across adult and
315 developmental panels for 15 traits, respectively, isoTWAS prioritized 724 and 385 constrained genes
316 compared with 106 and 200 with TWAS, a significant increase (adult: $P=0.048$, developmental: $P = 1.23 \times$
317 10^{-5} , Fisher's Exact test). Altogether, these results emphasize that isoTWAS not only recovers the vast
318 majority of TWAS associations but also greatly increases discovery of candidate GWAS mechanisms,
319 particularly for genes intolerant to protein-truncating variation⁷⁶.

320
321 In our evaluation of methods in real data, we sought to compare discovery using isoTWAS to discovery
322 using local splicing-event based trait mapping. Briefly, for the developmental brain dataset, we first
323 calculated intron usage using Leafcutter⁷⁷ and transformed these usage percentages to M-values⁷⁸. Then,
324 for all introns mapped to a given gene, we used all SNPs within 1 Mb of a splicing-event to predict its
325 usage and mapped trait associations for these splicing events using isoTWAS's multivariate framework
326 (**Methods**). Overall, when aggregated to the gene-level, across 15 traits, we found that isoTWAS
327 prioritized features at ~40% more independent GWAS loci (167 loci) than splicing-event based trait
328 mapping (119 loci), with 108 loci (90.7%) jointly identified (**Figure 5f**), using the same developmental
329 brain reference panel. Taken together, isoTWAS's specific focus on modeling isoforms of a gene
330 provided considerable gains over considering only total gene expression or intron usage in identifying trait
331 associations for genes and their transcript-isoforms.

332
333 To investigate whether this increase in trait mapping discovery reflected true biological signal, rather than
334 test statistic inflation due to the increased number of tests (~4-fold increase in number of tests), we next
335 compared the null distributions across methods for results across the 15 traits (**Supplemental Figure**
336 **S27**). As the genomic inflation factor is not a reliable measure in TWAS settings⁷⁹, we estimated inflation

337 in gene-level test statistics using an empirical Bayes approach (**Methods**). Collapsing across all 15 traits,
338 there were no significant differences between TWAS and isoTWAS in the 95% credible intervals for test
339 statistic inflation (**Figure 5d**). Using a heuristic to estimate increases in effective sample size (**Methods**),
340 we observed an approximate increase in effective sample size of 10-20% when using isoTWAS
341 compared to TWAS (**Figure 5e, Supplemental Table S9**). These analyses indicate that isoTWAS
342 discovery is both well-calibrated to the null and facilitates increased discovery in real data compared to
343 gene-level TWAS.

344
345 We also empirically compared probabilistic fine-mapping⁵⁵ of results from isoTWAS and gene-level TWAS
346 (**Methods**). Here, we conducted fine-mapping on significant trait-associated genes/isoforms (adjusted $P <$
347 0.05 and permutation $P < 0.05$) and are within 1 Mb of one another; we term a locus with overlapping
348 genes/isoforms a risk region. Overall, the mean number of genes in a risk region using TWAS was 3.15
349 compared to 3.90 using isoTWAS (**Supplemental Figure 28a**); the mean number of genes in a 90%
350 credible set using TWAS was 1.33 compared to 1.25 using isoTWAS (**Supplemental Figure 28a**). On
351 average, there were 1.54 isoforms per gene in a risk region and 1.27 isoforms per gene in a 90% credible
352 set (**Supplemental Figure 28b**). Isoform-centric modeling presents unique challenges for fine-mapping
353 due to potentially high levels of horizontal pleiotropy, and remains an important and open question for the
354 field. Nevertheless, isoTWAS identified a relatively comparable number of genes in risk regions compared
355 with TWAS, and the combination of conservative two-step trait mapping, permutation testing, and
356 probabilistic fine-mapping were critical for maintaining a narrow credible set size.

357

358 ***isoTWAS identifies biologically-meaningful trait associations undetectable by TWAS***

359 Overall, isoTWAS prioritized dozens of novel candidate risk genes and mechanisms in the developing
360 and adult brain for 15 neuropsychiatric traits. These isoTWAS-prioritized genes were enriched for multiple
361 relevant pathways consistent with the biology of the underlying trait, including cell proliferation pathways
362 for brain volume, calcium channel activity for schizophrenia and neuroticism, and proteasome regulation
363 in Alzheimer's Disease, among others (**Supplemental Figure S29**). We highlight below several
364 examples of novel trait associations in which isoTWAS prioritized a highly constrained gene within a

365 GWAS locus (**Supplemental Table S5-8**). For example, we identified an association for SCZ with
366 developmental expression of ENST00000519133, an isoform of *SNAP91* (adjusted $P = 6.06 \times 10^{-7}$ in
367 isoTWAS, $pLI = 0.99$, chromosomal location of 6q14.2), with a GWAS-significant variant rs217291 within
368 the gene body. *SNAP91* has been predicted to affect clathrin and phosphatidylinositol binding activity and
369 synaptic vesicle recycling and has previously shown to impact synaptic development⁸⁰. In addition, we
370 found that imputed developmental cortex expression of ENST00000476671, an isoform of *KMT2E*, was
371 associated with CDG risk (adjusted $P = 7.63 \times 10^{-3}$, $pLI = 1$, chromosomal location of 7q22.3); the GWAS
372 SNP rs2385537, associated in a meta-analysis of ADHD, ASD, BP, and SCZ⁸¹, is within the gene body.
373 *KMT2E* regulates post-translational histone methylation of histone 3 on lysine 4A, and *KMT2E*
374 heterozygous variants are associated with risk of neurodevelopmental disorders⁸². Imputed adult brain
375 expression of ENST00000492146 (adjusted $P = 1.14 \times 10^{-6}$, $pLI = 0.94$, chromosomal location of 3p21.1),
376 an isoform of *SFMBT1*, showed a strong association with SCZ risk and contains a GWAS SNP within the
377 gene body (rs2071044). In a recent gene-based analysis of GWAS data for SCZ and BD, decreased
378 expression of *SFMBT1* was associated with increased risk of both disorders⁸³, consistent with the effects
379 for the isoform we identify in this isoTWAS analysis. Lastly, our analysis implicated ENST00000537270,
380 an isoform of *KMT5A* ($pLI = 0.99$, found in 12q24.31), in an association with SCZ risk. *KMT5A*, a H4K20
381 methyltransferase, has been previously implicated in GWAS for SCZ but cis-eQTLs of the gene have not
382 be colocalized with the GWAS signal previously^{74,84,85}.

383

384 As evidenced by our simulation analyses, a main advantage of isoTWAS over TWAS is the identification
385 of trait associations for isoforms of genes, where the gene itself is not associated with the trait. We
386 illustrate several examples of isoforms prioritized by isoTWAS, all in adult cortex, for genes with limited or
387 distinct expression QTLs (**Figure 6, Supplemental Figure S30, Supplemental Data 14**). First, we
388 detected a strong SCZ association with ENST00000492957, an isoform of *AKT3* ($pLI = 1$) at 1q43-q44,
389 which encodes a serine/threonine-protein kinase that regulated many processes like metabolism and cell
390 growth, proliferation, and survival. *AKT3* has shown effects on anxiety, spatial and contextual memory,
391 and fear extinction in mice, such that loss-of-function of *AKT3* causes learning and memory deficits^{86,87}.
392 Within the GWAS locus, there was a strong overlapping isoQTL signal ($P < 10^{-50}$); however, the eQTL

393 signal includes only a single SNP that reaches $P < 10^{-6}$, which is in low LD with the GWAS-significant
394 SNPs in the locus (**Figure 6a**). The lead isoQTL (rs4430311) showed a significant, negative association
395 with ENST00000492957, but not *AKT3* expression, with increasing SCZ risk alleles. However, this SNP
396 has only a nominally significant positive association with *AKT3*, as the number of alternative alleles at the
397 SNP increased. Interestingly, a different isoform of *AKT3* (ENST00000681794) was prioritized in an
398 association with BV, which also has a GWAS hit at this locus (**Supplemental Figure S30**). The two
399 distinct isoforms of *AKT3* have distinct 3' transcript structures, particularly close to the lead isoQTL of
400 ENST00000681794. These results suggest a complex role of *AKT3* isoforms with brain-related traits to be
401 explored further computationally and experimentally.

402
403 Similarly, we found a strong isoQTL signal for ENST00000409096 but a weak eQTL signal of *CUL3* in the
404 2q36.2 locus (pLI = 0.99), in another association with SCZ (**Figure 6b**). *CUL3* is a component of Cullin-
405 RING E3 ubiquitin ligase complexes, involved in protein ubiquitylation, cell cycle regulation, protein
406 trafficking, signal transduction, and transcription. Previous work has implicated *CUL3* dysregulation as a
407 pathological mechanism for both SCZ and ASD risk⁸⁸. Next, an isoform ENST00000678969 of *HSPD1*,
408 encoding a mitochondrial heat shock protein, was associated with SCZ risk (adjusted $P = 8.79 \times 10^{-12}$, pLI
409 = 0.99, 2q33.1) and showed a similar pattern across GWAS, eQTL, and isoQTL signals (**Figure 6c**).
410 *HSPD1* was identified in 3 independent secondary analyses of SCZ GWAS data, is among multiple non-
411 MHC immune genes implicated in SCZ, and has roles in brain hypomyelination⁸⁹. Our work adds to these
412 previous studies by implicating a specific isoform of *HSPD1* at the locus. Lastly, ENST00000423517, an
413 isoform of *PCLO*, was associated with multiple traits in the CDG analysis (meta-analysis of ADHD, BP,
414 MDD, and SCZ; adjusted $P = 1.83 \times 10^{-4}$, pLI = 1). Again, we found a strong isoQTL but not eQTL signal,
415 with the CDG risk allele negatively associated with isoform expression. *PCLO* is involved in the
416 presynaptic cytoskeletal matrix, establishing active synaptic zones, and synaptic vesicle trafficking; rare
417 variants of *PCLO* in diverse populations have been recently implicated in risk of SCZ and ASD^{90,91}.
418 Altogether, these results highlight the importance of incorporating isoform-level regulation for prioritizing
419 novel candidate GWAS risk mechanisms, as implemented in our isoTWAS framework.

420

421 **DISCUSSION**

422 In this work, we introduce isoTWAS, a scalable framework that integrates genetic and isoform-level
423 transcriptomic variation with GWAS to not only identify genes whose expression are associated with a
424 trait but also prioritize a set of isoforms of the gene that best explains the association. We provide an
425 extensive set of isoform-level predictive models (<https://zenodo.org/record/6795947>⁹²) and software to
426 conduct isoform-level trait mapping with GWAS summary statistics ([https://github.com/bhattacharya-a-
427 bt/isoTWAS](https://github.com/bhattacharya-a-bt/isoTWAS)).

428
429 As demonstrated above, isoTWAS presents several advantages over traditional gene-level TWAS or
430 simple univariate modeling of isoform expression. First, modeling expression at the isoform-level can
431 detect isoQTL architectures that vary across isoforms and thus may not be captured by eQTLs of a gene.
432 Second, joint multivariate isoform-level modeling improved predictive accuracy of isoform and total gene
433 expression, both in simulated and real data with independent replication. Third, aggregating isoform-level
434 associations to the gene-level through ACAT substantially increased power to detect trait associations
435 over traditional TWAS. We attribute this increase in power to several key features: first, isoform-level
436 modeling in isoTWAS increases the number of imputable genes by >2-fold; second, isoTWAS models
437 improve accuracy of gene-level prediction by as much as 35%; third, isoTWAS jointly models both
438 expression and splicing regulation, thereby capturing multiple potential underlying complex trait
439 mechanisms. Finally, as genetic control of isoform expression and usage are often much more tissue-
440 and cell-type-specific than eQTLs^{26,37}, we hypothesize that isoTWAS is more capable of uncovering such
441 context-specific trait associations.

442
443 Recent work has highlighted the promise of alternative splicing as a biological mechanism underlying
444 complex traits not fully captured through eQTLs^{24,26,27,93}. Splicing-QTL integration is a promising vehicle
445 to prioritize alternative exons or splice sites that may explain the genetic association with a trait, but a
446 single exon or splice site can correspond to multiple isoforms of the same gene. Mapping genetic
447 regulation at the exon- rather than gene-level often leads to more detected signal⁹⁴. However, most of
448 these analyses have focused on local splicing events or exon-level inclusion, rather than the combined

449 consequences of these events – namely, different isoforms of the same gene. Local splicing events can
450 be computationally intensive to measure and are difficult to systematically integrate across multiple
451 distinct large-scale datasets, which is necessary for achieving sample sizes sufficient for interrogation of
452 population-level allelic effects^{27,28}. Furthermore, multiple splicing changes are often coordinated across a
453 gene, yielding many non-independent features that increase multiple testing burden. Our results
454 demonstrate that isoform-centric trait mapping with isoTWAS increases discovery by ~40% compared
455 with a matched local splicing event-based analysis, although these methods may recover some
456 independent signal. Future work should consider integrating both reference-guided and annotation-free
457 approaches for quantification and detection of isoform and splicing patterns to develop more nuanced
458 mechanistic hypotheses for GWAS loci.

459

460 We conclude with some limitations of and future considerations for isoTWAS. First, we note that isoform-
461 level expression quantifications are maximum-likelihood estimates, due to the inherent limitations of
462 short-read RNA-seq. These estimates are generally guided by existing transcriptome annotations (e.g.,
463 GENCODE) and thus are dependent on the completeness and accuracy of these genomic annotations.
464 Further, many dataset-specific sequencing factors may will affect the accuracy of these estimates,
465 especially sequencing depth, read length and library preparation (mRNA vs total RNA sequencing). The
466 continued emergence of long-read sequencing platforms, including PacBio IsoSeq and Oxford Nanopore
467 Technologies, will be instrumental for improving reference transcriptome annotations, particularly with
468 tissue specificity, which will in turn improve isoTWAS. Further, as these methods continue to gain
469 scalability and cost effectiveness, they will ultimately replace short-read sequencing (and isoform
470 estimation) for population-scale datasets. As isoTWAS is agnostic to the method of isoform expression
471 quantification, this framework will continue to be applicable as we approach this long read sequencing
472 era. Recent analyses have shown that larger sample sizes may outweigh sequencing coverage for eQTL
473 mapping, but this relationship has not been investigated for isoform-level expression⁹⁵. Thus, for optimal
474 discovery with isoTWAS, the appropriate balance between sample size and reference panel sequencing
475 depth remains to be determined.

476

477 Second, while inferential replicates from RNA-seq quantification can provide measures of technical
478 variation, they are not incorporated into the predictive models. Our analyses of prediction across
479 inferential replicates suggest a methodological opportunity: leveraging these inferential replicates as a
480 measure of quantification error may help in estimating the robustness of isoform prediction and,
481 potentially, of the precision of these SNP effects. A more flexible predictive model that estimates standard
482 errors for SNP effects by model-averaging across the replicate datasets may help with trait mapping by
483 providing a prediction interval for both isoform- and gene-level imputed expression. Third, just as TWAS
484 can be cast as a differential gene expression analysis conducted with imputed expression, isoform-level
485 trait mapping is akin to differential transcript expression analysis. An analogous goal of isoTWAS can be
486 to detect trait associations with genetically-regulated transcript usage. However, it is unclear if the
487 compositional nature of transcript usage data needs to be accounted for at the prediction step or the trait
488 mapping steps⁹⁶. Lastly, as we show, this framework can suffer from reduced power, inflated false
489 positives, and reduced sensitivity in fine-mapping in the presence of SNP horizontal pleiotropy, where the
490 genetic variants in the isoform expression model affect the trait, independent of isoform expression, or
491 when multiple SNPs affect expression of isoforms^{97,98}. For pathways that are not observed or accounted
492 for in the reference expression panel and GWAS, accounting for horizontal pleiotropy may improve trait
493 mapping. We motivate further methodological extensions of probabilistic fine-mapping to reconcile
494 pleiotropy for SNPs shared across models for multiple isoforms at the same genetic locus, as summary-
495 statistic based methods that control for horizontal pleiotropy are not yet effective⁹⁹.

496

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

502

503 **METHODS**

504 isoTwas consists of three steps: (1) training predictive models of isoform expression, (2) imputing
505 isoform-specific expression into a separate GWAS panel, and (3) association testing between imputed
506 expression and a phenotype (**Figure 1b**). isoTwas contrasts with TWAS as it models correlations
507 between the expression of isoforms of the same gene. We outline each step below, with further
508 mathematical details in **Supplemental Methods**.

509

510 **Training predictive models of isoform expression**

511 *Model and assumptions*

512 Assume a given gene G has M isoforms with expression levels across N samples, with each sample
513 having R inferential replicates. Let Y_G^* be the $N \times M$ matrix of mean isoform expression (log-scale TPM) for
514 the N samples and M isoforms, using the point estimates from the Expectation-Maximization algorithm of
515 a pseudo-mapping quantification algorithm, like Salmon or kallisto^{32,33}. We can jointly model isoform
516 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
517 G , and replicate $r \in \{1, \dots, R\}$, we have:

$$518 \quad y_{nmr} = x_n \beta_m + \epsilon_{nmr} \quad (1),$$

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

$$529 \quad Y_G^* = X_G B_G + E_G \quad (2)$$

530 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
531 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

532 the variance-covariance matrix in the errors (with precision matrix $\Omega = \Sigma^{-1}$), following the aforementioned
533 independence assumptions.

534

535 *Estimating SNP effects on isoform expression*

536 We apply 5 different methods to estimate \hat{B}_G , the matrix of SNP effects on isoform expression. The first
537 four are multivariate methods that model the isoforms jointly; the last method models each isoform
538 separately using univariate methods. The goal of this SNP effect estimation is marginal prediction, i.e.,
539 leveraging the correlation between isoforms to improve prediction of each isoform separately. The \hat{B}_G
540 matrix that gives the largest adjusted R^2 in 5-fold cross-validation across the 5 methods is selected as the
541 final model to predict isoform expression for a given gene. In settings where we are interested in
542 predicting gene-level expression from these predicted isoforms, isoTWAS trains an elastic net penalized
543 linear regression that predicts gene-level expression from genetically-predicted isoform-level expression;
544 this model training is conducted across the same 5 folds to prevent data leakage¹⁰⁰. We provide an
545 overview of the methods, with mathematical details in **Supplemental Methods**:

- 546 1. *Multivariate elastic net (MVEnet) regression*: This is an extension of elastic net, where the response is
547 a matrix of correlated responses⁴³. Here, the absolute penalty is imposed on each single coefficient
548 by a group-lasso penalty on each vector of SNP effects across isoforms (rows of B_G); accordingly, a
549 SNP can only have a non-zero effect on an isoform if it has a non-zero effect on all isoforms.
- 550 2. *Multivariate LASSO Regression with covariance estimation (MRCE)*: We adapt Rothman et al's
551 proposed procedure to simultaneously and iteratively estimate both \hat{B}_G , the SNP effects matrix, and
552 $\hat{\Omega}$, the precision matrix⁴⁴. This procedure accounts for the correlation between isoforms but does not
553 impose the group-lasso penalty as in MVEnet. As such, a single SNP need not have a non-zero effect
554 on all isoforms.
- 555 3. *Multivariate elastic net with stacked generalization (joinet)*: We use Rauschenberger and Glaab's
556 joinet method that uses a two-step prediction⁴⁵. In the first step, the design matrix of SNPs is used to
557 generate a cross-validated prediction of each isoform. In the second step, the matrix of predicted
558 isoform expression is used to predict each isoform.

- 559 4. *Sparse partial least squares (SPLS)*: This is an implementation of partial least squares with a sparsity
560 penalty, that attempts to find an optimal latent decomposition for the linear relationship between the
561 matrix of isoform expression and the design matrix of SNPs. We use the Chun and Keles's
562 implementation from the *spls* R package⁴⁶.
- 563 5. *Univariate FUSION*: the simplest implemented method is the univariate predictive modelling used in
564 FUSION². We disregard the correlation structure between isoforms and train a univariate elastic net⁴³,
565 estimation of the best linear unbiased predictor (BLUP) in a linear mixed model¹⁰¹, and SuSiE⁴⁷
566 predictive model for each isoform separately. The model with the largest adjusted R² out of these
567 three models is outputted. This approach serves as a baseline measurement for prediction of each
568 isoform independently.

569

570 **Trait association and step-wise hypothesis testing**

571 We note that the tests of association in isoTWAS are similar to tests in differential transcript expression
572 analyses, as TWAS tests of association are analogous to tests in differential gene expression analyses.
573 isoTWAS and TWAS are distinct, as these methods consider imputed isoform and gene expression,
574 respectively, as predicted by the trained expression models. If individual-level genotypes are available in
575 the external GWAS panel, isoform expression can be directly imputed by multiplying the SNP weights
576 from the predictive model with the genotype dosages in the GWAS panel. If only summary statistics are
577 available, we adopt the weighted burden test from Gusev et al with an ancestry-matched linkage
578 disequilibrium panel^{4,98}. Compared to TWAS, a main distinction for isoTWAS association testing is the
579 increased number of tests (approximately 4-fold the number of isoforms than genes)²¹ and the potential
580 correlation in test statistics for isoforms of the same gene.

581

582 Accordingly, we perform a two-step hypothesis testing framework (**Supplemental Figure S2**)¹⁰². In the
583 first step, for every isoform with a trained model, we generate a TWAS test statistic using either linear
584 regression for GWAS with individual-level genotypes or the weighted burden test for GWAS with only
585 summary statistics⁴. Given the t test statistics T_1, \dots, T_t for a single gene, we used an omnibus test to
586 aggregate the t test statistics into a single P-value for a gene. We benchmark different omnibus tests in

587 simulations, but the default omnibus test in isoTWAS is aggregated Cauchy association test (ACAT)⁴⁹.
588 We control for false discovery across all genes via the Benjamini-Hochberg procedure, but the Bonferroni
589 procedure can also be applied for more conservative false discovery control. In the second step, for
590 isoforms for genes with an adjusted omnibus $P < 0.05$, we employ Shaffer's modified sequentially
591 rejective Bonferroni (MSRB) procedure to control the within-gene family-wide error rate. At the end of
592 these two steps, we identify a set of genes and their isoforms that are associated with the trait.

593

594 **Control for false positives within GWAS loci**

595 In TWAS and related methods, association statistics have been shown to be well calibrated under the null
596 of no GWAS association. However, within loci harboring significant GWAS signal, false positive
597 associations can result when eQTLs and GWAS coincide within overlapping LD blocks. To address this,
598 we have adopted two conservative approaches to control for type 1 error within GWAS loci, namely (1)
599 permutation testing and (2) probabilistic fine mapping. The permutation testing approach, adopted from
600 Gusev et al⁴, is a highly conservative test of the signal added by the SNP-transcript effects from the
601 predictive models, conditional on the GWAS architecture of the locus. Briefly, we shuffle the SNP-
602 transcript effects in the predictive models 10,000 times and generate a null distribution for the TWAS test
603 statistic for each isoform. We then use this null distribution to generate a permutation-based P-value for
604 the original test statistic for each isoform. Additionally, we aggregate these null distributions using an
605 omnibus test and compare the omnibus P-value to this distribution to generate permutation P-value for
606 each gene. Finally, we can employ isoform-level probabilistic fine-mapping using methods from FOCUS⁵⁵
607 to generate credible set of isoforms that explain the trait association at a locus. We only run isoform-level
608 fine-mapping for significantly-associated isoforms in overlapping 1 Megabase windows.

609

610 **Simulation framework**

611 We adopt techniques from Mancuso et al's *twas_sim* protocol¹⁰³ to simulate multivariate isoform
612 expression based on randomly simulated genotypes and environmental random noise. First, for n
613 samples, we generate a matrix of genotype dosages for the SNPs within 1 Megabase of 22 different

614 genes (1 per chromosome) using an LD reference panel of European subjects from 1000 Genomes
615 Project⁵¹.

616
617 Next, we generate a matrix of SNP-isoform effects across different causal SNP proportions p_c , numbers
618 of isoforms t , and p_s proportion of the SNP-isoform effects being shared across isoforms of the same
619 gene. We then add two matrices of random noise U and ϵ . The first matrix U noise represents non cis-
620 genetic effects on isoforms that are correlated between samples and isoforms; we control the proportion
621 of variance explained in isoform expression attributed to U using a parameter σ_h . The second matrix ϵ is a
622 matrix of random noise that is independent for each isoform, such that $\epsilon_i \sim N(0, \sigma_e^2 I)$ where $\sigma_e^2 = 1 -$
623 $\sigma_h - h_g^2$. We generate 10,000 simulations for each configuration of the simulation parameters, varying $n \in$
624 $\{200, 500\}$, $p_c \in \{0.001, 0.01, 0.05\}$, $h_g^2 \in \{0.05, 0.10, 0.25\}$, $p_s \in \{0, 0.5, 1\}$, and $\sigma_h \in \{0.1, 0.25\}$. Full
625 mathematical details are provided in **Supplemental Methods** and summarized in **Figure 2**.

626
627 We also generate traits under three distinct scenarios, with a continuous trait with heritability $h_t^2 \in$
628 $\{0.01, 0.05, 0.10\}$ and a GWAS sample size of 50,000 (**Supplemental Methods**):

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

641

642 Lastly, to estimate the approximate false positive rate (FPR), we followed the same simulation framework
643 to generate eQTL data and GWAS data. In the GWAS data, however, we set the effect of gene- and
644 isoform-level imputed expression to 0 to generate a simulated trait under the null, where the gene and
645 isoforms are not associated with the trait. We then estimate the FPR by calculating the proportion of
646 gene-trait associations at $P < 0.05$ under this null across 20 sets of 1,000 simulated GWAS panels. We
647 also assess isoform-level fine-mapping using FOCUS in a scenario with a gene with 5 or 10 isoforms and
648 a single effect isoform. We compute the sensitivity of 90% credible sets of isoforms (proportion of credible
649 sets that contain the effect isoform) and the number of isoforms in the 90% credible set.

650

651 **GTEx processing and model training**

652 We quantified GTEx v8³⁷ RNA-Seq samples for 48 tissues using Salmon v1.5.2³² in mapping-based
653 mode. We first built a Salmon index for a decoy-aware transcriptome consisting of GENCODE v38
654 transcript sequences and the full GRCh38 reference genome as decoy sequences²¹. Salmon was then
655 run on FASTQ files with mapping validation and corrections for sequencing and GC bias. We computed
656 50 inferential bootstraps for isoform expression using Salmon's Expectation-Maximization algorithm. We
657 then imported Salmon isoform-level quantifications and aggregated to the gene-level using tximeta⁴¹.
658 Using edgeR, gene and isoform-level quantifications underwent TMM-normalization, followed by
659 transformation into a log-space using the variance-stabilizing transformation using DESeq2^{104,105}. We
660 then residualized isoform-level and gene-level expression (as log-transformed CPM) by all tissue-specific
661 covariates (clinical, demographic, genotype PCs, and expression PEER factors) used in the original QTL
662 analyses in GTEx. We calculated the quantification variance across inferential replicates using the
663 `computeInfRV()` function from the fishpond package¹⁰⁶. We computed the isoform fraction using the
664 `isoformToIsoformFraction()` function from the IsoformSwitchAnalyzeR package⁵⁴.

665

666 SNP genotype calls were derived from Whole Genome Sequencing data for samples from individuals of
667 European ancestry, filtering out SNPs with minor allele frequency (MAF) less than 5% or that deviated
668 from Hardy-Weinberg equilibrium at $P < 10^{-5}$. We further filtered out SNPs with MAF less than 1%
669 frequency among the European ancestry samples in 1000 Genomes Project⁵¹.

670

671 Details of the model training pipeline for GTEx are similar to those summarized in **Supplemental Figure**
672 **S10**. Gene-level univariate models were trained using elastic net regression⁴³, best linear unbiased
673 predictor (BLUP) in a linear mixed model⁴⁸, and SuSiE⁴⁷, using all SNPs within 1 Mb of the gene
674 body^{4,43,47,101}. For each gene, the best performing model was chosen based on McNemar's adjusted 5-
675 fold cross-validation (CV) R^2 . We selected only genes with $CV R^2 \geq 0.01$. We applied multivariate
676 modeling outlined in isoTwas to train isoform-level predictive models, selecting only those isoform
677 models with $CV R^2 \geq 0.01$ (**Supplemental Figure S2**). All isoTwas models generated are publicly
678 available (see **Data Availability**).

679

680 **Developmental brain reference panel processing and model training**

681 We quantified developmental frontal cortex²⁴ (N = 205) RNA-Seq samples using Salmon v1.8.0³² in
682 mapping-based mode. We used the same indexed transcriptome as in the GTEx analysis and ran
683 Salmon with mapping validation and corrections for sequencing and GC bias. We computed 50 inferential
684 bootstraps for isoform expression using Salmon's Expectation-Maximization algorithm. We then imported
685 Salmon isoform-level quantifications and aggregated to the gene-level using tximeta⁴¹. Using edgeR,
686 gene and isoform-level quantifications underwent TMM-normalization, followed by transformation into a
687 log-space using the variance-stabilizing transformation using DESeq2^{104,105}. We then residualized
688 isoform-level and gene-level expression (as log-transformed CPM) by covariates (age, sex, 10 genotype
689 PCs, 90 and 70 hidden covariates with prior, respectively). Typed SNPs with non-zero alternative alleles,
690 $MAF > 1\%$, genotyping rate $> 95\%$, HWE $P < 10^{-6}$ were first imputed to TOPMed Freeze 5 using
691 minimac4 and eagle v2.4^{107,108}. We then retained biallelic SNPs with imputation accuracy $R^2 > 0.8$, with
692 rsIDs. Finally, we filtered out SNPs with $MAF < 0.05$ or that deviated from Hardy-Weinberg equilibrium at
693 $P < 10^{-6}$.

694

695 **Adult brain reference panel processing and model training**

696 Matched genotype and RNAseq data from adult brain cortex tissue from $n = 2,365$ individuals were
697 compiled and processed from the PsychENCODE Consortium²² and the Accelerating Medicines

698 Partnership Program for Alzheimer's Disease (AMP-AD)⁵⁶, consisting of the individual studies BipSeq,
699 BrainGVEX, CommonMind Consortium (CMC), CommonMind Consortium's National Institute of Mental
700 Health Human Brain Collection Core (CMC HBCC), Lieber Institute for Brain Development-szControl
701 (LIBD_szControl), UCLA-ASD, Religious Orders Study and the Memory and Aging Project (ROSMAP),
702 Mount Sinai Brain Bank (MSBB) and MayoRNAseq.

703

704 Typed genotypes were lifted over to the GRCh38 build using CrossMap v.0.6.3¹⁰⁹ and then filtered to
705 remove variants where the reference allele matched any of the alternate alleles. Genotype data from
706 whole genome sequencing (BrainGVEX, UCLA-ASD, ROSMAP, MSBB, and MayoRNAseq) was further
707 filtered to variants present on the Infinium Omni5-4 v1.2 array in order to satisfy the imputation server's
708 maximum limit of 20,000 typed variants per 20Mb. All genotype data was further processed with PLINK
709 v1.90b6.21¹¹⁰, removing variants with HWE $P < 10^{-6}$, $MAF < 0.01$ or missingness rate > 0.05 , and
710 removing samples with missingness rate > 0.1 across typed variants or missingness rate > 0.5 on any
711 individual chromosome. Genotype data was prepared for imputation using the McCarthy Group's HRC-
712 1000G-check-bim-v4.3.0 tool against freeze 8 of the Trans-Omics for Precision Medicine (TOPMed)
713 reference panel¹¹¹. The tool removes A/T and G/C SNPs with $MAF > 0.4$, variants with alleles that differ
714 from the reference panel, variants with an allele frequency difference > 0.2 from the reference panel and
715 variants not in the reference panel. Additionally, the tool updates strand, position and reference/alternate
716 allele assignment to match the reference panel.

717

718 Genotypes were then passed into the TOPMed Imputation Server by individual array batch¹¹². The
719 genotypes were phased with Eagle v2.4 and imputed with Minimac4 using the TOPMed reference
720 panel^{107,108}. Further QC was performed on the imputed genotypes using bcftools v1.11 and PLINK. The
721 imputed genotypes were filtered to well-imputed variants with $R^2 > 0.8$. The arrays were merged after
722 filtering to variants that were well-imputed in all arrays to be merged. Only arrays with at least 400,000
723 variants after pre-imputation QC were merged in order to prevent too many variants from dropping out.
724 The merged genotype data was then converted to PLINK 1 binary format and further processed with
725 PLINK, removing variants with duplicates, HWE $P < 10^{-6}$, $MAF < 0.01$ or missingness rate > 0.05 and

726 removing samples with missingness rate > 0.1. Samples from the same individual were identified by
727 calculating the genetic relatedness matrix using SnpArrays.jl and finding sets of samples with relatedness
728 > 0.75. From each set of replicates, only the genotyped sample from the array with the most variants after
729 pre-imputation QC was kept. For model training, only SNPs annotated in HapMap3 were retained¹¹³.

730
731 RNAseq paired reads from each study were sorted by name and then converted to FASTQ format using
732 samtools v1.14¹¹⁴. The reads were then quantified using salmon v1.8.0 in mapping-based mode using a
733 full decoy indexed from GENCODE v38 transcriptome and GRCh38 patch 13 assembly³². Quantification
734 was run using standard EM algorithm with library type automatically inferred and estimates adjusted for
735 sequence-specific and fragment-level GC biases. Bootstrapped abundance estimates were calculated
736 using 50 bootstrap samples. Isoform-level expression was summarized to the gene-level using tximeta¹¹⁵.
737 Only isoforms with 0.1 TPM for more than 75% of samples were retained. The resulting expression was
738 normalized using the variance-stabilizing transformation from DESeq2¹⁰⁵. Samples with WGCNA network
739 connectivity scores of less than -3 were removed as outliers, resulting in a total of 2,115 samples¹¹⁶.
740 Isoform- and gene-level expression was then batch-corrected using ComBat, using study site as the
741 batch¹¹⁷. Lastly, age, sex, squared age, 10 genotype PCs, and 200 (for gene expression) and 175 (for
742 isoform expression) hidden covariates with prior (using sequencing metrics from PicardTools to estimate
743 the prior) were removed from the expression^{118,119}. The number of hidden covariates with prior (HCP)
744 were selected by optimizing the number of nominal cis-gene-level expression QTLs and cis-isoform-level
745 expression QTLs at Bonferroni-corrected $P < 0.05$, respectively, on a grid from 100 to 300 HCPs, as
746 detected by QTLtools¹²⁰.

747
748 Details of the model training pipeline are summarized are equivalent to those used to train models in
749 GTEx data. All isoTWAS models generated are available at <https://zenodo.org/record/6795947>⁹².

750

751 **Gene- and isoform-level trait mapping**

752 We conducted gene- and isoform-level trait mapping for 15 neuropsychiatric traits: attention-deficit
753 hyperactivity disorder (ADHD, $N_{\text{cases}} = 20,183/N_{\text{controls}} = 35,191$)⁵⁹, Alzheimer's disease (ALZ,

754 90,338/1,036,225)⁶⁰, anorexia nervosa (AN, 16,992/55,525)⁷², autism spectrum disorder (ASD,
755 18,381/27,969)⁵⁸, bipolar disorder (BP, 41,917/371,549)⁶¹, brain volume (BV, N = 47,316)⁶², cross-
756 disorder (CDG, 232,964/494,162)⁶³, cortical thickness (CortTH, N = 51,665)⁶⁴, intracranial volume (ICV, N
757 = 32,438)⁶⁵, major depressive disorder (MDD, 246,363/561,190)⁶⁶, neuroticism (NTSM, N = 449,484)⁶⁷,
758 obsessive compulsive disorder (OCD, 2,688/7,037)⁶⁸, panic and anxiety disorders (PANIC,
759 2,248/7,992)⁶⁹, post-traumatic stress disorder (PTSD, 32,428/174,227)⁷⁰, and schizophrenia (SCZ,
760 69,369/236,642)⁷¹. For gene-level trait mapping, we used the weighted burden test, followed by the
761 permutation test, as outlined by Gusev et al⁴. For isoform-level trait mapping, we used the stage-wise
762 testing procedure outlined in the isoTWAS method. In-sample LD from the QTL reference panels was
763 used to calculate the standard error in the weighted burden test. For isoforms, irrespective of their
764 corresponding genes, passing both stage-wise tests and the permutation test, we employed isoform-level
765 probabilistic fine-mapping using FOCUS with default parameters⁵⁵. These methods are summarized in
766 **Supplemental Figure S24.**

767
768 We estimated the percent increase in effective sample size by employing the following heuristic. We
769 convert gene-level association P-values into χ^2 test statistics with 1 degree-of-freedom. For $\chi^2 > 1$, we
770 then calculate the percent increase for isoTWAS-based associations versus TWAS-based associations.
771 As the mean of the χ^2 distribution is linearly related to power and sample size¹²¹, we can use this percent
772 increase in test statistic as a measure of power or effective sample size. We defined independent
773 genome-wide significant SNPs in GWAS by LD clumping with lead GWAS SNP $< 5 \times 10^{-8}$ with P-value
774 used for ranking and a R^2 threshold of 0.2.

775 776 **DATA AVAILABILITY**

777 GTEx genetic, transcriptomic, and covariate data were obtained through dbGAP approval at accession
778 number phs000424.v8.p2. Linkage disequilibrium reference data from the 1000 Genomes Project were
779 obtained at this link: <https://www.internationalgenome.org/data-portal/sample>. GWAS summary statistics
780 were obtained at the following links: ADHD (<https://www.med.unc.edu/pgc/download-results/>), ALZ
781 (https://ctg.cncr.nl/software/summary_statistics/), AN (<http://www.med.unc.edu/pgc/results-and->

782 downloads), ASD (<https://www.med.unc.edu/pgc/download-results/>), BP
783 (<https://www.med.unc.edu/pgc/download-results/>), BV (https://ctg.cncr.nl/software/summary_statistics),
784 CDG (<https://www.med.unc.edu/pgc/results-and-downloads>), CortTH
785 (<https://enigma.ini.usc.edu/research/download-enigma-gwas-results/>), ICV
786 (<https://enigma.ini.usc.edu/research/download-enigma-gwas-results/>), MDD
787 (<http://dx.doi.org/10.7488/ds/2458>), NTSM
788 (https://ctg.cncr.nl/software/summary_statistics/neuroticism_summary_statistics), OCD
789 (<https://www.med.unc.edu/pgc/download-results/>), PANIC ([https://www.med.unc.edu/pgc/download-
790 results/](https://www.med.unc.edu/pgc/download-
790 results/)), PTSD (<https://www.med.unc.edu/pgc/results-and-downloads/>), and SCZ
791 (<https://www.med.unc.edu/pgc/download-results/>). The Developmental Brain RNA-seq and genotype
792 dataset from Walker et al is available at dbGAP with accession number phs001900. The subset of Adult
793 Brain RNA-seq and genotype data from the PsychENCODE Consortium is available at
794 <https://psychencode.synapse.org/DataAccess> and from AMP-AD is available at
795 <https://adknowledgeportal.synapse.org/Data%20Access>. GWAS summary statistics and accession
796 numbers to genotype and RNA-seq data are provided in **Supplemental Table S10**. isoTWAS models for
797 48 tissues from GTEx are available at <https://zenodo.org/record/8047940>¹²², adult brain cortex from
798 PsychENCODE and AMP-AD are available at <https://zenodo.org/record/8048198>¹²³, and developmental
799 brain cortex from Walker et al are available at <https://zenodo.org/record/8048137>¹²⁴.

800

801 **CODE AVAILABILITY**

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

804

805 **REFERENCES**

- 806 1. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *The American journal of*
807 *psychiatry* **175**, 15–15 (2018).
- 808 2. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **20**, 467–
809 484 (2019).

- 810 3. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
811 transcriptome data. *Nature Genetics* **47**, 1091–1098 (2015).
- 812 4. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies.
813 *Nature Genetics* **48**, 245–252 (2016).
- 814 5. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression
815 variation inferred from GWAS summary statistics. *Nature Communications* **9**, 1–20 (2018).
- 816 6. Hu, Y. *et al.* A statistical framework for cross-tissue transcriptome-wide association analysis. *Nature*
817 *Genetics* **51**, 568–576 (2019).
- 818 7. Zhou, D. *et al.* A unified framework for joint-tissue transcriptome-wide association and Mendelian
819 randomization analysis. *Nature Genetics* **52**, 1239–1246 (2020).
- 820 8. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association
821 detection. *PLoS Genetics* **15**, e1007889 (2019).
- 822 9. Bhattacharya, A., Li, Y. & Love, M. I. MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide
823 Association Studies. *PLOS Genetics* **17**, e1009398 (2021).
- 824 10. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields
825 mechanistic disease insights. *Nature Genetics* **50**, 538–548 (2018).
- 826 11. Wu, L. *et al.* A transcriptome-wide association study of 229,000 women identifies new candidate
827 susceptibility genes for breast cancer. *Nature genetics* **50**, 968–978 (2018).
- 828 12. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association
829 Studies Using Summary Statistics. *PLoS Genetics* **10**, e1004383–e1004383 (2014).
- 830 13. Gleason, K. J., Yang, F., Pierce, B. L., He, X. & Chen, L. S. Primo: Integration of multiple GWAS and
831 omics QTL summary statistics for elucidation of molecular mechanisms of trait-associated SNPs and
832 detection of pleiotropy in complex traits. *Genome Biology* **21**, 236–236 (2020).
- 833 14. He, X. *et al.* Sherlock: Detecting gene-disease associations by matching patterns of expression QTL
834 and GWAS. *American Journal of Human Genetics* **92**, 667–680 (2013).
- 835 15. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *American*
836 *Journal of Human Genetics* **99**, 1245–1260 (2016).

- 837 16. Umans, B. D., Battle, A. & Gilad, Y. Where Are the Disease-Associated eQTLs? *Trends in Genetics*
838 **37**, 109–124 (2021).
- 839 17. Barrera, L. O. *et al.* Genome-wide mapping and analysis of active promoters in mouse embryonic
840 stem cells and adult organs. *Genome Res* **18**, 46–59 (2008).
- 841 18. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–
842 476 (2008).
- 843 19. Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665
844 (2015).
- 845 20. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform
846 regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
- 847 21. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic*
848 *Acids Res* **47**, D766–D773 (2019).
- 849 22. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and
850 bipolar disorder. *Science* **362**, (2018).
- 851 23. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human
852 brain. *Science* **362**, (2018).
- 853 24. RL, W. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease
854 Mechanisms. *Cell* **179**, 750-771.e22 (2019).
- 855 25. Leung, S. K. *et al.* Full-length transcript sequencing of human and mouse cerebral cortex identifies
856 widespread isoform diversity and alternative splicing. *Cell Rep* **37**, 110022 (2021).
- 857 26. Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**,
858 600–604 (2016).
- 859 27. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and
860 bipolar disorder. *Science* **362**, (2018).
- 861 28. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human
862 brain. *Science* **362**, (2018).
- 863 29. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci.
864 *Genome Biology* 2021 22:1 **22**, 1–24 (2021).

865 30. MM, S. & MS, S. RNA mis-splicing in disease. *Nature reviews. Genetics* **17**, 19–32 (2016).

866 31. Akula, N. *et al.* Deep transcriptome sequencing of subgenual anterior cingulate cortex reveals cross-
867 diagnostic and diagnosis-specific RNA expression changes in major psychiatric disorders.
868 *Neuropsychopharmacol.* **46**, 1364–1372 (2021).

869 32. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware
870 quantification of transcript expression. *Nature Methods* **14**, 417–419 (2017).

871 33. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification.
872 *Nat Biotechnol* **34**, 525–527 (2016).

873 34. Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate
874 schizophrenia pathogenesis. *Nat Neurosci* **21**, 1117–1125 (2018).

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

878 36. Jaffe, A. E. *et al.* Profiling gene expression in the human dentate gyrus granule cell layer reveals
879 insights into schizophrenia and its genetic risk. *Nat Neurosci* **23**, 510–519 (2020).

880 37. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues.
881 *Science* **369**, 1318–1330 (2020).

882 38. Breiman, L. & Friedman, J. H. Predicting Multivariate Responses in Multiple Linear Regression.
883 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**, 3–54 (1997).

884 39. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies.
885 *Nature Genetics* **48**, 245–252 (2016).

886 40. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference
887 transcriptome data. *Nature Genetics* **47**, 1091–1098 (2015).

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

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

- 892 43. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via
893 Coordinate Descent. *Journal of Statistical Software* **33**, 1–22 (2010).
- 894 44. Rothman, A. J., Levina, E. & Zhu, J. Sparse multivariate regression with covariance estimation.
895 *Journal of Computational and Graphical Statistics* **19**, 947–962 (2010).
- 896 45. Rauschenberger, A. & Glaab, E. Predicting correlated outcomes from molecular data. *Bioinformatics*
897 **37**, 3889–3895 (2021).
- 898 46. Chun, H. & Keleş, S. Sparse partial least squares regression for simultaneous dimension reduction
899 and variable selection. *J R Stat Soc Series B Stat Methodol* **72**, 3–25 (2010).
- 900 47. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in
901 regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B*
902 **82**, 1273–1300 (2020).
- 903 48. Endelman, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP.
904 *The Plant Genome* **4**, 250–255 (2011).
- 905 49. Liu, Y. *et al.* ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in
906 Sequencing Studies. *Am J Hum Genet* **104**, 410–421 (2019).
- 907 50. Van den Berge, K., Sonesson, C., Robinson, M. D. & Clement, L. stageR: a general stage-wise
908 method for controlling the gene-level false discovery rate in differential expression and differential
909 transcript usage. *Genome Biology* 2017 18:1 **18**, 1–14 (2017).
- 910 51. Auton, A. *et al.* A global reference for human genetic variation. *Nature* vol. 526 68–74 Preprint at
911 <https://doi.org/10.1038/nature15393> (2015).
- 912 52. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical
913 protocol choice. *PLoS Genetics* **17**, e1009405–e1009405 (2021).
- 914 53. Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical
915 protocol choice. *PLoS Genetics* **17**, e1009405 (2021).
- 916 54. Vitting-Seerup, K. & Sandelin, A. IsoformSwitchAnalyzeR: analysis of changes in genome-wide
917 patterns of alternative splicing and its functional consequences. *Bioinformatics* **35**, 4469–4471
918 (2019).

- 919 55. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nature*
920 *Genetics* **51**, 675–682 (2019).
- 921 56. Vialle, R. A., de Paiva Lopes, K., Bennett, D. A., Crary, J. F. & Raj, T. Integrating whole-genome
922 sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the
923 human brain. *Nat Neurosci* **25**, 504–514 (2022).
- 924 57. RL, W. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease
925 Mechanisms. *Cell* **179**, 750-771.e22 (2019).
- 926 58. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature*
927 *genetics* **51**, 431–444 (2019).
- 928 59. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention
929 deficit/hyperactivity disorder. *Nature Genetics* **51**, 63–75 (2019).
- 930 60. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways
931 influencing Alzheimer’s disease risk. *Nature Genetics* **51**, 404–413 (2019).
- 932 61. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases
933 provides new insights into the underlying biology. *Nat Genet* **53**, 817–829 (2021).
- 934 62. Jansen, P. R. *et al.* Genome-wide meta-analysis of brain volume identifies genomic loci and genes
935 shared with intelligence. *Nat Commun* **11**, 5606 (2020).
- 936 63. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders.
937 *Cell* **179**, 1469-1482.e11 (2019).
- 938 64. Grasby, K. L. *et al.* The genetic architecture of the human cerebral cortex. *Science* **367**, eaay6690
939 (2020).
- 940 65. Adams, H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through
941 genome-wide association. *Nat Neurosci* **19**, 1569–1582 (2016).
- 942 66. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants
943 and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343–352
944 (2019).
- 945 67. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484
946 individuals identifies novel genetic loci and pathways. *Nat Genet* **50**, 920–927 (2018).

- 947 68. Arnold, P. D. *et al.* Revealing the complex genetic architecture of obsessive-compulsive disorder
948 using meta-analysis. *Molecular Psychiatry* **23**, 1181–1188 (2018).
- 949 69. Forstner, A. J. *et al.* Genome-wide association study of panic disorder reveals genetic overlap with
950 neuroticism and depression. *Mol Psychiatry* **26**, 4179–4190 (2021).
- 951 70. Nievergelt, C. M. *et al.* International meta-analysis of PTSD genome-wide association studies
952 identifies sex- and ancestry-specific genetic risk loci. *Nat Commun* **10**, 4558 (2019).
- 953 71. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia.
954 *Nature* **604**, 502–508 (2022).
- 955 72. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-
956 psychiatric origins for anorexia nervosa. *Nat Genet* **51**, 1207–1214 (2019).
- 957 73. Prive, F., Aschard, H., Ziyatdinov, A. & Blum, M. G. B. Efficient analysis of large-scale genome-wide
958 data with two R packages: Bigstatsr and bigsnpr. *Bioinformatics* **34**, 2781–2787 (2018).
- 959 74. Consortium, T. S. W. G. of the P. G., Ripke, S., Walters, J. T. & O'Donovan, M. C. Mapping genomic
960 loci prioritises genes and implicates synaptic biology in schizophrenia. 2020.09.12.20192922 Preprint
961 at <https://doi.org/10.1101/2020.09.12.20192922> (2020).
- 962 75. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291
963 (2016).
- 964 76. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456
965 humans. *Nature* **581**, 434–443 (2020).
- 966 77. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151–158
967 (2018).
- 968 78. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by
969 microarray analysis. *BMC Bioinformatics* **11**, 587–587 (2010).
- 970 79. van Iterson, M., van Zwet, E. W., Heijmans, B. T. & Heijmans, B. T. Controlling bias and inflation in
971 epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome*
972 *Biology* **18**, 19–19 (2017).
- 973 80. Schrode, N. *et al.* Synergistic effects of common schizophrenia risk variants. *Nat Genet* **51**, 1475–
974 1485 (2019).

- 975 81. Bhattacharjee, S. *et al.* A Subset-Based Approach Improves Power and Interpretation for the
976 Combined Analysis of Genetic Association Studies of Heterogeneous Traits. *Am J Hum Genet* **90**,
977 821–835 (2012).
- 978 82. O'Donnell-Luria, A. H. *et al.* Heterozygous Variants in KMT2E Cause a Spectrum of
979 Neurodevelopmental Disorders and Epilepsy. *Am J Hum Genet* **104**, 1210–1222 (2019).
- 980 83. Reay, W. R. & Cairns, M. J. Pairwise common variant meta-analyses of schizophrenia with other
981 psychiatric disorders reveals shared and distinct gene and gene-set associations. *Transl Psychiatry*
982 **10**, 1–11 (2020).
- 983 84. Nishioka, K. *et al.* PR-Set7 Is a Nucleosome-Specific Methyltransferase that Modifies Lysine 20 of
984 Histone H4 and Is Associated with Silent Chromatin. *Molecular Cell* **9**, 1201–1213 (2002).
- 985 85. Schmidt-Kastner, R., Guloksuz, S., Kietzmann, T., van Os, J. & Rutten, B. P. F. Analysis of GWAS-
986 Derived Schizophrenia Genes for Links to Ischemia-Hypoxia Response of the Brain. *Front Psychiatry*
987 **11**, 393 (2020).
- 988 86. Wong, H. *et al.* Isoform-specific roles for AKT in affective behavior, spatial memory, and extinction
989 related to psychiatric disorders. *eLife* **9**, e56630 (2020).
- 990 87. Howell, K. R., Floyd, K. & Law, A. J. PKBy/AKT3 loss-of-function causes learning and memory
991 deficits and deregulation of AKT/mTORC2 signaling: Relevance for schizophrenia. *PLoS One* **12**,
992 e0175993 (2017).
- 993 88. Chen, H.-Y. & Maher, B. J. Lost in Translation: Cul3-Dependent Pathological Mechanisms in
994 Psychiatric Disorders. *Neuron* **105**, 398–399 (2020).
- 995 89. Pouget, J. G. The Emerging Immunogenetic Architecture of Schizophrenia. *Schizophrenia Bulletin*
996 **44**, 993–1004 (2018).
- 997 90. Liu, D. *et al.* Rare schizophrenia risk variant burden is conserved in diverse human populations.
998 2022.01.03.22268662 Preprint at <https://doi.org/10.1101/2022.01.03.22268662> (2022).
- 999 91. The GIT Family of Proteins Forms Multimers and Associates with the Presynaptic Cytomatrix Protein
1000 Piccolo* - Journal of Biological Chemistry. [https://www.jbc.org/article/S0021-9258\(20\)86625-3/fulltext](https://www.jbc.org/article/S0021-9258(20)86625-3/fulltext).
- 1001 92. Bhattacharya, A. isoTwas models using 48 GTEx tissues and PsychENCODE data (07.04.22).
1002 (2022) doi:10.5281/zenodo.6795947.

- 1003 93. Qi, T. *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat Genet*
1004 (2022) doi:10.1038/s41588-022-01154-4.
- 1005 94. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nature*
1006 *Communications* **8**, (2017).
- 1007 95. Schwarz, T. *et al.* Powerful eQTL mapping through low coverage RNA sequencing.
1008 2021.08.08.455466 Preprint at <https://doi.org/10.1101/2021.08.08.455466> (2021).
- 1009 96. Doose, G., Bernhart, S. H., Wagener, R. & Hoffmann, S. DIEGO: detection of differential alternative
1010 splicing using Aitchison's geometry. *Bioinformatics* **34**, 1066–1068 (2018).
- 1011 97. Veturi, Y. & Ritchie, M. D. How powerful are summary-based methods for identifying expression-trait
1012 associations under different genetic architectures? in *Pacific Symposium on Biocomputing* vol. 0
1013 228–239 (World Scientific Publishing Co. Pte Ltd, 2018).
- 1014 98. Bhattacharya, A. *et al.* Best practices for multi-ancestry, meta-analytic transcriptome-wide association
1015 studies: lessons from the Global Biobank Meta-analysis Initiative. *medRxiv* **3**, 2021.11.24.21266825-
1016 2021.11.24.21266825 (2021).
- 1017 99. Zhu, A. *et al.* MRLocus: Identifying causal genes mediating a trait through Bayesian estimation of
1018 allelic heterogeneity. *PLOS Genetics* **17**, e1009455–e1009455 (2021).
- 1019 100. Whalen, S., Schreiber, J., Noble, W. S. & Pollard, K. S. Navigating the pitfalls of applying
1020 machine learning in genomics. *Nat Rev Genet* **23**, 169–181 (2022).
- 1021 101. Endelman, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package
1022 rrBLUP. *The Plant Genome* **4**, 250–255 (2011).
- 1023 102. Van den Berge, K., Soneson, C., Robinson, M. D. & Clement, L. stageR: a general stage-wise
1024 method for controlling the gene-level false discovery rate in differential expression and differential
1025 transcript usage. *Genome Biology* 2017 18:1 **18**, 1–14 (2017).
- 1026 103. Wang, X., Lu, Z., Bhattacharya, A., Pasaniuc, B. & Mancuso, N. twas_sim, a Python-based tool
1027 for simulation and power analysis of transcriptome-wide association analysis. *Bioinformatics* **39**,
1028 btad288 (2023).
- 1029 104. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
1030 microarray studies. *Nucleic Acids Research* **43**, e47–e47 (2015).

- 1031 105. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
1032 seq data with DESeq2. *Genome Biology* **15**, 550–550 (2014).
- 1033 106. Zhu, A., Srivastava, A., Ibrahim, J. G., Patro, R. & Love, M. I. Nonparametric expression analysis
1034 using inferential replicate counts. *Nucleic Acids Research* **47**, e105 (2019).
- 1035 107. Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed)
1036 Consortium whole genome sequences improves imputation quality and detection of rare variant
1037 associations in admixed African and Hispanic/Latino populations. *PLoS Genetics* **15**, e1008500–
1038 e1008500 (2019).
- 1039 108. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel.
1040 *Nature Genetics* **48**, 1443–1448 (2016).
- 1041 109. Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome assemblies.
1042 *Bioinformatics* **30**, 1006–1007 (2014).
- 1043 110. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based
1044 Linkage Analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).
- 1045 111. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.
1046 *Nature* **590**, 290–299 (2021).
- 1047 112. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature Genetics* **48**,
1048 1284–1287 (2016).
- 1049 113. Belmont, J. W. *et al.* The international HapMap project. *Nature* **426**, 789–796 (2003).
- 1050 114. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
1051 (2009).
- 1052 115. Love, M. I. *et al.* Tximeta: Reference sequence checksums for provenance identification in RNA-
1053 seq. *PLoS Computational Biology* **16**, e1007664–e1007664 (2020).
- 1054 116. WGCNA: an R package for weighted correlation network analysis | BMC Bioinformatics | Full
1055 Text. <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559>.
- 1056 117. Leek, J. T. & Storey, J. D. Capturing Heterogeneity in Gene Expression Studies by Surrogate
1057 Variable Analysis. *PLoS Genetics* **3**, e161–e161 (2007).

- 1058 118. Mostafavi, S. *et al.* Normalizing RNA-Sequencing Data by Modeling Hidden Covariates with Prior
1059 Knowledge. *PLoS ONE* **8**, e68141–e68141 (2013).
- 1060 119. Picard toolkit. (2018).
- 1061 120. Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nature*
1062 *Communications* **8**, (2017).
- 1063 121. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological
1064 mechanisms mediating susceptibility to complex traits. *Nature Communications* **10**, 3834–3834
1065 (2019).
- 1066 122. Bhattacharya, A. isoTWAS models for 48 GTEx models (06/2023). (2023)
1067 doi:10.5281/zenodo.8047940.
- 1068 123. Bhattacharya, A. isoTWAS models for adult brain cortex (06/2023). (2023)
1069 doi:10.5281/zenodo.8048198.
- 1070 124. Bhattacharya, A. isoTWAS models for developmental brain cortex (06/2023). (2023)
1071 doi:10.5281/zenodo.8048137.

1072

1073

1074 **FIGURE LEGENDS**

1075 **Figure 1:** *Isoform-centric approach for complex trait mapping and prioritization of disease mechanisms at*
1076 *a genetic locus. (a)* Motivation for isoTWAS. Gene G has three isoforms associated with a trait but only
1077 one has an effect on the trait. Gene G itself does not show an association with the trait. Studying genetic
1078 associations with an isoform-centric perspective will prioritize Gene G, but not with a gene-centric
1079 perspective. **(b)** Schematic comparison of isoTWAS and TWAS. First, isoTWAS differs from TWAS by
1080 training a multivariate model of isoform expression, while TWAS models total gene expression, the sum of
1081 isoform expression. Second, isoTWAS maps isoform-trait associations through a step-wise hypothesis
1082 testing framework that provides gene-level false discovery control and isoform-level family-wide error rate
1083 control. TWAS only maps gene-trait association.

1084

1085 **Figure 2:** *IsoTwas models predict gene expression with more accuracy than TWAS models in simulated*
1086 *data. (a)* Simulation setup to generate isoform expression with specified isoQTL architecture, controlled
1087 expression heritability, number of isoforms, and inter-isoform correlation structure. **(b)** Proportion of
1088 simulations where the isoTwas model has the maximum adjusted R^2 for marginal isoform prediction (Y-
1089 axis) across shared isoQTL proportion (X-axis), colored by isoTwas method, faceted by causal isoQTL
1090 proportion (top margin) and proportion of isoform expression variance attributed to shared non cis-genetic
1091 effects (right margin). **(c)** Boxplots of difference in adjusted R^2 in predicting gene expression between
1092 isoTwas and TWAS models from simulations with sample size 500 where isoform and gene expression
1093 heritability are set to 0.05, across causal isoQTL proportion (X-axis) and colored by number of transcripts
1094 per gene, faceted by proportion of shared isoQTLs (top margin) and proportion of variance explained by
1095 shared non cis-genetic effects (right margin).

1096
1097 **Figure 3:** *Multivariate isoform-level predictive models improve upon gene-level predictive models in*
1098 *predictive gene-level expression. (a)* Barplot showing the number of isoforms with $CV R^2 > 0.01$ (Y-axis)
1099 using multivariate (cream) and univariate (blue) modelling methods across brain tissues (X-axis). **(b)**
1100 Boxplot of median percent difference in predicting isoform expression (Y-axis) using multivariate
1101 compared to univariate method by brain and other tissue (X-axis). P-value corresponds to difference in
1102 median across groups, adjusted for sample size. **(c)** Barplot showing the number of genes passing CV
1103 thresholds (Y-axis) using TWAS (red) and isoTwas (blue) across brain tissues. **(d)** Barplot showing the
1104 number of genes with $CV R^2 > 0.01$ (Y-axis) using TWAS (red) and isoTwas (blue) across brain tissues
1105 (X-axis). **(e)** Boxplot of percent difference in R^2 (Y-axis) of predicting gene expression (isoTwas –
1106 TWAS) in external datasets. X-axis shows the training and imputation datasets. The median percent
1107 difference is labelled.

1108
1109 **Figure 4:** *IsoTwas improves power to detect gene-trait associations, especially when genetic effects*
1110 *differ across isoforms, in simulations. (a)* Power to detect gene-trait association (proportion of tests with P
1111 $< 2.5 \times 10^{-6}$, Y-axis) across causal proportion of isoQTLs (X-axis), faceted by proportion of shared
1112 isoQTLs (top margin) and proportion of variance explained by shared non cis-genetic effects (right

1113 margin). **(b)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across
1114 proportion of gene expression explained by effect isoform (X-axis). **(c)** Power to detect gene-trait
1115 association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across ratio of effect sizes for 2 effect isoforms.
1116 Across all plots, isoform and total gene expression heritability is set to 0.05 and causal proportion of 0.01
1117 in **(b-c)**. Points are colored by method and shaped by the number of isoforms per gene.

1118

1119 **Figure 5:** *Isoform-level trait mapping increases discovery of genetic associations over gene-level trait*
1120 *mapping.* **(a)** Schematic diagram for trait mapping using gene-level TWAS and isoTWAS using
1121 PsychENCODE data. **(b)** Number of gene-trait associations overlapping GWAS risk SNPs within 0.5 Mb
1122 using gene-level TWAS (red), isoTWAS (blue), or either (green) at conservative permutation-based
1123 significance thresholds. **(c)** Scatterplot of standardized effect sizes (Z-scores) using isoTWAS and gene-
1124 level TWAS, considered associations with nominal $P < 0.05$ using both TWAS and isoTWAS. Gray line
1125 shows the 45-degree line and the green line shows an ordinary least squares regression. **(d)** Empirical
1126 Bayes estimate of test statistic inflation (X-axis) for TWAS (red) and isoTWAS (red) gene-level
1127 associations across 15 traits (Y-axis). **(e)** Mean percent increase in approximate χ^2 -test statistic (squared
1128 Z-score), which is proportional to increase in effective sample size, for isoTWAS trait associations over
1129 TWAS trait associations. **(f)** Number of gene-trait associations overlapping GWAS risk SNPs within 0.5
1130 Mb using splicing-event-based TWAS (purple), isoTWAS (blue), or either (green) at conservative
1131 permutation-based significance thresholds.

1132

1133 **Figure 6:** *isoTWAS implicates isoforms of AKT3, CUL3, HSPD1, and PCLO in genetic associations with*
1134 *psychiatric traits.* In each plot, (top) Manhattan plots of GWAS risk, total gene eQTLs, and isoQTLs,
1135 colored by LD to the lead isoQTL and lead isoQTL shown in a triangle and labelled. LD is based on the
1136 1000 Genomes European reference. (bottom) Boxplots of gene and isoform expression by genotype of
1137 the lead isoQTL SNP and forest plot of the lead isoQTL's effect and 95% confidence interval on the trait,
1138 gene, and isoform, with P-values labelled. **(a)** SCZ risk, *AKT3* gene expression, and ENST00000492957
1139 isoform expression. **(b)** SCZ risk, *CUL3* gene expression, ENST00000409096 isoform expression. **(c)**

1140 SCZ risk, *HSPD1* gene expression, ENST00000678969 isoform expression. **(d)** CDG risk, *PCLO* gene
1141 expression, ENST00000423517 isoform expression.

1142

1143 SUPPLEMENTAL INFORMATION

1144 Supplemental Methods

1145 Supplemental Methods are provided in the attached SupplementalMethods.pdf.

1146

1147 Supplemental Figure Legends

1148 **Figure S1:** *Directed acyclic graph (DAG) illustrating causal assumptions in isoTWAS.* We assume that
1149 the local genetic variants within 1 Megabase of a gene have direct effects on the expression of a gene G
1150 and its isoforms; these genetic effects need not be shared across isoforms and the gene. Furthermore,
1151 we assume that the abundance of a gene is the sum of abundances of its isoforms. Lastly, we assume
1152 that the isoform and gene need not affect the complex trait through the same path. We also acknowledge
1153 that genetic variants may have effects on the trait through pathways independent of gene and isoform
1154 expression.

1155

1156 **Figure S2:** *Step-wise hypothesis testing in isoTWAS.* First, isoform-trait associations are estimated either
1157 using appropriate regressions of imputed expression in individual-level GWAS or the weighted burden
1158 test in Gusev et al in GWAS summary statistics. Then, associations for isoforms of the same gene are
1159 aggregated to the gene-level using the Aggregated Cauchy Association Test (ACAT). These aggregated
1160 gene-level associations are adjusted for multiple testing burden to control the false discovery rate (FDR)
1161 using either a Bonferroni or Benjamini-Hochberg procedure. Lastly, for isoforms of genes that pass gene-
1162 level testing, we control the family-wide error rate (FWER) using Shaffer's modified sequentially rejective
1163 procedure.

1164

1165 **Figure S3:** *Comparison of predictive performance of 6 models implemented in isoTWAS.* Across 1,000
1166 simulations of 5 isoforms with isoform heritability (h_i^2) set to 0.05 or 0.10. Boxplots of adjusted R^2 of
1167 prediction of isoform expression (Y-axis) across shared isoQTL proportion (X-axis)

1168

1169 **Figure S4:** *IsoTWAS models predict gene expression with more accuracy than TWAS models.* Boxplots
1170 of percent difference in adjusted R^2 in predicting gene expression between isoTWAS and TWAS models
1171 from simulations with sample size 200 (compare with sample size 500 in **Figure 2**), where isoform and
1172 gene expression heritability are set to (top) 0.05 and (bottom) 0.10.

1173

1174 **Figure S5:** Across 48 GTEx tissues (Y-axis), the number of multivariate (cream) and univariate (blue)
1175 models predicting isoform expression at $CV R^2 > 0.01$ (X-axis).

1176

1177 **Figure S6:** Across 48 GTEx tissues (Y-axis), percent difference in $CV R^2$ (X-axis) of prediction of isoform
1178 expression models using multivariate models versus univariate models. The label shows the proportion of
1179 isoforms with improved performance using multivariate models.

1180

1181 **Figure S7:** Number of isoforms with $CV R^2 > 0.01$ (Y-axis) using the baseline univariate model (teal, best
1182 univariate) and 4 multivariate models.

1183

1184 **Figure S8:** Across 48 GTEx tissues (Y-axis), number of genes that pass TWAS (blue) and isoTWAS (red)
1185 $CV R^2$ cutoffs to be available for testing in the trait-mapping step (X-axis)

1186

1187 **Figure S9:** Across 48 GTEx tissues (Y-axis), percent difference in $CV R^2$ (X-axis) of prediction of isoform
1188 expression models using multivariate models versus univariate models. The label shows the proportion of
1189 isoforms with improved performance using multivariate models.

1190

1191 **Figure S10:** Across 48 GTEx tissues (Y-axis), number of genes predicted at $CV R^2 > 0.01$ using TWAS
1192 (blue) and isoTWAS (red)

1193

1194 **Figure S11:** Number of genes (left) and isoforms (right) predicted at $CV R^2 > 0.01$ using isoTWAS across
1195 48 GTEx tissues.

1196

1197 **Figure S12:** Performance of isoTWAS across number of isoforms per gene across 48 GTEx tissues. **(a)**

1198 Ratio of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus univariate prediction. **(b)**

1199 Ratio of number of genes passing CV threshold using isoTWAS versus TWAS. **(c)** Median number of

1200 isoforms predicted at CV $R^2 > 0.01$ in isoTWAS models across increasing number of isoforms per gene.

1201 The red line shows the line $Y = X + 1$. **(d)** Ratio of number of genes with CV $R^2 > 0.01$ using isoTWAS

1202 versus TWAS.

1203

1204 **Figure S13:** Performance of isoTWAS across increasing maximum isoform fraction across 48 GTEx

1205 tissues. **(a)** Ratio of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus univariate

1206 prediction. **(b)** Ratio of number of genes passing CV threshold using isoTWAS versus TWAS. **(c)** Ratio of

1207 number of genes with CV $R^2 > 0.01$ using isoTWAS versus TWAS.

1208

1209 **Figure S14:** Performance of isoTWAS across increasing gene length across 48 GTEx tissues. **(a)** Ratio

1210 of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus univariate prediction. **(b)** Ratio of

1211 number of genes passing CV threshold using isoTWAS versus TWAS. **(c)** Ratio of number of genes with

1212 CV $R^2 > 0.01$ using isoTWAS versus TWAS. $R^2 > 0.01$ using isoTWAS versus TWAS.

1213

1214 **Figure S15:** Performance of isoTWAS across increasing SNP density across 48 GTEx tissues. **(a)** Ratio

1215 of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus univariate prediction. **(b)** Ratio of

1216 number of genes passing CV threshold using isoTWAS versus TWAS. **(c)** Ratio of number of genes with

1217 CV $R^2 > 0.01$ using isoTWAS versus TWAS.

1218

1219 **Figure S16:** Performance of isoTWAS across increasing sample size across 48 GTEx tissues. **(a)** Ratio

1220 of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus univariate prediction. **(b)** Ratio of

1221 number of genes passing CV threshold using isoTWAS versus TWAS. **(c)** Ratio of number of genes with

1222 CV $R^2 > 0.01$ using isoTWAS versus TWAS.

1223

1224 **Figure S17:** Performance of isoTWAS across proportion of shared isoTWAS model effect SNPs across
1225 48 GTEx tissues. **(a)** Ratio of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus
1226 univariate prediction. **(b)** Ratio of number of genes passing CV threshold using isoTWAS versus TWAS.
1227 **(c)** Ratio of number of genes with CV $R^2 > 0.01$ using isoTWAS versus TWAS.

1228

1229 **Figure S18:** Performance of isoTWAS across increasing mean counts of isoforms and genes across 48
1230 GTEx tissues. **(a)** Ratio of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus univariate
1231 prediction across increasing mean counts of isoforms. **(b)** Ratio of number of genes with CV $R^2 > 0.01$
1232 using isoTWAS versus TWAS across increasing mean counts of genes.

1233

1234 **Figure S19:** Performance of isoTWAS across increasing quantification variance of isoforms and genes
1235 across 48 GTEx tissues. **(a)** Ratio of number of isoforms predicted at $R^2 > 0.01$ using multivariate versus
1236 univariate prediction across increasing quantification variance of isoforms. **(b)** Ratio of number of genes
1237 with CV $R^2 > 0.01$ using isoTWAS versus TWAS across increasing quantification variance of genes.

1238

1239 **Figure S20:** **(a)** Median percent difference in R^2 of predicting original isoform expression using
1240 multivariate versus univariate models across increasing number of isoforms per gene, colored by models
1241 trained in the original dataset (pink) and the leave-one-out dataset (teal) **(b)** Median percent difference in
1242 R^2 of predicting original gene expression using isoTWAS versus TWAS models across increasing number
1243 of isoforms per gene, colored by models trained in the original dataset (pink) and the leave-one-out
1244 dataset (teal)

1245

1246 **Figure S21:** IsoTWAS and gene-level TWAS show relatively similar false positive rates. Across 20
1247 iterations of 1,000 simulations, boxplots of false positive rate to detect a gene-trait association using
1248 Cauchy-aggregated P-values of isoform-trait associations (red) and gene-level TWAS (blue). In these
1249 simulations, we simulate 200 and 5,000 samples in the eQTL and GWAS panels, 5 isoforms, and isoform
1250 and gene expression heritability of 0.10. We vary the causal isoQTLs proportion (p_c , shown on X-axis),
1251 the QTL architecture (right margin), and correlation between isoforms (top margin). For each iteration, we

1252 simulate one eQTL panel and 1,000 GWAS panels where genetically-regulated gene and isoform
1253 expression have no effect on the trait. We calculate the false positive rate as the proportion of the 1,000
1254 tests that give $P > 0.05$.

1255

1256 **Figure S22:** Power comparison between TWAS and isoTWAS in detecting gene-trait association across 3
1257 scenarios **(Methods)**. **(a)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-
1258 axis) across number of total isoforms per gene (X-axis), faceted by proportion of shared isoQTLs (top
1259 margin) and proportion of expression heritability attributed to shared non-genetic effects across isoforms
1260 (right margin). Points are shaped by causal isoQTL proportion and colored by method. **(b)** Power to
1261 detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across proportion of gene
1262 expression explained by effect isoform (X-axis), faceted by proportion of shared isoQTLs (top margin)
1263 and proportion of expression heritability attributed to shared non-genetic effects across isoforms (right
1264 margin). Points are shaped by number of isoforms per gene and colored by method. **(c)** Power to detect
1265 gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across ratio of effect sizes of 2 effect
1266 isoforms (X-axis), faceted by proportion of shared isoQTLs (top margin) and proportion of expression
1267 heritability attributed to shared non-genetic effects across isoforms (right margin). Points are shaped by
1268 number of isoforms per gene and colored by method. Here, expression heritability is set of 0.05, trait
1269 heritability is set to 0.1, and causal proportion of **(b-c)** is set of 0.01.

1270

1271 **Figure S23:** Sensitivity **(a)** and mean set size **(b)** of 90% credible set using FOCUS to finemap isoform-
1272 trait associations for a single gene, across causal isoQTL proportion (X-axis). Points are colored by trait
1273 heritability and shaped by the number of isoforms per gene. Plots are faceted by proportion of shared
1274 isoQTLs (top margin) and proportion of expression heritability attributed to shared non-genetic effects
1275 across isoforms (right margin). Line-range shows a 95% jackknife confidence interval.

1276

1277 **Figure S24:** Schematic diagram for analysis using adult and developmental frontal cortex data from
1278 *PsychENCODE* and *AMP-AD*. Data sources for eQTL reference data, GWAS cohorts, and reference LD

1279 data are provided on the left (black). The full gene-level TWAS (red) and isoTWAS (blue) are summarized
1280 on the right.

1281

1282 **Figure S25:** Number of gene-trait associations (Y-axis) using TWAS (red) and isoTWAS (blue) across
1283 trait (X-axis), faceting by tissue (top margin) and threshold (right margin: adjusted $P < 0.05$ and
1284 permutation $P < 0.05$, top; in 90% credible set using FOCUS fine-mapping, bottom).

1285

1286 **Figure S26:** Number of isoform-trait associations (Y-axis) using isoTWAS across trait (X-axis), faceting by
1287 tissue (top margin) and threshold (right margin: adjusted $P < 0.05$ and permutation $P < 0.05$, top; in 90%
1288 credible set using FOCUS fine-mapping, bottom).

1289

1290 **Figure S27:** QQ-plots of gene-level P-values using TWAS (red) and isoTWAS (blue) across 15 traits.

1291

1292 **Figure S28: (a)** Distribution of number of genes in risk region (left) and in 90% credible set (right) using
1293 TWAS and isoTWAS. **(b)** Distribution of number of isoforms per gene in risk region (left) and in 90%
1294 credible set (right) using isoTWAS.

1295

1296 **Figure S29:** Lollipop plot of enrichment ratio (X-axis) of ontologies (Y-axis) for isoTWAS-prioritized genes
1297 associated at adjusted $P < 0.05$ and permutation $P < 0.05$. Points are shaped by tissue type (adult or
1298 developmental) and colored by ontology type (biological process, cell component, molecular function).

1299

1300 **Figure S30:** For ENST00000681794 association with BV **(a)** and ENST00000492957 with BV **(b)**,
1301 Manhattan plots of GWAS, eQTL, and isoQTL signal colored by LD (top), boxplots of gene (red) and
1302 isoform (blue) expression (Y-axis) by genotype (X-axis) (bottom left), and forest plot of lead isoQTL
1303 association with isoform (blue), gene (red), and trait (black) (bottom right). **(c)** Comparison of exon and
1304 intron structure of ENST00000681794 and ENST00000492957, based on Gencode v38 reference.

1305

1306 **Supplemental Tables**

1307 **Table S1:** Sample size, source, and tissues for functional genomics reference panels
1308

1309 **Table S2:** Number of genes/models that pass cross-validation prediction cutoffs using TWAS and
1310 isoTWAS feature selection criteria and prediction methods
1311

1312 **Table S3:** Distribution of predictive external R² of observed total gene expression vs. predicted total gene
1313 expression (isoTWAS -TWAS)
1314

1315 **Table S4:** Number of GWAS loci with an isoTWAS-, TWAS-, and splice-TWAS-prioritized gene within 0.5
1316 Mb
1317

1318 **Table S5:** TWAS and fine-mapping results for genes with adjusted P < 0.05 and permutation P < 0.05
1319 across 15 traits using adult brain cortex models
1320

1321 **Table S6:** isoTWAS and fine-mapping results for genes with adjusted P < 0.05 and permutation P < 0.05
1322 across 15 traits using adult brain cortex models
1323

1324 **Table S7:** TWAS and fine-mapping results for genes with adjusted P < 0.05 and permutation P < 0.05
1325 across 15 traits using developmental brain cortex models
1326

1327 **Table S8:** isoTWAS and fine-mapping results for genes with adjusted P < 0.05 and permutation P < 0.05
1328 across 15 traits using developmental brain cortex models
1329

1330 **Table S9:** Empirical bayes estimates of test statistic inflation and increase in χ^2 test statistics of gene-
1331 level associations using isoTWAS and TWAS
1332

1333 **Table S10:** Accession numbers and URLs for data access
1334

1335 **Supplemental Data**

1336 **Data 1:** Predictive performance comparison of isoTWAS multivariate methods in simulated data across a
1337 variety of genetic architecture settings

1338

1339 **Data 2:** Predictive performance comparison of isoTWAS and TWAS gene expression prediction in
1340 simulated data across a variety of genetic architecture settings

1341

1342 **Data 3:** Isoform expression prediction metrics across a variety of factors, using 48 GTEx datasets

1343

1344 **Data 4:** Gene expression prediction metrics across a variety of factors, using 48 GTEx datasets

1345

1346 **Data 5:** False positive rates using isoTWAS and TWAS to detect a gene-trait association at $P < 0.05$
1347 across a variety genetic architecture parameters

1348

1349 **Data 6:** Power to detect trait association at $P < 2.5 \times 10^{-6}$ across 1,000 simulations each for 22 genes
1350 using TWAS and isoTWAS across various simulations. These simulations are under Scenario 1 in Figure
1351 4a (gene has a true effect on the trait, but none of the isoforms have a true effect on the trait)

1352

1353 **Data 7:** Power to detect trait association at $P < 2.5 \times 10^{-6}$ across 1,000 simulations each for 22 genes
1354 using TWAS and isoTWAS (ACAT) across various simulations. These simulations are under Scenario 2
1355 in Figure 4b (a gene has multiple isoforms, only one has an effect on the trait, and we vary the usage of
1356 this effect isoform)

1357

1358 **Data 8:** Power to detect trait association at $P < 2.5 \times 10^{-6}$ across 1,000 simulations each for 22 genes
1359 using TWAS and isoTWAS (ACAT) across various simulations. These simulations are under Scenario 3
1360 in Figure 4c (a gene has two isoforms with differing effects on the trait, and we vary the effect size of one
1361 of the isoforms)

1362

1363 **Data 9:** Sensitivity and mean set size of 90% credible sets determined by FOCUS in simulated data
1364 across a variety of genetic architecture parameters

1365

1366 **Data 10:** Raw TWAS results across 15 neuropsychiatric traits using adult brain cortex expression models

1367

1368 **Data 11:** Raw isoTWAS results across 15 neuropsychiatric traits using adult brain cortex expression
1369 models

1370

1371 **Data 12:** Raw TWAS results across 15 neuropsychiatric traits using developmental brain cortex
1372 expression models

1373

1374 **Data 13:** Raw isoTWAS results across 5 neuropsychiatric traits using developmental brain cortex
1375 expression models

1376

1377 **Data 14:** GWAS and nominal eQTL and isoQTL summary statistics corresponding to isoTWAS isoform-
1378 trait association examples shown in Figure 7 and Supplemental Figure S30.

1379

1380

1381 **ACKNOWLEDGEMENTS**

1382 We thank Kangcheng Hou, Tommer Schwarz, Vidhya Venkateswaran, Pan Zhang, Leanna Hernandez,
1383 Nathan LaPierre, Harold Pimentel, Mike Love, and Achal Patel for engaging discussion during the
1384 research process. We thank Kanishka Patel for her aesthetic advice for figures. We thank the Psychiatric
1385 Genomics Consortium and Complex Trait Genomics Lab for their publicly available GWAS summary
1386 statistics.

1387

1388 BP was partially supported by NIH awards R01 HG009120, R01 MH115676, R01 CA251555, R01
1389 AI153827, R01 HG006399, R01 CA244670, and U01 HG011715. MJG was supported by SFARI Bridge
1390 to Independence Award, NIMH R01-MH121521, NIMH R01-MH123922, and NICHD-P50-HD103557.

1391

1392 **AUTHOR INFORMATION**

1393 **Contributions**

1394 Conceptualization: AB, BP, MJG; Methodology: AB, BP, MJG; Software: AB; Validation: AB; Formal
1395 analysis: AB, MK, CW, CJ; Investigation: AB, MJG; Resources: BP, MJG; Data curation: AB, MK, CW,
1396 CJ, MJG; Writing – original draft: AB, MJG; Writing – reviewing and editing: AB, MK, CW, CJ, BP, MJG;
1397 Visualization: AB; Supervision: BP, MJG; Project administration: BP, MJG; Funding acquisition: BP, MJG

1398

1399 **Correspondence**

1400 Please direct all correspondence to Arjun Bhattacharya (abtbhatt@ucla.edu).

1401

1402 **ETHICS DECLARATIONS**

1403 **Competing Interests**

1404 The authors declare no competing interests.

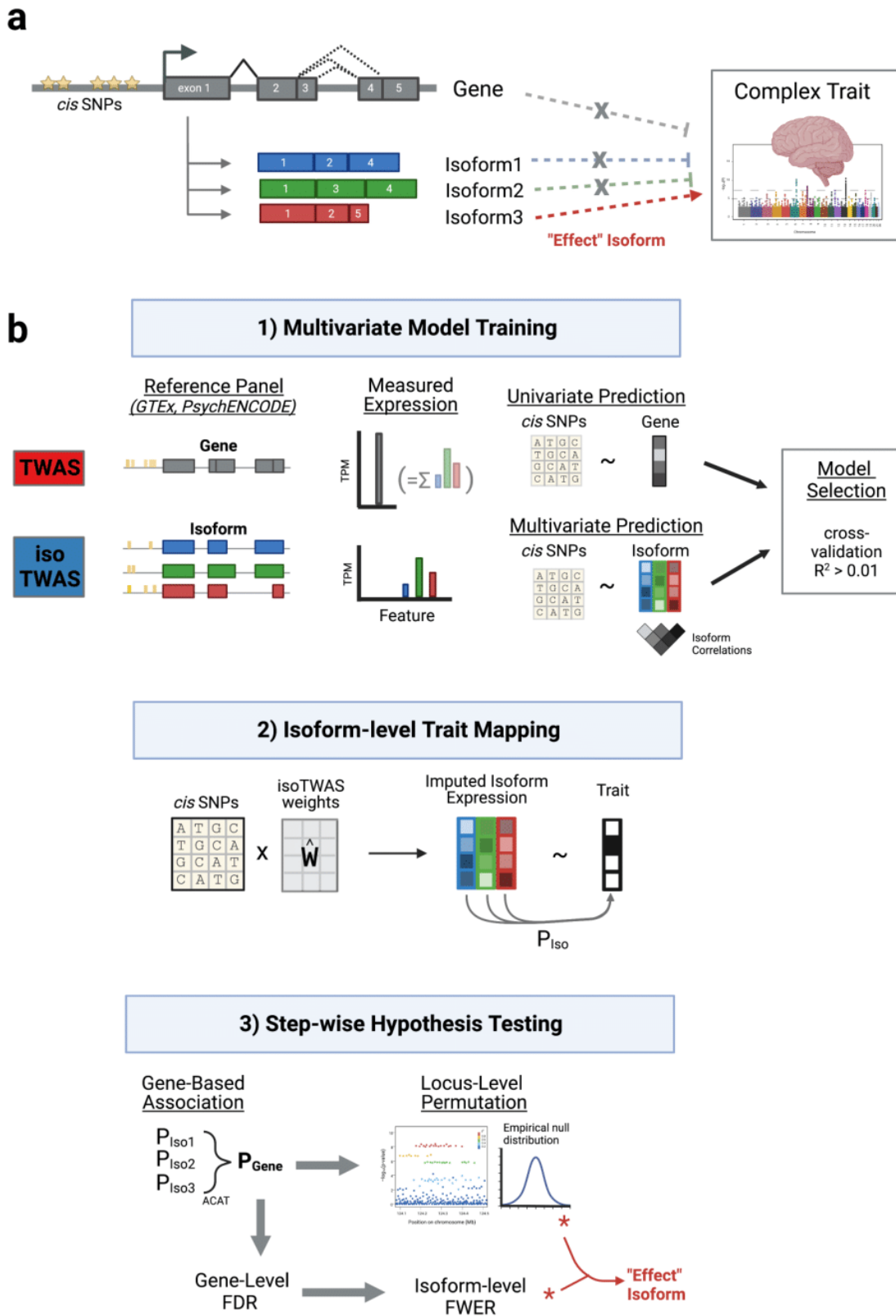
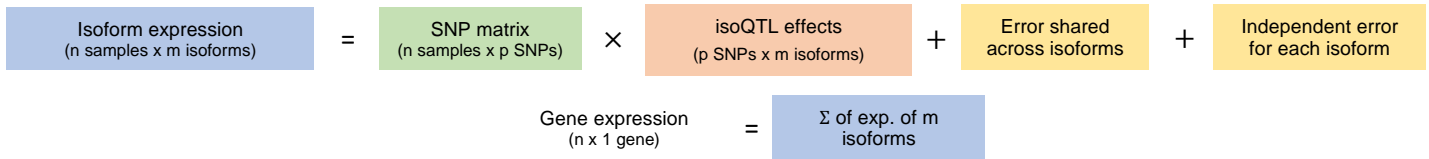


Figure 1: Isoform-centric approach for complex trait mapping and prioritization of disease mechanisms at a genetic locus. (a) Motivation for isoTWAS. Gene G has three isoforms associated with a trait but only one has an effect on the trait. Gene G itself does not show an association with the trait. Studying 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 association.

a I. Simulate isoform and gene expression using 1000 Genomes genotypes



II. Predict total expression using isoTAS and TWAS

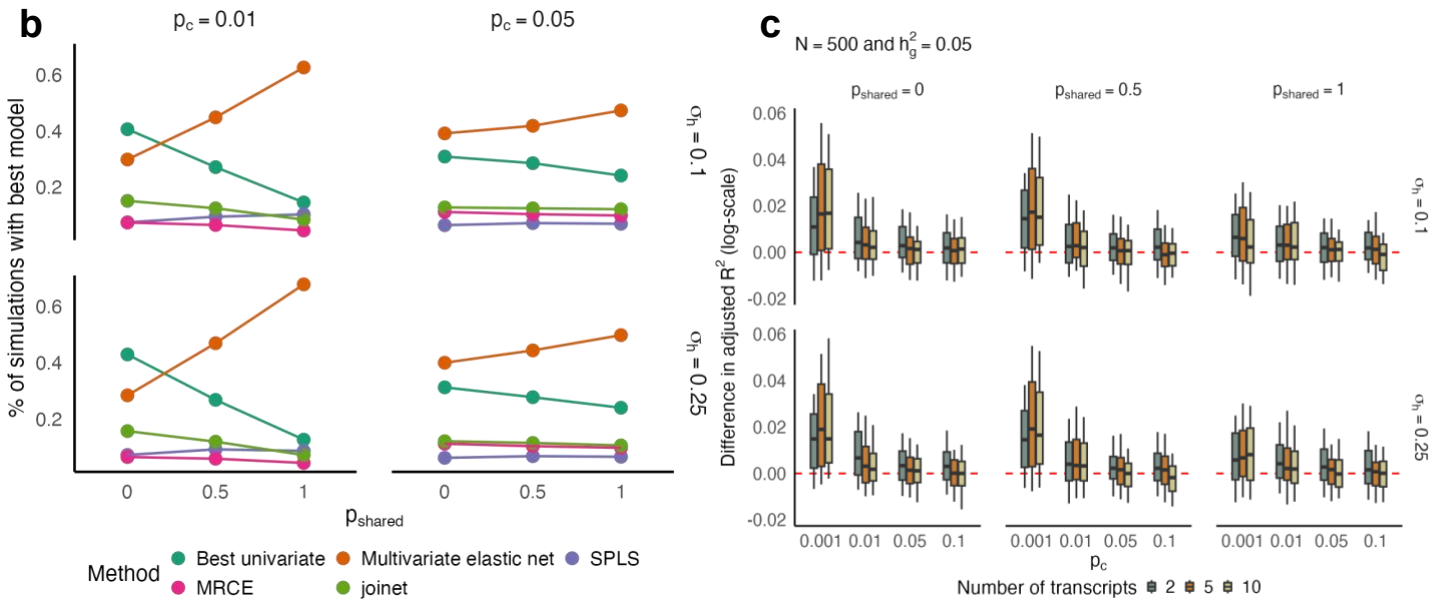
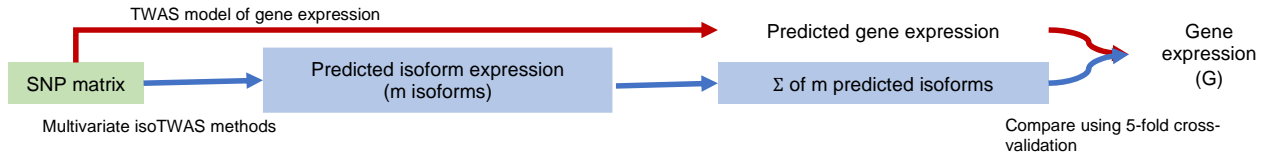


Figure 2: IsoTAS models predict gene expression with more accuracy than TWAS models in simulated data. (a) Simulation setup to generate isoform expression with specified isoQTL architecture, controlled expression heritability, number of isoforms, and inter-isoform correlation structure. **(b)** Proportion of simulations where the isoTAS model has the maximum adjusted R^2 for marginal isoform prediction (Y-axis) across shared isoQTL proportion (X-axis), colored by isoTAS method, faceted by causal isoQTL proportion (top margin) and proportion of isoform expression variance attributed to shared non cis-genetic effects (right margin). **(c)** Boxplots of difference in adjusted R^2 in predicting gene expression between isoTAS and TWAS models from simulations with sample size 500 where isoform and gene expression heritability are set to 0.05, across causal isoQTL proportion (X-axis) and colored by number of transcripts per gene, faceted by proportion of shared isoQTLs (top margin) and proportion of variance explained by shared non cis-genetic effects (right margin).

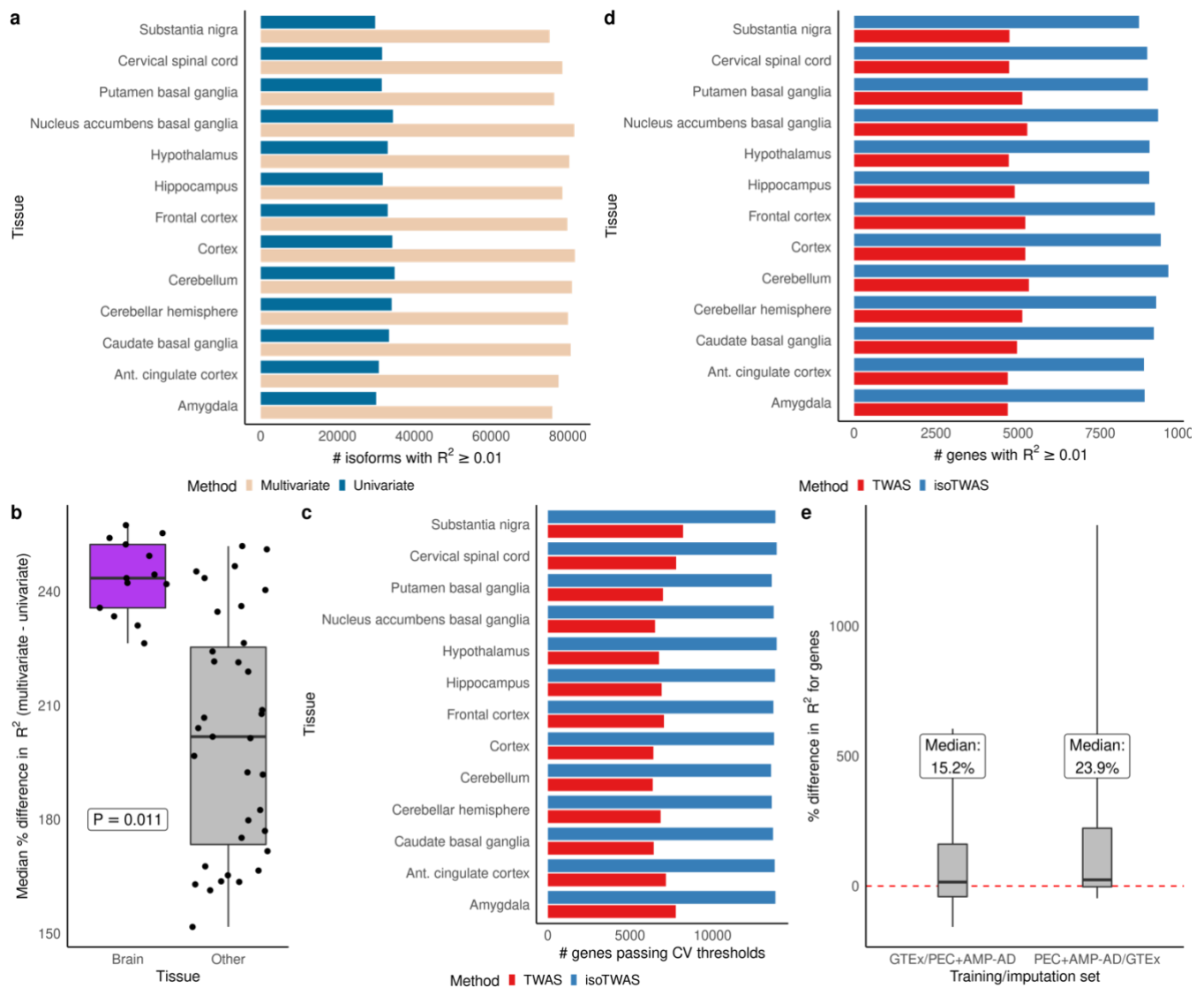


Figure 3: Multivariate isoform-level predictive models improve upon gene-level predictive models in predictive gene-level expression. (a) Barplot showing the number of isoforms with $CV R^2 > 0.01$ (Y-axis) using multivariate (cream) and univariate (blue) modelling methods across brain tissues (X-axis). (b) Boxplot of median percent difference in predicting isoform expression (Y-axis) using multivariate compared to univariate method by brain and other tissue (X-axis). P-value corresponds to difference in median across groups, adjusted for sample size. (c) Barplot showing the number of genes passing CV thresholds (Y-axis) using TWAS (red) and isoTWAS (blue) across brain tissues. (d) Barplot showing the number of genes with $CV R^2 > 0.01$ (Y-axis) using TWAS (red) and isoTWAS (blue) across brain tissues (X-axis). (e) Boxplot of percent difference in R^2 (Y-axis) of predicting gene expression (isoTWAS - TWAS) in external datasets. X-axis shows the training and imputation datasets. The median percent difference is labelled.

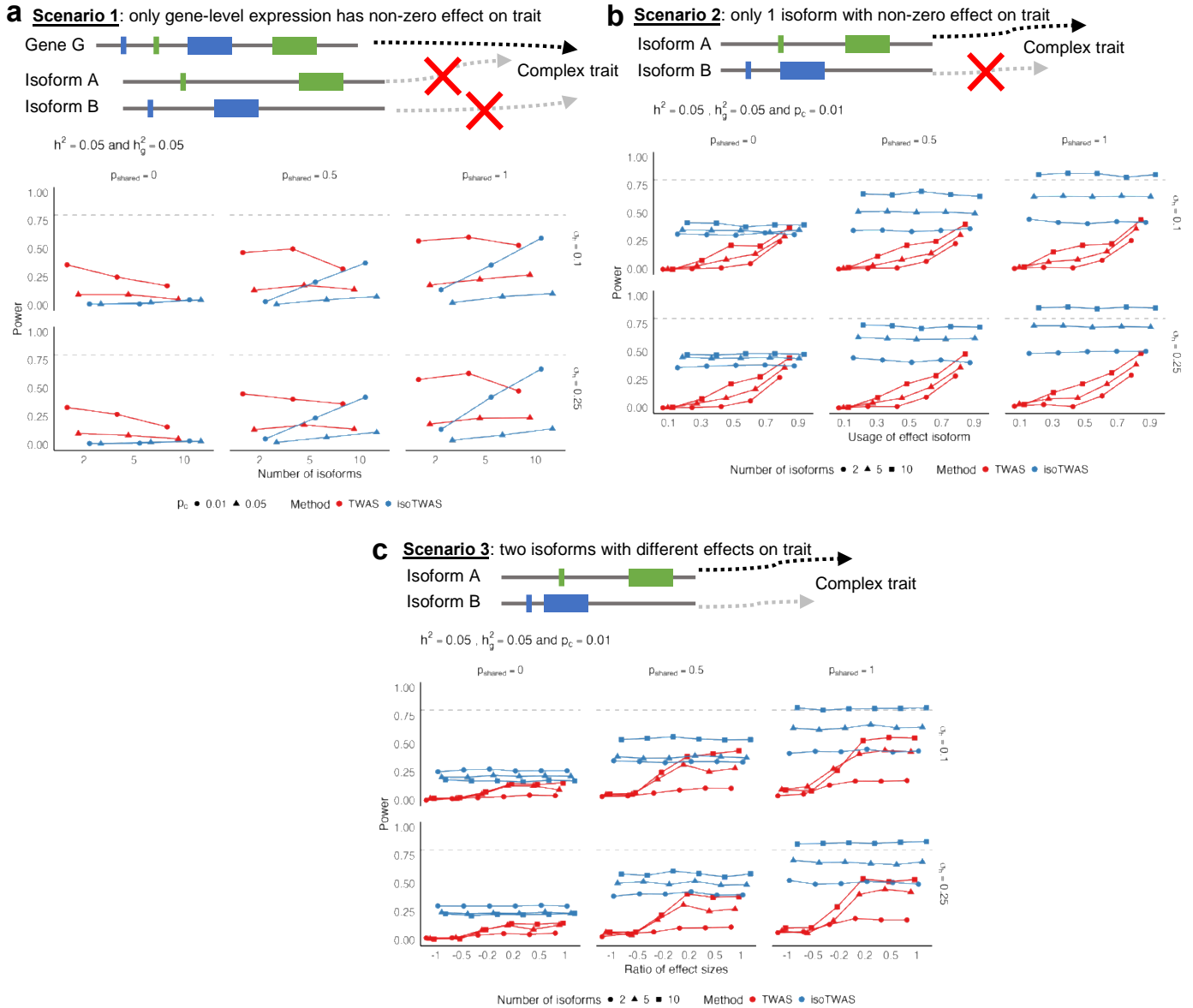


Figure 4: IsoTWAS improves power to detect gene-trait associations, especially when genetic effects differ across isoforms, in simulations. **(a)** Power to detect gene-trait association (proportion of tests with $P < 2.5 \times 10^{-6}$, Y-axis) across causal proportion of isoQTLs (X-axis), faceted by proportion of shared isoQTLs (top margin) and proportion of variance explained by shared non cis-genetic effects (right 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). **(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. Across all plots, isoform and total gene expression heritability is set to 0.05 and causal proportion of 0.01 in **(b-c)**. Points are colored by method and shaped by the number of isoforms per gene.

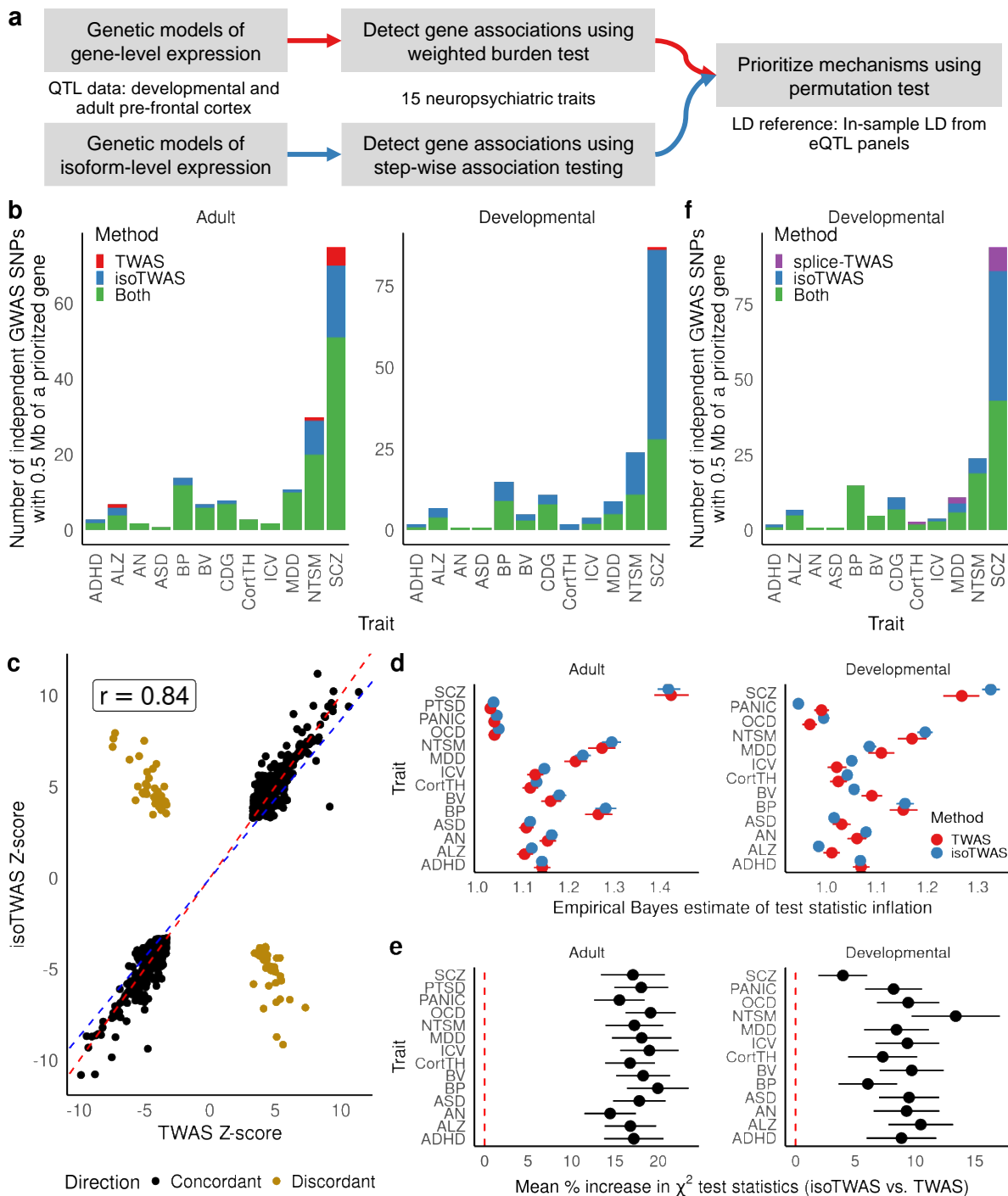


Figure 5: Isoform-level trait mapping increases discovery of genetic associations over gene-level trait mapping. (a) Schematic diagram for trait mapping using gene-level TWAS and isoTwas using PsychENCODE data. **(b)** Number of gene-trait associations overlapping GWAS risk SNPs within 0.5 Mb using gene-level 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)** Empirical Bayes estimate of test statistic inflation (X-axis) for TWAS (red) and isoTwas (red) gene-level associations across 15 traits (Y-axis). **(e)** 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. **(f)** Number of gene-trait associations overlapping GWAS risk SNPs within 0.5 Mb using splicing-event-based TWAS (purple), isoTwas (blue), or either (green) at conservative permutation-based significance thresholds.

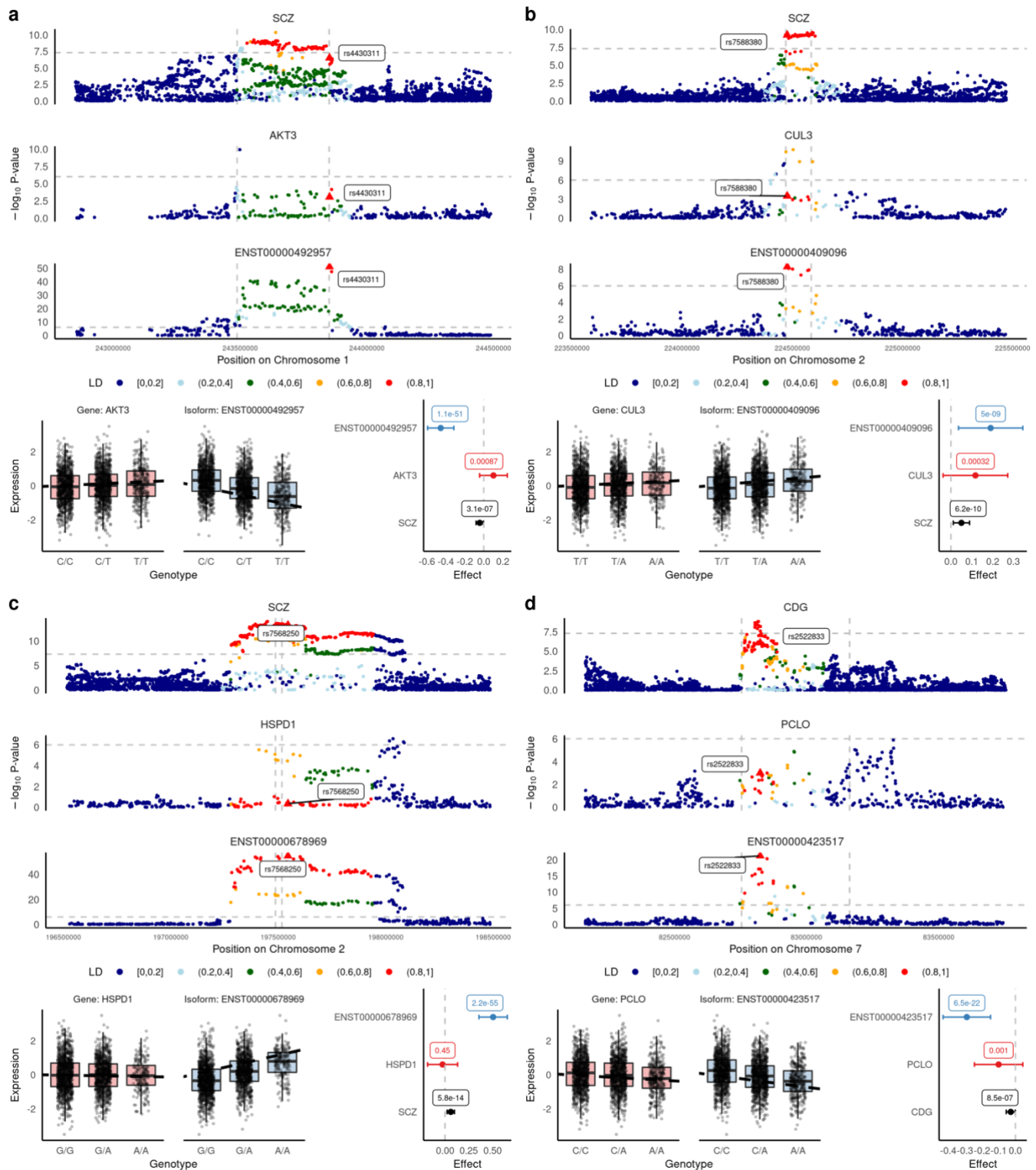


Figure 6: *isoTAS* implicates isoforms of *AKT3*, *CUL3*, *HSPD1*, and *PCLO* in genetic associations with psychiatric traits. In each plot, (top) Manhattan plots of GWAS risk, total gene eQTLs, and isoQTLs, colored by LD to the lead isoQTL and lead isoQTL shown in a triangle and labelled. LD is based on the 1000 Genomes European reference. (bottom) Boxplots of gene and isoform expression by genotype of the lead isoQTL SNP and forest plot of the lead isoQTL's effect and 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, *CUL3* gene expression, ENST00000409096 isoform expression. **(c)** SCZ risk, *HSPD1* gene expression, ENST00000678969 isoform expression. **(d)** CDG risk, *PCLO* gene expression, ENST00000423517 isoform expression.