Supplementary Note

CMC STRATIFIED BY SCZ CASES AND CONTROLS

Half of the CMC samples were SCZ/BIP cases, and we evaluated h_g^2 in the case/control subgroups separately. We observed apparent differences between cases and controls in both cis and trans estimates, with controls exhibiting significantly more cis-regulation (permuted P=2×10^{-04}). While little correlation is expected in estimates of cis-h_g^2, the trans components for each gene will be nearly identical and so any co-expressed genes will introduce additional variance that is not accounted for by the simple standard error. We permuted case/control labels and found only the cis-h_g^2 estimates to be significantly different (P=2×10^{-04}).

To ensure that results from the CMC data were not biased by ascertainment for SCZ/BIP cases, we performed a separate TWAS in which cis SNP-expression effect sizes were stratified on CMC case-control status and meta-analyzed; we observed no significant differences compared to using the full sample (Supplementary Table 4).

h_g^2 OF SPLICING EVENTS

Average cis and trans h_g^2 for splicing events (that mapped to canonical exon junctions) was 0.017 (se 0.001) and 0.046 (0.003) respectively, highly significant though lower than the average h_g^2 for total gene expression in the CMC data set (Supplementary Table 1). After Bonferroni correction, 2,847 introns (in 1,453 unique genes) were significantly cis-heritable, comparable to the 2,702 genes with significant total expression in the CMC (Supplementary Table 1). Interestingly, only 531 of the 1,453 unique genes with a significantly heritable intron also had significantly heritable total expression, confirming previous findings that heritable splice variants tend to be independent of cis-effects on total expression[1]. In contrast, we observed little independent transcript-level variation when analyzing individual probes in the NTR expression array data, and few additionally significant transcripts (Supplementary Table 1).

SPLICING EVENTS WITH COMPENSATORY EFFECTS

To demonstrate the complex relationship between multiple alternative exons within a single isoform, we consider *SDCCAG8*, which harbored six significant intron-level TWAS associations with effects in either direction (Supplementary Figure 7). We used step-wise model selection to identify two SNPs that were jointly significantly associated with alterantive splicing of multiple introns. We observe significant and opposite effects for the first SNP on introns 2 and 3 (Supplementary Figure 7, top); likewise, we observe significant and opposite effects for the second SNP on introns 1 and 3 (Supplementary Figure 7, bottom). Importantly, all effects remained significant after including total gene expression as a covariate, and the second SNP was not significantly associated with total expression. This example highlights the compensatory effect of splicing - the inclusion of intron 2 is inversely correlated with the inclusion of intron 3 - and the complexity of interpreting splicing TWAS effects.

TWAS REPLICATION WITH INDIVIDUAL-LEVEL PGC

We used the individual-level PGC data[2] to perform replication analyses and compare to summary-based results (we had permission to access individual genotypes for 58,246 of 79,845 samples). Cis SNP-expression effect sizes were computed using the BSLMM software (as in the main analysis) and predicted into each PGC sub-cohort to produce individual-level gene expression predictions. We first verified that TWAS using the raw GWAS data produced similar results as TWAS using summary statistics[3], observing a correlation of Z-scores ranging from 0.85-0.90 despite the somewhat different set of GWAS samples and independent measures of LD (Supplementary Table 5).

Next, we down-sampled the PGC data into GWAS discovery samples of various sizes (10,000-50,000) to quantify the power and out-of-sample replication of the SCZ TWAS associations (where the expression panel size was always held constant; Supplementary Figure 3A). We estimated TWAS effect-size precision as the slope of a regression of replication vs. discovery TWAS effect sizes across transcriptome-wide significant associations (Online Methods), where a slope below 1 represents over-estimated effect sizes in the discovery data due to winner's curse. Across random down-samples, average effect-size precision was 0.93 at a discovery

sample size of 50,000, indicating minimal winner's curse and projecting highly accurate effect-size estimates in the full TWAS. The number of transcriptome-wide significant associated genes increased linearly with GWAS sample size and was similar across all four expression reference panels (with more associations in brain expression at the largest discovery size), consistent with a polygenic architecture and many undiscovered associations.

TWAS REPLICATION IN CMC

We replicated the TWAS associations using case-control phenotypes from the CMC cohort (SCZ+BIP cases and controls; we included BIP cases due to the high genetic correlation between these two diseases [4] and insufficient power to distinguish differences at this sample size). We observed significant replication across all four expression panels, with no significant bias (Supplementary Figure 3B; Supplementary Figure 5). We note that even though the same CMC samples were used for the TWAS brain expression reference panel and replication using case-control status, this is an independent replication because CMC case-control status was never used in the discovery TWAS. Next, using all transcriptome-wide significant TWAS genes and splicing events identified in the PGC and their corresponding discovery effect sizes, we constructed gene-based risk scores (GE-PRS) from their predicted expression in the CMC (SCZ+BIP) case-control samples (Online Methods). The GE-PRS from each expression panel were significantly associated to case-control status, with the strongest association coming from the CMC/brain expression GeRS, which explained 1.4× more SCZ variance (and was more significant; $P = 2.0 \times 10^{-4}$ vs. 1.2×10^{-3}) than a genetic risk score computed using the 104 (non-HLA, autosomal) published genome-wide significant GWAS associations [2] (Supplementary Figure 3C). The GE-PRS remained significant in a joint model with the published GWAS predictor ($P = 0.01$, Supplementary Table 6).

ANALYSES OF BRAINSPAN TEMPORAL EXPRESSION

We sought to investigate temporal differences within the CMC/brain TWAS signal using individual transcriptomes collected by the BRAINSPAN study (see Web Resources) across developmental periods in the same brain sub-region (PFC) ranging from fetal to adult. For each of 19 developmental periods, we estimated differential expression relative to the other periods as an indicator of temporal specificity. The "RNA-Seq Gencode v10 summarized to genes" (RPKM) data was downloaded from the BRAINSPAN web-site and the following quality control performed: Samples without meta-data or with RIN scores < 8 were removed; data was quantile normalized to remove transcriptome-wide differences; $log_2(RPKM + 1)$-transformed to stabilize the variance; genes where transformed values were < 1 for more than 50% of samples were removed as under-expressed; co-expression was computed across all pairs of individuals and any samples more than 2.5 standard deviations away from the mean removed as outliers; finally, we restricted to the pre-frontal cortex subregion.

For each developmental period i and gene j, preferential expression for that period was computed across all genes as a Z-score $[GE_{i,j} - \mu(GE_{*,j})]/\sigma(GE_{*,j})$. The correlation matrix across of these Z-scores roughly divided the transcriptomes into two clusters split at 26 weeks post-conception (pcw), though non-monotonic correlation across some developmental periods were also observed (Supplementary Figure 11). For each developmental period, a signed Spearman rank correlation was computed between these Z-scores and the SCZ TWAS χ^2 statistics across all genes tested in the corresponding TWAS, and significance assessed (Supplementary Figure 12). A high correlation indicates that genes preferentially expressed in that developmental period also tend to be genes with a significant TWAS effect on schizophrenia.

We observed TWAS χ^2 statistics to be significantly positively correlated with differential expression from the mid-fetal developmental period ($P < 0.05/19$), and corresponding significant negative correlation for differential expression from post-fetal periods (Supplementary Figure 11, 12). The effect was most significant in the CMC/brain TWAS, less significant in the two blood TWAS, and non-significant in the adipose TWAS.

BRAINSPAN transcriptomes:
`http://www.brainspan.org/static/download.html`

ESTIMATING h^2_g FROM GENE-BASED POLYGENIC RISK SCORES

We used classical theory from risk prediction to estimate the h^2_g explained by the SNP and GeRS predictors [5,6]. The liability-scale prediction $R^2 = h^2_g h^2_g/(h^2_g + M_e/N)$ where N is the effective number of samples and M_e is the effective number of SNPs (estimated at 60,000 genotyped markers genomewide [7]). Given prediction R^2, M_e, and N, this equation can be used to back-calculate the h^2_g of the corresponding predictor [2]. For the published PGC 102,627 SNP predictor, the liability-scale prediction R^2 was 0.125, corresponding to an h^2_g of 0.37 under an $M_e = 60,000$. This h^2_g is higher than previous estimates, which could be explained by the presence of more extreme cases in the CMC data. After similarly pruning at $r^2 < 0.3$, the TWAS GeRS from all tissues jointly retained 14,023 genes, from which we computed an $M_e = 14,023 \times 60,000/102,627 = 8,413$. The liability-scale prediction R^2 was 0.046, corresponding to an h^2_g of 0.096 under an $M_e = 8,413$. The fraction of SNP-h^2_g explained by genes across all tissues was then $0.096/0.37 = 26\%$.

We separately estimated these parameters using the AVENGEME method of ref. [5], which uses a maximum likelihood model over the distribution of prediction R^2 values from multiple TWAS/GWAS p-value cutoffs. This model additionally infers the fraction of variants that are non-causal (π_0). The h^2_g of the PGC SNP predictor was 0.33 (95% CI of 0.30 - 0.36) and the π_0 was 0.78 (0.71 - 0.84). The h^2_g of the TWAS predictor from all tissues was 0.086 (0.060 - 0.099) and the π_0 was 0.67 (0.55 - 0.88). This fraction of SNP-h^2_g explained by genes across all tissues was then $0.086/0.33 = 26\%$.

We note that the SCZ variance explained by gene expression can be estimated under a fixed-effects model (ex: ref. [8]) or random-effects model (ex: ref. [9]) and may yield different estimates. We consider the estimate above to be a fixed-effects-based upper-bound, akin to asking how much of the variance in the phenotype can be explained by regressing out the predicted expression of all genes. Importantly, this estimate does not distinguish between (1) cis-expression that causally effects SCZ, and (2) pleiotropic effects of the same variants on expression and SCZ that are independent.

GENETIC CORRELATION OF CHROMATIN ELEMENTS

Genetic correlation of regulatory features and expression were estimated using Haseman-Elston (HE) regression [10]. HE regression was used to estimate cis-heritability σ^2_{g} by regressing the product of standardized phenotypes on the off-diagonal GRM entries using the linear model $y_i y_j \sim \sigma^2_{\mathrm{g}} K(i,j) + e$. Similarly, HE regression was used to estimate $\sigma_{\mathrm{g,A:B}}$, the cis-genetic covariance between traits A and B in the linear regression $A_i B_j \sim \sigma_{\mathrm{g,A:B}} K(i,j) + e$. Each value was computed as the mean from HE regressions across all target phenotypes and genetic correlation was then estimated from these mean values as $r_g = \sigma_{\mathrm{g,A:B}}/(\sigma^2_{\mathrm{g,A}} \sigma^2_{\mathrm{g,B}})^{1/2}$. The standard error for each h^2_g estimate was computed as the standard deviation over the square root of the number of estimates, and the standard error for r_g computed using the Falconer approximation $SE(r_g) \approx (1 - r^2_g)[SE(h^2_A) SE(h^2_B)/(2h^2_A h^2_B)]^{1/2}$. For a given chromatin phenotype (e.g. H3k27ac), we iterated over each gene and computed genetic correlation to all peaks ±500kb, with results averaged across all gene-peak pairs and reported as a function of distance. The same process was repeated for every other phenotype, substituting peaks for genes.

SIMULATIONS TO ASSESS CHROMATIN TWAS POWER

Using real genotypes from the UK10K [11] study, we simulated a model where SNPs X are causal for a chromatin peak ($C = X\beta_X + e_C$) and chromatin is causal for expression ($E = C\beta_C + e_E$). 100 1MB loci were randomly selected across the genome, and causal SNPs in each locus were then randomly selected from common variants (MAF > 1%). Environmental noise was set such that SNPs explain 30% of the variance in a chromatin peak, and chromatin explains 30% of the variance in expression (consistent with our observations that expression h^2_g is ∼ 10%). The TWAS was then performed using cis SNP-expression effect sizes computed by BSLMM either predicting the chromatin phenotype (with increasing sample size) into 1,000 independent individuals with expression or vice versa, and then performing an association between the predicted and measured phenotype. Separately, eQTL and cQTL association was computed for every common SNP in the locus using individuals with both expression and chromatin measured. To evaluate power at genome-wide significance, a TWAS association was reported as significant if it had $P < 0.05$ after

correcting for (average 30 peaks per locus) $\times$ (20,000 genes) = 600,000 tests. For the QTL-based approach, given M SNPs in the locus, a SNP was reported as significant if it had an eQTL $P < 0.05$ after correcting for $M \times 20,000$ tests and a cQTL $P < 0.05$ after correcting for $M \times 30 \times 20,000$ tests. For a given chromatin sample size, the simulation was then performed at 100 random loci and 5 random seeds each, with the fraction of loci reported as significant by each method taken as the power (Supplementary Figure 20). The TWAS simulation was separately performed under the null, using non-heritable chromatin and expression, and shown to be well-calibrated under the null (Supplementary Figure 19).

We confirmed by simulation that this chromatin TWAS strategy is well-calibrated (Supplementary Figure 19) and better powered to identify SNP $\rightarrow$ chromatin $\rightarrow$ expression associations compared to the conventional approach of testing each SNP for a significant association to both expression (eQTL) and nearby chromatin peaks (cQTL)[12,13] (Supplementary Figure 18A, 20).

SIMULATIONS TO EVALUATE TWAS SCATTERPLOTS

We performed simulations to explore the relationship between the TWAS predicted phenotype and marginal GWAS/QTL associations shown in Figure 4 and 5. Under the TWAS causal model and multivariate-normal assumptions, the Z-score of each SNP is (in expectation) the product of the TWAS phenotype Z-score and the SNP-TWAS phenotype correlation (labeled "Corr to GE_{pred}" in figures). This visualization therefore shows how well each SNP association follows the distribution that would be expected if the TWAS phenotype completely explained the genetic association at the locus.

To simulate the functional phenotypes, the same chromatin/expression model was used as described in the previous section: Using real genotypes from the UK10K[11] study, we simulated a model where SNPs X are causal for a chromatin peak ($C = X\beta_X + e_C$) and chromatin is causal for expression ($E = C\beta_C + e_E$). 100 1MB loci were randomly selected across the genome, and causal SNPs in each locus were then randomly selected from common variants (MAF > 1%). Environmental noise was set such that SNPs explain 30% of the variance in a chromatin peak, and chromatin explains 30% of the variance in expression (consistent with our observations that expression h^2_g is $\sim$ 10%). We then selected a GWAS causal variant to be: (1) a common SNP with a specified correlation to the genetic component of expression (i.e. the eQTL), corresponding to a model where expression is partially independent of GWAS; or (2) the actual genetic component of expression, corresponding to a model where expression is causal for the GWAS phenotype. The GWAS phenotype was simulated for a held-out set of 3,000 individuals by adding Gaussian environmental noise to the causal variant such that it explained 0.005% of the trait (consistent with a polygenic GWAS architecture). GWAS summary statistics for a simulated set of 90,000 individuals were then computed by regenerating the Gaussian environmental noise 30$\times$ and meta-analyzing the 30 resulting marginal statistics. After this process, any instances that did not contain a genomewide significant variant were dropped from the analysis. A total of 500 simulations for each architecture were performed.

Supplementary Figures 33, 34 show the results of these simulations using either chromatin or GWAS as the TWAS outcome. These simulations demonstrate that: (1, left panels) when the TWAS phenotype is non-causal there is no relationship between the SNP association and TWAS phenotype correlation; (2, right panels) when the TWAS phenotype is causal the relationship is linear across all of the marginal associations in the locus; (3, middle panels) when the TWAS phenotype is partially correlated to the causal GWAS/chromatin variant the relationship is partially linear but the most significant GWAS/chromatin associations are typically not the variants most strongly correlated to the TWAS phenotype. In general, as the correlation between expression and the chromatin/GWAS causal variant increases, the linear $y \sim x$ relationship emerges, but only becomes completely linear when correlation is > 0.8 (right two panels). Absence of the linear trend and/or presence of outlier SNPs that are highly GWAS significant (y values away from zero) but not strongly correlated to the predicted expression (x values near zero) should be treated as indicators of a coincidental TWAS association.

EQTL/CQTL OVERLAP ANALYSIS USING ALL SIGNIFICANT EQTLS

We considered an alternative eQTL/cQTL approach using all significant QTLs as used in refs.[12,13]. For each population and given distance to TSS, we performed this analysis in two stages. Stage 1: We identified all

eQTLs (by standard linear regression) that were significant after Bonferroni correction for the total number of SNP-gene pairs tested. When distance to TSS was the maximal allowed (500kb), this resulted in 355 eQTLs in the YRI and 579 eQTLs in the CEU data. Stage 2: From the set of significant eQTLs, we then looked for those that were also significantly associated (by standard linear regression) with a given chromatin phenotype (for peaks within the given distance to TSS), after Bonferroni correction for the number of eQTL-peak pairs tested. In both stages the tests were only counted for the given chromatin phenotype (e.g. H3K27ac in CEU). This was compared to the chromatin TWAS analysis where each gene was tested against any peak within the given distance (by standard linear regression), and number of significant results reported after Bonferroni correction for total number of gene-peak pairs tested in that phenotype (Supplementary Figure 18B).

PLEIOTROPY/ASSOCIATION BETWEEN CHROMATIN AND SCZ

For loci that were identified in both the SCZ TWAS and the chromatin TWAS, we sought to establish a direct connection between chromatin and SCZ using two approaches.

First, we performed a TWAS-like analysis for SCZ using the chromatin phenotype as the molecular measure instead of expression, which we call a CWAS (chromatin-wide association study). Given BSLMM estimates of the chromatin-QTL effect-size α, SCZ Z-scores Z, and an LD matrix D taken from the 1000 Genomes, the CWAS statistic is then $Z_{\text{CWAS}} = \alpha Z/(\alpha D \alpha)^{1/2}$. This differs from the formal TWAS analysis in that the numerator does not involve a matrix inversion to estimate the causal QTL effect-sizes, which we elected not to perform due to numeric instability at this sample size (see below). Of the 72 unique chromatin peaks which were implicated for SCZ, 67 had a significant ($P < 0.05$) CWAS association to SCZ and 47 were Bonferroni significant (Supplementary Table 21).

Next, we performed the single-variant SMR test [14] for pleiotropy between the chromatin phenotype and SCZ. The locus window was expanded to ±500kb of the gene and peak boundary to ensure that all SNPs that had been used for TWAS were evaluated. SMR uses the effect-size at the top e/c-QTL and the corresponding effect-size in the disease GWAS to estimate the causal or pleiotropic effect of molecular phenotype on disease. We used default parameters but included all SNPs in the cis locus and allowed any QTL with $P < 0.05$ to be selected as the instrument. Of the 42 SCZ TWAS genes with a chromatin TWAS association, 39 (93%) had a significant ($P < 0.05$) SMR effect of cQTL on SCZ and 24 were Bonferroni significant (Supplementary Table 21). Only 11 genes (out of 26 tested: 42%) had a significant ($P < 0.05$) SMR effect of eQTL on SCZ in the same samples (Supplementary Table 21), indicative of more power to detect pleiotropy for SCZ using cQTL rather than eQTL molecular associations. To investigate the difference in power further, we retested the loci identified in expression panels for which raw eQTL data was available (CMC/NTR) and compared to using eQTLs in the CEU/YRI samples (Supplementary Table 22, 23). All eQTLs yielded significant SMR effects on SCZ in the corresponding reference tissues (Supplementary Table 23). Notably the SMR effect sizes from the chromatin data were comparable to the SMR effect sizes from the (much larger) expression reference panels, again underscoring greater power to detect pleiotropy in the chromatin data (at the corresponding sample size).

ALTERNATIVE TESTS OF GENETIC COVARIANCE

The TWAS methodology applied here is fundamentally a test for non-zero genetic covariance between expression and a second phenotype. In this note we consider alternative frameworks in which this quantity could be estimated and their respective advantages/disadvantages.

We elected not to perform the TWAS test using variance-components methods due to computational intractability. Where raw GWAS data is available (as it was here for most of the SCZ samples) multi-variate variance-component estimators (typically multi-variate REML) have been shown to be more powerful than moments-based estimators such as the TWAS used here, Haseman-Elston regression, or LD-score regression [15,16]. Multi-variate REML does not require all phenotypes to be measured in all samples, and is often applied to completely disjoint phenotype pairs [4]. In principle, one can merge the chromatin, expression, and GWAS datasets to estimate the corresponding genetic correlations and perform statistical testing in a

variance-component framework. However, the computational burden of computing genetic relatedness matrices and corresponding matrix inversions for inference over 60,000 individuals for hundreds of thousands of chromatin peaks was too great to perform in practice. Though methods exist to estimate genetic correlation using Monte Carlo sampling without the need to construct a relatedness matrix[17], we are not aware of such methods for the disjoint phenotype case. Lastly, no multi-variate methods exist that incorporate the BSLMM non-infinitessimal framework, and current BSLMM methods are only computationally feasible for $\sim 10,000$ individuals[18]. We have previously demonstrated a substantial gain from using the BSLMM to predict expression[3] and we observed a similar gain in our chromatin TWAS simulations here.

We elected not to perform the TWAS directly from chromatin phenotypes (aside from replication analyses) because the sample size was too small to estimate predictor accuracy. The TWAS can be performed from any molecular phenotype into the SCZ GWAS cohort and the prediction accuracy will follow the canonical equation[5,6] $r^2 = (h_g^2 h_g^2)/(h_g^2 + M_e/N)$ where M_e and N is the number of SNPs and individuals, respectively. For a small number of SNPs (where $M_e/N < h_g^2$) accuracy will primarily be driven by the phenotype h_g^2. For example, assuming 20 effective SNPs in a 1MB locus[7] and chromatin h_g^2 of 0.30, the prediction r^2 from 50 samples would be 0.13, equivalent to the prediction r^2 from 1,000 samples (a 20× increase) with h_g^2 of 0.15 (a 2× decrease). For genes where a causal chromatin phenotype explains less than half of the variance in expression, TWAS from a 50 sample chromatin cohort will therefore be theoretically comparable to TWAS from a 1,000 sample expression cohort (which is consistent with our simulations). However, 50 samples is not sufficient to estimate that the feature is predictive, either using cross-validation or model likelihood. So although the chromatin predictors would be accurate, we would have no way of knowing which of the predictors are accurate/heritable and restricting the analyses to those. In contrast, expression from several hundred individuals allows us to restrict to a large number of heritable genes and be confident that these predictions are not null. Predicting from gene expression also ensures that every association is anchored by a gene, which remains a more interpretable molecular unit than a chromatin peak.

Lastly, although cross-trait LD-score regression (LDSC) can be used to estimate genetic covariance using summary statistics, it does not provide for a formal statistical test of local covariance[15]. Cross-trait LDSC uses the block jackknife to estimate standard errors and perform statistical testing. However, the independence assumptions of the block jackknife are violated within a single cis locus, where SNPs are typically highly correlated and cannot be cleanly divided into independent blocks. We have also previously shown empirically that forcing cross-trait LDSC estimates on local regions yields substantially lower power than the TWAS approach[3]. We therefore use cross-trait LDSC as an aggregate measure of local genetic covariance across many loci and compare to random loci to evaluate significance, but not as a formal statistical test for individual loci.

ANALYSES OF HI-C AT PREDICTED CHROMATIN ELEMENTS

Hi-C data measured in the GM12878 LCL was downloaded after post-processing performed by Grubert et al[12]. Briefly, for each pair of fragments, a covariance score was computed as the fraction of fragments that interact with both pairs relative to the number of fragments interacting with only one. Based on concordance between replicates, a threshold of 0.4 was selected which we also used here to define "contact". For each significant chromatin TWAS gene-peak pair (at 10% FDR) greater than 10kb apart, we counted the number of contacts between ±2kb of the TSS and the corresponding chromatin peak. This was compared to the number of contacts at all non-associated chromatin TWAS gene-peak pairs greater than 10kb apart, with enrichment and significance estimated by Fisher's exact test (Supplementary Figure 24).

Grubert et al. Hi-C quantification:
`http://mitra.stanford.edu/kundaje/portal/chromovar3d/hic.html`

ESTIMATING SUPPORT FOR MEDIATION BY EXPRESSION/CHROMATIN

We evaluated whether the SCZ and chromatin TWAS associated loci were more consistent with a chromatin mediating (SNP → chromatin → expression) or expression mediating (SNP → expression → chromatin) causal model leading to SCZ susceptibility. The observed relationship between expression, chromatin, and

disease can arise from multiple models and we sought to evaluate the evidence supporting each. The default model:

$$X \rightarrow H \rightarrow T \rightarrow Y$$
$$H = X\beta_h + Ie_h : \hat{\beta}_h \sim N(\beta_h, 1/N_h)$$
$$T = H(1-e_t) + Ie_t : \hat{\beta}_t \sim N(\sqrt{1-e_t^2}\beta_h, 1/N_t)$$
$$Y = T(1-e_y) + Ie_y : \hat{\beta}_y \sim N(\sqrt{1-e_t^2}\sqrt{1-e_y^2}\beta_h, 1/N_y)$$
$$E[\hat{\beta}_h\hat{\beta}_y] = E[\sigma_{h,y}] = \sqrt{1-e_t^2}\sqrt{1-e_y^2}\beta_h^2$$
$$E[\hat{\beta}_t\hat{\beta}_y] = E[\sigma_{t,y}] = (1-e_t^2)\sqrt{1-e_y^2}\beta_h^2$$
$$E[\sigma_{h,y}/\sigma_{t,y}] = \sqrt{1-e_t^2}/(1-e_t^2) = 1/\sqrt{1-e_t^2} \geq 1$$

an alternative model where chromatin is correlated with transcript but conditionally independent of trait:

$$X \rightarrow T \rightarrow Y; T \rightarrow H$$
$$T = X\beta_t + Ie_t : \hat{\beta}_t \sim N(\beta_t, 1/N_t)$$
$$Y = T(1-e_y) + Ie_y : \hat{\beta}_y \sim N(\sqrt{1-e_y^2}\beta_t, 1/N_y)$$
$$H = T(1-e_h) + Ie_h : \hat{\beta}_h \sim N(\sqrt{1-e_h^2}\beta_t, 1/N_h)$$
$$E[\sigma_{h,y}] = \sqrt{1-e_h^2}\sqrt{1-e_y^2}\beta_t^2$$
$$E[\sigma_{t,y}] = \sqrt{1-e_y^2}\beta_t^2$$
$$E[\sigma_{h,y}/\sigma_{t,y}] = \sqrt{1-e_h^2} \leq 1$$

In practice we want to aggregate the estimates from many loci to measure the average effect. Because we cannot guarantee that the sign of the covariance will be the same for every locus, we instead seek to estimate $E[\sigma_{h,y}^2]/E[\sigma_{t,y}^2]$ and solve using squared terms in the above equations. By replacing the expectation of a ratio with the ratio of an expectation we incur a small bias on the order of $1/N$. Additionally, both the numerator and the denominator of $\hat{\sigma}_{h,y}^2/\hat{\sigma}_{t,y}^2$ have an upwards bias corresponding to $(1-\sigma^2)/N$ which will bias the ratio towards 1, yielding an estimate of causal differences that is conservative.

To compare these two models without bias from sample size or assay, we estimated the genetic covariance (cov_g) using the PGC SCZ summary statistics and molecular data from the CEU/YRI individuals which had both chromatin and gene expression measured. For each SCZ/chromatin TWAS gene, we defined the target locus as ±50kb of the union of gene and peak boundary and estimated SCZ-chromatin and SCZ-expression cov_g using cross-trait LD score regression[15], restricting to the well-imputed HapMap3 SNPs and using in-sample LD. To estimate significance, background cov_g was computed using the same procedure over 200 randomly gene-peak pairs within 500kb of the TSS. The observed and background estimates were then compared using the non-parametric Kolmogorov-Smirnov test. We note that for the YRI samples LD in the GWAS summary statistics is expected to differ from LD in the eQTL/cQTL data, and confound the raw estimate of cov_g; however, because the random gene-peak pairs are also computed from the same population, we do not expect significance measured against this null to be inflated. We separately considered a partial correlation analysis, where each expression measurement was transformed to the residual of the associated chromatin peak in a standard linear regression (and likewise for each chromatin measurement). The two estimates of cov_g were again computed from these partial phenotypes as described above.

ALLELE-SPECIFIC ANALYSES AT MAPK3 LOCUS

RNA-seq reads at the MAPK3 locus were processed using the WASP pipeline[19], which re-maps reads containing polymorphisms to account for mapping bias. First, CMC genotypes were phased and imputed using EAGLE2[20]. Next, every RNA-seq read spanning a polymorphism was identified and remapped to the reference with the alternative allele substituted; any reads that did not map to the same location were then dropped form the analysis. Finally, any duplicate reads were removed randomly to prevent bias due to allele-specific read quality.

Allele-specific analyses were then carried out using the WASP Combined Haplotype Test (CHT) to evaluate imbalance of RNA-seq reads at each cis-SNP while accounting for over-dispersion with a beta-binomial model. WASP first estimates a beta-binomial overdispersion parameter from all reads (for the allelic test), as well as a beta-negative-binomial overdispersion parameter from the tested peak (for the QTL test). The overdispersion parameters are then incorporated into a single likelihood model to test for allelic imbalance at each SNP in the locus. We report both results using only allelic information at heterozygous sites (which is independent of the eQTL-based TWAS) as well as the combined effect of heterozygous and homozygous variants.

FUNCTIONAL ACTIVITY AT MAPK3 LOCUS

We used brain tissue chromatin data from the psychENCODE project to identify epigenetic peaks near the MAPK3 SCZ/chromatin TWAS association. This included ChIP-seq from H3k27ac and H3k4me3 in 17 control adult brain samples (prefrontal cortex and anterior cingulate cortex), and ATAC-seq from 12 adult SCZ case/control brains (prefrontal cortex). BWA[21] aligned bam files and MaCS[22] narrow peak calls (q-value threshold of 0.01) were downloaded from the psychENCODE data portal and merged within each tissue to identify peaks in the locus and create coverage plots (Supplementary Figure 36). Human-gain neuro-developmental enhancers were downloaded from ref.[23] and overlapped with the locus: two overlapping peaks were observed in H3k4me2 from 8.5 post-conception-week (pcw) and 120pcw fetal tissues.

psychENCODE data portal:
`https://www.synapse.org/#!Synapse:syn4566010`
`https://www.synapse.org/#!Synapse:syn5321694`

Reilly et al. human-gain developmental enhancers:
`http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63648`

PSYCHIATRIC GENOMICS CONSORTIUM

Collaborators from the Psychiatric Genomics Consortium were, in alphabetical order: Devin Absher, Rolf Adolfsson, Ingrid Agartz, Esben Agerbo, Huda Akil, Margot Albus, Madeline Alexander, Farooq Amin, Ole A Andreassen, Adebayo Anjorin, Richard Anney, Dan Arking, Philip Asherson, Maria H Azevedo, Silviu A Bacanu, Lena Backlund, Judith A Badner, Tobias Banaschewski, Jack D Barchas, Michael R Barnes, Thomas B Barrett, Nicholas Bass, Michael Bauer, Monica Bayes, Martin Begemann, Frank Bellivier, Judit Bene, Sarah E Bergen, Thomas Bettecken, Elizabeth Bevilacqua, Joseph Biederman, Tim B Bigdeli, Elisabeth B Binder, Donald W Black, Douglas HR Blackwood, Cinnamon S Bloss, Michael Boehnke, Dorret I Boomsma, Anders D Borglum, Elvira Bramon, Gerome Breen, Rene Breuer, Richard Bruggeman, Nancy G Buccola, Randy L Buckner, Jan K Buitelaar, Brendan Bulik-Sullivan, William E Bunner, Margit Burmeister, Joseph D Buxbaum, William F Byerley, Sian Caesar, Wiepke Cahn, Guiqing Cai, Murray J Cairns, Dominique Campion, Rita M Cantor, Vaughan J Carr, Noa Carrera, Miquel Casas, Stanley V Catts, Aravinda Chakravarti, Kimberley D Chambert, Raymond CK Chan, Eric YH Chen, Ronald YL Chen, Wei Cheng, Eric FC Cheung, Siow Ann Chong, Khalid Choudhury, Sven Cichon, David St Clair, C Robert Cloninger, David Cohen, Nadine Cohen, David A Collier, Edwin Cook, Hilary Coon, Bru Cormand, Paul Cormican, Aiden Corvin, William H Coryell, Nicholas Craddock, David W Craig, Ian W Craig, Benedicto Crespo-Facorro, James J Crowley, David Curtis, Darina Czamara, Mark J Daly, Ariel Darvasi, Susmita Datta, Michael Davidson, Kenneth L Davis, Richard Day, Franziska Degenhardt, Lynn E DeLisi, Ditte Demontis, Bernie Devlin, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Enrico Domenici, Gary Donohoe, Alysa E Doyle, Elodie Drapeau, Jubao Duan, Frank Dudbridge, Naser Durmishi, Howard J Edenberg, Hannelore Ehrenreich, Peter Eichhammer, Amanda Elkin, Johan Eriksson, Valentina Escott-Price, Tonu Esko, Laurent Essioux, Bruno Etain, Ayman H Fanous, Stephen V Faraone, Kai-How Farh, Anne E Farmer, Martilias S Farrell, Jurgen Del Favero, Manuel A Ferreira, I Nicol Ferrier, Matthew Flickinger, Tatiana Foroud, Josef Frank, Barbara Franke, Lude Franke, Christine Fraser, Robert Freedman, Nelson B Freimer, Marion Friedl, Joseph I Friedman, Louise Frisen, Menachem Fromer, Pablo V Gejman, Giulio Genovese, Lyudmila Georgieva, Elliot S Gershon, Eco J De Geus, Ina Giegling, Michael Gill, Paola Giusti-Rodriguez, Stephanie Godard, Jacqueline I Goldstein, Vera Golimbet, Srihari Gopal, Scott D Gordon, Katherine Gordon-Smith, Jacob Gratten, Elaine K Green, Tiffany A Greenwood, Gerard Van Grootheest, Magdalena Gross, Detelina Grozeva, Weihua Guan, Hugh Gurling, Omar Gustafsson, Lieuwe de Haan, Hakon Hakonarson, Steven P Hamilton, Christian Hammer, Marian L Hamshere, Mark Hansen, Thomas F Hansen, Vahram Haroutunian, Annette M Hartmann, Martin Hautzinger, Andrew C Heath, Anjali K Henders, Frans A Henskens, Stefan Herms, Ian B Hickie, Maria Hipolito, Joel N Hirschhorn, Susanne Hoefels, Per Hoffmann, Andrea Hofman, Mads V Hollegaard, Peter A Holmans, Florian Holsboer, Witte J Hoogendijk, Jouke Jan Hottenga, David M Hougaard, Hailiang Huang, Christina M Hultman, Masashi Ikeda, Andres Ingason, Marcus Ising, Nakao Iwata, Assen V Jablensky, Stephane Jamain, Inge Joa, Edward G Jones, Ian Jones, Lisa Jones, Erik G Jonsson, Milan Macek Jr, Richard A Belliveau Jr, Antonio Julia, Tzeng JungYing, Anna K Kahler, Rene S Kahn, Luba Kalaydjieva, Radhika Kandaswamy, Sena KarachanakYankova, Juha Karjalainen, David Kavanagh, Matthew C Keller, Brian J Kelly, John R Kelsoe, Kenneth S Kendler, James L Kennedy, Elaine Kenny, Lindsey Kent, Jimmy Lee Chee Keong, Andrey Khrunin, Yunjung Kim, George K Kirov, Janis Klovins, Jo Knight, James A Knowles, Martin A Kohli, Daniel L Koller, Bettina Konte, Ania Korszun, Robert Krasucki, Vaidutis Kucinskas, Zita Ausrele Kucinskiene, Jonna Kuntsi, Hana Kuzelova-Ptackova, Phoenix Kwan, Mikael Landen, Niklas Langstrom, Mark Lathrop, Claudine Laurent, Jacob Lawrence, William B Lawson, Marion Leboyer, Phil Hyoun Lee, S Hong Lee, Sophie E Legge, Todd Lencz, Bernard Lerer, Klaus-Peter Lesch, Douglas F Levinson, Cathryn M Lewis, Jun Li, Miaoxin Li, Qingqin S Li, Tao Li, KungYee Liang, Paul Lichtenstein, Jeffrey A Lieberman, Svetlana Limborska, Danyu Lin, Chunyu Liu, Jianjun Liu, Falk W Lohoff, Jouko Lonnqvist, Sandra K Loo, Carmel M Loughland, Jan Lubinski, Susanne Lucae, Donald MacIntyre, Pamela AF Madden, Patrik KE Magnusson, Brion S Maher, Pamela B Mahon, Wolfgang Maier, Anil K Malhotra, Jacques Mallet, Sara Marsal, Nicholas G Martin, Manuel Mattheisen, Keith Matthews, Morten Mattingsdal, Robert W McCarley, Steven A McCarroll, Colm McDonald, Kevin A McGhee, James J McGough, Patrick J McGrath, Peter McGuffin, Melvin G McInnis, Andrew M McIntosh, Rebecca McKinney, Alan W McLean, Francis J McMahon, Andrew McQuillin, Helena Medeiros, Sarah E Medland, Sandra Meier, Carin J Meijer, Bela Melegh, Ingrid Melle, Fan Meng, Raquelle I Mesholam-Gately, Andres Metspalu, Patricia T Michie, Christel M Middeldorp, Lefkos Middleton, Lili Milani, Vihra Milanova, Philip B Mitchell,

Younes Mokrab, Grant W Montgomery, Jennifer L Moran, Gunnar Morken, Derek W Morris, Ole Mors, Preben B Mortensen, Valentina Moskvina, Bryan J Mowry, Pierandrea Muglia, Thomas W Muehleisen, Walter J Muir, Bertram Mueller-Myhsok, Kieran C Murphy, Robin M Murray, Richard M Myers, Inez Myin-Germeys, Benjamin M Neale, Michael C Neale, Mari Nelis, Stan F Nelson, Igor Nenadic, Deborah A Nertney, Gerald Nestadt, Kristin K Nicodemus, Caroline M Nievergelt, Liene Nikitina-Zake, Ivan Nikolov, Vishwajit Nimgaonkar, Laura Nisenbaum, Willem A Nolen, Annelie Nordin, Markus M Noethen, John I Nurnberger, Evaristus A Nwulia, Dale R Nyholt, Eadbhard OCallaghan, Michael C ODonovan, Colm ODushlaine, F Anthony ONeill, Robert D Oades, Sang-Yun Oh, Ann Olincy, Line Olsen, Edwin JCG van den Oord, Roel A Ophoff, Jim Van Os, Urban Osby, Hogni Oskarsson, Michael J Owen, Aarno Palotie, Christos Pantelis, George N Papadimitriou, Sergi Papiol, Elena Parkhomenko, Carlos N Pato, Michele T Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Brenda P Penninx, Michele L Pergadia, Diana O Perkins, Roy H Perlis, Tune H Pers, Tracey L Petryshen, Hannes Petursson, Benjamin S Pickard, Olli Pietilainen, Jonathan Pimm, Joseph Piven, Andrew J Pocklington, Porgeir Porgeirsson, Danielle Posthuma, James B Potash, John Powell, Alkes Price, Peter Propping, Ann E Pulver, Shaun M Purcell, Vinay Puri, Digby Quested, Emma M Quinn, Josep Antoni Ramos-Quiroga, Henrik B Rasmussen, Soumya Raychaudhuri, Karola Rehnstrom, Abraham Reichenberg, Andreas Reif, Mark A Reimers, Marta Ribases, John Rice, Alexander L Richards, Marcella Rietschel, Brien P Riley, Stephan Ripke, Joshua L Roffman, Lizzy Rossin, Aribert Rothenberger, Guy Rouleau, Panos Roussos, Douglas M Ruderfer, Dan Rujescu, Veikko Salomaa, Alan R Sanders, Susan Santangelo, Russell Schachar, Ulrich Schall, Martin Schalling, Alan F Schatzberg, William A Scheftner, Gerard Schellenberg, Peter R Schofield, Nicholas J Schork, Christian R Schubert, Thomas G Schulze, Johannes Schumacher, Sibylle G Schwab, Markus M Schwarz, Edward M Scolnick, Laura J Scott, Rodney J Scott, Larry J Seidman, Pak C Sham, Jianxin Shi, Paul D Shilling, Stanley I Shyn, Engilbert Sigurdsson, Teimuraz Silagadze, Jeremy M Silverman, Kang Sim, Pamela Sklar, Susan L Slager, Petr Slominsky, Susan L Smalley, Johannes H Smit, Erin N Smith, Jordan W Smoller, Hon-Cheong So, Erik Soderman, Edmund Sonuga-Barke, Chris C A Spencer, Eli A Stahl, Matthew State, Hreinn Stefansson, Kari Stefansson, Michael Steffens, Stacy Steinberg, Hans-Christoph Steinhausen, Elisabeth Stogmann, Richard E Straub, John Strauss, Eric Strengman, Jana Strohmaier, T Scott Stroup, Mythily Subramaniam, Patrick F Sullivan, James Sutcliffe, Jaana Suvisaari, Dragan M Svrakic, Jin P Szatkiewicz, Peter Szatmari, Szabocls Szelinger, Anita Thapar, Srinivasa Thirumalai, Robert C Thompson, Draga Toncheva, Paul A Tooney, Sarah Tosato, Federica Tozzi, Jens Treutlein, Manfred Uhr, Juha Veijola, Veronica Vieland, John B Vincent, Peter M Visscher, John Waddington, Dermot Walsh, James TR Walters, Dai Wang, Qiang Wang, Stanley J Watson, Bradley T Webb, Daniel R Weinberger, Mark Weiser, Myrna M Weissman, Jens R Wendland, Thomas Werge, Thomas F Wienker, Dieter B Wildenauer, Gonneke Willemsen, Nigel M Williams, Stephanie Williams, Richard Williamson, Stephanie H Witt, Aaron R Wolen, Emily HM Wong, Brandon K Wormley, Naomi R Wray, Adam Wright, Jing Qin Wu, Hualin Simon Xi, Wei Xu, Allan H Young, Clement C Zai, Stan Zammit, Peter P Zandi, Peng Zhang, Xuebin Zheng, Fritz Zimprich, Frans G Zitman, and Sebastian Zoellner.

SUPPLEMENTARY TABLES

Supplementary Table 1: **Expression SNP-heritability in CMC/NTR panels.** Average cis and trans h_g^2 across all genes is reported, followed by the number of genes with significant cis-h_g^2 at $P < 0.05$ after Bonferroni (family-wise error rate, FWER) correction as well as $P < 0.01$. Significance was established by likelihood ratio test after dropping the cis component. Standard error on h_g^2 estimates reported in parenthesis as standard deviation over square-root of number of genes.

Expression	# genes	cis-h_g^2	trans-h_g^2	FWER < 0.05	$P < 0.01$
CMC genes	14,201	0.067 (0.001)	0.077 (0.006)	2,702	5,472
CMC genes (cases)	13,894	0.067 (0.001)	0.208 (0.011)	1,344	NA
CMC genes (controls)	13,959	0.077 (0.001)	0.304 (0.014)	1,229	NA
CMC introns†	86,817	0.017 (0.001)	0.046 (0.003)	(1,453) 2,847	(3,908) 9,009
NTR genes	16,044	0.012 (0.001)	0.056 (0.002)	1,117	2,692
NTR genes (HRC common)	17,068	0.011 (0.001)	0.034 (0.002)	1,122	NA
NTR genes (HRC cis common*)	16,419	0.011 (0.001)	NA	NA	NA
NTR genes (HRC cis rare*)	16,419	0.001 (0.000)	NA	NA	NA
NTR transcripts (all probes)	NA	0.010	0.050	1,368	NA
METSIM genes[3]	12,598	0.055 (0.001)	0.015 (0.001)	1,985	4,583
YFS genes[3]	10,619	0.074 (0.001)	0.06 (0.001)	3,836	5,337

*: Both components estimated jointly.
†: Total gene expression included as fixed effect. Number of unique genes shown in parenthesis at left.

Supplementary Table 2: **Associated genes with experiment-wide Bonferroni and study-wide FDR-based multiple test correction.** Top panel shows number of genes significantly associated after Bonferroni correction for total number of tests within a given molecular experiment (e.g. H3k27ac peaks measured in CEU). These results are identical to the primary primary results in Table 1. Bottom panel shows number of genes significant associated after 5% FDR control across all 9 molecular experiments.

per-phenotype Bonferroni 0.05	CMC-introns*	CMC	NTR	YFS	METSIM	Unique
Chromatin associated	(224) 125	244	182	346	232	806
SCZ and chromatin associated	(10) 8	11	10	13	7	42
all-phenotype FDR 0.05	CMC-introns*	CMC	NTR	YFS	METSIM	Unique
Chromatin associated	(779) 429	854	554	1070	1015	2837
SCZ and chromatin associated	(12) 10	14	12	17	7	52

* Number of splicing events shown in parenthesis.

Supplementary Table 3: **Transcriptome-wide significant SCZ TWAS associations.** Provided in Supplementary Excel Spreadsheet

Supplementary Table 4: **TWAS gene associations in CMC data stratified by case/control status.** CMC expression was tested for TWAS association to schizophrenia using expression in cases and controls separately, followed by sample-size weighted meta-analysis. This should obviate biases due to having cases and controls in the expression reference. Z-scores from the full reference (Z_{all}) and the stratified reference (Z_{strat}) are compared here for genome-wide significant genes identified in the former. Out of 5,511 converged genes, one gene - SDCCAG8 - was genome-wide significant and significantly more associated in the meta-analysis than in the full analysis (Z of -5.7 vs -1.4, respectively). This gene may harbor complimentary effects on expression in cases and controls (GxE) that confound the full-sample analysis.

Gene	Z_{all}	Z_{strat}
ALMS1P	-4.8	-5.1
C2orf47	7.6	7.8
CEP170	-6.1	-5.9
CORO7	5.3	4.9
CPNE7	5.9	6.4
DOC2A	-5.2	-6.0
DRG2	-5.4	-5.9
ELOVL7	-4.9	-4.7
EMB	-5.0	-5.5
FAM109B	-5.3	-6.1
FAM53C	4.5	4.7
FANCL	5.1	4.6
FURIN	5.2	4.8
GATAD2A	4.9	5.3
GLT8D1	5.5	6.0
HAPLN4	-4.7	-5.2
LOC100507140	4.8	4.3
LOC388152	-5.0	-5.6
MAPK3	-4.8	-4.9
MAU2	4.7	5.3
MCHR1	4.8	5.3
MED8	4.7	4.5
MPHOSPH9	5.6	6.4
NAGA	-5.9	-6.4
NDUFA2	4.6	4.8
NDUFAF2	-4.5	-5.5
NEK4	-6.0	-6.6
NT5C2	6.4	6.7
PCCB	6.3	6.8
PCDHA2	-4.5	-4.1
PRMT7	-4.5	-4.9
REXO1	-4.6	-3.4
SEPT10	-4.6	-4.6
SETD8	5.4	5.4
SF3B1	-5.7	-6.1
SLC45A1	5.5	5.5
SNAP91	-4.8	-5.0
TEK	5.1	4.9
TMEM110	4.5	5.4
TMEM81	-4.8	-4.2
TSNAXIP1	-4.8	-5.0
XPNPEP3	-4.6	-5.0
ZMAT2	-4.7	-5.0
ZSCAN2	-5.1	-5.3

Supplementary Table 5: **Correlation between genotype and summary based TWAS in sub-cohorts.** For each PGC sub-cohort, TWAS Z-scores (using CMC as reference) were estimated for 1,000 random genes using (a) summary statistics and (b) genotype-based prediction (BLUP) and Pearson correlation between the two computed. Each line reports the correlation when estimated either using array SNPs or imputed HapMap3 SNPs. Bottom section reports Pearson correlation between TWAS Z-scores from summary-based prediction and individual-based prediction across four expression reference panels. Note that summary data includes additional restricted samples that were not available for individual-based prediction.

Cohort	Array	Imputed HapMap3
CMC into sub-cohorts		
aber	0.89	0.92
ajsz	0.98	0.98
asrb	0.91	0.92
buls	0.94	0.96
cati	0.67	0.68
caws	0.80	0.82
clm2	0.97	0.97
clo3	0.97	0.98
cou3	0.84	0.86
denm	0.95	0.96
dubl	0.94	0.96
edin	0.94	0.96
egcu	0.92	0.92
ersw	0.90	0.90
gras	0.89	0.89
irwt	0.94	0.96
lacw	0.78	0.79
lie2	0.94	0.95
lie5	0.96	0.97
mgs2	0.91	0.92
munc	0.85	0.86
pewb	0.82	0.83
pews	0.87	0.90
port	0.85	0.87
s234	0.88	0.91
swe1	0.85	0.90
swe5	0.90	0.91
swe6	0.85	0.87
top8	0.92	0.95
ucla	0.95	0.96
uclo	0.87	0.90
umeb	0.90	0.90
umes	0.96	0.97
zhh1	0.82	0.84
into all samples		
CMC	NA	0.90
METSIM	NA	0.89
NTR	NA	0.90
YFS	NA	0.85

Supplementary Table 6: **GE-PRS prediction accuracy in CMC/brain.** Prediction R^2 trained in PGC and predicted into the CMC SCZ/BIP cohort shown for three predictors: (a) gene risk score (GE-PRS) from TWAS with CMC gene expression and splicing, with associations in the HLA region excluded; (b) GWAS genetic risk score (GRS) compiled by the PGC; (c) same GRS with associations in the HLA region excluded. 3 rightmost columns report P-values from a single/joint linear regression with the specified predictors (with results from the single predictor for each row on the diagonal). Prediction R^2 was computed after subtracting out the R^2 from two principal components, and converting to the liability scale at 1% prevalence (which is a linear transformation). Clumping of variants with maximum LD $r^2 = 0.1$ was used for all three predictors.

Predictor	liability R^2	P-value with (a)	P-value with (b)	P-value with (c)
(a) CMC/brain TWAS GE-PRS (no HLA)	0.016	2.4×10^{-4}	1.4×10^{-2}	1.3×10^{-2}
(b) PGC GWAS GRS	0.014	3.4×10^{-2}	5.7×10^{-4}	NA
(c) PGC GWAS GRS (no HLA)	0.013	6.3×10^{-1}	NA	1.2×10^{-3}

Supplementary Table 7: **Conditional analysis of TWAS loci.** For each significant TWAS gene, the maximum GWAS χ^2 statistic across all SNPs, and the lead GWAS SNP χ^2 statistic was measured in conditional models with the average across all loci reported. The models evaluated were: (row 1) the standard marginal GWAS statistics; (row 2) conditioning the locus on the most significant eQTL in the reference expression; (row 3) conditioning the locus on the full TWAS prediction from the reference expression; (row 4) conditioning the locus on the lead GWAS SNP (which is an overfit due to selection performed in the same sample). All conditional analyses were performed using the COJO method from ref.[24], with reference LD from all 1000 Genomes European samples. LD to the TWAS predicted expression was likewise computed by predicting into the 1000G EUR samples.

Feature	avg max χ^2	lead SNP χ^2
TWAS-associated locus	42.1	42.1
Locus conditioned on best eQTL	22.3	13.0
Locus conditioned on TWAS predictor	20.0	10.3
Locus conditioned on best GWAS SNP	16.6	NA

Supplementary Table 8: **Conditional analysis of GWAS SNPs in LD with TWAS genes.** The 108 PGC-SCZ GWAS significant SNPs were conditioned on all transcriptome-wide significant TWAS associations. Results are reported for the 43 SNPs that were in LD ($r^2 > 0.05$) with at least one TWAS gene and considered for conditioning. Mean SNP association was $\chi^2 = 42.8$ and $\chi^2 = 6.1$ before/after conditioning. Median P-value was $P = 1.2 \times 10^{-10}$ and $P = 0.028$ before/after conditioning. All joint/conditional analyses were performed using the COJO method from ref.[24], with reference LD from the CMC samples. LD to the TWAS predicted expression was likewise computed by predicting into the CMC samples.

SNP	chr	position	Z-score	PV	cond. Z-score	cond. PV
rs34269918	1	8424984	-6.0	3.0e-09	-2.04	2.5e-02
rs11210892	1	44100084	6.2	8.7e-10	3.56	3.5e-04
rs1702294	1	98501984	8.4	6.5e-17	4.36	1.5e-05
rs140505938	1	150031490	6.1	1.3e-09	0.11	2.0e-01
rs11682175	2	57987593	7.0	4.6e-12	3.01	2.2e-03
rs6434928	2	198304577	-6.7	3.0e-11	-0.78	1.5e-01
rs199532108	2	200825237	7.7	2.9e-14	1.13	1.1e-01
rs6704768	2	233592501	7.0	6.0e-12	4.09	4.6e-05
rs4330281	3	17859366	-5.5	6.9e-08	-1.96	2.9e-02
rs2535627	3	52845105	6.6	6.0e-11	0.70	1.6e-01
rs7432375	3	136288405	-6.6	9.6e-11	-0.42	1.8e-01
rs33972009	3	180594593	6.6	8.9e-11	3.01	2.2e-03
rs10520163	4	170626552	-5.7	1.4e-08	-2.84	3.5e-03
rs1501357	5	45364875	-5.7	1.7e-08	-2.87	3.2e-03
rs4391122	5	60598543	-7.4	3.1e-13	-2.69	5.3e-03
rs4388249	5	109036066	5.3	1.3e-07	2.45	9.9e-03
rs3849046	5	137851192	-5.9	7.0e-09	-1.50	6.5e-02
rs111896713	5	140143664	-5.1	4.6e-07	-0.22	1.9e-01
rs202141053	6	84280274	6.1	1.5e-09	1.48	6.6e-02
rs10650434	7	2025096	-7.5	9.5e-14	-4.08	4.8e-05
rs149009306	7	24747494	-5.1	4.7e-07	-2.93	2.7e-03
rs13240464	7	110898915	7.2	1.1e-12	1.96	2.9e-02
rs11191419	10	104612335	8.6	1.8e-17	0.42	1.8e-01
rs61126341	11	46350213	-6.7	3.3e-11	-2.21	1.7e-02
rs2514218	11	113392994	6.2	7.7e-10	4.15	3.6e-05
rs10791097	11	130718630	7.0	5.6e-12	1.76	4.2e-02
rs4766428	12	110723245	6.2	1.1e-09	2.01	2.7e-02
rs2851447	12	123665113	7.6	4.0e-14	1.76	4.2e-02
rs12887734	14	104046834	7.4	1.8e-13	0.48	1.8e-01
rs8042374	15	78908032	-7.1	3.1e-12	-0.85	1.4e-01
rs950169	15	84706461	6.5	1.4e-10	1.32	8.3e-02
rs4702	15	91426560	7.0	4.6e-12	3.09	1.7e-03
rs12691307	16	29939877	-6.4	2.3e-10	-1.35	8.0e-02
rs8044995	16	68189340	-5.5	5.1e-08	-0.89	1.3e-01
rs4523957	17	2208899	6.1	1.7e-09	4.08	4.8e-05
rs8082590	17	17958402	-5.8	1.0e-08	-0.73	1.5e-01
rs2905426	19	19478022	-5.8	1.1e-08	-1.14	1.0e-01
rs56873913	19	50091199	5.2	2.8e-07	1.82	3.8e-02
rs6065094	20	37453194	6.6	8.7e-11	2.66	5.8e-03
rs199694726	22	39987017	6.7	3.2e-11	4.47	9.2e-06
rs9607782	22	41587556	-6.9	1.3e-11	-4.12	4.0e-05
rs6002655	22	42603814	-6.0	2.3e-09	-2.20	1.8e-02

Supplementary Table 9: **Heritability of chromatin phenotypes in CEU.** Estimates of chromatin phenotype cis-h_g^2 averaged across all corresponding peaks in the CEU . Estimation performed by Haseman-Elston regression and standard error reported as the standard deviation over the square-root of the number of observations.

Assay	# of peaks	cis-h_g^2 (se)
RNA-seq	19,666	0.048 (0.002)
H3K27AC	44,019	0.043 (0.001)
H3K4ME1	77,070	0.030 (0.001)
H3K4ME3	18,139	0.029 (0.002)
PU1	50,215	0.026 (0.001)
RPB2	46,985	0.027 (0.001)

Supplementary Table 10: **Chromatin TWAS associations vs eQTL/cQTL overlap associations.** Unique number of genes significantly associated with a chromatin phenotype peak in YRI/CEU, using TWAS or eQTL/cQTL overlap, reported as a function of minimum distance around the gene. Associations were significant after experiment-wide Bonferroni correction for tests across all chromatin phenotypes in the population. Corresponds to results represented graphically in Supplementary Figure 18.

CEU eQTL/cQTL	100bp	1kb	2kb	10kb	50kb	100kb	250kb	500kb
H3K4ME1	1	2	2	5	7	8	9	8
H3K4ME3	0	3	3	8	10	12	12	12
H3K27AC	0	3	6	10	11	11	12	11
PU1	0	0	0	1	3	4	6	6
RPB2	5	8	8	11	13	14	15	15
CEU TWAS	100bp	1kb	2kb	10kb	50kb	100kb	250kb	500kb
H3K4ME1	20	23	23	30	43	44	40	35
H3K4ME3	29	30	30	29	29	36	39	35
H3K27AC	51	55	58	63	72	77	69	63
PU1	37	38	39	44	64	62	66	61
RPB2	54	62	60	71	74	81	84	75
YRI eQTL/cQTL	100bp	1kb	2kb	10kb	50kb	100kb	250kb	500kb
H3K4ME1	1	3	11	21	27	26	27	28
H3K4ME3	4	19	25	26	28	31	33	33
H3K27AC	3	22	25	29	31	31	34	34
DHS	3	7	9	10	8	9	8	8
YRI TWAS	100bp	1kb	2kb	10kb	50kb	100kb	250kb	500kb
H3K4ME1	138	142	155	202	277	311	297	284
H3K4ME3	89	121	151	183	239	260	260	250
H3K27AC	102	129	142	183	250	282	272	263
DHS	27	33	33	36	41	46	41	39

Supplementary Table 11: **Number of chromatin TWAS associations at 5% FWER.** Left section shows analysis of chromatin peaks (and corresponding multiple-test correction) within 2kb of the TSS; right panel shows analysis of peaks within 500kb of the TSS.

A: CEU	±2kb					±500kb				
Chromatin pheno	YFS	MET	NTR	CMC	Union	YFS	MET	NTR	CMC	All
PU1	27	25	15	9	51	35	28	25	18	77
RPB2	41	50	32	26	91	41	49	34	32	92
H3K4ME1	14	15	5	12	30	19	19	14	14	49
H3K4ME3	10	25	13	9	41	14	25	13	9	42
H3K27AC	39	33	19	23	82	36	37	19	22	84
Union	105	103	67	60	216	112	99	67	71	235
B: YRI	±2kb					±500kb				
Chromatin pheno	YFS	MET	NTR	CMC	Union	YFS	MET	NTR	CMC	All
DHS	20	17	9	11	43	24	16	8	12	51
H3K4ME1	90	81	42	55	197	150	145	74	96	363
H3K4ME3	96	69	44	60	185	127	126	67	97	313
H3K27AC	89	69	37	52	173	147	127	56	99	333
Union	176	148	87	110	360	272	257	131	193	652

Supplementary Table 12: **Number of chromatin TWAS associations at 10% FDR.** Left section shows analysis of chromatin peaks (and corresponding multiple-test correction) within 2kb of the TSS; right panel shows analysis of chromatin peaks within 500kb of the TSS.

A: CEU	±2kb					±500kb				
Chromatin pheno	YFS	MET	NTR	CMC	Union	YFS	MET	NTR	CMC	Union
PU1	73	85	46	40	177	122	118	93	71	304
RPB2	224	312	136	155	569	270	339	143	175	667
H3K4ME1	90	90	66	59	236	100	121	63	74	274
H3K4ME3	99	132	67	56	249	65	118	68	52	229
H3K27AC	259	233	136	139	549	225	231	152	164	574
Union	520	597	310	347	1252	574	653	354	404	1454
B: YRI	±2kb					±500kb				
Chromatin pheno	YFS	MET	NTR	CMC	Union	YFS	MET	NTR	CMC	Union
dhs	91	54	18	41	157	87	54	33	44	174
H3K4ME1	546	475	227	352	1197	974	937	449	715	2425
H3K4ME3	464	374	186	325	983	699	662	295	571	1717
H3K27AC	535	433	205	340	1112	818	770	388	571	1985
Union	1003	828	377	685	2108	1561	1428	698	1216	3682

Supplementary Table 13: **Chromatin TWAS using CMC/brain splicing events.** For each chromatin phenotype and distance to TSS ($\pm 2kb$ or $\pm 500kb$), columns report (1) the number of genes associated with chromatin identified using gene-chromatin TWAS; (2) the number of splicing events associated with chromatin peaks identified using splicing-chromatin TWAS; (3) the number of unique genes corresponding to those splicing events; (4) the number of unique genes that was not identified using the gene-chromatin TWAS. CMC splice events (at 5% FDR) were significantly less likely to have a chromatin association in YRI than in CEU compared to CMC expression phenotypes (OR=6.5; $P < 1 \times 10^{-20}$), explained either by increased population specificity for splicing or systematic differences in LD at sQTLs relative to eQTLs.

	±2kb				±500kb			
Chromatin phenotype	Genes	Splice variants	Spliced genes	New spliced genes	Genes	Splice variants	Spliced genes	New spliced genes
H3K4ME3	68	18	10	10	104	60	35	31
H3K4ME1	62	24	13	12	105	71	36	27
H3K27AC	65	29	14	12	112	78	42	31
dhs	11	3	3	3	12	2	2	2
PU1	9	17	8	7	18	43	25	23
RPB2	26	27	16	11	32	50	28	19

Supplementary Table 14: **Genes with multiple chromatin peaks haves stronger top eQTL.** Genes in each reference panel were stratified by the number of 5% FDR significant chromatin TWAS associations they had, and the ratio of variance explained by top eQTL (h^2_{eQTL}) divided by the total cis-h^2_g was computed. Table reports the # of genes for each peak count, the mean cis-h^2_g, the mean fraction of cis-h^2_g explained by the top eQTL, and the standard error of the mean in paranethsis.

# of TWAS associated peaks	# of genes	cis-h^2_g (se)	h^2_{eQTL}/cis-h^2_g (se)
0	11910	0.12 (0.001)	0.57 (0.003)
1	1927	0.12 (0.003)	0.68 (0.009)
2	889	0.12 (0.004)	0.78 (0.014)
3	448	0.12 (0.005)	0.84 (0.021)
> 3	589	0.14 (0.006)	0.89 (0.019)

Supplementary Table 15: **Accuracy of predicted expression in CEU/YRI samples.** Mean adjusted R^2 of predicted expression with measured RNA-seq in the in the CEU/YRI samples, reported for NTR/CMC reference panels. For each training/testing combination, the humber of significant genes observed at 5% FWER (Bonferroni correction); 10% FDR; and the mean cis-h^2_g in reference panel is also reported. The NTR : CEU analysis is from a single tissue (blood) and population, yielding prediction R^2 comparable to the h^2_g; the CMC : CEU analysis is a cross-tissue prediction (from brain into blood) that may explain relatively lower R^2/h^2_g; and the two YRI analyses represent cross-ethnic predictions which are also expected to have lower accuracy due to differences in LD.

	NTR : CEU	CMC : CEU	NTR : YRI	CMC : YRI
Mean R^2 (se = 0.001)	0.036	0.032	0.017	0.018
# Genes at 5% FWER	43	54	23	48
# Genes at 10% FDR	253	532	129	288
Mean cis-h^2_g in reference	0.056	0.148	0.056	0.148

Supplementary Table 16: **Variance in gene expression explained by chromatin peak for chromatin TWAS associations.** For each significant chromatin TWAS association detected in CEU using the NTR blood reference, the variance explained in corresponding measured RNA-seq expression by the chromatin phenotype is reported, as well as the cis-h^2_g of expression in the training cohort.

	±2kb		±500kb	
Category	R^2 (se)	h^2_g (se)	R^2 (se)	h^2_g (se)
PU1	7.4% (3.5%)	5.6% (0.8%)	2.1% (1.4%)	6.8% (1.4%)
RPB2	27.2% (4.0%)	11.3% (2.1%)	28.4% (3.2%)	11.3% (2.4%)
H3K4ME1	0.4% (1.9%)	5.8% (0.3%)	16.6% (5.2%)	24.8% (7.1%)
H3K4ME3	35.7% (7.9%)	19.8% (5.5%)	23.7% (7.4%)	21.1% (5.8%)
H3K27AC	30.6% (6.6%)	12.1% (3.7%)	27.2% (5.7%)	18.0% (4.6%)

Supplementary Table 17: **Cross-cohort meta-analysis of chromatin TWAS associations.** Number of unique genes with significant ±500kb chromatin TWAS associations shown for each population and meta-analyzed across populations. Meta column reports significant associations (after Bonferroni correction) from meta-analysis of chromatin peaks that physically overlapped in CEU and YRI. New column reports genes that were identified in the meta-analysis but not in the population-specific analyses. Union of all other sets reported in last column.

Chromatin phenotype	CEU	YRI	Meta	New	Union
H3K4ME1	25	158	186	79	254
H3K4ME3	15	154	141	70	237
H3K27AC	37	147	217	107	279

Supplementary Table 18: **Splicing events with SCZ TWAS and chromatin TWAS associations.** For each significant CMC/brain splicing events with a SCZ TWAS association, the number of significant chromatin TWAS associations are reported (FWER 5% among chromatin TWAS associations, by Bonforroni correction).

Gene	Intron	DHS	H3K27AC	H3K4ME1	H3K4ME3	PU1	RPB2
CCDC90B	chr11:82985783:82991184	0	1	0	0	0	0
CCDC90B	chr11:82989872:82991184	0	1	1	0	0	0
SBNO1	chr12:123821038:123825535	0	0	0	0	1	0
KLC1	chr14:104145855:104151323	0	0	1	1	0	0
RTN1	chr14:60074210:60097164	0	0	1	0	1	0
RTN1	chr14:60074210:60193637	0	0	1	0	0	0
TAOK2	chr16:29997825:29998165	0	0	0	0	0	1
PPP4C	chr16:30094168:30094715	0	0	0	0	0	1
TBC1D5	chr3:17255862:17279655	0	0	1	0	0	0
NEK4	chr3:52800010:52800194	0	0	1	0	0	0

Supplementary Table 19: **Significant SCZ and chromatin TWAS associations.** Provided in Supplementary Excel Spreadsheet

Supplementary Table 20: **Enrichment of chromatin association for SCZ TWAS associated genes.** Enrichment of chromatin associations at SCZ TWAS genes was tested against randomly selected null genes matched on heritability features. Each row reports ratio of real data chromatin associations to permuted chromatin associations, as well as corresponding p-value in 20,000 permutations. Null genes were selected as (1st row) random heritable; (2nd row) matched on gene cis-heritability (an estimate of maximum linear prediction accuracy); (3rd row) matched on REML likelihood from heritability estimate (an estimate of model fit); (4th row) fraction of cis-heritability explained by top eQTL (an estimate of sparsity). Matching was done by (1) computing the feature distribution for the SCZ TWAS set; (2) dividing into quintiles; (3) downsampling the null genes to have the same frequencies in each SCZ TWAS quintile.

	CMC splice	CMC	NTR	YFS	MET	Total
No matching	1.59 (0.0461)	1.82 (0.0124)	1.66 (0.0265)	1.68 (0.0160)	0.98 (0.4223)	1.53 (0.0004)
Matched on cis-h^2_g	1.40 (0.0988)	1.78 (0.0142)	1.59 (0.0350)	1.62 (0.0211)	1.01 (0.3882)	1.47 (0.0015)
Matched on cis-h^2_g likelihood	1.29 (0.1447)	1.77 (0.0156)	1.53 (0.0488)	1.53 (0.0336)	0.87 (0.5704)	1.38 (0.0053)
Matched on % cis-h^2_g from top eQTL	1.30 (0.1404)	1.71 (0.0194)	1.49 (0.0537)	1.40 (0.0640)	0.84 (0.6114)	1.33 (0.0114)

Supplementary Table 21: **SMR replication rate at chromatin TWAS loci.** Each column shows the number of SCZ/chromatin TWAS implicated loci that also had significant QTL-based associations using the SMR test and CEU/YRI eQTL and cQTL data. Column 1 reports the number of SCZ/chromatin TWAS genes that had significant eQTL SMR associations for SCZ, for the 26/35 genes (excluding introns) that had post-QC RNA-seq expression in the CEU/YRI samples. Column 2 reports the number of SCZ/chromatin TWAS genes that had significant cQTL SMR associations for SCZ in at least one peak. Column 3 reports the number of SCZ/chromatin TWAS peaks that had significant cQTL SMR associations for SCZ. Column 4 reports the number of peaks with a that had a significant chromatin-SCZ association using a TWAS-like approach on the chromatin phenotype (see Supplementary Materials). Note: there are 72 unique peaks as 4 peaks were significantly associated in multiple tissues.

	# Genes (SMR eQTLs)	# Genes (SMR cQTLs)	# Peaks (SMR cQTLs)	# Peaks (CWAS)
Total unique	35	42	72	72
Tested	26	42	72	72
$P < 0.05$	11	39	66	67
Bonferroni $P < 0.05$	3	24	31	47

Supplementary Table 22: **Specific chromatin TWAS loci with SMR replication (in chromatin data).** For each gene with a schizophrenia TWAS and chromatin TWAS association, the best eQTL/cQTL is listed as well as the causal effect on schizophrenia estimated by mendelian randomization.

Gene	Phenotype	SNP	β_{SMR}	s.e	P-value
GRAP	RNA-seq	rs2453610	0.02	0.02	1.3e-01
	RPB2	rs1624825	0.04	0.01	* 1.4e-03
KIAA0391	RNA-seq	rs147548181	-0.00	0.01	8.9e-01
	H3K27AC	rs7159324	0.06	0.01	* 1.7e-04
	H3K27AC	rs78426265	0.05	0.01	* 6.2e-04
	H3K4ME1	rs7159324	0.05	0.01	* 1.8e-05
MAP7D1	RNA-seq	rs6667450	-0.00	0.02	8.0e-01
	RPB2	rs10752586	-0.04	0.01	* 4.6e-04
MAPK3	RNA-seq	rs3935873	0.10	0.03	† 1.6e-03
	H3K27AC	rs3924855	0.08	0.02	* 1.2e-05
	RPB2	rs6565175	0.05	0.01	* 1.3e-05
MPHOSPH9	RNA-seq	rs4930719	-0.05	0.02	5.3e-03
	PU1	rs7299943	-0.07	0.02	* 6.5e-04
PCCB	RNA-seq	rs696081	-0.08	0.02	† 1.5e-04
	H3K4ME1	rs480330	0.05	0.01	* 3.3e-04
PPP2R3C	RNA-seq	rs74970204	-0.03	0.01	5.7e-03
	H3K27AC	rs7159324	0.06	0.01	* 1.7e-04
	H3K27AC	rs78426265	0.05	0.01	* 6.2e-04
	H3K4ME1	rs75741665	0.05	0.02	* 1.5e-03
	H3K4ME1	rs7159324	0.05	0.01	* 1.8e-05
PRMT7	RNA-seq	rs28570171	0.02	0.01	1.0e-01
	H3K4ME1	rs57434514	0.05	0.02	4.8e-03
RERE	RNA-seq	rs139118270	-0.06	0.03	3.7e-02
	RPB2	rs301816	0.06	0.01	* 9.2e-05

Supplementary Table 23: **Specific chromatin TWAS loci with SMR replication (in expression reference data).** The SMR test requires eQTL effect-sizes, which were unavailable for the YFS and METSIM data, restricting this analysis to TWAS genes that were identified in NTR and CMC. The test was performed using either 46 randomly sub-sampled individuals (corresponding to the size of the CEU/Waszak et al. expression data) or the full dataset. CBR3 and SLC45A1 were unavailable in the CEU reference and were not evaluated in Supplementary Table 22. (*) indicates significance after Bonferroni correction.

		46 samples				all samples			
Gene/Mark	expression	SNP	β_{SMR}	s.e	P-value	SNP	β_{SMR}	s.e	P-value
CBR3	NTR	rs2835266	-0.01	0.02	5.5e-01	rs2835281	-0.11	0.03	* 1.1e-04
GRAP	NTR	rs8080953	0.02	0.02	3.7e-01	rs7501702	0.20	0.06	* 5.1e-04
MAP7D1	NTR	NA	NA	NA	NA	rs795047	-0.30	0.10	* 1.9e-03
MAPK3	CMC	rs9933310	0.00	0.02	8.8e-01	rs11865086	0.09	0.03	* 4.1e-04
MPHOSPH9	CMC	rs596940	-0.04	0.02	7.8e-02	rs7296418	-0.24	0.07	* 3.2e-04
PCCB	CMC	rs1279831	-0.09	0.03	* 2.4e-03	rs7618871	-0.11	0.02	* 3.4e-08
PPP2R3C	NTR	rs12878912	0.02	0.03	3.3e-01	rs4594164	0.19	0.05	* 4.3e-05
PRMT7	CMC	rs8046501	0.06	0.03	1.9e-02	rs3785113	0.05	0.02	* 8.3e-04
RERE	NTR	rs1294040	0.00	0.01	9.1e-01	rs4908760	0.15	0.05	* 2.9e-03
SLC45A1	CMC	rs10399665	0.05	0.02	7.8e-03	rs10399665	-0.11	0.04	* 2.4e-03

Supplementary Table 24: **Posterior probabilities of colocalization.** Posterior probabilities of a single shared underlying variant between pairs of phenotypes computed from COLOC (with default parameters) for each locus highlighted in the Results. Each section first lists the results from a eQTL-SCZ colocalization for the SCZ TWAS gene, followed by cQTL-eQTL and cQTL-SCZ colocalization for all chromatin TWAS associated peaks. [*] Chromatin peaks identified in YRI, which are a model violation for COLOC due to trans-ethnic LD but included for completeness.

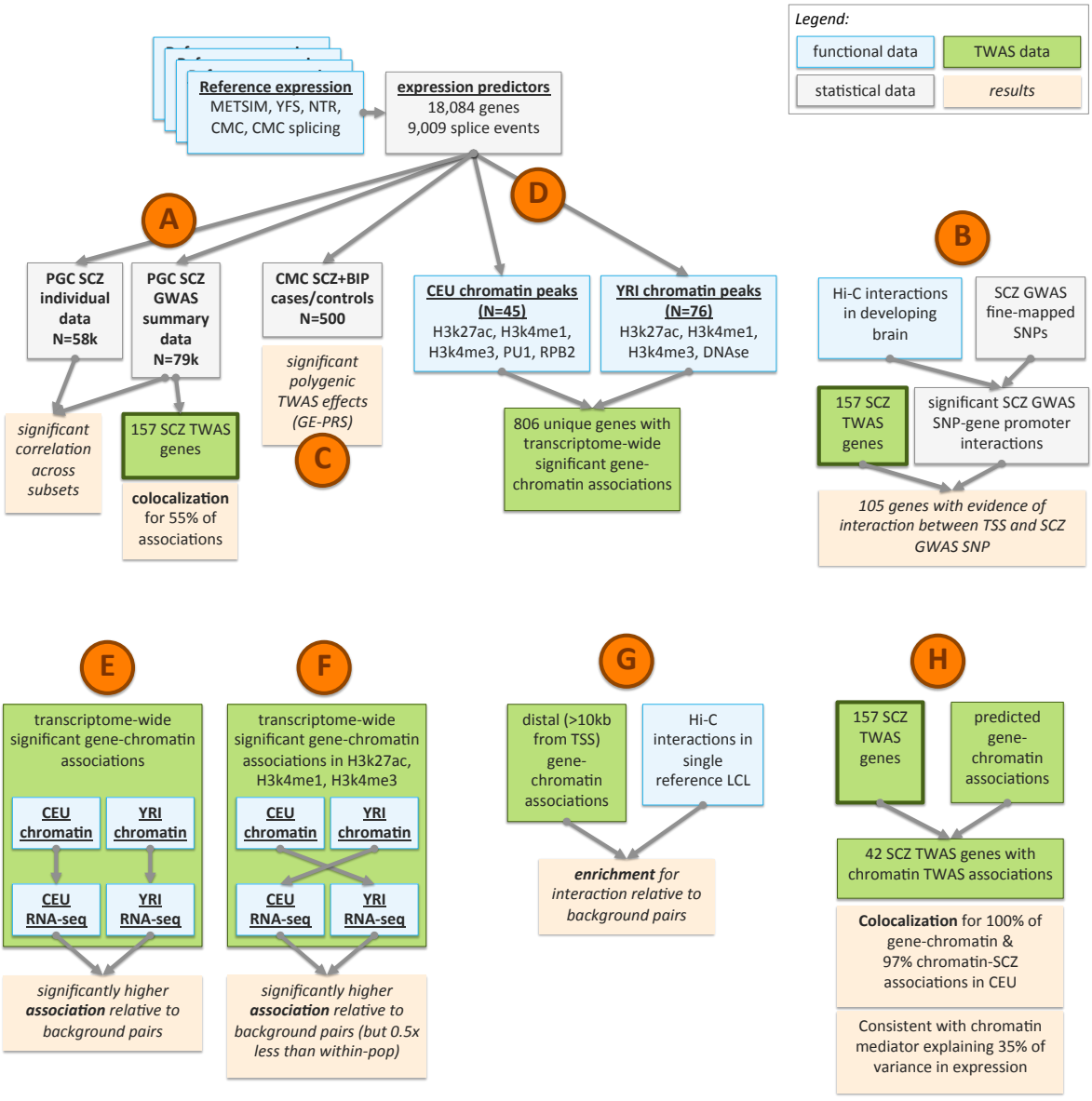
Feature	eQTL	GWAS
PPP2R3C NTR	-	99%
H3k27ac CEU	98%	98%
H3k27ac CEU	99%	99%
H3k4me1 CEU	79%	79%
H3k4me1 CEU	99%	99%
KLC1 CMC splicing	-	58%
H3k4me1 YRI*	100%	100%
H3k4me3 YRI*	84%	84%
MAPK3 CMC	-	61%
H3k27ac CEU	89%	86%
RPB2 CEU	100%	100%

Supplementary Table 25: **Genetic correlation of gene expression.** Average cish^2_g and genetic correlation (ρ_g) between NTR/blood and CMC/brain studies (top row); and between cases and controls within the CMC/brain study. Both values were estimated for the cis locus at each gene using Haseman-Elston regression. The ρ_g standard error was approximated using the Falconer approximation (see Methods) on the reported standard errors of the cis-h^2_g and cov_g estimates, assuming zero correlation of errors. We note that the blood/brain correlation is substantially lower than reported analyses of tissues from a single RNA-seq analysis pipeline [25], likely due to assay-specific effects.

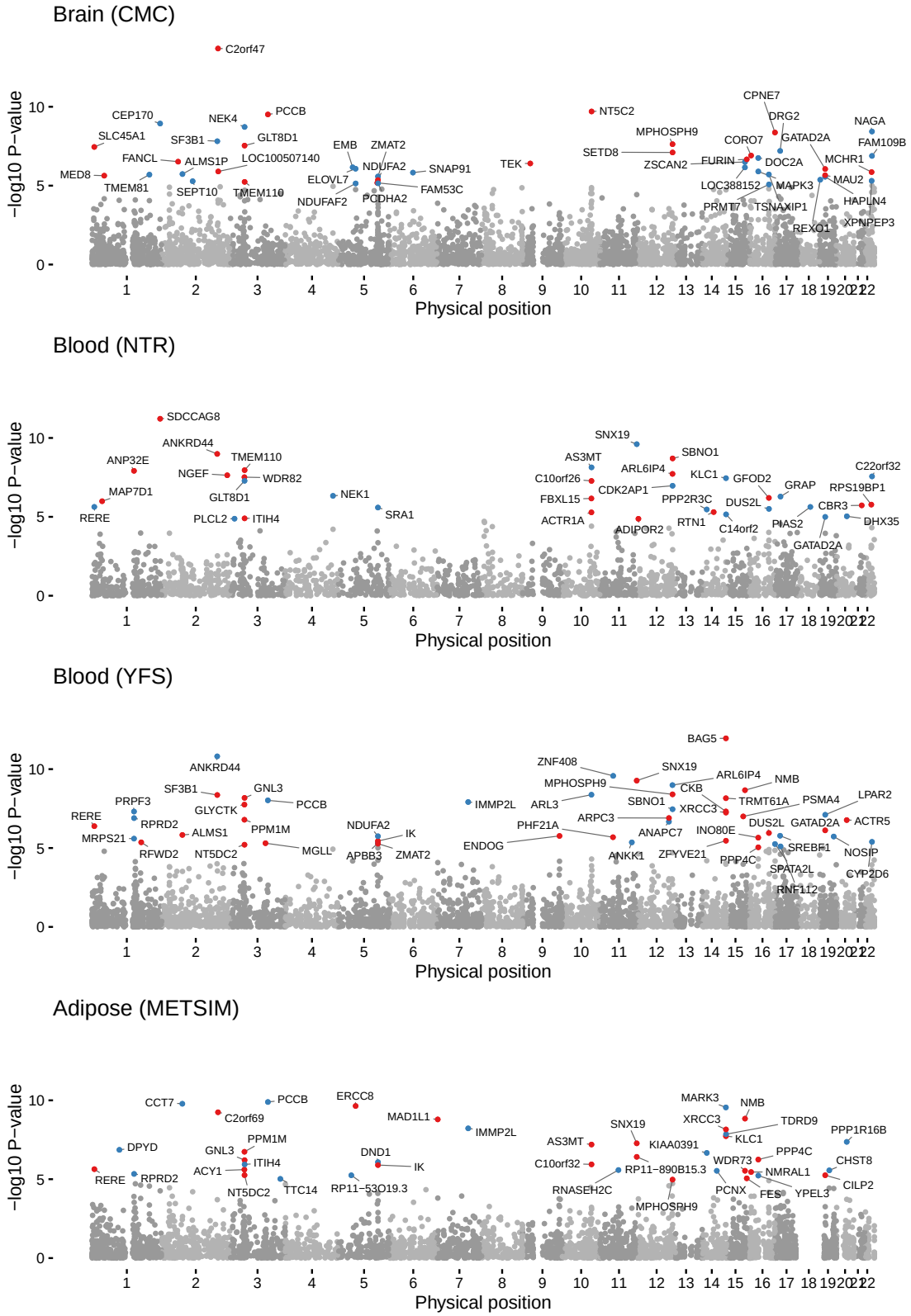
Study 1	Study 2	Study 1 cis-h^2_g	Study 2 cis-h^2_g	cov_g	ρ_g
NTR/blood	CMC/brain	0.0099 (0.0003)	0.0644 (0.0011)	0.0053 (0.0003)	0.21 (0.01)
CMC-case	CMC-control	0.0770 (0.0013)	0.0863 (0.0014)	0.0819 (0.0013)	1.00 (0.02)

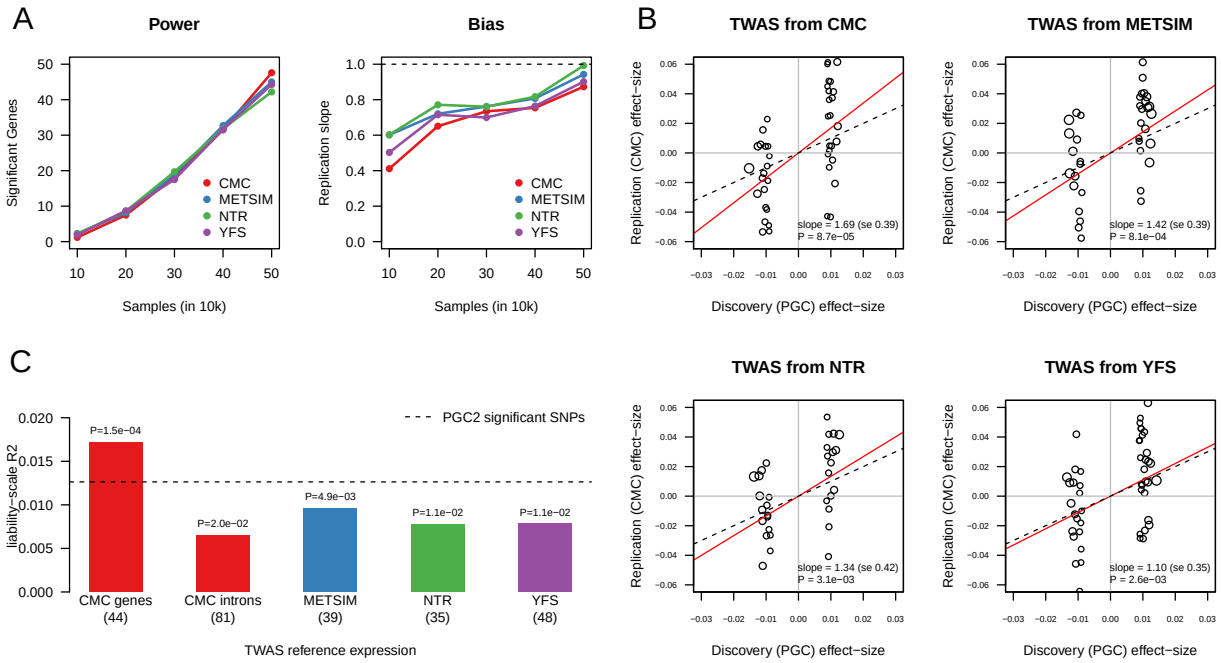
SUPPLEMENTARY FIGURES

Supplementary Figure 1: **Analysis Flowchart.**



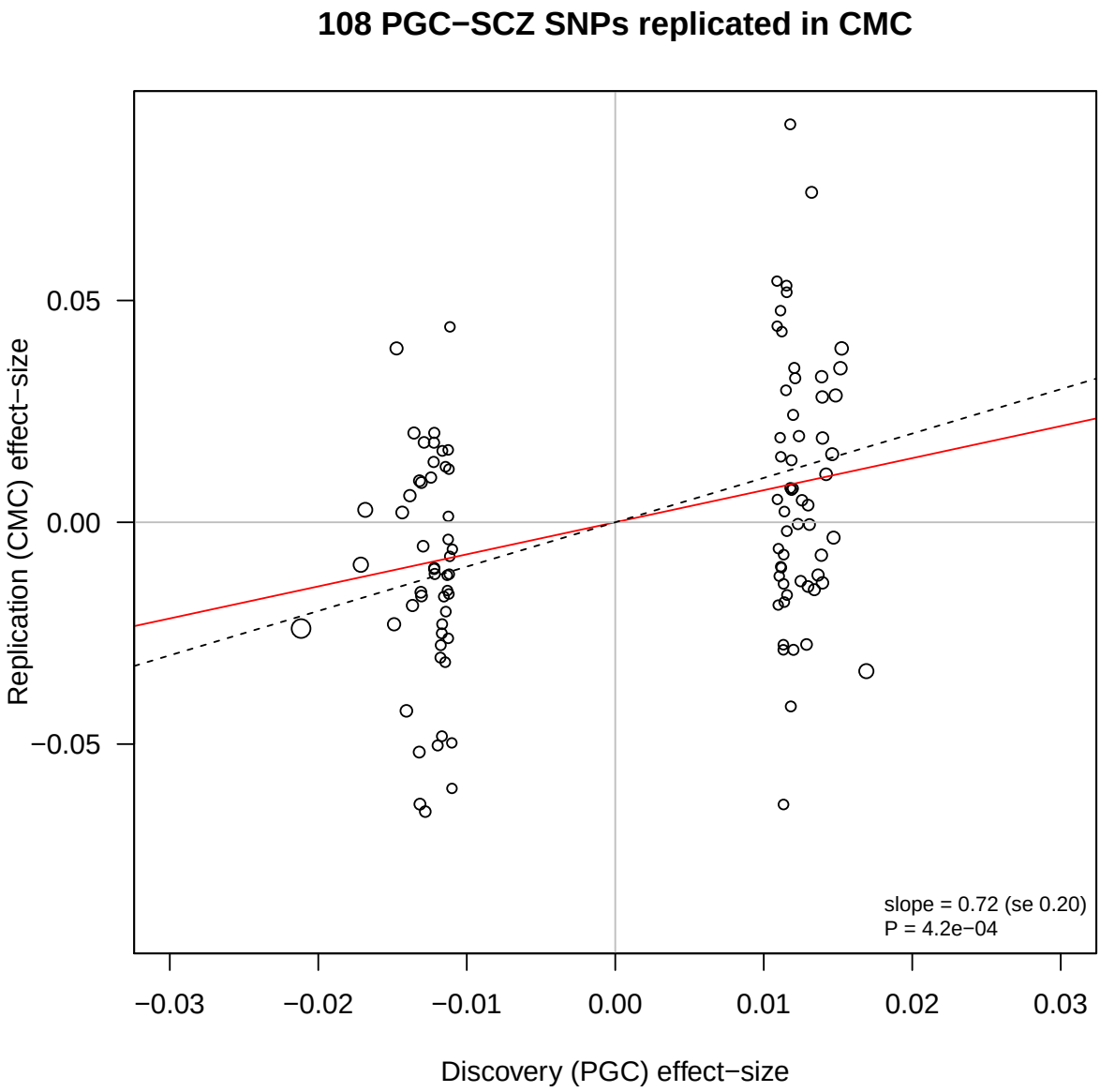
Supplementary Figure 2: **Individual-tissue Manhattan plots.**



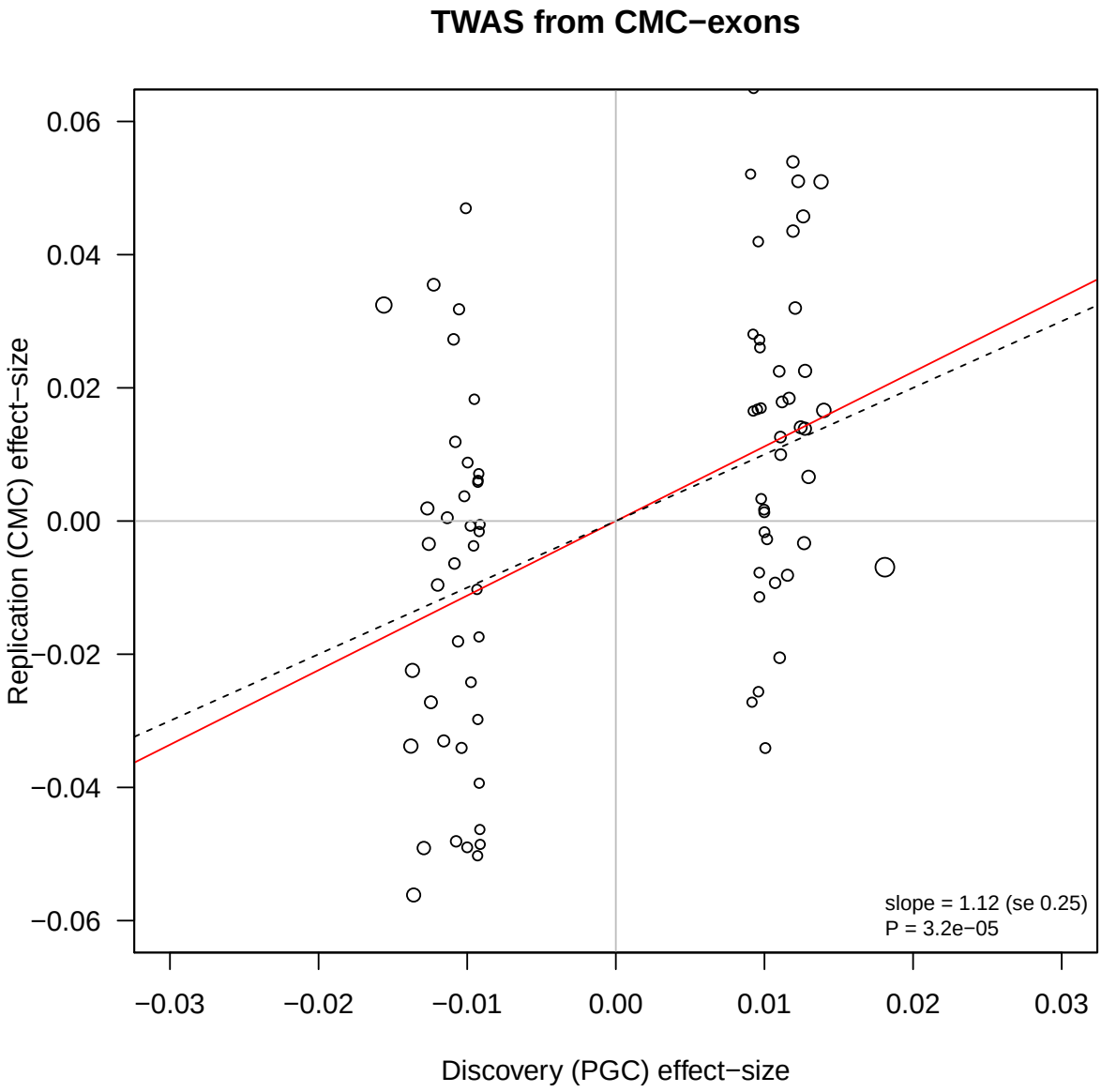


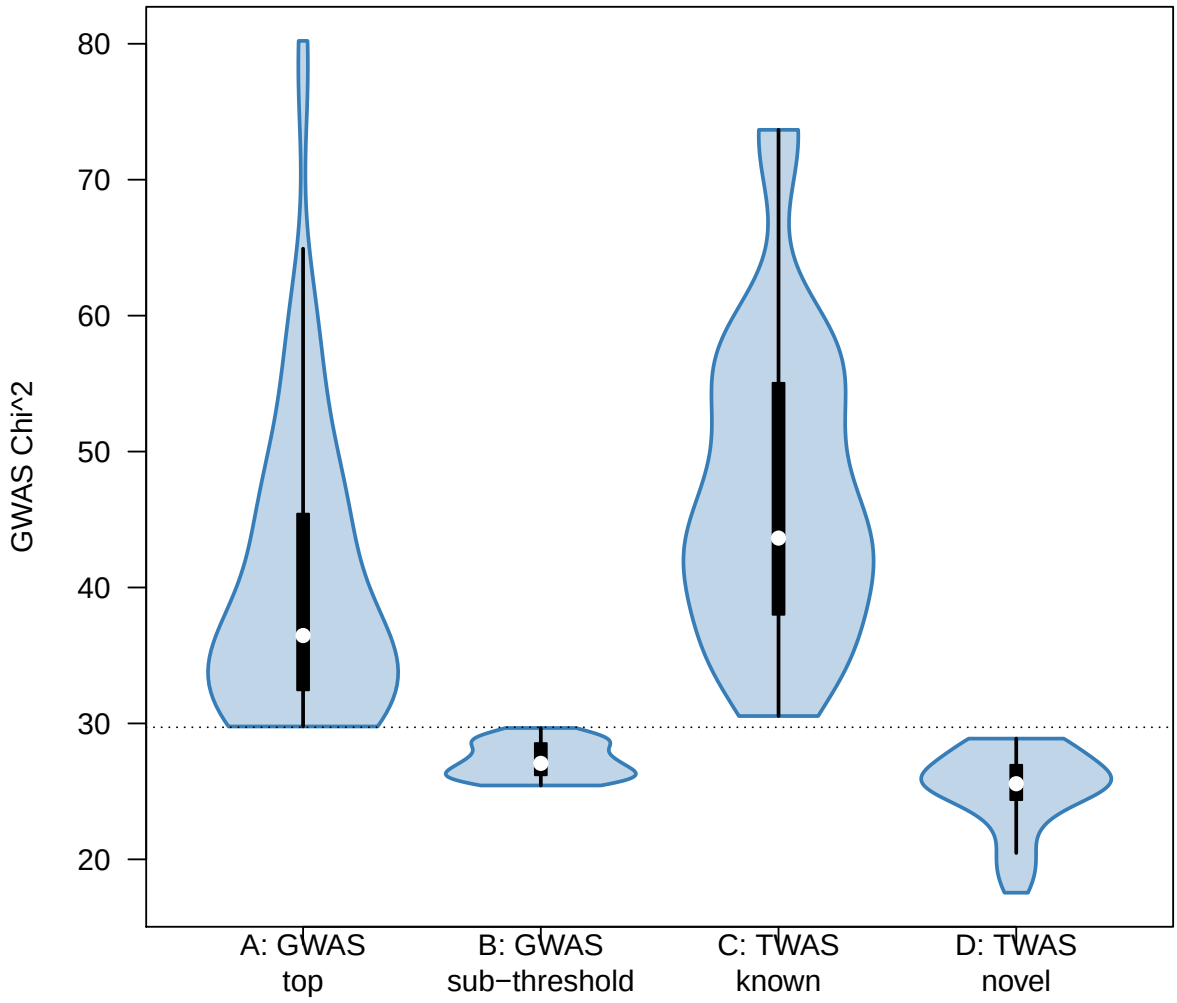
Supplementary Figure 3: **Replication of TWAS associations.** (**a**) The PGC data was randomly split into increasingly large discovery samples (size on the x-axis) and TWAS statistics were estimated from each reference panel. Left panel reports the number of significant genes (after 5% FWER (Bonferroni) correction) for a given GWAS sample size. Right panel reports the slope from a regression of $\beta_{\text{replication}} \sim \beta_{\text{discovery}}$ for significant genes identified at each sample size (where all $56,000 - x$ remaining samples were used as replication). (**b**) TWAS effect-sizes for association to schizophrenia identified in the PGC (x-axis) compared to corresponding estimates in the CMC (y-axis). Dotted line corresponds to $y = x$; red line corresponds to the slope from a (Z-score weighted) regression of CMC $\sim$ PGC, with estimate and p-value shown in bottom right. (**c**) Schizophrenia risk prediction R^2 shown for risk scores constructed from significant TWAS genes (bars) and PGC2 GWAS SNPs (dashed line, comparable to the $1\% - 3\%$ reported in ref[2]). Number of genes used in each score reported in parenthesis. Linear-regression R^2 of phenotype on predictor (after subtracting R^2 from jointly fit ancestry PCs) was transformed to the liability scale assuming schizophrenia prevalence of 1%, a linear transformation consistent across all predictors. P-value for the predictor in a joint model with the ancestry PCs shown above each bar.

Supplementary Figure 4: **Validation of significant PGC-SCZ SNPs in CMC.** SNP effect-sizes for association to schizophrenia identified in the PGC (x-axis) compared to corresponding estimates in the CMC (y-axis). Effect-sizes are log(odds-ratio) after accounting for ancestry PCs. Results were identical when using effect-sizes estimates by converting association Z-scores to liability-scale β, as in Figure 3. Dotted line corresponds to $y = x$, red line corresponds to the slope from a (Z-score weighted) regression of CMC $\sim$ PGC, with estimate and p-value shown in bottom right.



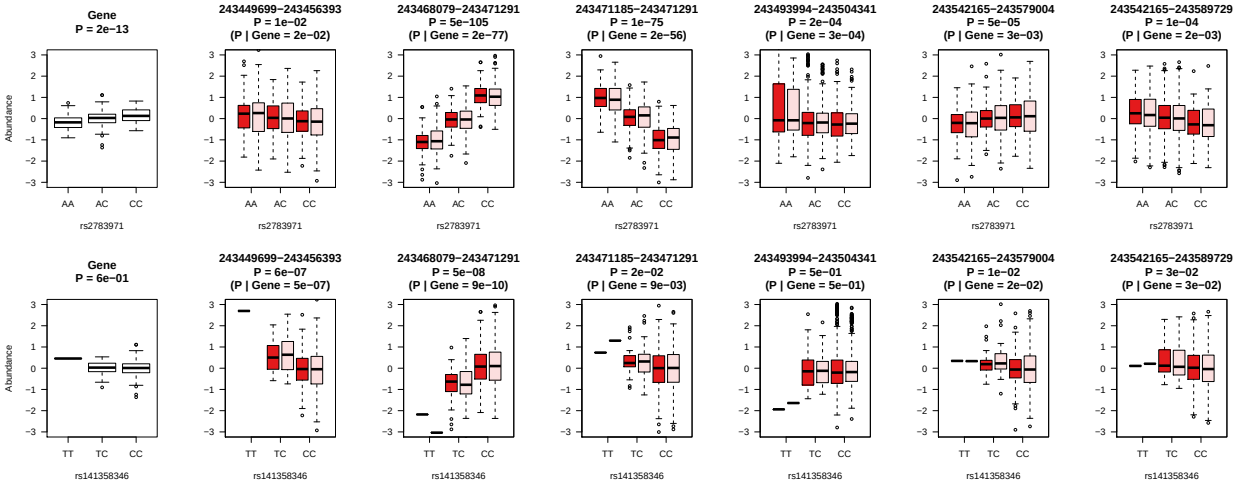
Supplementary Figure 5: **Validation of significant splice-TWAS associations in CMC.**

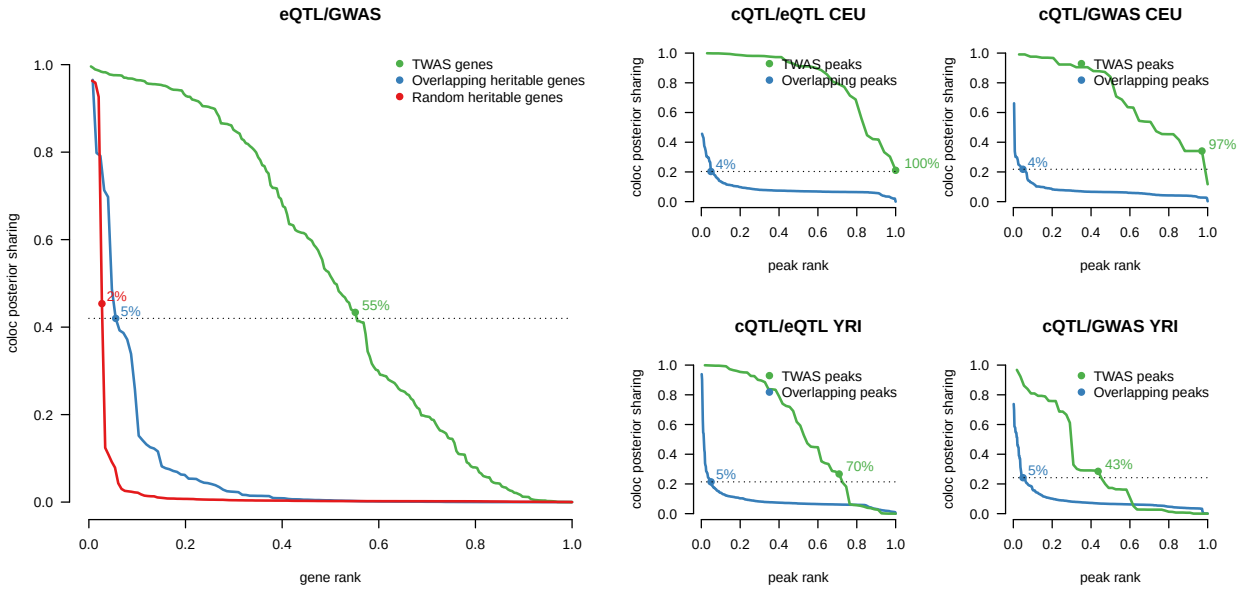




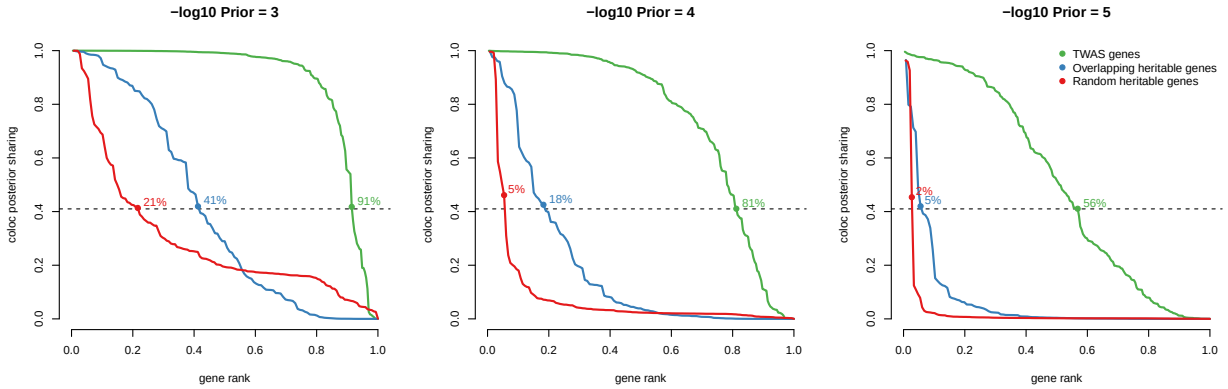
Supplementary Figure 6: **GWAS statistics at SCZ TWAS loci.** Density plots of best GWAS χ^2 statistics for (A) 108 PGC SCZ published loci containing genome-wide significant SNPs; (B) 56 most significant 1MB loci that did not overlap a genome-wide significant SNP (for comparison with D); (C) 190 TWAS loci that overlapped genome-wide significant SNPs; (D) 56 TWAS loci that did not overlap a genome-wide significant SNP.

Supplementary Figure 7: **Independent effects on alternative splicing at SDCCAG8.** Top (bottom) panels show the effect of the best and second best splice-QTL after stepwise model selection across all phenotypes. Effect on total expression shown at left, followed by the effect on each of the six inferred splicing events. The two variants have independent effects on splicing, and the second variant is also independent of total gene expression. Red box plots show intron abundance and pink box plots show same after regressing out the total gene expression.



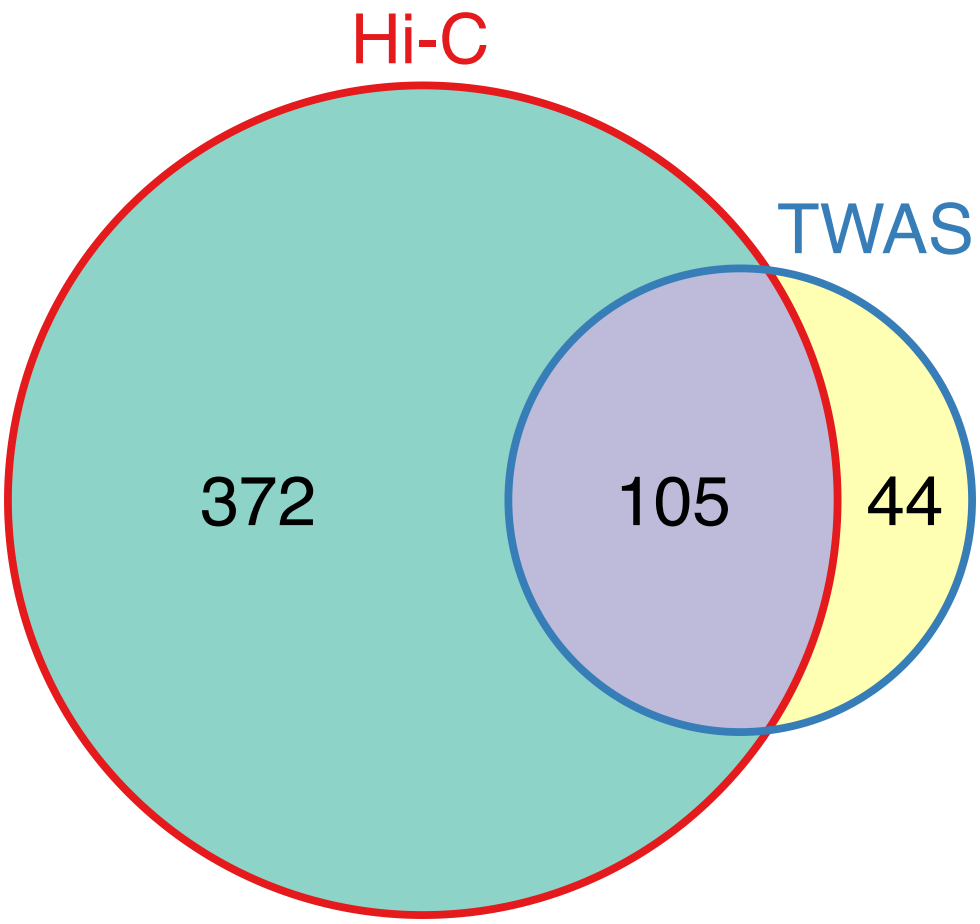


Supplementary Figure 8: **COLOC results at SCZ/chromatin TWAS loci.** Distribution of COLOC posterior probabilities for SCZ TWAS associations (left) and subset with chromatin TWAS associations (right). Dashed line (and labeled points) shows threshold calibrated so that $< 5\%$ of overlapping features (blue line) are considered colocalized.

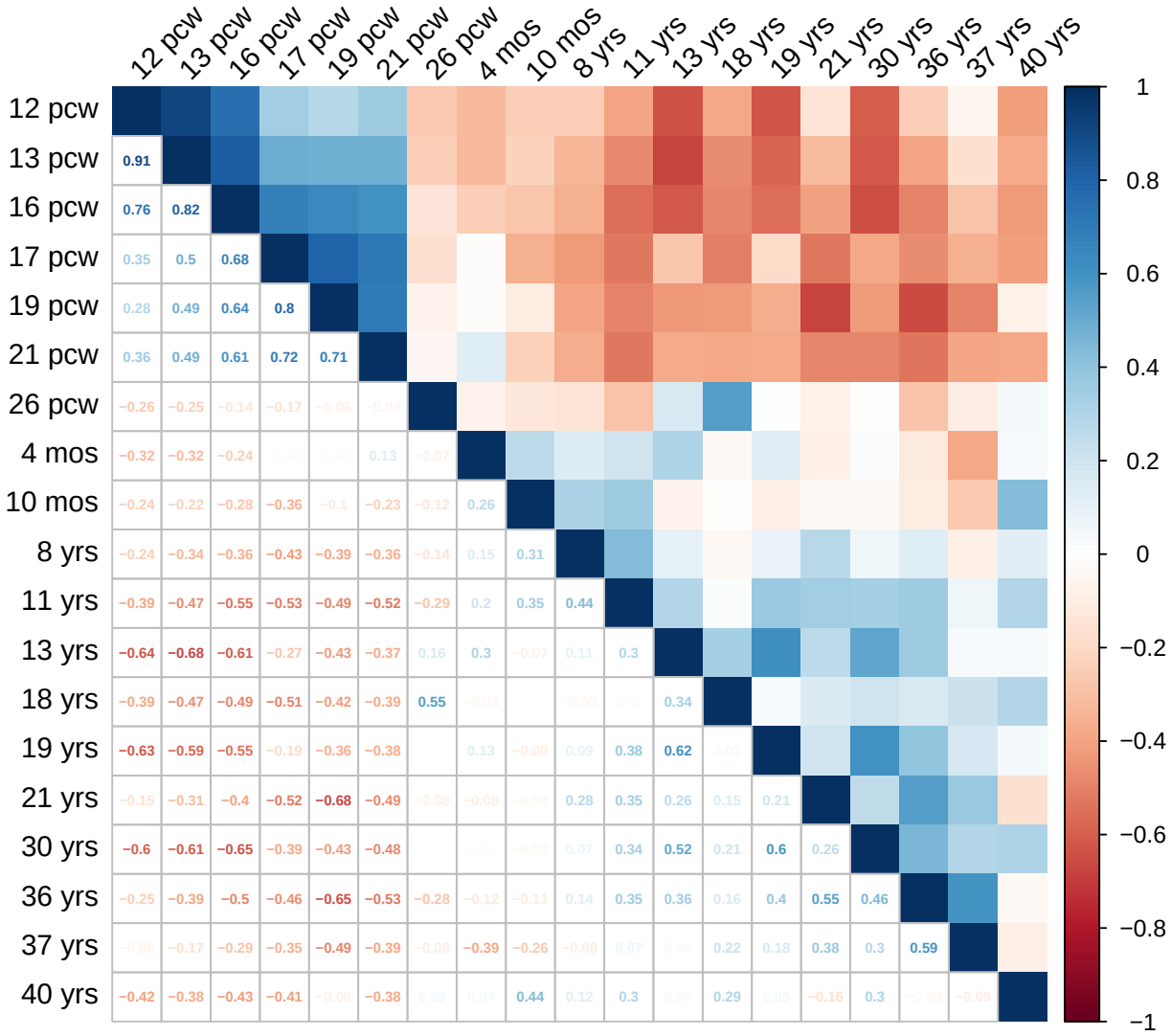


Supplementary Figure 9: **COLOC results at SCZ/chromatin TWAS loci.** Distribution of COLOC posterior probabilities for SCZ TWAS associations with three different priors on a shared genetic effect. (right) panel shows default prior (as in Supplementary Figure 8). Dashed line (and labeled points) shows threshold calibrated from default prior (as in Supplementary Figure 8) so that $< 5\%$ of overlapping features (blue line) are considered colocalized.

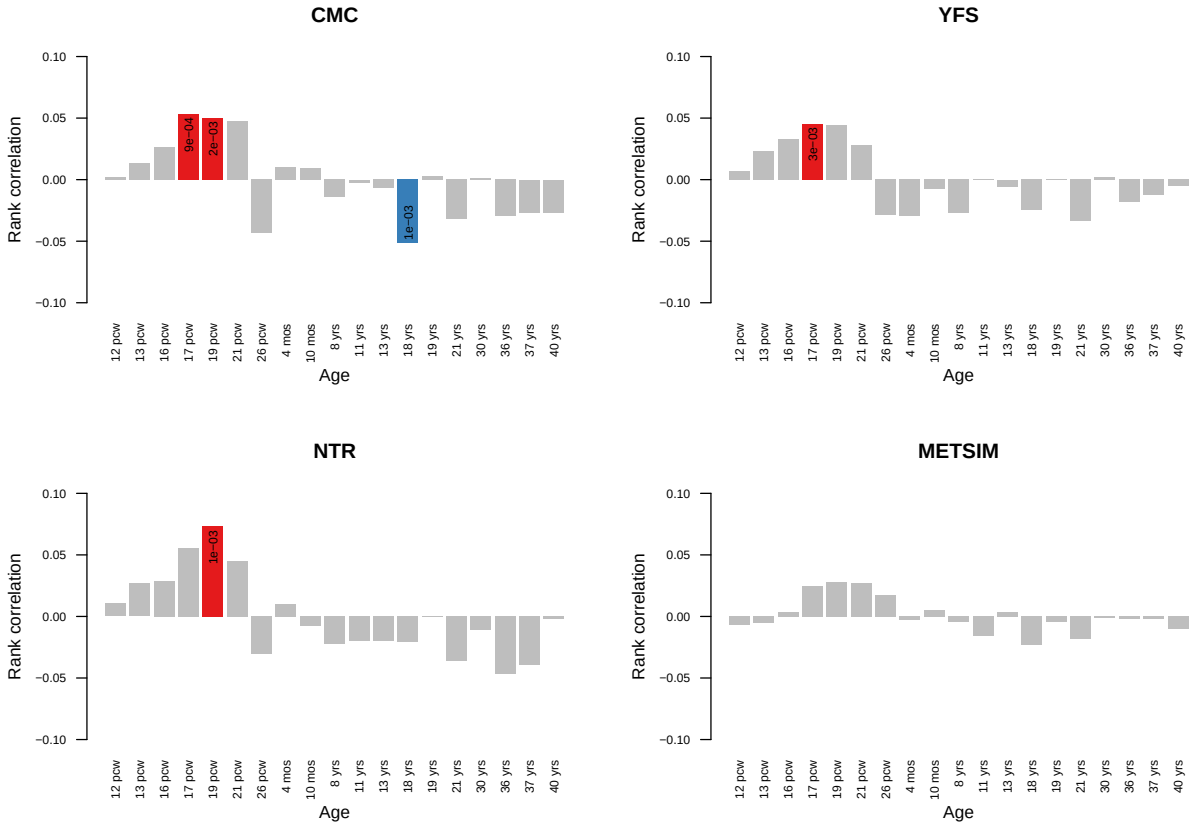
Supplementary Figure 10: **Overlap of genes prioritized by TWAS and Hi-C contacts.** (Hi-C) Genes with Hi-C contact to lead GWAS SNP in fetal brain; (TWAS) genes with SCZ TWAS association; and intersection.



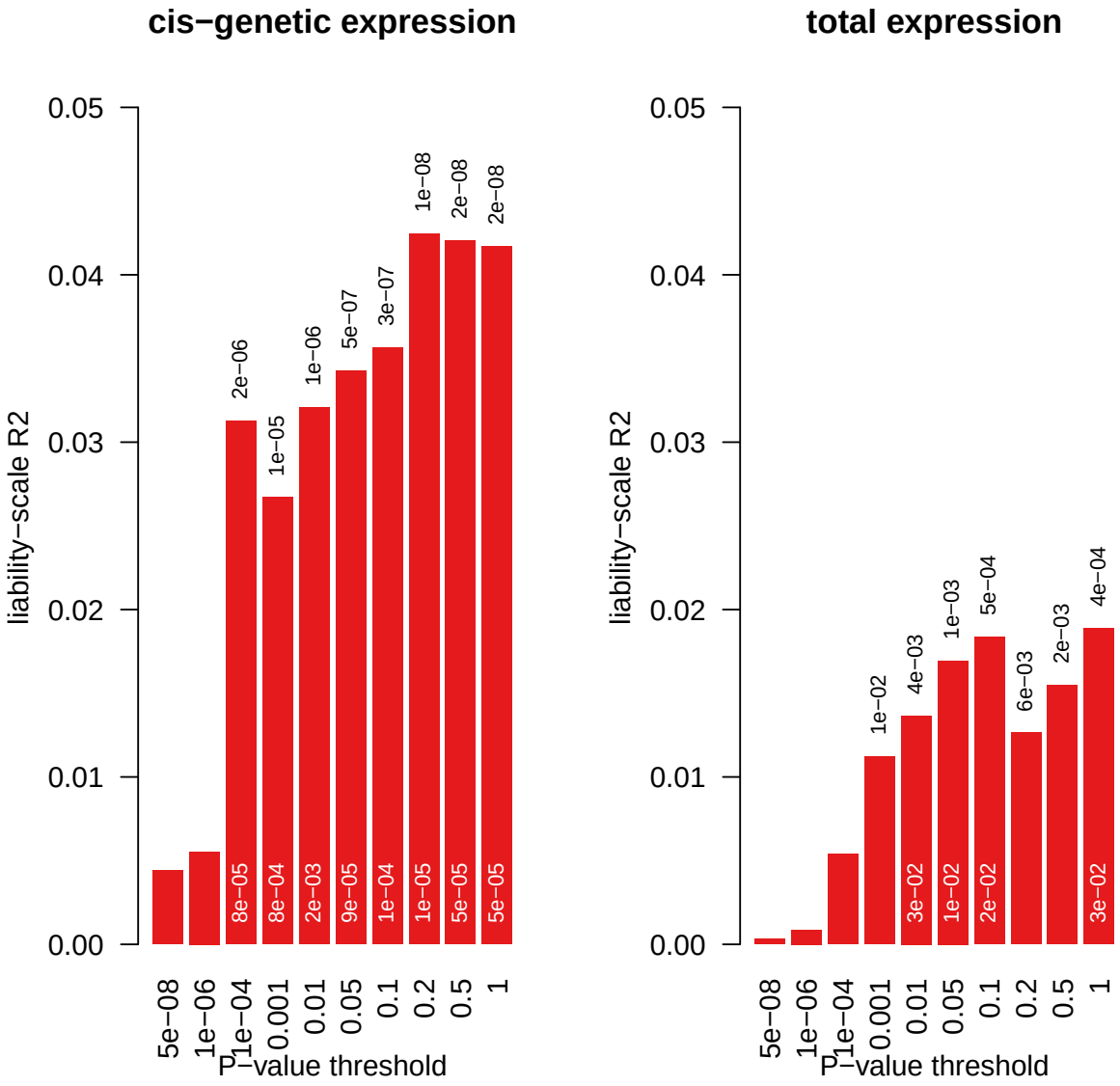
Supplementary Figure 11: **BRAINSPAN differential expression correlation matrix.** Correlation of differential expression Z-scores across developmental periods within the BRAINSPAN pre-frontal cortex RNA-seq. Upper right shows correlation with color; lower left shows correlation coefficient.



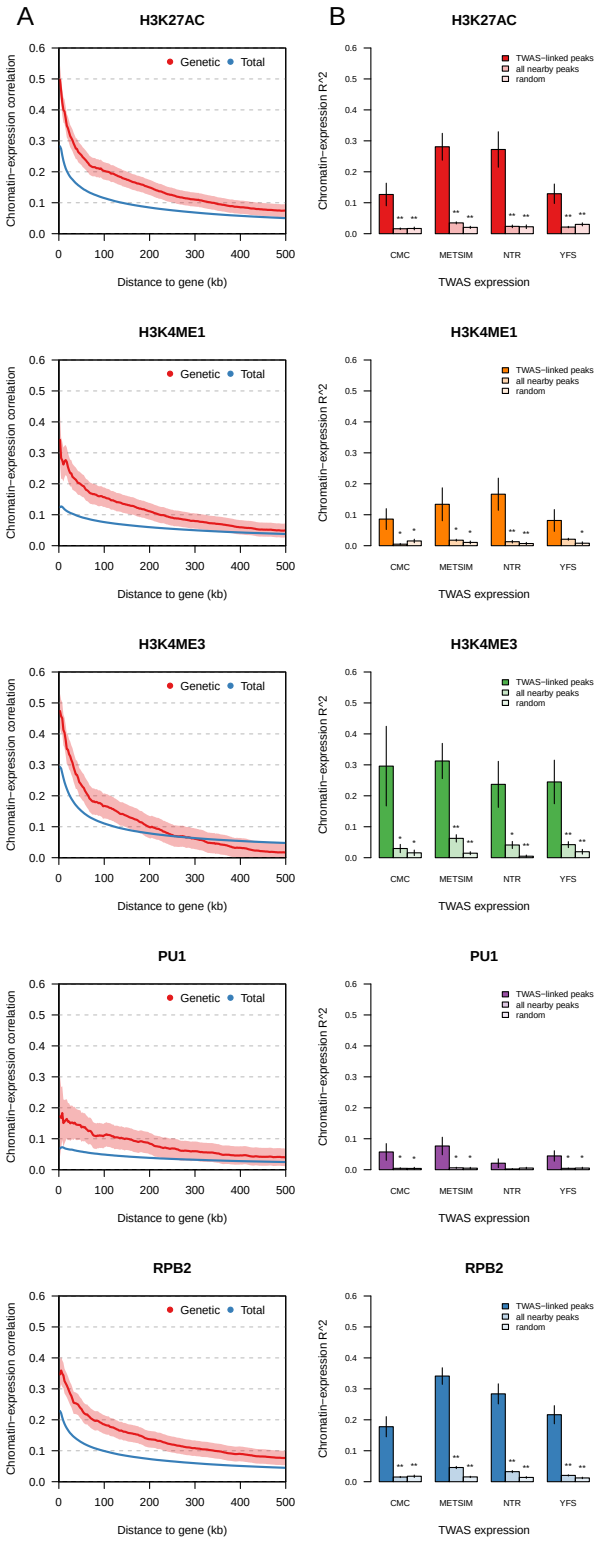
Supplementary Figure 12: **SCZ TWAS correlation with BRAINSPAN temporal differential expression.** For each TWAS reference panel, the Spearman rank correlation is reported between SCZ TWAS effect sizes (Z-scores) and BRAINSPAN differential expression Z-scores across developmental periods in pre-frontal cortex RNA-seq. Correlations that are significant after Bonferroni correction for number of periods are color-coded and labeled with p-values.



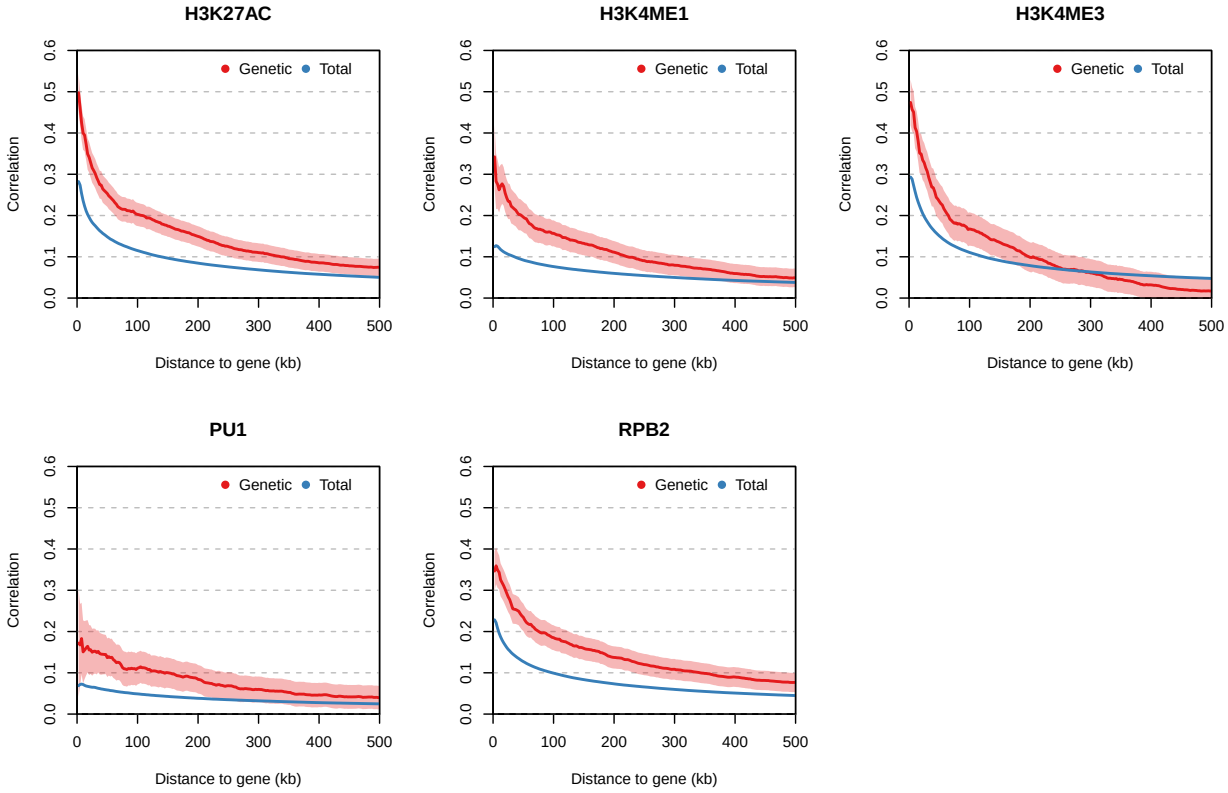
Supplementary Figure 13: **Gene risk score from CMC/brain TWAS.** Liability-scale R^2 between gene risk score and schizophrenia status in CMC for different discovery (PGC) significance thresholds. Left panel shows score constructed from genetic component of expression/splicing; right panel shows score from overall expression/splicing. In both cases, the R^2 is measured from the model $y \sim GeRS_{gene} + GeRS_{intron} + PC1 + PC2$ after subtracting R^2 from PCs only and transforming to liability scale at 1% prevalence (a linear transformation). P-value in black corresponds to 2-dof ANOVA test against model with PC's only. P-value in white corresponds to $GeRS_{intron}$ regression coefficient. P-values only reported where < 0.05.



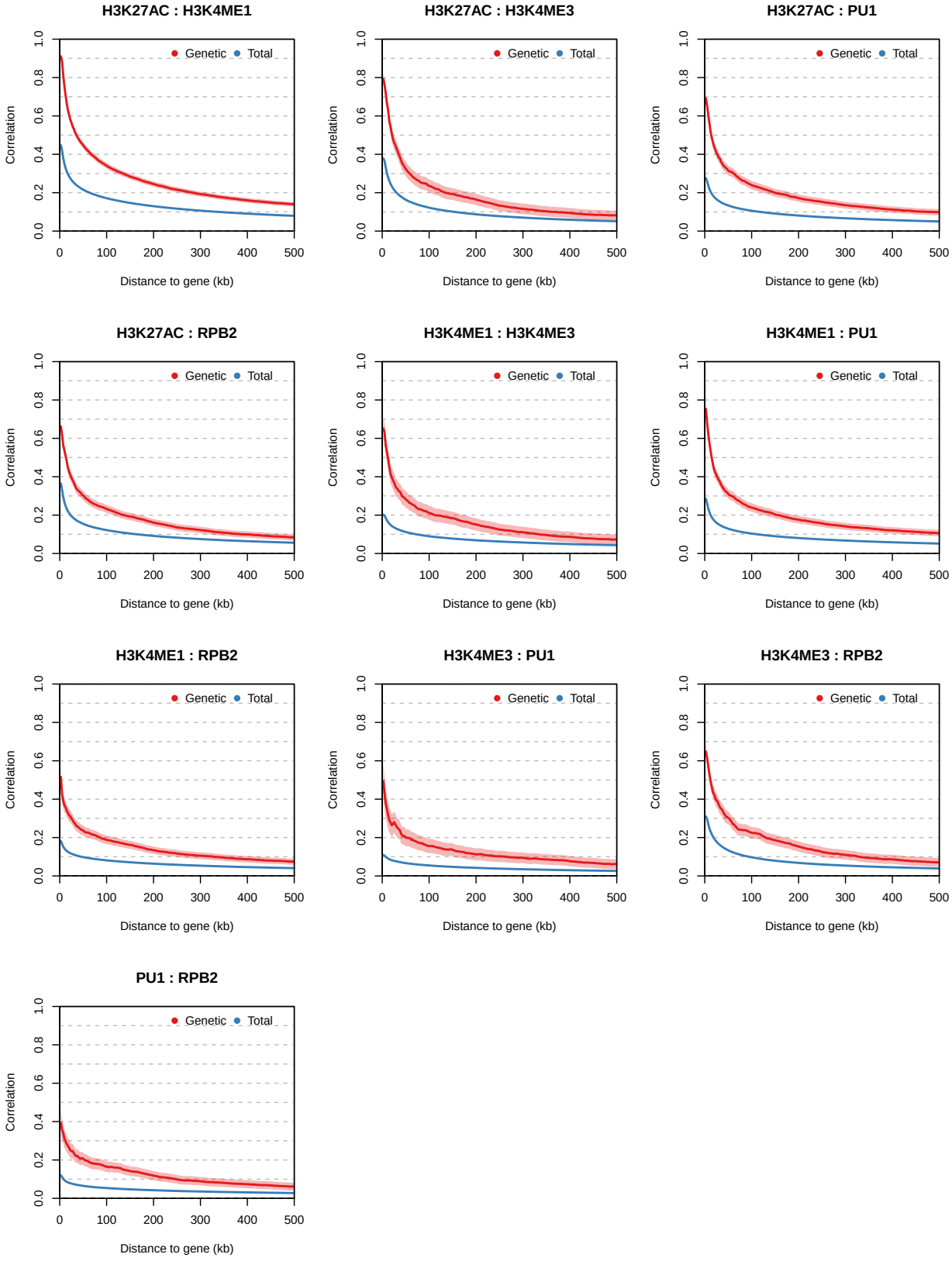
Supplementary Figure 14: **Correlation between expression and chromatin phenotypes.** (**A**) For each gene and cis-locus (±500kb) the average total and genetic correlation between expression of the gene and nearby chromatin state is plotted as function of distance from chromatin mark to gene center. (**B**) For every significant TWAS-linked gene-mark combination, the adjusted R^2 between mark and corresponding RNA-seq levels was measured, with dark bars reporting the mean by chromatin category and reference panel. The same quantity was computed for all peaks ±500kb to the significant TWAS genes ("all nearby peaks" bar); as well as for 10 samplings of any random peak-gene combination no more than 500kb apart ("random" bar). Error bars indicate standard error of the mean. * (**) indicate instances where the null is significantly lower than TWAS-gene estimate by T-test at $P < 0.05$ ($P < 0.05/5$).

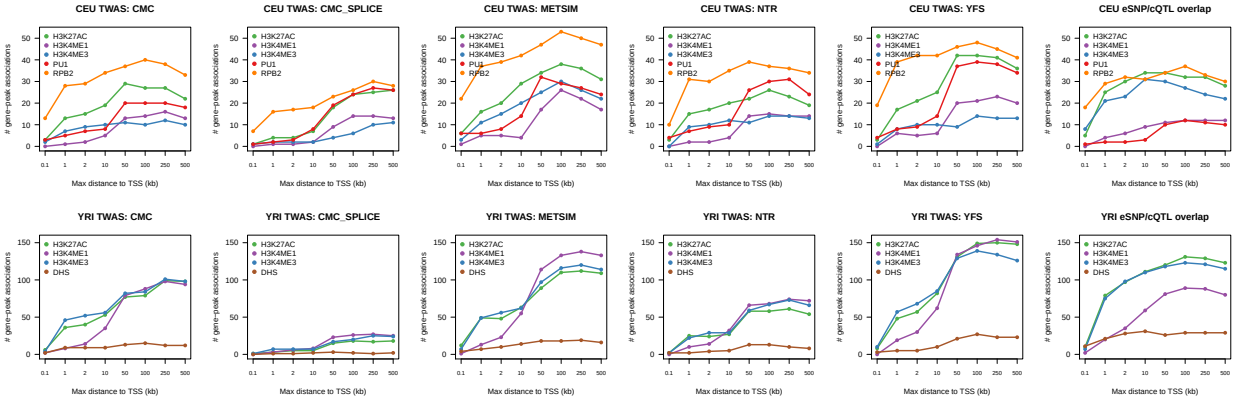


Supplementary Figure 15: **Genetic correlation between chromatin and cis-expression in CEU.** For each gene and cis-locus (±500kb) the average total and genetic correlation between expression of the gene and nearby chromatin state is plotted as function of distance from chromatin peak to gene center. Genetic correlation was computed using the mean cis-h_g^2 and cis-cov_g across all peaks closer than the specified distance (x-axis). Total correlation is the Pearson correlation between expression and chromatin state across individuals.

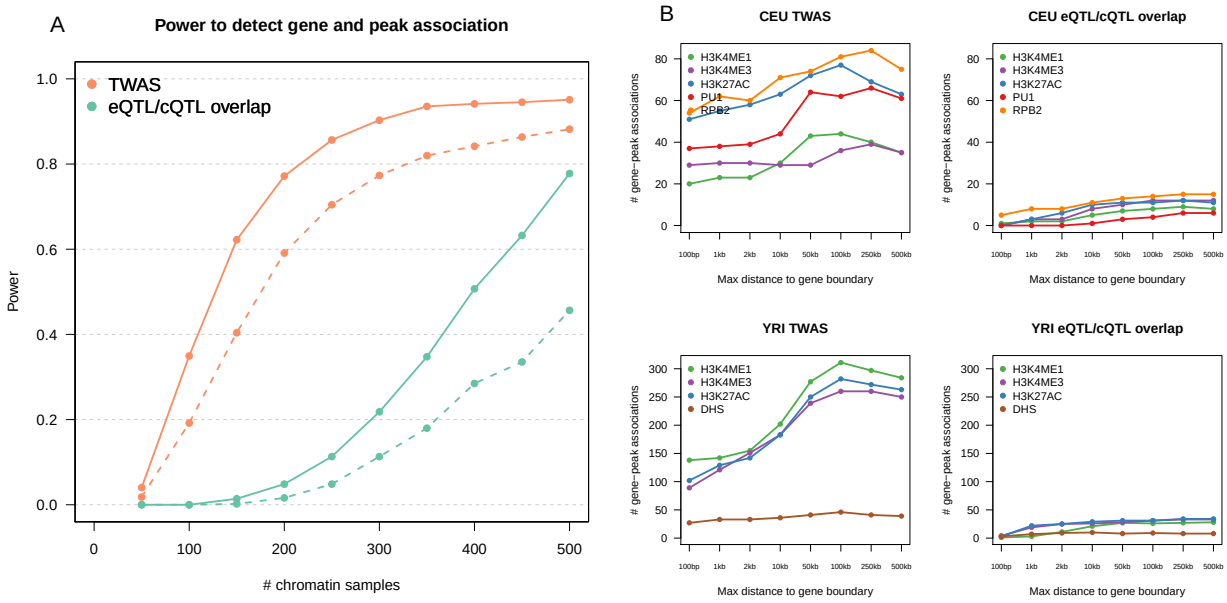


Supplementary Figure 16: **Genetic correlation between chromatin phenotypes in CEU.** For each chromatin peak and cis-locus (±500kb) the average total and genetic correlation between nearby peaks is plotted as function of maximum distance between peaks (x-axis). Genetic correlation was computed using the mean cis-h_g^2 and cis-cov_g across all peaks in the window. Total correlation is the Pearson correlation between peaks across individuals.



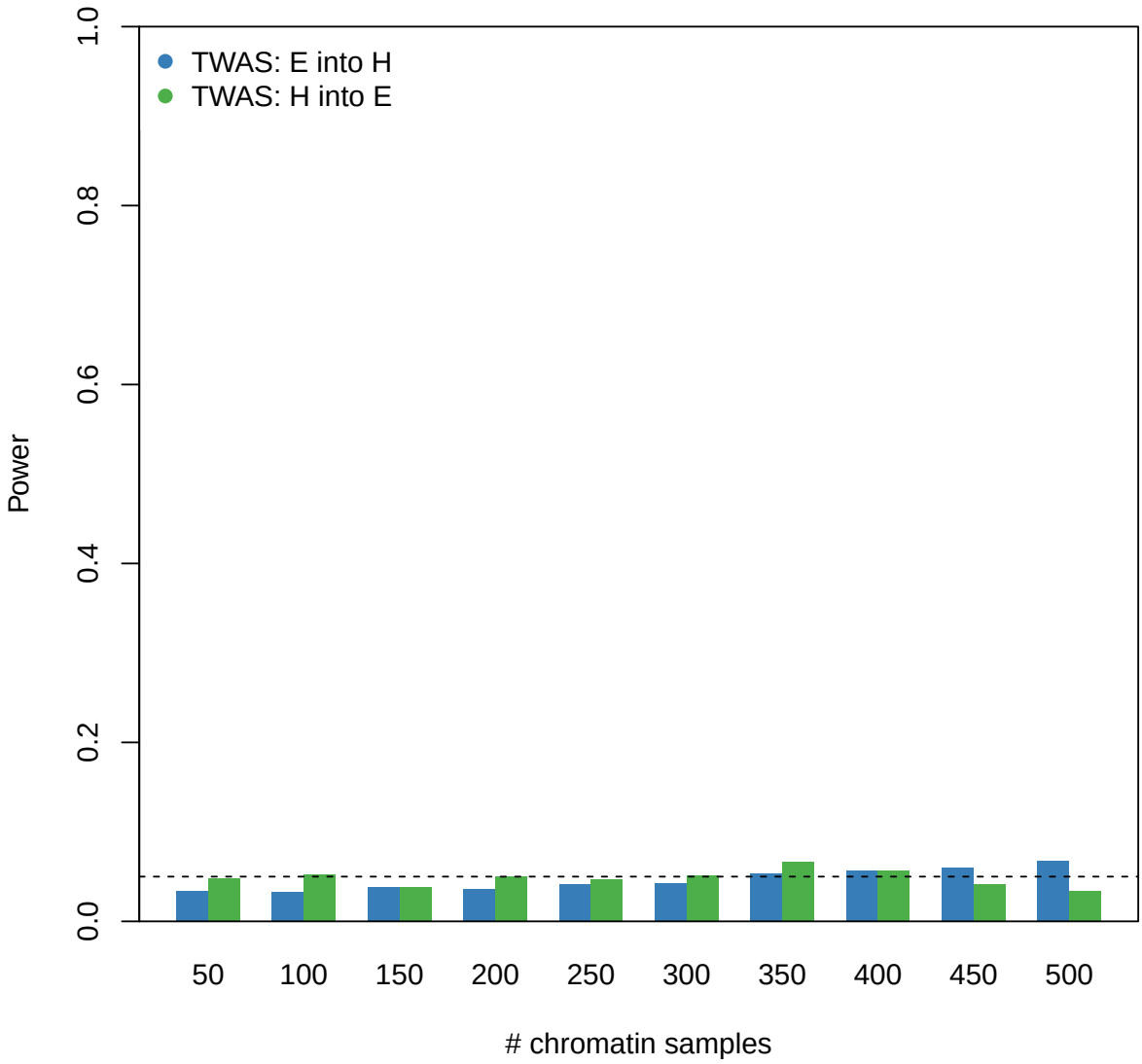


Supplementary Figure 17: **Chromatin TWAS associations compared to top eSNP/cQTL associations.** Number of unqiue genes significantly associated with a chromatin peak after Bonferroni correction for a given distance from the gene (x-axis): (right) using top eSNP in chromatin cohort; (left) using TWAS from each reference panel. Results from CEU (YRI) populations shown in top (bottom) panels.

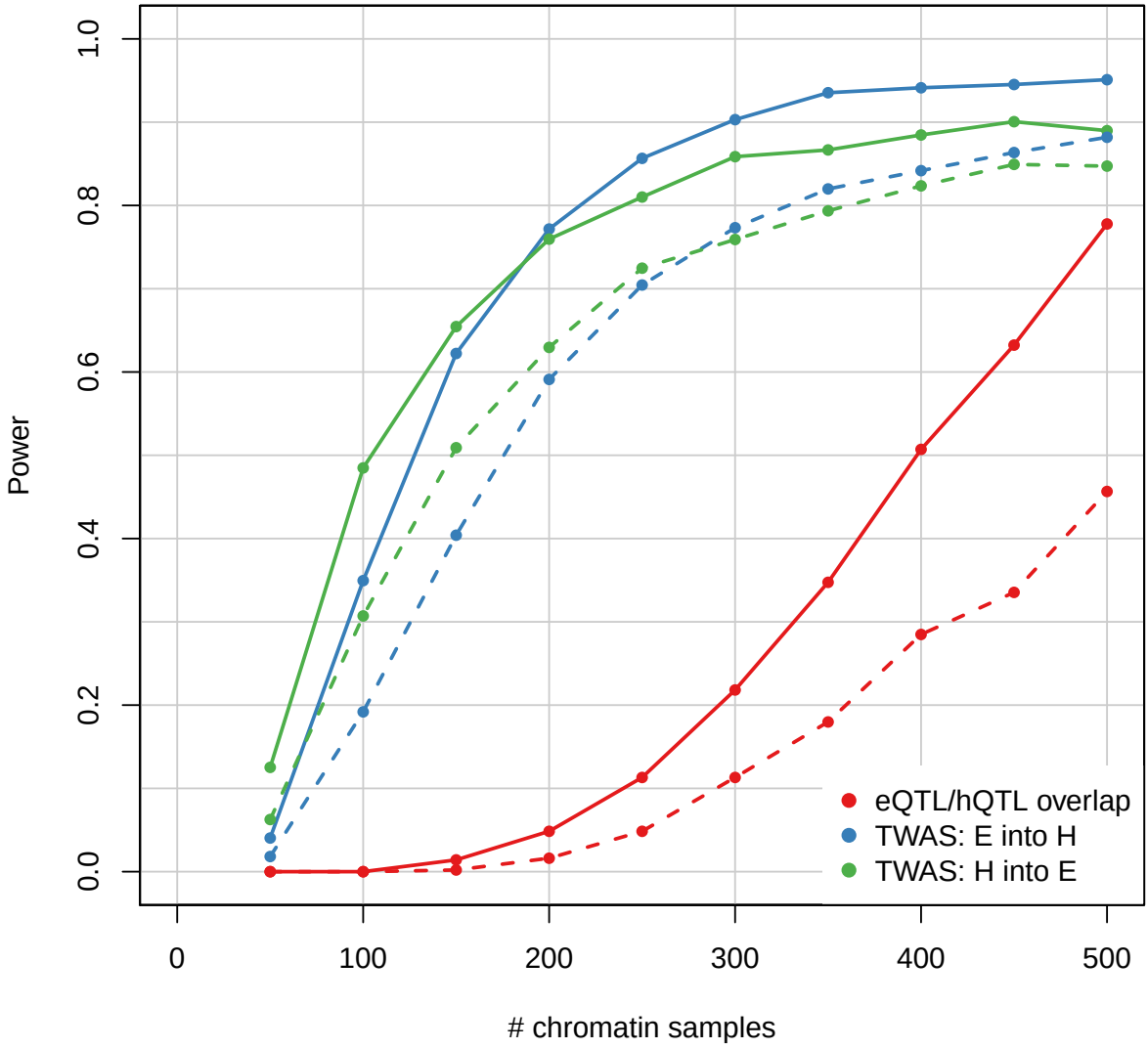


Supplementary Figure 18: **Power and detection of significant gene-mark associations.** (**a**) Molecular phenotypes were simulated under the SNP → chromatin → expression model and two methods to detect gene-mark associations evaluated. (TWAS, orange) corresponds to predicting expression from a held-out reference panel with 1,000 individuals and testing each proximal chromatin peak for association. (eQTL/cQTL, green) corresponds to identifying SNPs that are significantly associated with both chromatin and expression at the locus. For a given chromatin phenotype sample size (x-axis), power was measured as the number of instances where locus was deemed significant after accounting for number of gene-mark pairs tested (TWAS) or number of SNPs and gene-mark pairs tested (eQTL/cQTL). Solid (dashed) lines correspond to 1 (2) chromatin-causing variants in the simulation. (**b**) Genes significantly associated with a chromatin phenotype peak in YRI/CEU. For a given distance from the gene (x-axis), the number of unique genes is reported after experiment-wide Bonferroni correction for tests within each chromatin phenotype. Results from TWAS prediction-based approach shown on left; results from overlapping QTL approach on the right (see Methods).

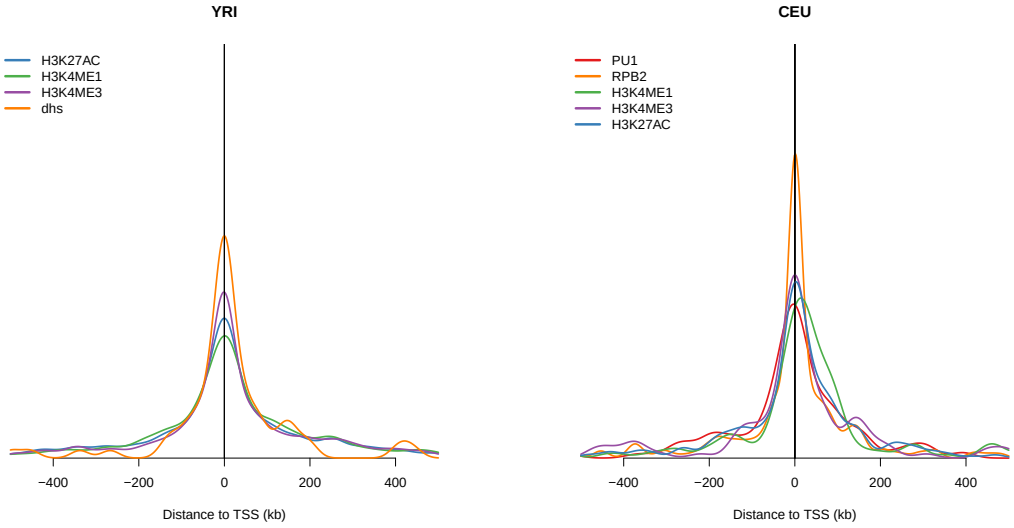
Supplementary Figure 19: **Calibration of chromatin/expression TWAS under the null.** Chromatin phenotypes and expression were simulated as non-heritable and a chromatin/expression TWAS performed. Bars correspond to the fraction of genes significant at $P < 0.05$, with the dotted line corresponding to the expected 5%.



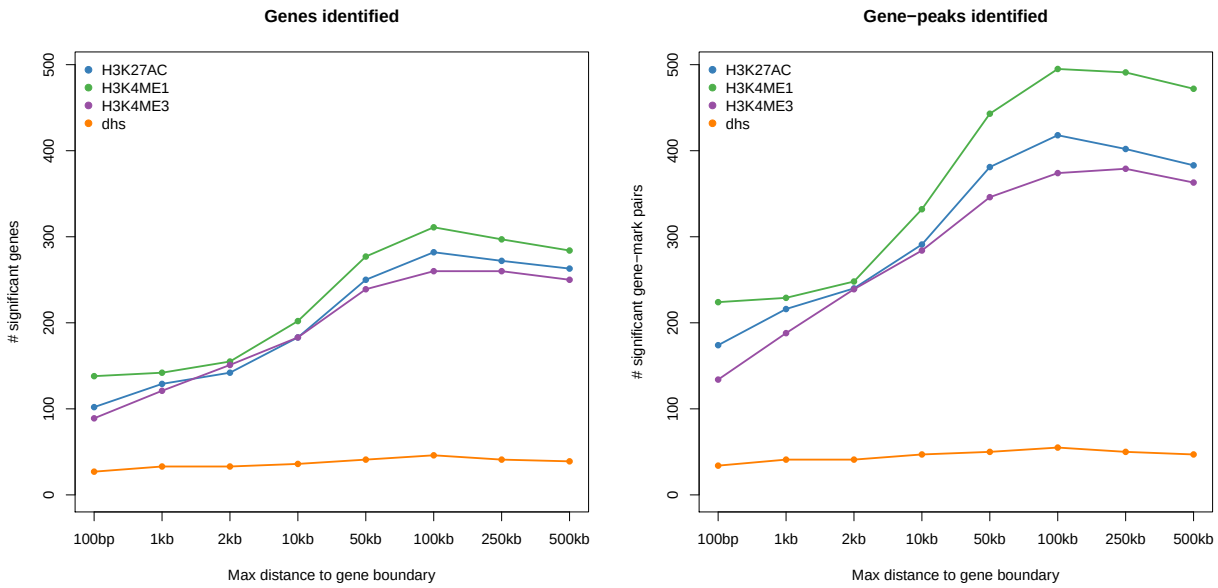
Supplementary Figure 20: **Chromatin/expression TWAS performance in simulation.** Molecular phenotypes were simulated under the SNP → chromatin → expression model and methods to detect gene-chromatin-peak associations evaluated. (TWAS: E to H, blue) corresponds to predicting expression from a reference panel with 1,000 individuals and testing against each proximal chromatin peak for association at a given chromatin phenotype size (x-axis). (TWAS: H to E, green) corresponds to the same procedure but predicting from chromatin samples (x-axis) into the expression individuals. (eQTL/hQTL, red) corresponds to identifying SNPs that are significantly associated with both chromatin and expression at the locus. Power is measured as the number of instances where locus was deemed significant after accounting for number of gene-peak pairs tested (TWAS) or number of SNPs and gene-peak pairs tested (eQTL/cQTL). Solid (dashed) lines correspond to 1 (2) chromatin-causing variants in the simulation.



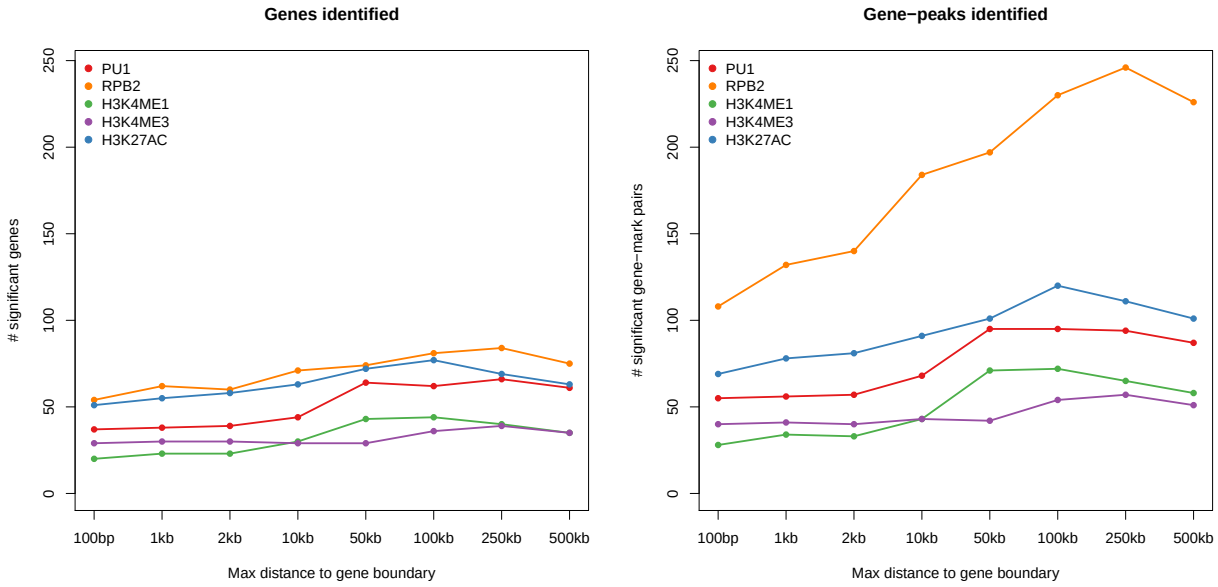
Supplementary Figure 21: **Distribution of chromatin TWAS associated peaks.** For each chromatin phenotype and cohort, the density plot of the distance between peak center and TSS of the associated gene is shown. Computed for chromatin TWAS associations at 10% FDR.



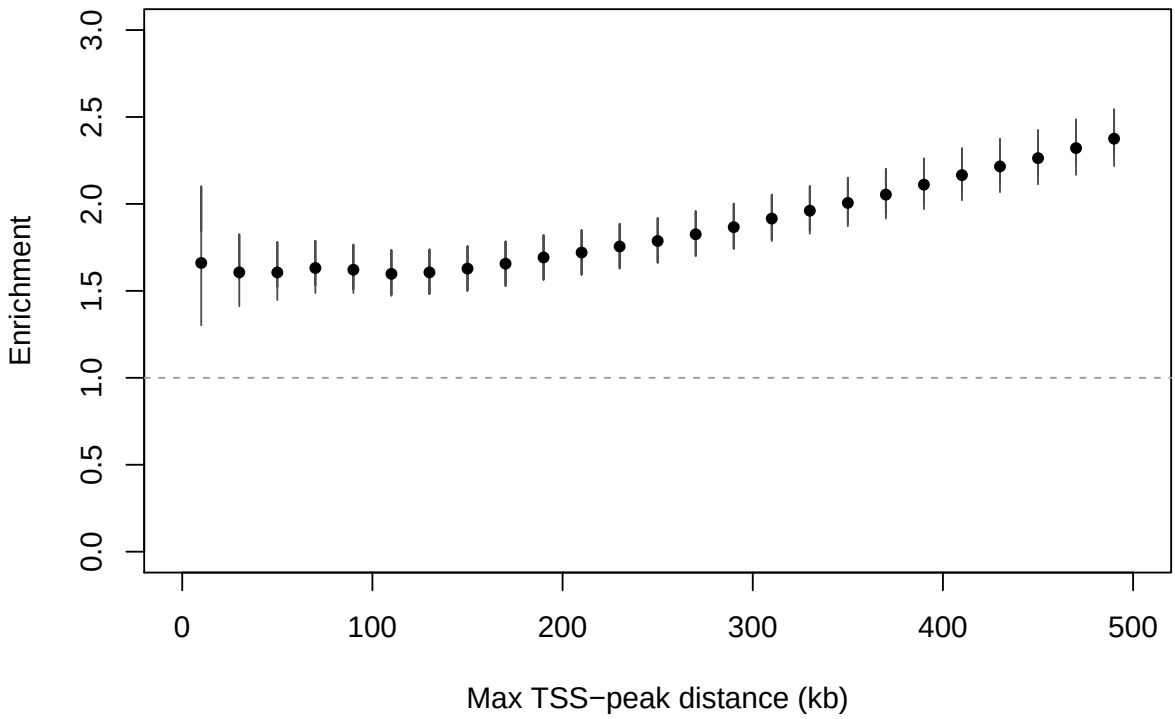
Supplementary Figure 22: **Significant chromatin TWAS associations in YRI.** For each chromatin phenotype, number of chromatin TWAS associations (left: unique genes; right: unique gene-peak pairs) significant at $P < 0.05$ after Bonferroni correction (y-axis) is shown for a given cis window size (x-axis). Note x-axis is non-linear.



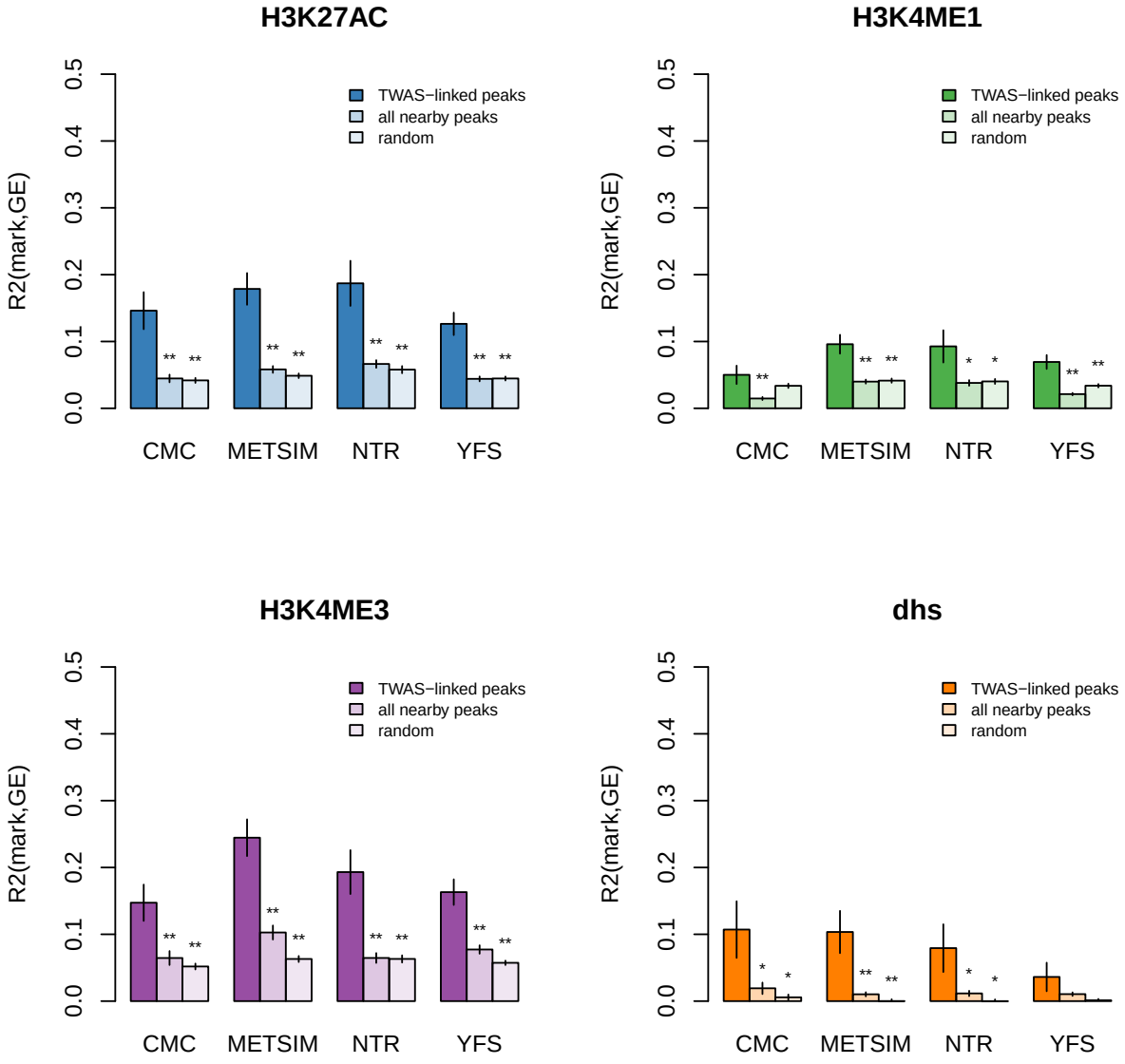
Supplementary Figure 23: **Significant chromatin TWAS associations in CEU.** For each chromatin phenotype, number of chromatin TWAS associations (left: unique genes; right: unique gene-peak pairs) significant at $P < 0.05$ after Bonferroni correction (y-axis) is shown for a given cis window size (x-axis). Note x-axis is non-linear. Note that, in contrast to the YRI data, chromatin peaks in the CEU samples were defined based on a small set of reference individuals, resulting in many fewer peaks and yielding fewer chromatin TWAS associations.



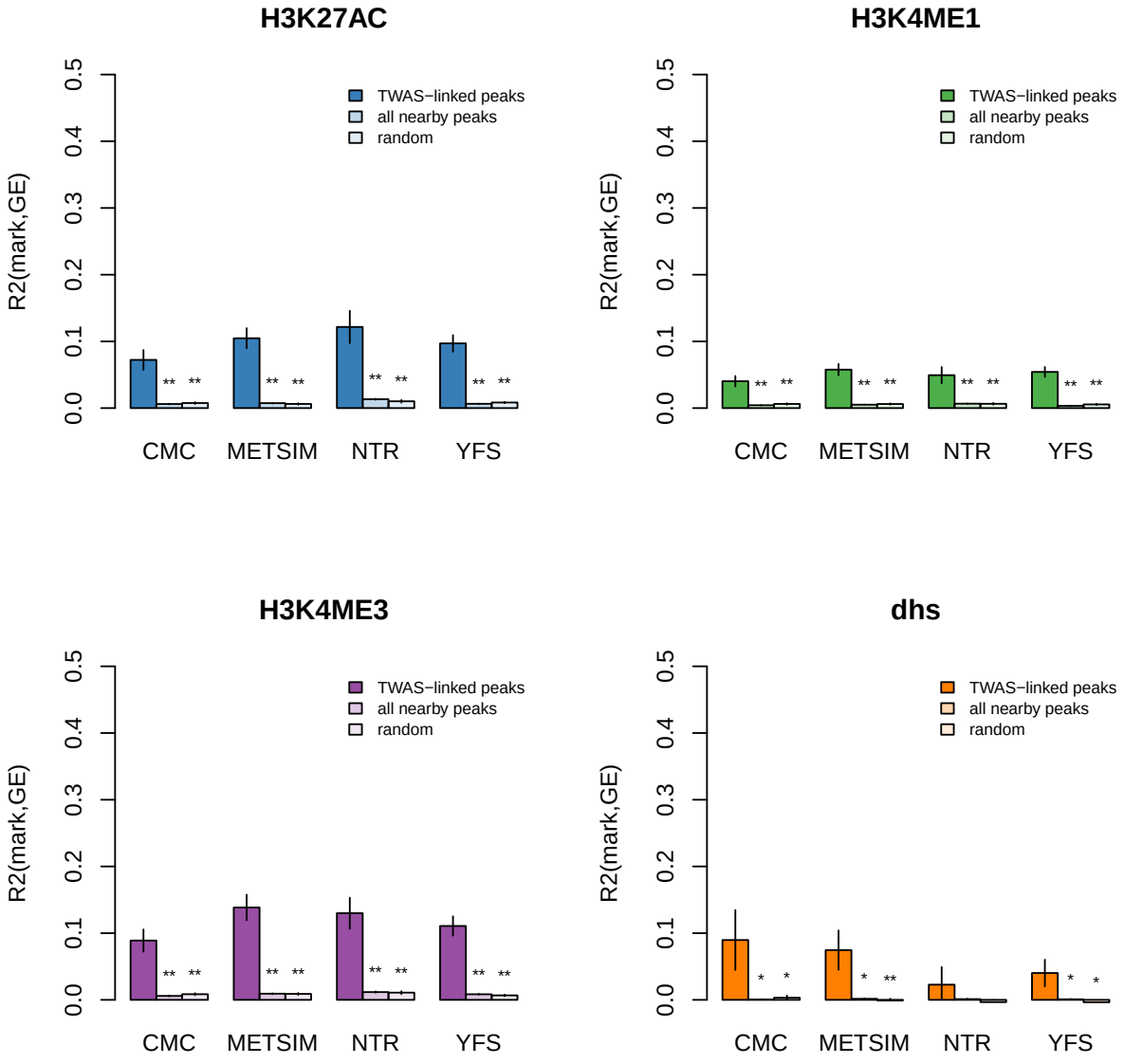
Supplementary Figure 24: **Enrichment of Hi-C loops at chromatin TWAS associated elements.** Across all TWAS associated gene-peak pairs (at 10% FDR) that are less than (x-axis) kb apart, the y-axis shows the odds ratio of Hi-C looping enrichment. Enrichment computed as the fraction of associated pairs with a Hi-C chromatin contact compared to the fraction of all gene-peak pairs at the same maximum distance with Hi-C contact. Minimal peak-gene distance for all tests was 10kb, to ascertain for distal associations. Note that successive points describe nested data.



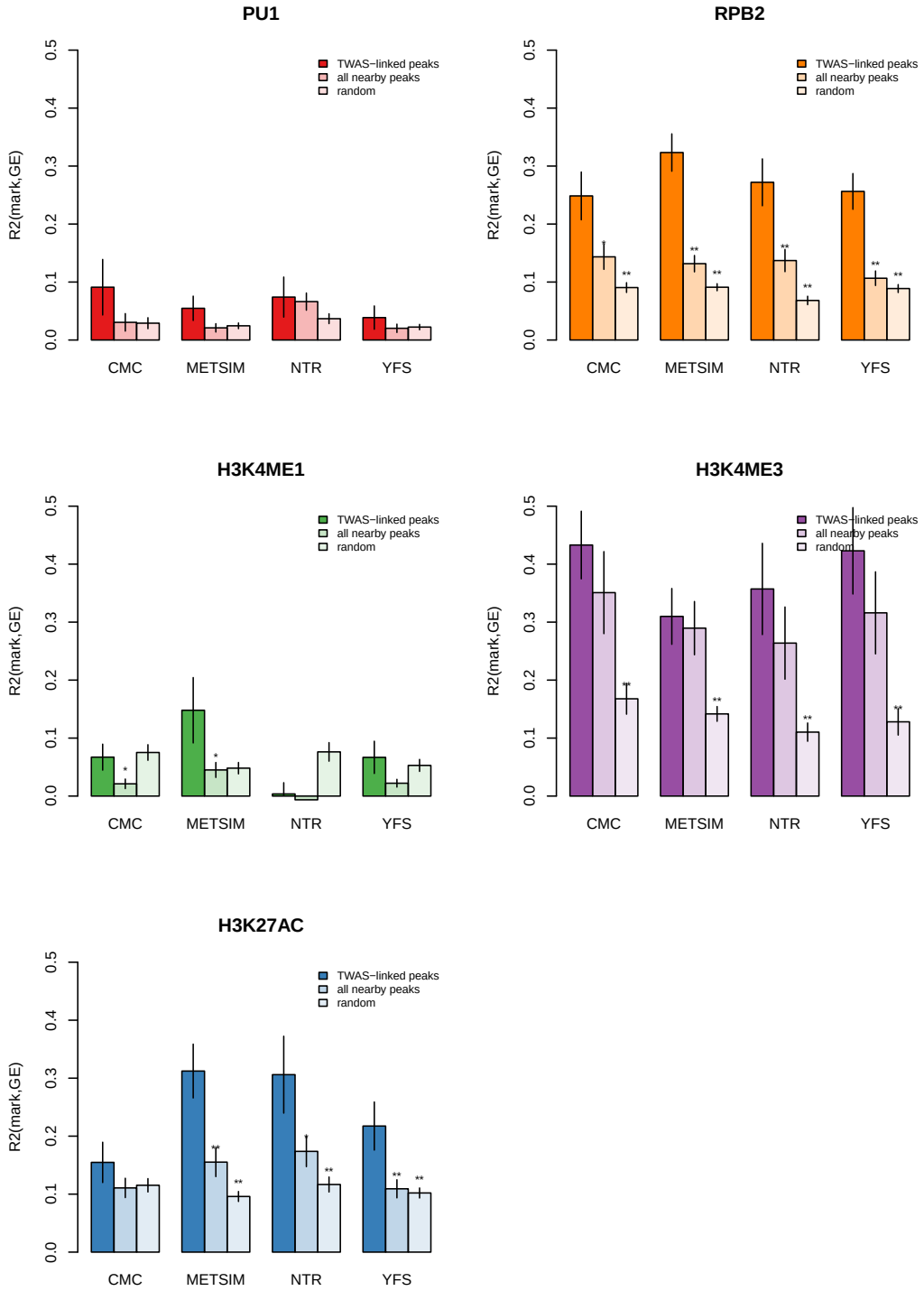
Supplementary Figure 25: **Variance in RNA expression explained by chromatin TWAS associations (±2kb) in YRI.** For every significant chromatin TWAS associations, the adjusted R^2 between the peak and corresponding measured expression levels is reported, with dark bars reporting the mean by chromatin category and reference panel. Note that the RNA-seq phenotype was not used to identify the gene-mark combination and the TWAS prediction is out-of-sample; this yielding an unbiased estimate of variance in expression explained by the associated mark. Because nearby peaks are expected to have some non-zero correlation to the gene we considered two nulls: the same quantity was computed for all peaks ±2kb to the significant TWAS genes ("all nearby peaks" bar); as well as for 10 samplings of any random peak-gene combination no more than 2kb apart ("random" bar). Adjusted R^2 was always used so as to yield an unbiased mean estimate. Error bars indicate standard error of the mean. * (**) indicate instances where the null is significantly lower than TWAS-gene estimate by T-test at $P < 0.05$ ($P < 0.05/4$).



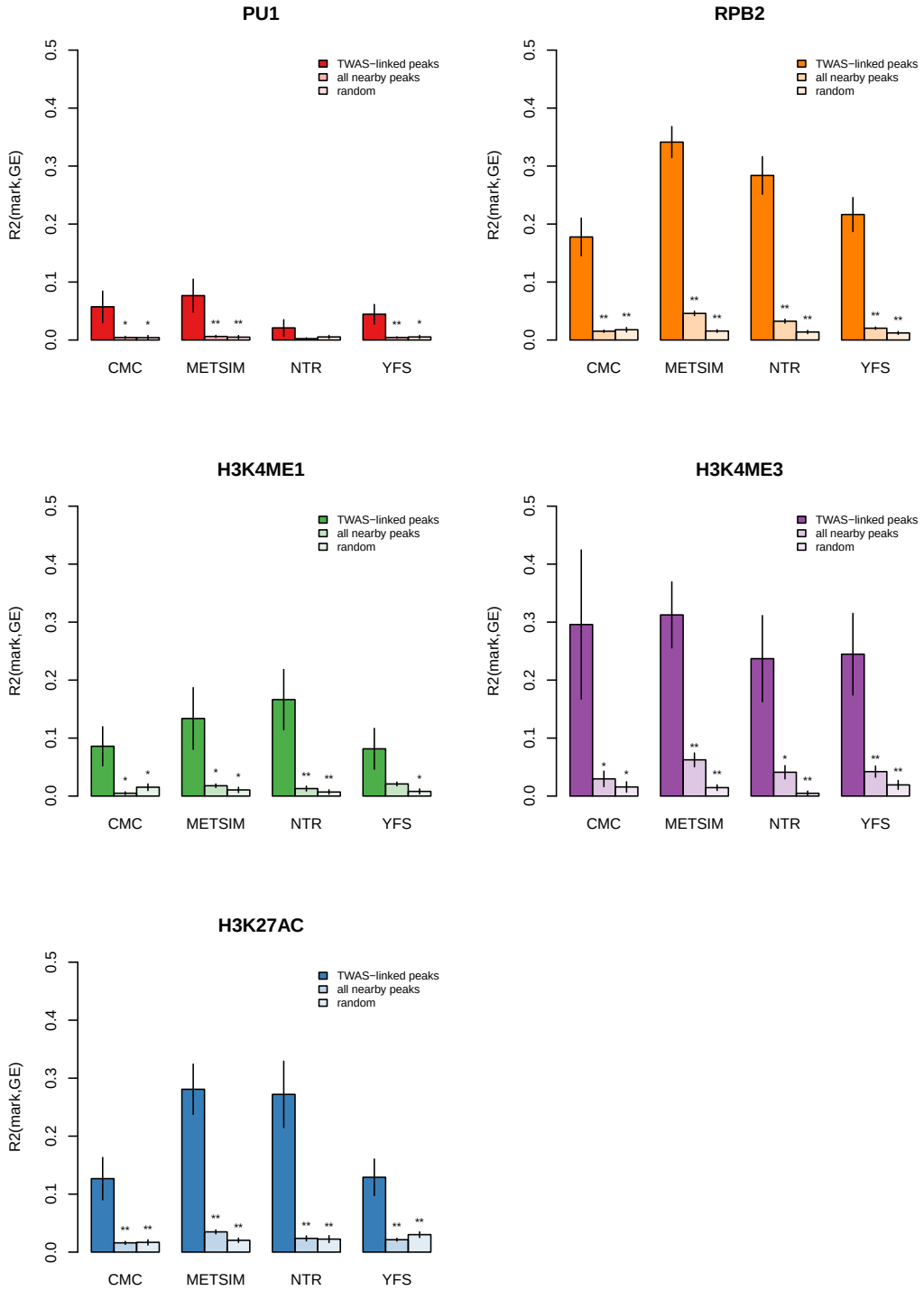
Supplementary Figure 26: **Variance in RNA expression explained by chromatin TWAS associations (±500kb) in YRI.** For every significant chromatin TWAS associations, the adjusted R^2 between the peak and corresponding measured expression levels is reported, with dark bars reporting the mean by chromatin category and reference panel. The same quantity was computed for all peaks ±500kb to the significant TWAS genes ("all nearby peaks" bar); as well as for 10 samplings of any random peak-gene combination no more than 500kb apart ("random" bar). Error bars indicate standard error of the mean. * (**) indicate instances where the null is significantly lower than TWAS-gene estimate by T-test at $P < 0.05$ ($P < 0.05/4$). See Supplementary Figure 25 for details.



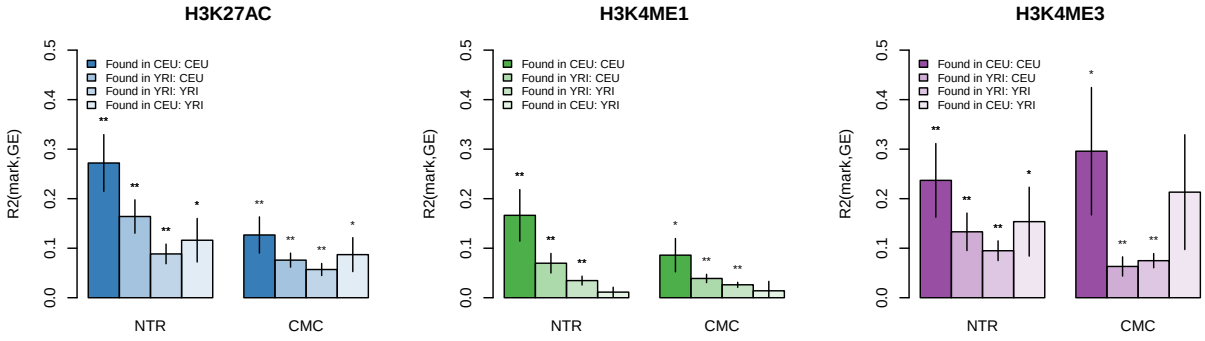
Supplementary Figure 27: **Variance in RNA expression explained by chromatin TWAS associations (±2kb) in CEU.** For every significant chromatin TWAS associations, the adjusted R^2 between the peak and corresponding measured expression levels is reported, with dark bars reporting the mean by chromatin category and reference panel. The same quantity was computed for all peaks ±500kb to the significant TWAS genes ("all nearby peaks" bar); as well as for 10 samplings of any random peak-gene combination no more than 500kb apart ("random" bar). Error bars indicate standard error of the mean. * (**) indicate instances where the null is significantly lower than TWAS-gene estimate by T-test at $P < 0.05$ ($P < 0.05/4$). See Supplementary Figure 25 for details.



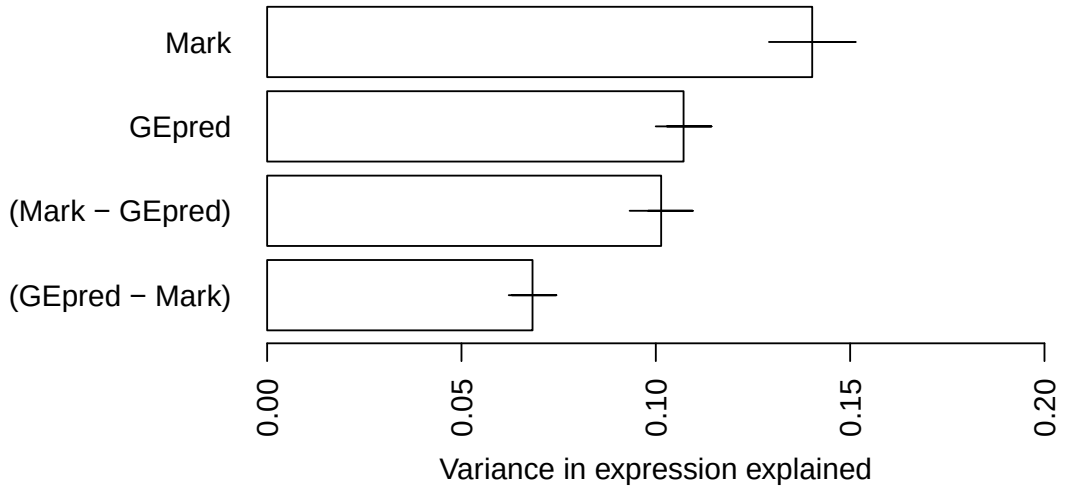
Supplementary Figure 28: **Variance in RNA expression explained by chromatin TWAS associations (±500kb) in CEU.** For every significant chromatin TWAS associations, the adjusted R^2 between the peak and corresponding measured expression levels is reported, with dark bars reporting the mean by chromatin category and reference panel. The same quantity was computed for all peaks ±500kb to the significant TWAS genes ("all nearby peaks" bar); as well as for 10 samplings of any random peak-gene combination no more than 500kb apart ("random" bar). Error bars indicate standard error of the mean. * (**) indicate instances where the null is significantly lower than TWAS-gene estimate by T-test at $P < 0.05$ ($P < 0.05/4$). See Supplementary Figure 25 for details.



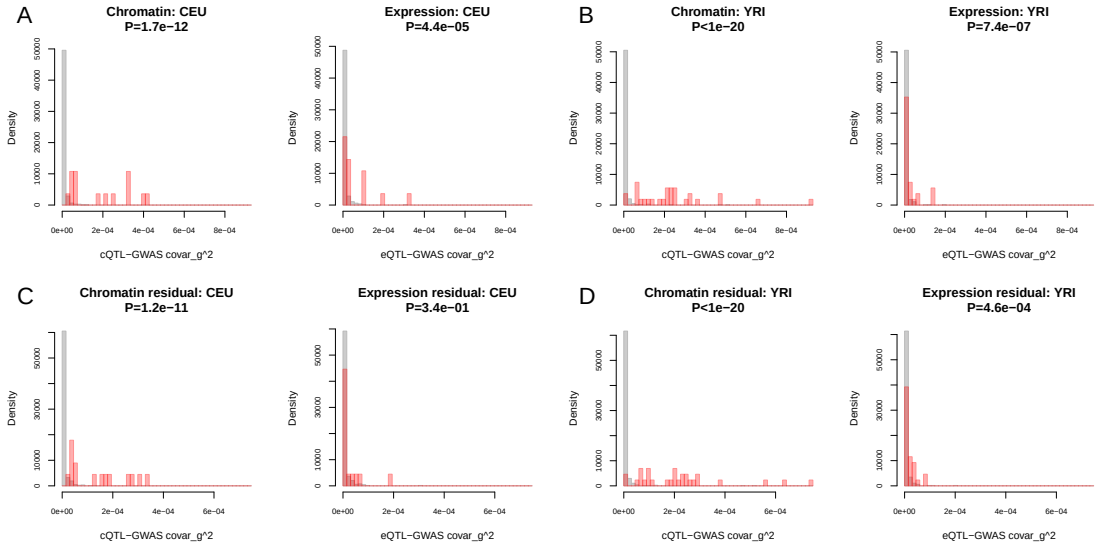
Supplementary Figure 29: **Variance in RNA expression explained by marks (±500kb) across populations.** For every significant chromatin TWAS association, the adjusted R^2 between mark and corresponding RNA-seq levels was measured using associations within and across the CEU/YRI populations. Note that both expression reference panels were of European origin, so only the CEU:CEU case involves no trans-ethnic prediction. * (**) indicate mean is significantly greater than zero by T-test at $P < 0.05$ ($P < 0.05/8$).

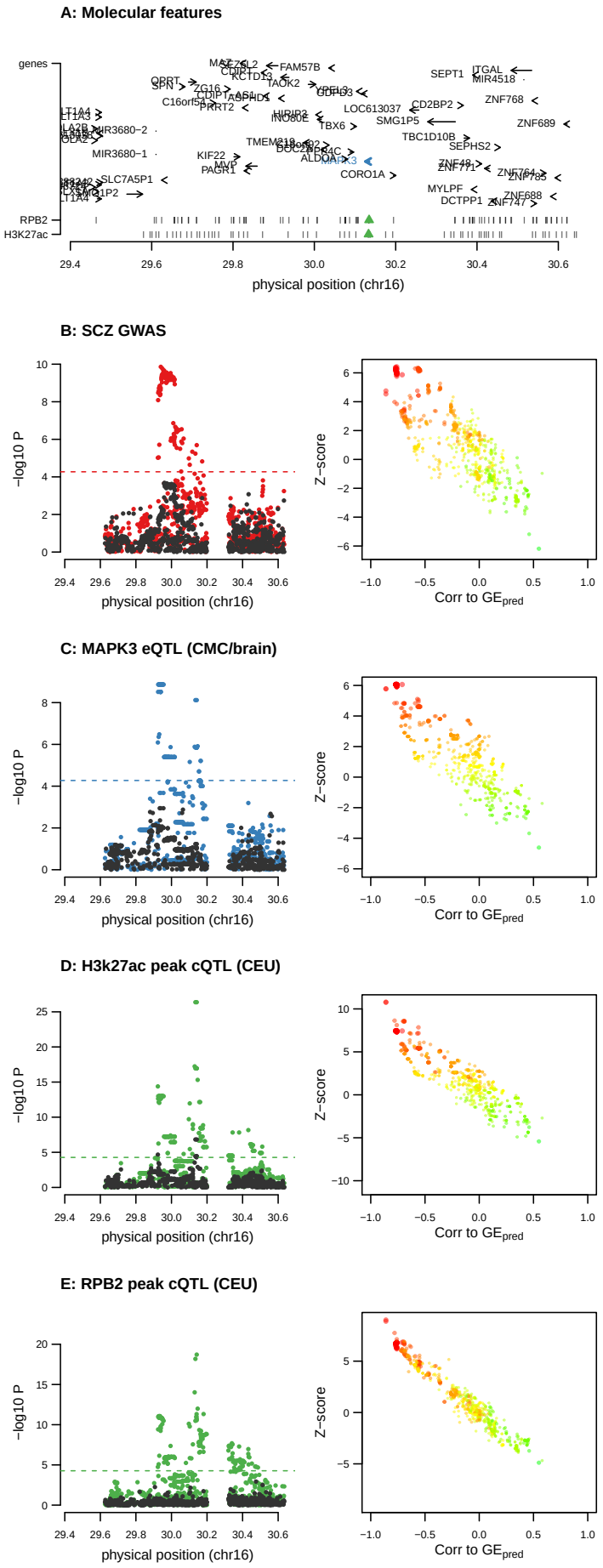


Supplementary Figure 30: **Mediation analysis of expression and chromatin phenotype.** For each chromatin TWAS association identified in YRI, a mediation analysis was performed in CEU for the overlapping peak quantifying variance in RNA-seq expression explained by: (a) the chromatin phenotype; (b) the predicted expression for that gene (from the NTR blood reference); (c) the chromatin phenotype after regressing out the predicted expression; and (d) the predicted expression after regressing out the chromatin phenotype. The mean and standard error of each variable is reported in the barplot in the above order. The mark residual explained more variance in expression than the prediction residual ($P = 0.001$) indicative of the chromatin phenotype mediating between cis-SNPs and expression.



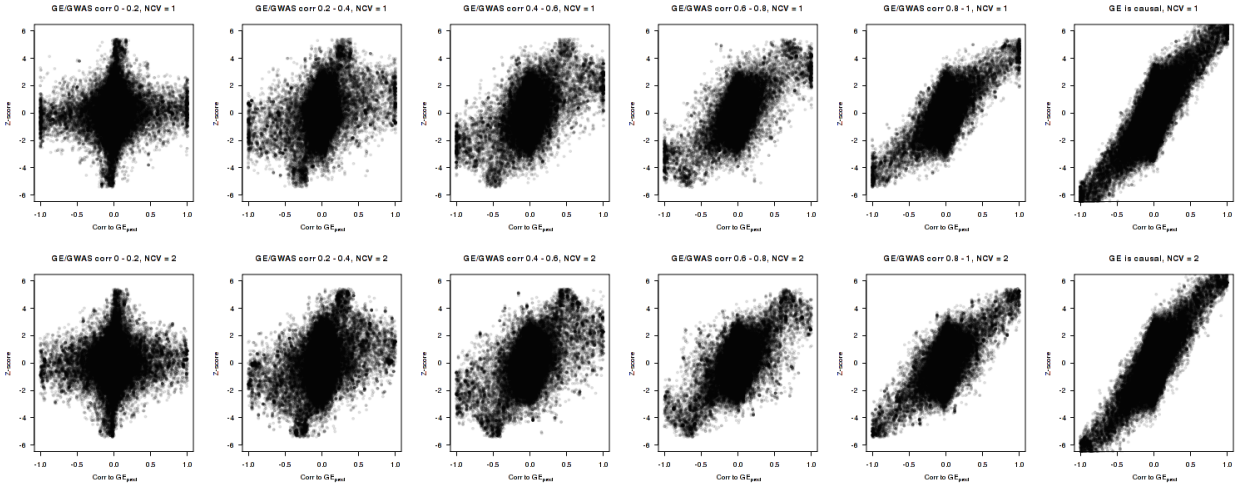
Supplementary Figure 31: **Genetic covariance of chromatin TWAS genes/peaks and SCZ.** For each TWAS gene associated with SCZ and chromatin, the histogram of cis genetic covariance between eQTL/cQTL and GWAS effect-sizes is shown in red. Distribution from randomly selected marks (near heritable TWAS genes) are shown in gray. (**a,b**) Show estimates from CEU and YRI populations. (**c,d**) Show estimates from CEU and YRI after conditioning on the complementary phenotype (i.e. chromatin residualized on expression, and vice-versa). P-value reported for the difference between focal and background distribution by non-parametric Kolomagorov-Smirnov test. Note: because the GWAS was performed in Europeans, the correlation from YRI QTLs will be dampened by cross-population heterogeneity. Averaging across the 42 chromatin TWAS associations at genes identified in the SCZ TWAS, both the expression-SCZ and chromatin-SCZ cov_g were significantly higher than that of a random background of gene-peak pairs less than 500kb apart, with chromatin-SCZ cov_g 2.5× greater on average than expression-SCZ cov_g and more significantly different from the background. This corresponds to the chromatin phenotype explaining 35% of the variance in expression under the model where it is the mediator, with the rest due to environmental or trans variance independent of disease.



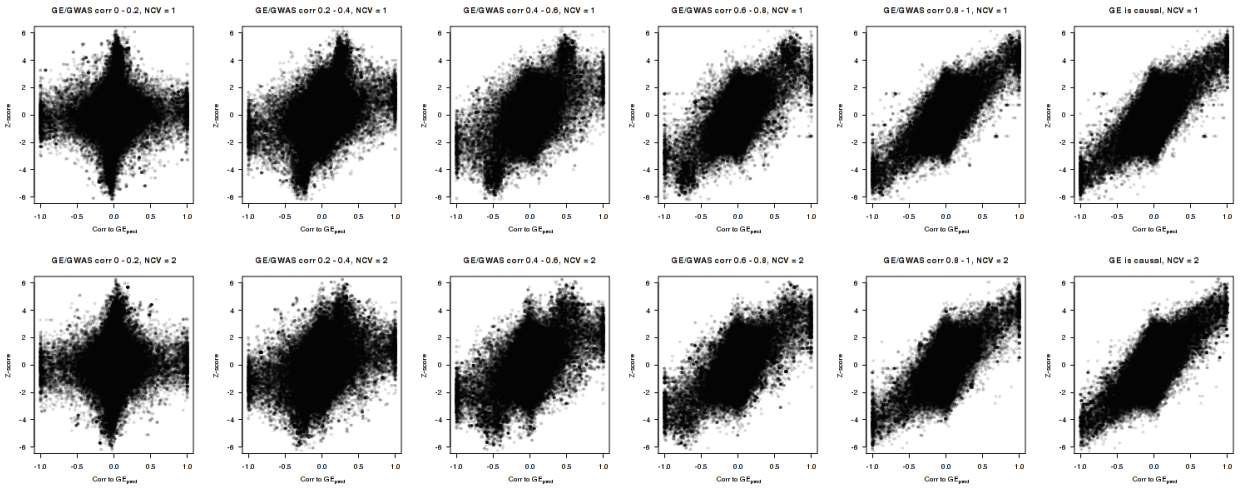


Supplementary Figure 32: **Chromatin and SCZ TWAS association at MAPK3.** Example association of *MAPK3* gene expression and SCZ and nearby chromatin peak. (**A**) locus schematic with all nearby genes and chromatin peaks; TWAS associated features highlighted in blue and green. (**B-D left**) Manhattan plots of marginal association statistics before and after conditioning on the TWAS predicted expression (colored/dark dots, respectively). Dashed line shows local significance threshold after Bonferroni correction for number of SNPs. (**B-E right**) Relationship between marginal GWAS/QTL association (y axis) and the correlation (x-axis) between TWAS predicted expression (GE_{pred} estimated in the 1000 Genomes reference) and marginal GWAS/QTL association. The color of each point reflects the eQTL effect size of the expression used for GE_{pred} and size of each point reflects absolute significance of the eQTL. (**B**) SCZ GWAS association; (**C**) MAPK3 expression phenotype used for TWAS prediction and associated with SCZ/chromatin; (**D**) TWAS associated H3k27ac2 peak in CEU; (**E**) TWAS associated RPB2 peak in CEU. See Supplementary Figure 33, S34 for additional examples and simulations.

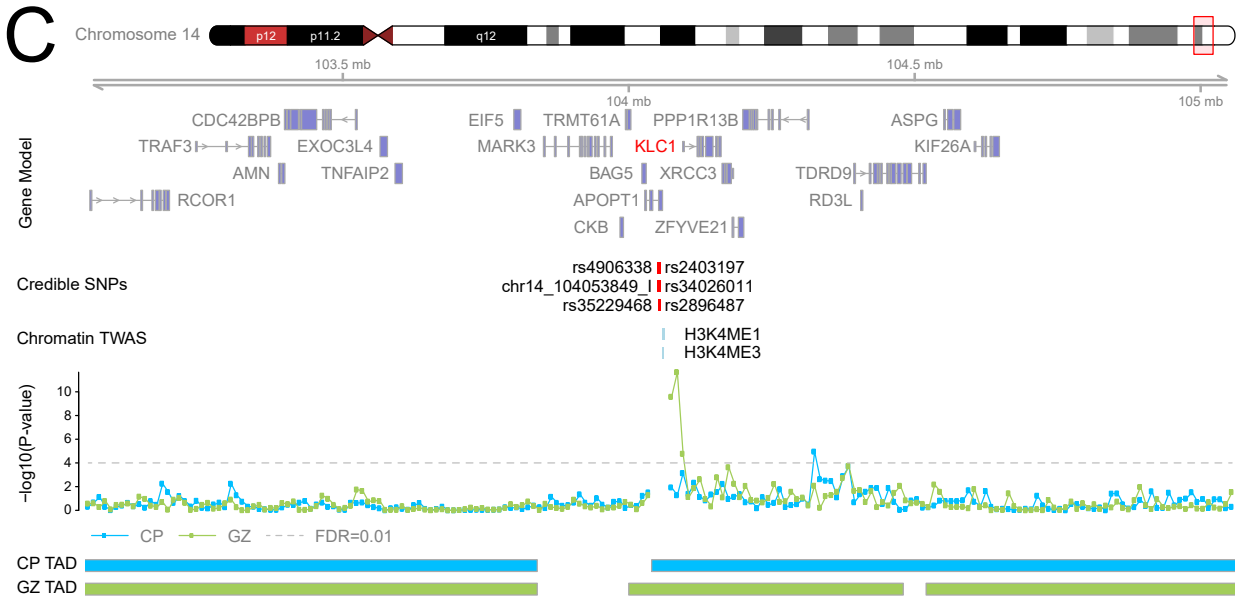
Supplementary Figure 33: **TWAS scatterplot for simulated GWAS loci.** Relationship between marginal GWAS association (y axis) and the correlation (x axis) between TWAS predicted expression ($\mathrm{GE}_{\mathrm{pred}}$) and marginal GWAS association shown for simulated GWAS loci. An expression phenotype was simulated from 900 individuals to have either a single common causal variant (top panels) or two common causal variants (bottom panels), and the genetic value of the phenotype inferred. The GWAS causal variant was then chosen to be: (right-most panels) the expression phenotype, corresponding to a model where expression is causal for trait; (left-most panels) a random common variant uncorrelated with expression, corresponding to a model where eQTL and GWAS effects in the locus are independent; (middle panels) increasingly more correlated with expression. GWAS association statistics were simulated for a study size of $N = 90,000$ and only loci which achieved genome-wide significance were retained. A total of 500 simulations for each architecture were performed and results in the format of Main Figure 4 and 5 overplotted as semi-transparent points. As the correlation between expression and the GWAS causal variant increases, the linear $y \sim x$ relationship emerges, but only becomes completely linear when correlation is > 0.8 (right two panels).



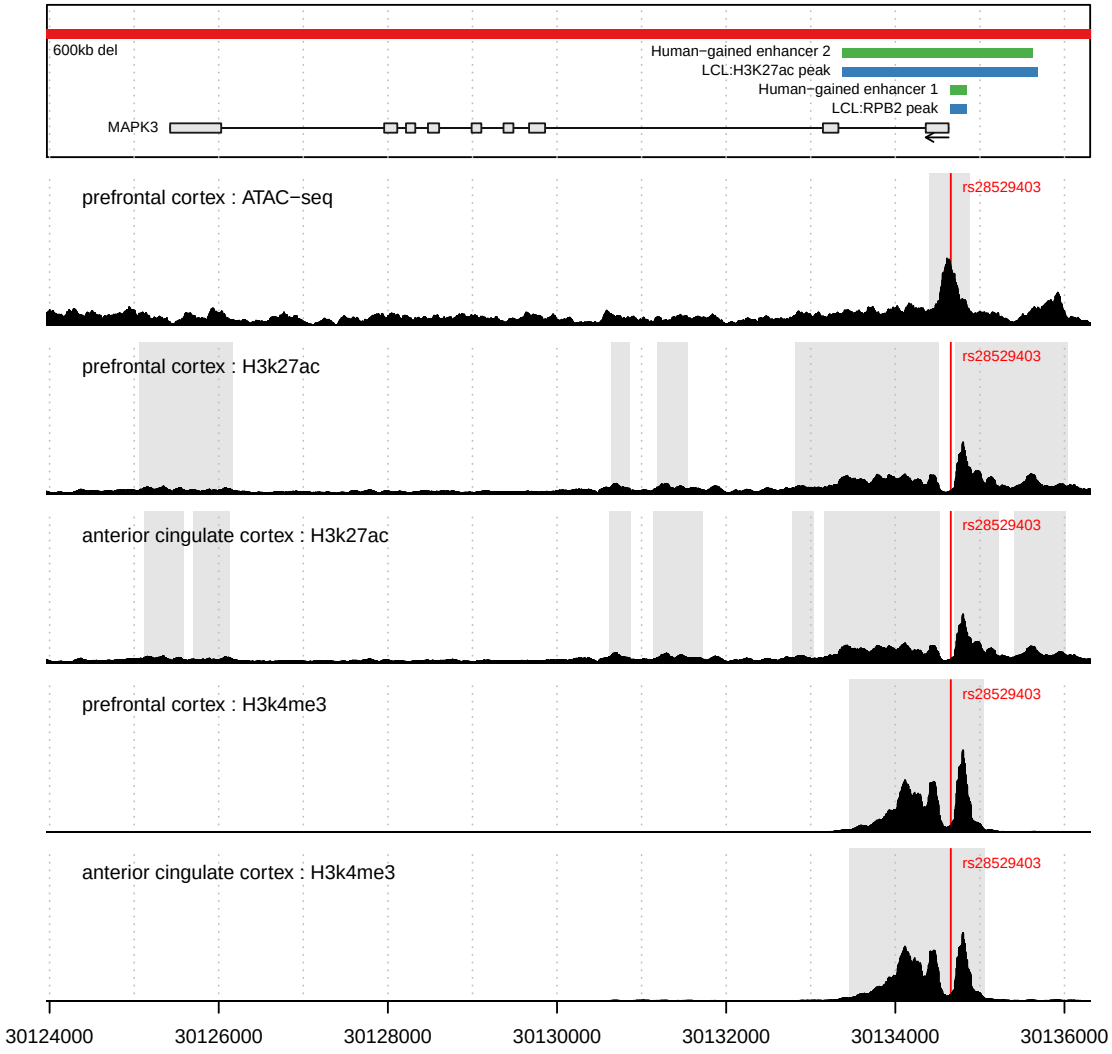
Supplementary Figure 34: **TWAS scatterplot for simulated chromatin association.** Relationship between marginal cQTL association (y axis) and the correlation (x axis) between TWAS predicted expression ($\mathrm{GE}_{\mathrm{pred}}$) and marginal cQTL association shown for simulated GWAS loci. An expression phenotype was simulated from 900 individuals to have either a single common causal variant (top panels) or two common causal variants (bottom panels), and the genetic value of the phenotype inferred. The cQTL causal variant was then chosen to be: (right-most panels) the expression phenotype, corresponding to a model where expression is causal for trait; (left-most panels) a random common variant uncorrelated with expression, corresponding to a model where eQTL and cQTL effects in the locus are independent; (middle panels) increasingly more correlated with expression. cQTL association statistics were simulated for a study size of $N = 70$ and only loci which achieved genome-wide significance were retained. A total of 500 simulations for each architecture were performed and results in the format of Main Figure 4 and 5 over-plotted as semi-transparent points. As the correlation between expression and the cQTL causal variant increases, the linear $y \sim x$ relationship emerges, but only becomes completely linear when correlation is > 0.8 (right two panels). Note that all distributions are significantly noisier than Supplementary Figure 33 due to much smaller sample size.



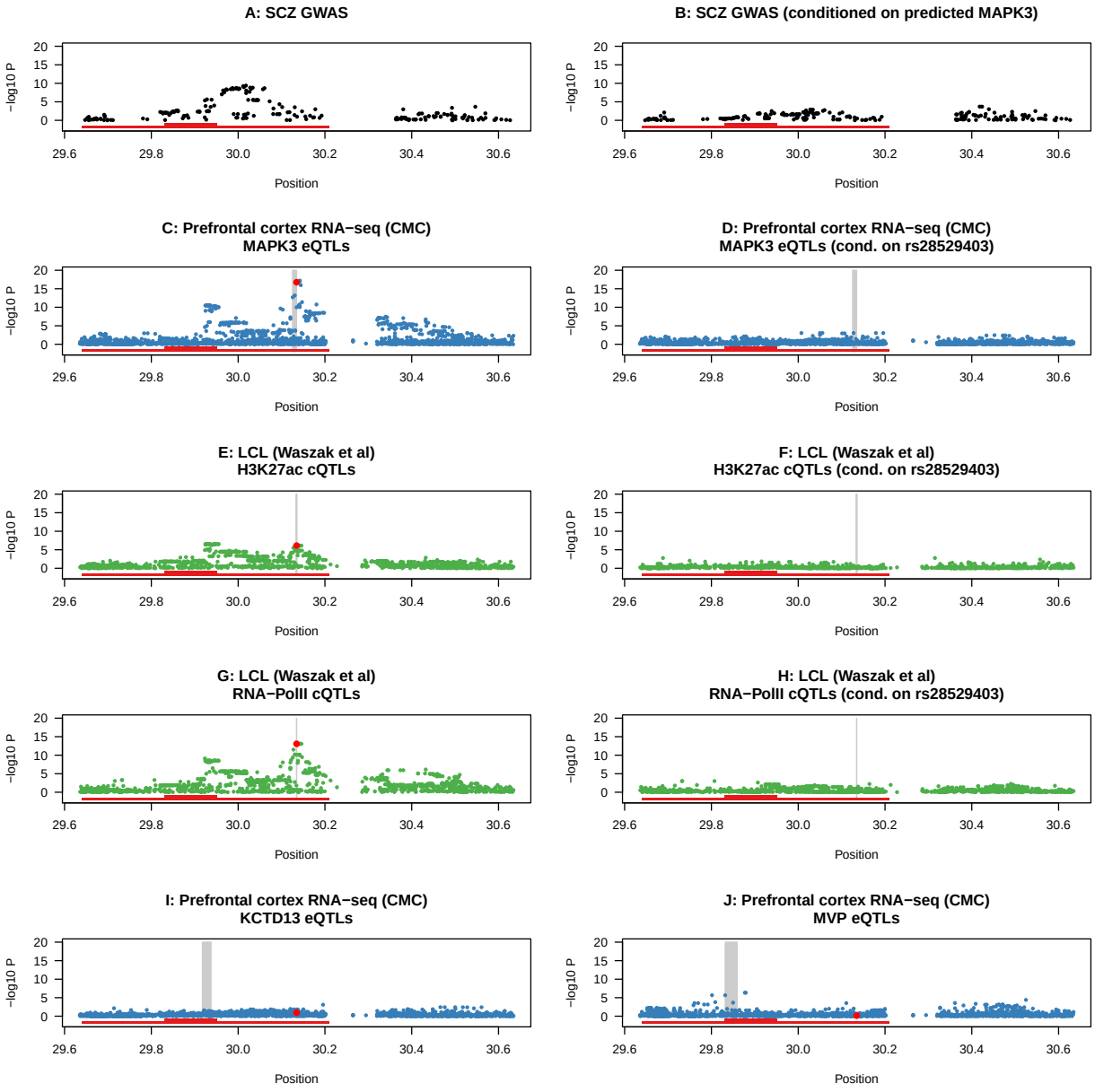
Supplementary Figure 35: **Hi-C interactions between SCZ finemappped SNPs and KLC1 promoter.** (Gene model) Shows KLC1 locus and known genes. (Credible SNPs) Shows positions of fine-mapped 95% credible set of SCZ GWAS markers at this locus. (Chromatin TWAS) Shows position of two chromatin peaks and (below) the significance of Hi-C chromatin interaction between 10kb region containing the chromatin TWAS peaks and each other 10kb region in the locus.



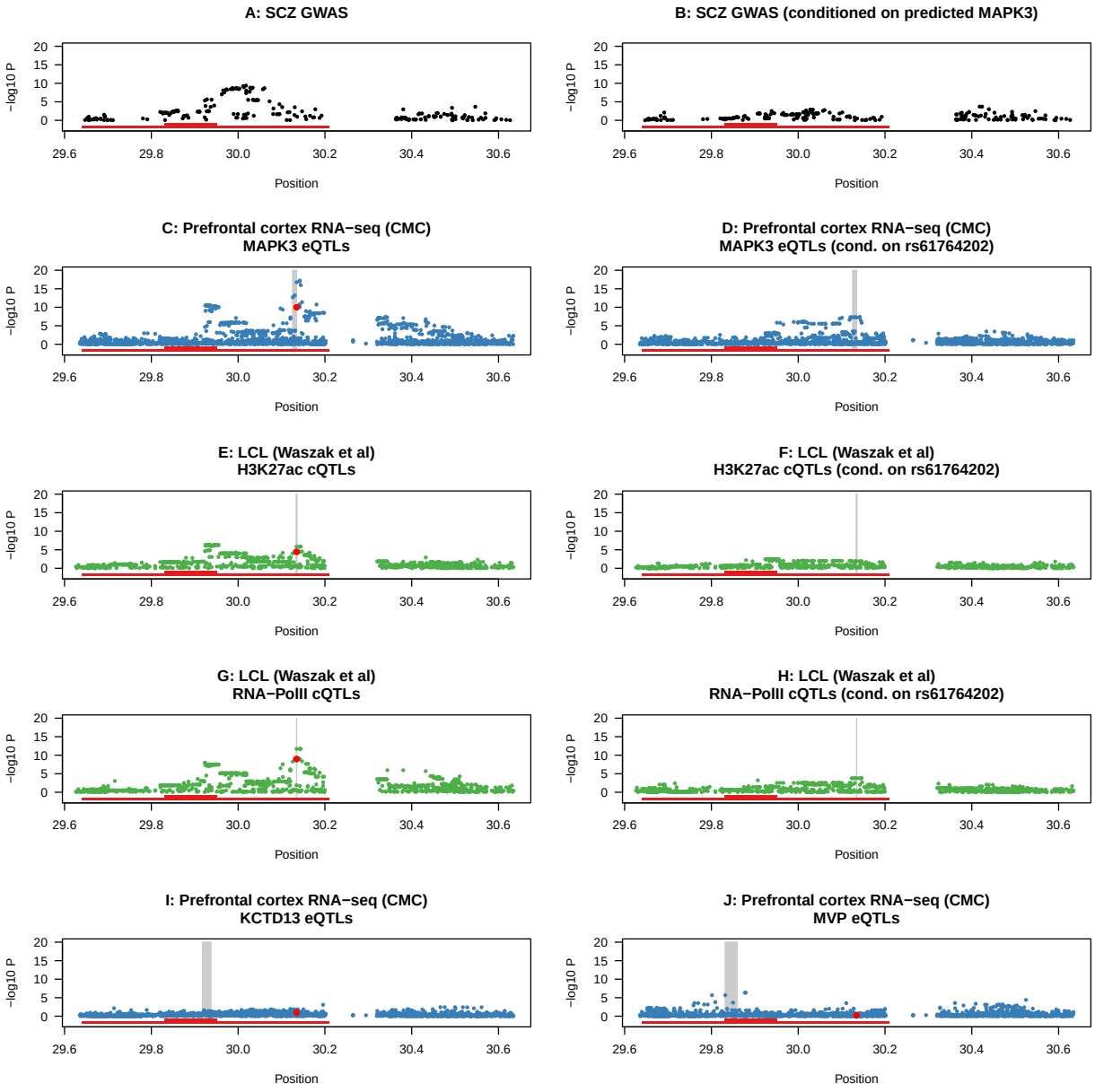
Supplementary Figure 36: **Locus schematic and coverage plot of MAPK3 epigenetic features in brain.** Top panel shows relevant features in the chromosome 16 MAPK3 locus: exons shown in gray; chromatin TWAS associated peaks shown in blue; human-gain developmental enhancers identified by ref. [23] shown in green (observed in H3k4me2 from 8.5pcw and 120pcw fetal tissue); and rare CNVs implicated in SCZ and ASD [26] shown in red. Bottom panels show coverage plots from ATAC-seq; as well as ChIP-seq analysis of H3k27c and H3k4me3 in two cortical tissues; with MaCS peak calls shaded in gray.



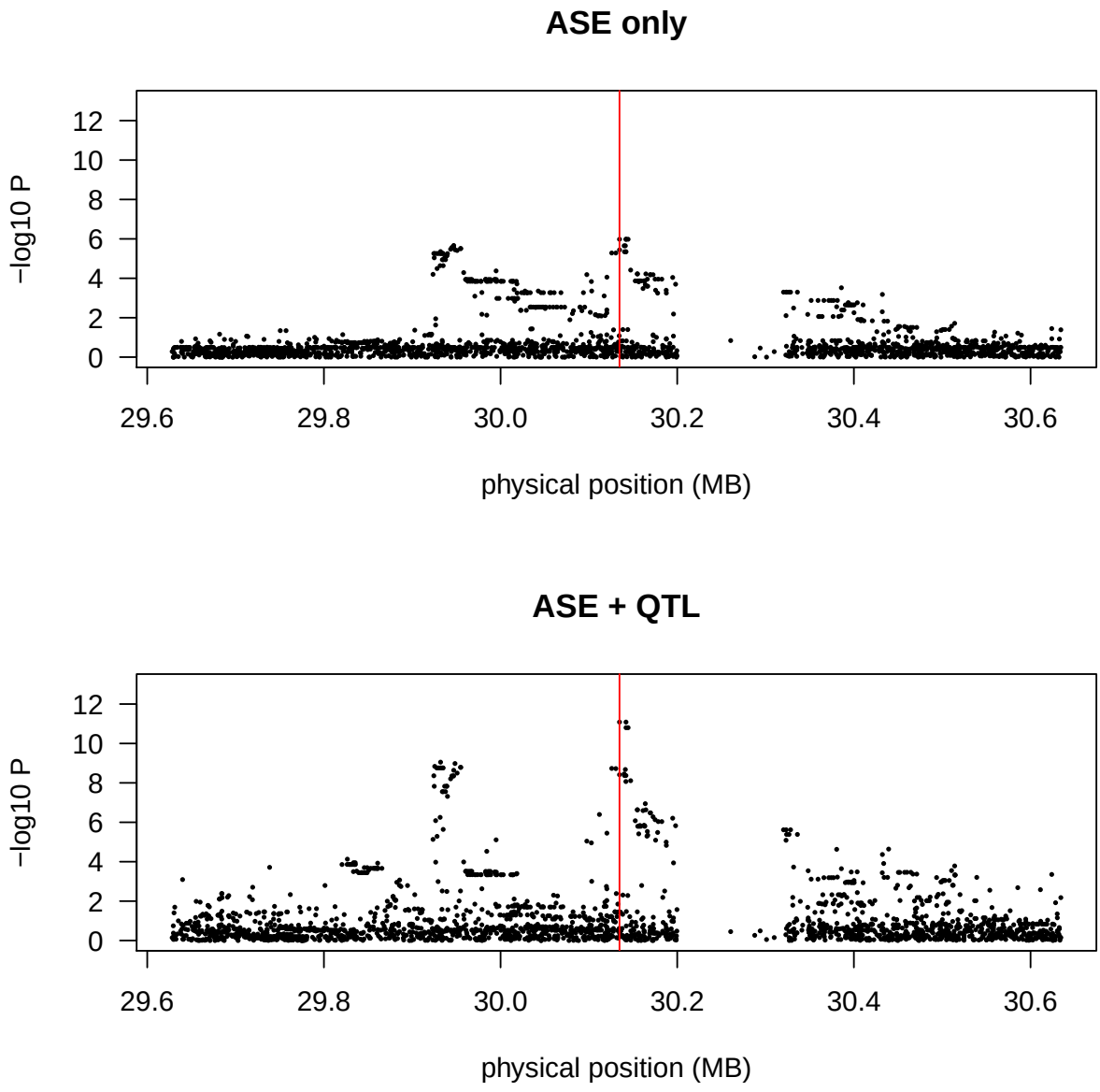
Supplementary Figure 37: **Conditional analysis at MAPK3 locus (rs28529403).** Marginal QTL associations before (left) and after (right) conditioning are shown for all molecular phenotypes associated at MAPK3. Nearby genes KCTD13 and MVP were previously implicated in neuroanatomical phenotypes [26,27] and corresponding eQTLs are shown here for reference, though no evidence of significant association was observed. Rare CNVs implicated in SCZ and ASD [26] indicated in red lines below points.



Supplementary Figure 38: **Conditional analysis at MAPK3 locus (rs61764202).** Marginal QTL associations before (left) and after (right) conditioning are shown for all molecular phenotypes associated at MAPK3. Nearby genes KCTD13 and MVP were previously implicated in neuroanatomical phenotypes [26,27] and corresponding eQTLs are shown here for reference, though no evidence of significant association was observed. Rare CNVs implicated in SCZ and ASD [26] indicated in red lines below points.



Supplementary Figure 39: **Allele-specific expression at MAPK3 locus.** Manhattan plots from tests for allele-specific haplotype expression of CMC RNA-seq at the MAPK3 locus. Top panel shows results from an allele-specific (heterozygous) only test, bottom panel shows results from combined test that uses heterozygous and homozygous sites. The center of the TWAS associated chromatin peaks shown with read line.



Bibliography

[1] Li, Y., van+de+Geijn, B., Raj, A., Knowles, D., Petti, A., Golan, D., Gilad, Y., and Pritchard, J. (2016). Rna splicing is a primary link between genetic variation and disease. Science *352*, 600–604.

[2] Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

[3] Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B., Jansen, R., de Geus, E., Boomsma, D., Wright, F., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics *48*, 245–252.

[4] Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide snps. Nature Genetics *45*, 984–994.

[5] Palla, L. and Dudbridge, F. (2015). A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. The American Journal of Human Genetics *97*, 250–259.

[6] Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS one *3*, e3395.

[7] Yang, J., Weedon, M. N., Purcell, S., Lettre, G., Estrada, K., Willer, C. J., Smith, A. V., Ingelsson, E., O'Connell, J. R., Mangino, M., et al. (2011). Genomic inflation factors under polygenic inheritance. European Journal of Human Genetics *19*, 807–812.

[8] Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. The American Journal of Human Genetics *99*, 139–153.

[9] Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjálmsson, B. J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., et al. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. The American Journal of Human Genetics *95*, 535–552.

[10] Haseman, J. and Elston, R. (1972). The investigation of linkage between a quantitative trait and a marker locus. Behavior Genetics *2*, 3–19.

[11] Consortium, U. et al. (2015). The uk10k project identifies rare variants in health and disease. Nature *526*, 82–90.

[12] Grubert, F., Zaugg, J., Kasowski, M., Ursu, O., Spacek, D., Martin, A., Greenside, P., Srivas, R., Phanstiel, D., Pekowska, A., et al. (2015). Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell *162*, 1051–1065.

[13] Waszak, S., Delaneau, O., Gschwind, A., Kilpinen, H., Raghav, S., Witwicki, R., Orioli, A., Wiederkehr, M., Panousis, N., Yurovsky, A., et al. (2015). Population variation and genetic control of modular chromatin architecture in humans. Cell *162*, 1039–1050.

[14] Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M., Powell, J., Montgomery, G., Goddard, M., Wray, N., Visscher, P., et al. (2016). Integration of summary data from gwas and eqtl studies predicts complex trait gene targets. Nature Genetics.

[15] Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Duncan, L., Perry, J. R., Patterson, N., Robinson, E. B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nature Genetics.

[16] Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., et al. (2010). Common snps explain a large proportion of the heritability for human height. Nature Genetics *42*, 565–569.

[17] Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J., de Candia, T. R., Lee, S. H., Wray, N. R., Kendler, K. S., et al. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nature Genetics.

[18] Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. PLoS Genetics *9*, e1003264.

[19] van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J. K. (2015). Wasp: allele-specific software for robust molecular quantitative trait locus discovery. Nat Meth *12*, 1061–1063.

[20] Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., et al. (2016). Reference-based phasing using the haplotype reference consortium panel. bioRxiv pp. 052308.

[21] Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. Nature methods *9*, 357–359.

[22] Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008). Model-based analysis of chip-seq (macs). Genome biology *9*, 1.

[23] Reilly, S. K., Yin, J., Ayoub, A. E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J. P. (2015). Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science *347*, 1155–1159.

[24] Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., et al. (2012). Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. Nature Genetics *44*, 369–375.

[25] Liu, X., Finucane, H. K., Gusev, A., Bhatia, G., Gazal, S., O'Connor, L., Bulik-Sullivan, B., Wright, F., Sullivan, P., Neale, B., et al. (2016). Functional partitioning of local and distal gene expression regulation in multiple human tissues. bioRxiv.

[26] Migliavacca, E., Golzio, C., Männik, K., Blumenthal, I., Oh, E. C., Harewood, L., Kosmicki, J. A., Loviglio, M. N., Giannuzzi, G., Hippolyte, L., et al. (2015). A potential contributory role for ciliary dysfunction in the 16p11. 2 600 kb bp4-bp5 pathology. The American Journal of Human Genetics *96*, 784–796.

[27] Golzio, C., Willer, J., Talkowski, M. E., Oh, E. C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J. F., et al. (2012). Kctd13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11. 2 copy number variant. Nature *485*, 363–367.