Multi-omic strategies for transcriptome-wide prediction and association studies

Arjun Bhattacharya[1], Michael I. Love[1,2,*]

Abstract

Traditional predictive models for transcriptome-wide association studies (TWAS) consider only single nucleotide polymorphisms (SNPs) local to genes of interest and perform parameter shrinkage with a regularization process. These approaches ignore the effect of distal-SNPs or possible effects underlying the SNP-gene association. Here, we outline multi-omic strategies for transcriptome imputation from germline genetics for testing gene-trait associations by prioritizing distal-SNPs to the gene of interest. In one extension, we identify mediating biomarkers (CpG sites, microRNAs, and transcription factors) highly associated with gene expression and train predictive models for these mediators using their local SNPs. Imputed values for mediators are then incorporated into the final model as fixed effects with local SNPs to the gene included as regularized effects. In the second extension, we assess distal-eSNPs (SNPs in eQTLs) for their mediation effect through mediators local to these distal-eSNPs. Highly mediated distal-eSNPs are then included in the eventual transcriptomic prediction model. We show considerable gains in percent variance explained of gene expression and TWAS power to detect gene-trait associations using simulation analysis and real data applications with TCGA breast cancer data and in ROS/MAP brain tissue data. This integrative approach to transcriptome-wide imputation and association studies aids in understanding the complex interactions underlying genetic regulation within a tissue and identifying important risk genes for various traits and disorders.

Keywords: TWAS, GWAS, eQTL, mediation analysis

1. Introduction

Genomic methods that borrow information from multiple data sources, or "omics" assays, offer advantages in interpretability, statistical efficiency, and opportunities to understand the flow of information in disease regulation [1, 2]. Transcriptome-wide association studies (TWAS) aggregate genetic information into functionally-relevant testing units that map to genes and their expression in a relevant tissue. This gene-based approach combines the effects of many regulatory variants into a single testing unit that increases study power and more interpretable trait-associated genomic loci [3, 4]. However, traditional TWAS methods like

*Corresponding author
Email address: milove@email.unc.edu (Michael I. Love)
[1]Department of Biostatistics, University of North Carolina at Chapel Hill
[2]Department of Genetics, University of North Carolina at Chapel Hill

PrediXcan and FUSION focus on local genetic regulation of transcription. These methods ignore significant portions of heritable expression that can be attributed to distal genetic variants which may indicate complex mechanisms that contribute to gene regulation.

Recent work in transcriptional regulation has estimated that distal genetic variants can account for up to 70% of the variance in gene expression [5, 6]. This result accords with Boyle *et al.*'s omnigenic model, proposing that regulatory networks are so interconnected that a majority of genetic variants in the genome, local or distal, have indirect effects on transcription of any particular gene [7, 6]. Many groups have leveraged this model to identify distal expression quantitative trait loci (eQTLs) by testing the effect of a distal-eSNP on an eGene mediated through a set of genes local to the SNP, concluding that many distal-eQTLs are often eQTLs for many local genes [8, 9, 10, 11, 12, 13]. It has been shown previously that distal-eQTLs found in regulatory hotspots are generally cell-type specific [8, 12, 14]. Deep learning methods have employed similar logic to link GWAS-identified variants to nearby regulatory mechanisms for functional hypothesis generation [15]. Furthermore, Zhang *et al.*'s recent EpiXcan method demonstrates that incorporating epigenetic information into transcriptomic prediction models generally improves predictive performance and power in detecting gene-trait associations in TWAS [16].

Here, we outline two extensions to TWAS, borrowing information from other omics assays to enrich or prioritize mediator relationships of eQTLs in expression models. Using simulations and data from the The Cancer Genome Atlas (TCGA) [17] and the Religious Orders Study and the Rush Memory and Aging Project (ROS/MAP) [18], we show improvements in transcriptomic prediction and power to detect gene-trait associations. These **M**ulti-**O**mic **S**trategies for **T**ranscriptome-**W**ide **A**ssociation **S**tudies are curated in the R package MOSTWAS, available freely at `www.github.com/bhattacharya-a-bt/MOSTWAS`.

2. Material and Methods

We first outline the two methods proposed in this work, (1) mediator-enriched transcriptome-wide prediction (MeTWAS) and (2) distal-eQTL prioritization via mediation analysis (DePMA). MeTWAS and DePMA are combined in the MOSTWAS R package, available freely at `https://bhattacharya-a-bt.github.io/MOSTWAS`.

2.1. Mediator-enriched TWAS (MeTWAS)

2.1.1. Transcriptomic prediction using MeTWAS

Here, we present mediator-enriched TWAS, or MeTWAS one of the two tools presented in the MOSTWAS R package. Across n samples, consider the vector Y_G, the expression of a gene G of interest, the matrix $\mathbf{X}_G$ of local SNPs dosages in a user-defined window around gene G (default to 0.5 Megabases), and m_G mediating biomarkers that we estimate to be significantly associated with the expression of gene G via a relevant one-way test of association. These mediating biomarkers can be DNA methylation sites, microRNAs, transcription factors, or any molecular profile that may be genetically heritable and affect transcription. Accordingly, let

the matrix $\mathbf{X}_{M_j}$ be the local-genotype dosages in a 500 kilobase window around mediator $j,\ 1 \leq j \leq m_G$. Furthermore, let M_j be the intensity of mediator j (methylation M-value if j is a CpG site, expression if j is a miRNA or a gene, etc). Prior to any modelling, we scale Y_G and all $M_j,\ 1 \leq j \leq m_G$ to zero mean and unit variance. We also residualize $M_j,\ 1 \leq j \leq m_G$ and Y_G with the covariate matrix $\mathbf{X}_C$ to account for population stratification using principal components of the global genotype matrix and relevant clinical covariates to obtain $\tilde{M}_j,\ 1 \leq j \leq m_G$ and $\tilde{Y}_G$.

Transcriptome prediction in MeTWAS draws from two-step regression, as summarized in **Figure 1**. First, in the training set for a given training-test split, for $1 \leq j \leq m_G$, we model the residualized intensity $\tilde{M}_j$ of training-set specific mediator j with the following additive model:

$$\tilde{M}_j = \mathbf{X}_{M_j,train} w_j + \varepsilon_m, \tag{1}$$

where w_j is the effect-sizes of the SNPs in $\mathbf{X}_{M_j,train}$ on $\tilde{M}_j$ in the training set. As in traditional transcriptomic imputation models [3, 4], we find $\hat{w}_j$ using one of the two following methods with the largest predicted adjusted R^2: (1) elastic net regression with mixing parameter $\alpha = 0.5$ and λ tuned over 5-fold cross validation using glmnet [19], or (2) linear mixed modelling assuming random effects for $\mathbf{X}_{M_j}$ using rrBLUP [20]. Only significantly heritable (default $P < 0.05$ for the likelihood ratio test) [21] and well-cross validated (default $R^2 \geq 0.01$) expression models are considered.

For all j, using these optimized predictive models for M_j as denoted by $\hat{w}_{M_j}$, we estimate the genetically regulated intensity (GRIn) of the mediator m_j, denoted GRIn_{m_j}, in the test set. Denote $\hat{\mathbf{M}}$ as the $n \times m$ matrix of estimated GRIn, such that the jth column of $\hat{\mathbf{M}}_j$ is GRIn_{m_j} across all n samples.

Next, we consider the following additive model for the residualized expression of gene G:

$$\begin{aligned} \tilde{Y}_G &= \mathbf{M}\beta_{\mathbf{M}} + \varepsilon_{Y_{G_1}}, \\ \tilde{Y}_G - \mathbf{M}\hat{\beta}_{\mathbf{M}} &= \mathbf{X}_G \beta_M + \varepsilon_{Y_{G_2}}. \end{aligned} \tag{2}$$

where β_M is the fixed effect-sizes of GRIn_{m_j} on $\tilde{Y}_G$, $\hat{\mathbf{M}}$ is the matrix of estimated GRIn for all m_j mediators, $\mathbf{X}_G$ are the local-SNPs to gene G, and w_G are the "random" or regularized effect sizes of the local-SNPs. We first estimate $\hat{\beta}_M$ by traditional ordinary least squares, where $\hat{\beta}_M = \left(\hat{\mathbf{M}}'\hat{\mathbf{M}}\right)^{-1} \hat{\mathbf{M}}' E_g$. Next, using one of the methods outlined above when estimating $\hat{w}_{M_j}$, we can estimate the effect sizes $\hat{w}_G$ of the local-SNPs on $\tilde{Y}_G$, residualized with $\hat{\mathbf{M}}$, using either elastic net or linear mixed models [19, 20].

2.1.2. Transcriptomic imputation with MeTWAS

In an external GWAS panel, if individual SNPs are available, we construct the mediator-enriched genetically regulated expression (MeGReX) of gene G directly using $\hat{w}_G$ and $(\hat{w}_j, \hat{\beta}_j),\ 1 \leq j \leq m_G$:

$$MeGReX_G = \sum_{j=two1}^{m_G} \mathbf{X}_{M_j,GWAS} \hat{w}_{M_j} \hat{\beta}_{M,j} + \mathbf{X}_{G,GWAS} \hat{w}_G,$$

where $\mathbf{X}_{M_j,GWAS}$ and $\mathbf{X}_{G,GWAS}$ are the SNPs in the GWAS panel local to mediator j and gene G, respectively. $MeGREX_G$ can be used in downstream tests of association.

2.2. Distal-eQTL prioritization via mediation analysis (DePMA)

2.2.1. Transcriptomic prediction using DePMA

Expression prediction in DePMA hinges on assessing distal-eSNPs for inclusion in the design matrix via mediation analysis, adopting methods from previous studies [10, 13, 11]. We first split data for gene expression, SNP dosages, and any potential mediators into k training-testing splits. Based on the minor allele frequencies of SNPs and sample size, we recommend a low number of splits (i.e. $k \leq 5$).

In the training set, we identify mediation test triplets that consist of (1) a gene of interest G with expression Y_G (scaled to zero mean and unit variance), (2) a distal-eSNP s in association with G at a user-defined P-value threshold (default of $P = 10^{-6}$) with dosages X_s, and (3) a set of m biomarkers local to s that are associated with s at a user-defined P-value threshold (default of FDR-adjusted $P = 0.05$) with intensities in the m columns of $\mathbf{M}$. The columns of $\mathbf{M}$ are scaled to zero mean and unit variance. Consider the following mediation model:

$$\begin{aligned} Y_G &= X_s\beta_s + \mathbf{M}\beta_{\mathbf{M}} + \mathbf{X}_C\beta_C + \varepsilon_{Y_G} \\ M_j &= X_s\alpha_{M_j} + \mathbf{X}_C\alpha_{C,j} + \varepsilon_{M_j}, \quad 1 \leq j \leq m. \end{aligned} \tag{3}$$

Here, we have $\beta_{\mathbf{M}}$ as the effects of the M mediators local to s on Y_G adjusting for the effects from s and the covariates $\mathbf{X}_C$, and $\alpha_{\mathbf{M}} = (\alpha_{M_1}, \ldots, \alpha_{M_m})'$ as the effects of s on mediators M_j, for $1 \leq j \leq m$. We assume that $\varepsilon_{Y_G} \sim N(0, \sigma^2)$ and $\varepsilon_{\mathbf{M}} \sim \mathbf{N}_m(0, \mathbf{\Sigma}_M)$, where $\mathbf{\Sigma}_M$ may have non-zero off-diagonal elements that represent non-zero covariance between mediator intensities. Further, we assume that ε_{Y_G} and $\varepsilon_{\mathbf{M}}$ are independent. We define the total mediation effect (TME) [22] of SNP s as

$$TME = \alpha_{\mathbf{M}}^T \beta_{\mathbf{M}}.$$

We are interested in SNPs with large absolute TME, which we assess with a two-sided test of $H_0 : TME = 0$. We assess this hypothesis with a permutation test to obtain a permutation P-value, as more direct methods of computing standard errors for the estimated TME are often biased [23, 13], We also provide an option to estimate an asymptotic approximation to the standard error of TME and conduct a Wald-type test for $H_0 : TME = 0$. This asymptotic option is significantly faster at the cost of inflated false positives (see **Supplemental Materials** and **Supplemental Figure S1**). Corresponding to the t testing triplets identified, we obtain vectors of length t of TMEs and P-values for each distal-eSNP to G. We estimate the q-value for each test to adjust for multiple testing [24]. For the predictive model, we select distal SNPs with $TME \neq 0$ at a given q-value threshold ($q < 0.10$ as a default) and include them with all local SNPs in a design matrix. We then find estimated SNP weights using either elastic net or weighted least squared regression.

2.2.2. Transcriptomic imputation with DePMA

In an external GWAS panel, if individual SNPs are available, we construct the genetically regulated expression (GReX) of gene G directly using $\hat{w}_G$ and $\hat{w}_t$:

$$GReX_G = \mathbf{X}_{t,GWAS}\hat{w}_t + \mathbf{X}_{G,GWAS}\hat{w}_G,$$

where $\mathbf{X}_{t,GWAS}$ is the matrix of dosages of the t distal-SNPs and $\mathbf{X}_{G,GWAS}$ is the matrix of dosages of the local SNPs to gene G in the external GWAS panel. $GREX_G$ can be used in downstream tests of association. If individual SNPs are not available, the weighted burden test can be employed using summary statistics with permutation follow-up test [4].

2.3. Tests of associations

If individual SNPs are not available, then the weighted burden Z-test proposed by Gusev *et al.* can be employed [4] using summary statistics. Briefly, we compute

$$\tilde{Z} = \frac{\mathbf{W}Z}{(\mathbf{W}\Sigma_{s,s}\mathbf{W}^T)^{1/2}}. \tag{4}$$

Here, Z is the vector of Z-scores of SNP-trait associations for SNPs used in estimating $\hat{w}_{M_j}$ and $\hat{w}_G$. The matrix $\mathbf{W}$ is defined as $\Sigma_{e,s}\Sigma_{s,s}^{-1}$, the product of the covariance matrix between all SNPs and the expression of gene G and the covariance matrix among all SNPs. These covariance matrices are estimated from the eQTL reference panel used to estimate $\hat{w}_{M_j}$ and $\hat{w}_G$. and $\hat{w}_G$. The test statistic $\tilde{Z}$ can be compared to the standard Normal distribution for inference. We implement a permutation test conditioning on the GWAS effect sizes to assess whether the same distribution of $\hat{w}_G$ effect sizes could yield a significant association by chance [4]. We permute $\hat{w}_G$ 1,000 times without replacement and recompute the weighted burden test statistic to generate a permutation null distribution for $\tilde{Z}$. This permutation test is only conducted for overall associations at a user-defined significance level (default to FDR-adjusted $P < 0.05$).

Lastly, we also implement a test to assess the information added from distal-eSNPs in the weighted burden test beyond what we find from local SNPs. This test is analogous to a group added-last test in regression analysis, applied here to GWAS summary statistics. Let Z_l and Z_d be the vector of Z-scores from GWAS summary statistics from local and distal-SNPs identified by a MOSTWAS model. The local and distal-SNP effects from the MOSTWAS model are represented in $\mathbf{w}_l$ and $\mathbf{w}_d$. Formally, we test whether the weighted Z-score $\tilde{Z}_d \equiv \mathbf{w}_d^T Z_d$ from distal-SNPs is significantly larger than 0 given the observed weighted Z-score from local SNPs $\tilde{Z}_l \equiv \mathbf{w}_l^T Z_l$, drawing from the assumption that $(\tilde{Z}_l, \tilde{Z}_d)$ follow a bivariate Normal distribution. Namely, we conduct a two-sided Wald-type test for the null hypothesis:

$$H_0 : \mathbf{w}_d^T \mathbf{Z}_d | \mathbf{w}_l^T \mathbf{Z}_l = \tilde{Z}_{l,\text{obs}} = 0.$$

Under the null hypothesis, we can derive that $\tilde{Z}_d|\tilde{Z}_l = \tilde{Z}_{l,\text{obs}}$ is normally distributed with mean and variance determined from the observed local $\tilde{Z}_l$-score, the SNP-effect size vectors $\mathbf{w}_l$ and $\mathbf{w}_d$, and components of the linkage disequilibrium as estimated from the reference panel [25]. Full details and derivation for this added-last test are given in **Supplemental Methods**.

2.4. Simulation framework

We first conducted simulations to assess the predictive capability and power to detect gene-trait associations under various phenotype heritability (h_p^2), local heritability of expression ($h_{e,l}^2$), distal heritability of expression ($h_{e,d}^2$), and proportion of causal local ($p_{c,l}$) and distal ($p_{c,e}$) SNPs for MeTWAS and DePMA. We considered two scenarios for each combination of $(h_p^2, h_{e,l}^2, h_{e,d}^2, p_{c,l}, p_{c,e})$: (1) the leveraged association between the distal-SNP and gene of interest exists in both the reference and imputation panel, and (2) the leveraged association between distal-SNP and gene of interest exists in the reference panel but is null in the imputation panel.

Using TCGA data, we extracted 2,592 SNPs local to the gene *ESR1* on Chromosome 6 and 1,431 SNPs local to the gene *FOXA1*. We generated (1) a reference panel with sample size 400 with simulated SNPs, expressions, and one mediator and (2) a GWAS panel of 1,500 samples with simulated SNPs and phenotypes using the following data generating process, modified from Mancuso *et al.*'s framework [26]:

We estimated the linkage disequilibrium LD matrix of the SNPs X_G with n samples and p SNPs, as follows with regularization to ensure LD is positive semi-definite:

$$LD = \frac{1}{n} X_G^T X_G + \frac{1}{10} I_p.$$

We computed the Cholesky decomposition of LD for faster sampling [26]. We simulated SNPs for a 400-sample reference panel $X_{g,ref}$ and 1,500-sample GWAS panel $X_{g,GWAS}$.

We then simulated effect sizes for $p_{c,l}$ of the 2,592 local SNPs $w_{g,l}$ from a standard Normal distribution. We generated locally heritable expression

$$E_{g,l} = X_{G,ref}^T w_{g,l} + \varepsilon_l,$$

with $\varepsilon_l \sim N(0, 1 - h_{e,l}^2)$ and $w_{g,l}$ scaled to ensure the given $h_{e,l}^2$. Similarly, we simulated effect sizes for $p_{c,d}$ of the 1,431 distal-SNPs $w_{g,d}$ and generated the distally heritable intensity of the mediator $M_{g,d}$. We constructed the distally heritable expression $E_{g,d}$ by scaling $M_{g,d}$ by $\beta \sim N(0,1)$ and adding random noise that scales distal heritability to $h_{e,d}^2$. We lastly formed the total expression $E_g = E_{g,l} + E_{g,d}$.

Next, we simulated the phenotype in the GWAS panel using the GREx as estimated from causal eQTLs to match the variance explained due to genetics. We drew a causal effect size from gene expression $\alpha \sim N(0,1)$. We computed the unobserved gene expression in the GWAS panel as

$$E_{g,GWAS}|X_{g,GWAS,local}, X_{g,GWAS,distal} = X_{g,GWAS,local}^T w_{g,l} + X_{g,GWAS,distal}^T w_{g,d}\beta + \varepsilon_{GWAS}.$$

Here, we also considered a "null" case as well, where the simulated distal eQTLs do not contribute to the simulated phenotype in the GWAS panel (i.e. $w_{g,d} = 0$ for all distal-SNPs). GWAS summary statistics were computed in this step for downstream weighted burden testing. We then fitted predictive models using MeTWAS, DePMA, and local-only models (i.e. FUSION [4]), computed the adjusted predictive R^2 in the reference panel, and tested the gene-trait association in the GWAS panel using a weighted burden test. The association study power was defined as the proportion of gene-trait association tests with $P < 2.5 \times 10^{-6}$, the Bonferroni-corrected significance threshold for testing 20,000 independent genes. With these simulated datasets, we also assessed the power of the distal added-last test by computing the proportion of significant distal associations conditional on the local association at FDR-adjusted $P < 0.05$.

2.5. Data acquisition

2.5.1. Multi-omic data from TCGA-BRCA

We retrieved genotype, RNA expression, miRNA expression, and DNA methylation data for breast cancer indications in The Cancer Genome Atlas (TCGA). Birdseed genotype files of 914 subjects were downloaded from the Genome Data Commons (GDC) legacy (GRCh37/hg19) archive. Genotype files were merged into a single binary PLINK file format (BED/FAM/BIM) and imputed using the October 2014 (v.3) release of the 1000 Genomes Project dataset as a reference panel in the standard two-stage imputation approach, using SHAPEIT v2.87 for phasing and IMPUTE v2.3.2 for imputation [27, 28, 29]. We excluded variants (1) with a minor allele frequency of less that 1% based on genotype dosage, (2) that deviated significantly from Hardy-Weinberg equilibrium ($P < 10^{-8}$) using appropriate functions in PLINK v1.90b3 [30, 31], and (3) located on sex chromosomes. Final TCGA genotype data was coded as dosages, with reference and alternative allele coding as in dbSNP.

TCGA level-3 normalized RNA-seq expression data, miRNA-seq expression data, and DNA methylation data collected on Illumina Infinium HumanMethylation450 BeadChip were downloaded from the Broad Institute's GDAC Firehose (2016/1/28 analysis archive) via FireBrowse [32]. We intersected to the subset of samples assayed for genotype (4,564,962 variants), RNA-seq (15,568 genes), miRNA-seq (1,046 miRNAs), and DNA methylation (485,578 CpG sites), resulting in a total of 563 samples. We only considered the autosome in our analyses. We adjusted gene and miRNA expression and DNA methylation by relevant covariates (10 principal components of the genotype matrix, tumor stage at diagnosis, and age).

2.5.2. Multi-omic data from ROS/MAP

We retrieved imputed genotype, RNA expression, miRNA expression, and DNA methylation data from The Religious Orders Study and Memory and Aging Project (ROS/MAP) Study for samples derived from human pre-frontal cortex [33, 34, 35, 36].We excluded variants (1) with a minor allele frequency of less that 1% based on genotype dosage, (2) that deviated significantly from Hardy-Weinberg equilibrium ($P < 10^{-8}$) using appropriate functions in PLINK v1.90b3 [30, 31], and (3) located on sex chromosomes. Final ROS/MAP genotype data was coded as dosages, with reference and alternative allele coding as in dbSNP. We intersected

to the subset of samples assayed for genotype (4,141,537 variants), RNA-seq (15,857 genes), miRNA-seq (247 miRNAs), and DNA methylation (391,626 CpG sites), resulting in a total of 370 samples. Again, we only considered the autosome in our analyses. We adjusted gene and miRNA expression and DNA methylation by relevant covariates (10 principal components of the genotype age at death, sex, and smoking status).

2.5.3. Summary statistics for downstream association studies

We conducted TWAS association tests using relevant GWAS summary statistics for breast cancer-specific survival, risk of late-onset Alzheimer's disease, and risk of major depressive disorder. We downloaded iCOGs GWAS summary statistics for breast cancer-specific survival for women of European ancestry [37]. All studies and funders as listed in Michailidou *et al.* [38, 39] and in Guo *et al.* [37] are acknowledged for their contributions. Furthermore, we downloaded GWAS summary statistics for risk of late-onset Alzheimer's disease from the International Genomics of Alzheimer's Project (IGAP) [40]. We also downloaded GWAS and genome-wide association by proxy (GWAX) summary statistics for risk of major depressive disorder (MDD) from the Psychiatric Genomics Consortium [41] and the UK Biobank [42], respectively. IGAP is a large two-stage study based on GWAS on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 7,055,881 single nucleotide polymorphisms (SNPs) to meta-analyze four previously-published GWAS datasets consisting of 17,008 Alzheimer's disease cases and 37,154 controls (The European Alzheimer's disease Initiative – EADI the Alzheimer Disease Genetics Consortium – ADGC The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium – CHARGE The Genetic and Environmental Risk in AD consortium – GERAD). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and 11,312 controls. Finally, a meta-analysis was performed combining results from stages 1 and 2.

2.6. Model training and association testing in TCGA-BRCA and ROS/MAP

Using both TCGA-BRCA and ROS/MAP multiomic data, we first identified associations between SNPs and mediators (transcription factor genes, miRNAs, and CpG methylation sites), mediators and gene expression, and SNPs and gene expression using MatrixEQTL [43]. These QTL analyses were adjusted for 10 principal components of the genotype matrix to account for population stratification, along with other relevant covariates (tumor stage and age for TCGA-BRCA, age and sex and smoking status for ROS/MAP). For MeTWAS modeling, we considered the top 5 mediators associated with the gene of interest, assessed by the smallest FDR-adjusted $P < 0.05$. For DePMA models, we considered all distal-SNPs associated with gene expression at raw $P < 10^{-6}$ and any local mediators at FDR-adjusted $P < 0.05$. Local windows for all models were set to 0.5 Mb. For association testing, we consider only genes with significant non-zero estimated total heritability by GCTA-LDMS [21] and cross-validation adjusted $R^2 > 0$ across 5 folds. The MeTWAS or DePMA model with larger cross-validation R^2 was considered as the final MOSTWAS model for a given gene. All other modeling options in MeTWAS and DePMA were set to the defaults provided by the MOSTWAS package.

Using TCGA-BRCA models, we conducted TWAS burden testing [25, 4] in iCOGs GWAS summary statistics for breast cancer-specific survival in a cohort of women of European ancestry. We subjected TWAS-identified loci at Benjamini-Hochberg [44] FDR-adjusted $P < 0.05$ to permutation testing, and any loci that persisted past permutation testing to distal variation added-last testing.

Using ROS/MAP models, we first conducted TWAS burden testing in GWAS summary statistics for late-onset Alzheimer's disease risk from IGAP, prioritized 14 known risk loci identified from literature [40, 45, 46, 47]. We subjected TWAS-identified loci at Benjamini-Hochberg [44] FDR-adjusted $P < 0.05$ to permutation testing, and any loci that persisted past permutation testing to distal variation added-last testing. We similarly conducted TWAS for risk of major depressive disorder (MDD) using GWAS summary statistics from PGC (excluding data from 23andMe and the UK Biobank) with the necessary follow-up tests. For any TWAS-identified loci that persisted permutation in PGC, we further conducted TWAS in GWAX summary statistics for MDD risk in the UK Biobank [42] for replication.

3. Results

3.1. Overview of MOSTWAS

MOSTWAS incorporates two methods to include distal-eQTLs in transcriptomic prediction: mediator-enriched TWAS (MeTWAS) and distal-eQTL prioritization via mediation analysis (DePMA). As large proportions of total heritable gene expression are explained by distal-eQTLs local to regulatory hotspots [6, 10, 12, 13], we used data-driven approaches to either identify mediating regulatory biomarkers (MeTWAS) or distal-eQTLs mediated by local biomarkers (DePMA) to increase predictive power and power to detect gene-trait associations. We assess in-sample predictive performance of MeTWAS and DePMA via cross-validation over a user-defined k folds of training and test sets, such that the expression model is estimated in the training set and imputed into the test set for cross-validation. These methods are summarized graphically in **Figure 1**.

[Figure 1 about here.]

- MeTWAS first identifies m potential mediators (e.g. CpG sites, miRNAs, genes coding for transcription factors, etc) such that the intensity (methylation, expression levels, etc) of these mediators are associated with the mRNA expression of the gene of interest. A model for the genetically regulated intensities (GRIn) is estimated using the local SNPs to these mediators, and the GRIn of these mediators are imputed into the training set. The final gene expression model is estimated by incorporating the GRIn of the mediators as fixed effects and the local SNPs to the gene as regularized effects (see **Methods** for more details).

- DePMA first identifies testing triplets of the gene of interest, a distal eSNP (SNP in an eQTL) to the gene, and any potential associated mediators local to the eQTL. The total mediation effect (TME)

of the eQTL on the gene through the set of mediators is estimated and the two-sided hypothesis of $H_0 : TME = 0$ is tested via one of two options (**Supplemental Figure S1**). Any distal-eSNP with a significant TME is included with the SNPs local to the gene of interest to form the final design matrix. A model including all local SNPs and all significant distal-eSNPs is fit to the expression of the gene using either elastic net or linear mixed modeling (see **Methods** for more details).

If individual genotype data is available in an external GWAS panel, using either a MeTWAS or DePMA model, we impute expression as a linear combination of the genotypes. If only summary statistics are available in the GWAS panel, the Imp-G weighted burden testing framework [25], as implemented in Gusev *et al.*[4], can be applied. We further implement a permutation test to assess whether the overall association is significant conditional on the GWAS effect sizes [4] and an added-last test that assesses the added information from distal-SNPs given the weighted Z-score from local SNPs (see **Methods** and **Supplemental Materials**).

3.2. Simulation analysis

We first conducted simulations to assess the predictive capability and power to detect gene-trait associations under various settings for phenotype heritability (h^2_p), local heritability of expression ($h^2_{e,l}$), distal heritability of expression ($h^2_{e,d}$), and proportion of causal local ($p_{c,l}$) and distal ($p_{c,e}$) SNPs for MeTWAS and DePMA. We considered two scenarios for each combination of ($h^2_p, h^2_{e,l}, h^2_{e,d}, p_{c,l}, p_{c,e}$): (1) the leveraged association between the distal-SNP and the gene of interest exists in both the reference and imputation panel, and (2) the leveraged association between distal-eSNP and the gene of interest exists in the reference panel but does not contribute to the heritable portion of the phenotype in the imputation panel. Using genetic data from TCGA-BRCA, we used the 2,592 SNPs local to the gene *ESR1* on Chromosome 6 to generate local eQTLs and the 1,431 SNPs local to the gene *FOXA1* to generate distal eQTLs for a 400-sample eQTL reference panel and 1,500-sample GWAS imputation panel (see **Methods**). Though the choice of these loci was arbitrary for generating the simulated data, there is evidence that *ESR1* and *FOXA1* are highly co-expressed in breast tumors and local-eQTLs of *FOXA1* have been shown to be distal-eQTLs of *ESR1* [48].

In these simulation studies, we found that MOSTWAS methods performed well in prediction across different causal proportions and local and distal mRNA expression heritabilities. Furthermore, across all simulation settings, we observed that MOSTWAS showed greater or equal power to detect gene-trait associations as local-only models. We found that, under the setting that distal variation contributes to trait heritability, the best MOSTWAS model has greater power to detect gene-trait associations than the local-only model, with the advantage in power over local-only models increasing with increased distal expression heritability (**Figure 2A**). Similarly, we found that as the proportion of total expression heritability that is attributed to distal genetic variation increased, the positive difference in predictive performance between the best MOSTWAS model and the local-only model increased (**Supplemental Figure S3**). Under the "null" case that distal variation influences expression in the reference panel but does not affect the trait in

the GWAS panel, we find that local-only and MOSTWAS models perform similarly, as we would expect. At low causal proportion ($p_c = 0.01$) and low trait heritability ($h_p^2 = 0.2$), local-only models have a modest advantage in TWAS power over MOSTWAS models (shown in **Supplemental Data**). This difference is mitigated at larger causal proportions and trait heritabilities (**Figure 2B**).

We also find that the power of the distal added-last increase significantly as both the sample sizes of the eQTL reference panel and the GWAS imputation panel increases. At a sample size of 10,000 in the GWAS panel with summary statistics (a suitably large GWAS), we obtain greater than 65% power to detect significant distal associations given the local association at eQTL sample sizes of greater than 200 (**Supplemental Figure S2**). Overall, these results demonstrated the advantages of MOSTWAS methods for modeling the complex genetic architecture of transcriptomes, especially when distal variation has a discernibly large effect on the heritability of both the gene and trait of interest. Full simulation results are provided in **Supplemental Data** at `www.github.com/bhattacharya-a-bt/mostwas_paper`. The MOSTWAS package also contains functions for this simulation framework.

[Figure 2 about here.]

3.3. Real data application in breast cancer tumors

We applied MOSTWAS in the context of breast tumor multi-omics and disease outcomes, motivated by recent GWAS and TWAS into breast cancer-specific survival [49, 38, 39, 37, 50]. Previous breast tumor eQTL studies have also revealed several signficant distal-eQTLs in trait-associated loci, many of which are in regulatory or epigenetic hotspots [50, 51], motivating our application of MOSTWAS for breast tumor expression modeling. Using TCGA-BRCA datasets comprising germline SNPs, tumor mRNA expression, DNA methylation, and miRNA expression, we trained MeTWAS, DePMA, and traditional local-only predictive models for the mRNA expression of all genes with germline heritability $h^2 > 0$ at $P < 0.05$. Estimates of heritability for genes were considerably larger when we considered distal variation using MOSTWAS methods (mean heritabilities in **Supplemental Table S1**). We also found that MeTWAS and DePMA perform better in cross-validation R^2, with larger numbers of models at $R^2 \geq 0.01$ and significant germline heritability at raw $P < 0.05$ using MOSTWAS methods than local-only models (**Figures 3A-C**). Mean predictive R^2 for local-only models was 0.011 (25% to 75% inter-quartile interval (0.0,0.013)), for MeTWAS models was 0.028 (0.013, 0.032), and for DePMA models was 0.051 (0.019, 0.068).

In addition to cross-validation, we used 351 samples in TCGA-BRCA with only genotype and mRNA expression data that were not used in model training to test the portability of MOSTWAS models in independent external cohorts. As shown in **Figure 4A and Supplementary Figure S4**, DePMA models obtain the highest predictive adjusted R^2 in the external cohort (mean 0.016, 25% to 75% inter-quartile interval (0.003, 0.018)), with local-only models (0.013, (0.00,0.013)) outperforming MeTWAS models (0.011, (0.002, 0.012)), considering only genes that attained cross-validation adjusted $R^2 \geq 0.01$ using a given method. Overall, among genes with cross-validation adjusted $R^2 \geq 0.01$, 37 out of 280 genes achieved external pre-

dictive $R^2 \geq 0.01$ using local-only models, 89 out of 709 using MeTWAS, and 787 out of 1,185 using DePMA (**Figure 3A-C**).

Lastly, we conducted association studies for breast cancer-specific survival using local-only and the MOSTWAS model with largest R^2 trained in TCGA-BRCA and summary-level GWAS data from iCOGs. Here, we constructed the weighted burden test, as described above and in Pasaniuc *et al.* and Gusev *et al.* [25, 4]. We prioritized genes with Benjamini-Hochberg [44] adjusted $P < 0.05$ for permutation testing. Of the 122 genes that had cross-validation $R^2 \geq 0.01$ in TCGA-BRCA using both local-only and MOSTWAS models, we found 2 survival associations with the same loci at Benjamini-Hochberg FDR-adjusted 0.05 using both local-only and MOSTWAS models, with the strength of association marginally larger with the MOSTWAS model in each case (**Supplemental Figure S5**). Furthermore, we find that 116 of these loci showed larger strengths of association with survival using the MOSTWAS model than the local-only model (**Supplemental Figure S5**). QQ-plots for TWAS Z-statistics are provided in **Supplemental Figure S6B** for both local-only and MOSTWAS models. These results in TCGA-BRCA demonstrated the improved transcriptomic prediction and power to detect gene-trait associations using MOSTWAS over local-only modeling.

3.3.1. Functional hypothesis generation with MOSTWAS

We next conducted TWAS for breast cancer-specific survival using all genes with significant germline heritability at $P < 0.05$ with the MOSTWAS model with larger cross-validation $R^2 > 0.01$. We identified 21 survival-associated loci at Benjamini-Hochberg FDR adjusted $P < 0.05$ Of these 21 loci, 11 persisted when subjected to permutation testing at a significance threshold of FDR-adjusted $P < 0.05$ (**Figure 4C** and **Supplemental Table S2**).

An advantage of MOSTWAS is its ability to aid in functional hypothesis generation for mechanistic follow-up studies. The added-last test allows users to identify genes where trait association from distal variation is significant above and beyond the contribution of the local component. For 8 of the TWAS-associated 11 loci, at FDR-adjusted $P < 0.05$, we found significant distal variation added-last associations (see **Supplemental Methods** and **Supplemental Table S2**), suggesting that distal variation may contribute to the gene-trait association. All 8 of these loci showed distal association with the gene of interest mediated through a set of four transcription factors (*NAA50*, *ATP6V1A*, *ROCK2*, *USF3*), all highly interconnected with the critical MAPK pathway [52, 53, 54, 55, 56, 57]. These regulatory sites serve as an example of how distal genomic regions can be prioritized for functional follow-up studies to elucidate the mechanisms underlying the SNP-gene-trait associations. These results show the strength of MOSTWAS to detect and prioritize gene-trait associations that are influenced by distal variation.

3.4. Real data application in brain tissue

We also applied MOSTWAS to transcriptomic data on samples of prefrontal cortex, a tissue that has been used previously in studying neuropsychiatric traits and disorders with TWAS [58, 59]. There has been ample evidence in brain tissue, especially the prefrontal cortex, that non-coding variants may regulate

distal genes [60, 58, 61]; in fact, an eQTL analysis by Sng *et al.* found that approximately 20-40% of eQTLs in the frontal cortex can be considered trans-acting [62]. Thus, the prefrontal cortex in the context of neuropsychiatric disorders provides a prime example to assess MOSTWAS. Using ROS/MAP data on germline SNPs, tumor mRNA expression, DNA methylation, and miRNA expression, we trained MeTWAS, DePMA, and traditional local-only predictive models for the mRNA expression of all genes with germline heritability $h^2 > 0$ at $P < 0.05$. Consistent with results in TCGA-BRCA, estimates of heritability for genes were considerably larger when we considered distal variation using MOSTWAS methods (**Supplemental Table S1**). We also find that MeTWAS and DePMA perform better in cross-validation R^2 than local-only models (**Figures 3D-F**). Mean predictive R^2 for local-only models was 0.029 (25% to 75% inter-quartile interval (0.0,0.015)), for MeTWAS models was 0.079 (0.019, 0.082), and for DePMA models was 0.045 (0.013, 0.037).

In addition to cross-validation, we used 87 samples in ROS/MAP with genotype and mRNA expression data that were not used in model training to test the portability of MOSTWAS models in independent external cohorts. As shown in **Figure 4A and Supplementary Figure S4**, DePMA models obtain the highest predictive adjusted R^2 in the external cohort (0.042 (25% quantile 0.009, 75% quantile 0.057)), with MeTWAS models (0.040 (0.010, 0.054)) also outperforming local-only models (0.031 (0.007, 0.039)), considering only genes that attained cross-validation adjusted $R^2 \geq 0.01$ using a given method. Overall, among genes with cross-validation adjusted $R^2 \geq 0.01$, 187 out of 267 genes achieved external predictive $R^2 \geq 0.01$ using local-only models, 683 out of 911 using MeTWAS, and 2,135 out of 2,934 using DePMA (**Figure 3D-F**).

We next conducted association tests for known Alzheimer's disease risk loci using local-only and the best MOSTWAS model (comparing MeTWAS and DePMA cross-validation R^2) trained in ROS/MAP and summary-level GWAS data from IGAP. From literature, we identified 14 known common and rare loci of late-onset Alzheimer's disease [40, 45, 46, 47], 11 of which had MOSTWAS models with cross-validation $R^2 \geq 0.01$. Five of these 11 loci (*APOE*, *CLU*, *PLCG2*, *SORL1*, *ZCWPW1*) showed significant association at Benjamini-Hochberg FDR-adjusted $P \leq 0.05$ (**Supplemental Table S3**). We also compared these 11 associations to those identified by local-only models and by latent Dirichlet process regression (DPR) as implemented in TIGAR [63], with raw P-values of association shown in **Figure 4B**. MOSTWAS showed stronger associations at 8 of these loci than both local-only and DPR models. We followed up on the 5 significantly associated loci using the permutation and added-last tests (see **Methods** and **Supplemental Materials**). Three of these loci (*APOE*, *SORL1*, *ZCWPW1*) showed significant associations, conditional on large GWAS effect sizes (permutation test significant at FDR-adjusted $P < 0.05$) These three loci also showed significant associations with distal variants, given the association with local variants, at FDR-adjusted $P < 0.05$ (**Supplemental Table S3**).

We then conducted a transcriptome-wide association study for risk of major depressive disorder (MDD) using summary statistics from the Psychiatric Genomics Consortium (PGC) genome-wide meta-analysis

that excluded data from the UK Biobank and 23andMe [41]. QQ-plots for TWAS Z-statistics are provided in **Supplemental Figure S6B** for both local-only and MOSTWAS models. Overall, using all heritable genes with cross-validation R^2 with the best MOSTWAS model in ROS/MAP, we identified 102 MDD risk-associated loci with FDR-adjusted $P < 0.05$ that persisted when subjected to permutation testing at an FDR-adjusted significance threshold of $P < 0.05$ (colored red in **Figure 4D**). We downloaded genome-wide association study by proxy (GWAX) summary statistics from the UK Biobank [42] for replication analysis of loci identified using PGC summary statistics. We found that 7 of these 102 loci (labelled in **Figure 4D**) also show an association in UK Biobank GWAX that is in the same direction as in PGC. In comparison, using local-only models, we identified 11 genes with significant association with MDD risk at FDR-adjusted $P < 0.05$ that persisted permutation testing; however, none of these loci showed significant associations in the UK Biobank GWAX in the same direction as in PGC. Summary statistics for MOSTWAS associations in PGC and UK Biobank are provided in **Supplemental Table S4**. Local-only results are provided in **Supplemental Data**. It is important to note the UK Biobank dataset is not a GWAS dataset as it defines a case of MDD as any subject who has the disorder or a first-degree relative with MDD, leading to lower power to detect associations in this dataset. Nonetheless, we believe that strong prediction of predictive models in independent cohorts and TWAS results across two independent cohorts provides an example of the robustness of models using MOSTWAS.

[Figure 3 about here.]

[Figure 4 about here.]

We observed that MOSTWAS models generally had higher predictive R^2 than local-only models both in training and independent cohorts. We also found that MOSTWAS has recapitulated 5 known Alzheimer's risk loci that were not detected by local-only modeling (both PrediXcan [3] and TIGAR [63]), 3 of which had significant distal associations using our added-last test. We also illustrated that the MOSTWAS detected MDD-risk loci that were replicable across independent GWAS and GWAX cohorts [41, 42].

3.5. Comparison of computational time

To assess the difference in computational burden between local-only, MeTWAS, and DePMA modelling, we randomly selected a set of 50 genes that are heritable across all three models from TCGA-BRCA and computed per-gene time for fitting using a 24-core, 3.0 GHz processor. We found that MeTWAS (mean of 225 seconds per gene) and DePMA (mean 312 seconds per gene) took approximately 6-10 times longer to fit than a traditional local-only model (mean 36 seconds) (**Supplemental Figure S7**). Model-fitting here includes heritability estimation, estimating the SNP-expression weights, and cross-validation. We have implemented parallel options to train an expression model for a single gene in MOSTWAS. We also recommend fitting an entire set of genes from an RNA-seq panel via a batch computing approach [64, 65]. Using parallel implementation with 5 cores and batch computing, we analyzed 15,568 genes from TCGA-BRCA in approximately 28 hours.

4. Discussion

Through a variety of simulations and real applications in two settings, we demonstrate that multi-omic methods that prioritize distal variation in TWAS have higher predictive performance and power to detect cell-specific gene-trait associations [8, 12, 14], especially when distal variation contribute substantially to trait heritability. We proposed two methods (MeTWAS and DePMA) for identifying and including distal genetic variants in gene expression prediction models. We have provided implementations of these methods in the MOSTWAS (Multi-omic Strategies for Transcriptome-Wide Association Studies) R package, available freely on Github. MOSTWAS contains functions to train expression models with both MeTWAS and DePMA and outputs models with 5-fold cross-validation $R^2 \geq 0.01$ and significant germline heritability. The package also contains functions and documentation for simulation analyses [26], the weighted burden [25, 4] and follow-up permutation [4] and distal-SNPs added-last tests for TWAS using GWAS summary statistics, and file-formatting. We also provide guidelines for parallelization to lessen computational time.

Not only does MOSTWAS improve transcriptomic imputation both in- and out-of-sample, it also provides a test for the identification of heritable mediators that affect eventual transcription of the gene of interest. These identified mediators can provide insight into the underlying mechanisms for SNP-gene-trait associations to improve detection of gene-trait associations and to prioritize biological units for functional follow-up studies. Using MOSTWAS and iCOGs summary-level GWAS statistics for breast cancer-specific survival [37], we identified 11 survival-associated loci that are enriched for p53 binding and oxidoreductase activity pathways [66, 67]. These loci include two genes (*MAP3K6* and *MAP4K5*) encoding mitogen-activated protein kinases, which are signalling transduction molecules involved in the progression of aggressive breast cancer hormone subtypes [68]. TWAS using MOSTWAS models was able to recapitulate 5 out of 14 known Alzheimer's disease risk loci in IGAP GWAS summary statistics [40], which were not recoverable with local-only models. We showed the utility of the distal-SNPs added-last test to prioritize significant distal SNP-gene-trait associations for follow-up mechanistic studies, which could not be identified using traditional local-only TWAS. In PGC GWAS summary-level data for major depressive disorder [41], we found 102 risk loci, 7 of which were replicated in independent GWAX summary statistics from the UK Biobank [42]. Three of these seven loci (*SYT1*, *CACNA2D3*, *ADAD2*) encode important proteins involved in synaptic transmission in the brain and RNA editing. Studies have shown that variation at these loci may lead to loss of function at synapses and RNA editing that lead to psychiatric disorders [69, 70, 71, 72, 73]. All survival- or risk-associated loci identified by MOSTWAS were not detected using local-only models.

A considerable limitation of MOSTWAS is the increased computational burden over local-only modelling, especially in DePMA's permutation-based mediation analysis for multiple genome-wide mediators. We believe a Monte-Carlo resampling method will aid in scalability by making some standard distributional assumptions on the effect sizes of SNPs and mediators in the DePMA mediation model [74]. Nevertheless, we believe that MOSTWAS's gain in predictive performance and power to detect gene-trait associations may outweigh this computational time. In addition, RNA-sequencing alignment errors can lead to false positives

in distal-eQTL detection [75], and in turn, bias the mediation modeling. Cross-mapping estimation, as described by Saha *et al.* [75], can be used to flag potentially falsely positive distal-QTLs that are detected as the first step in MeTWAS and DePMA. Another limitation of MOSTWAS is the general lack of rich multi-omic panels, like TCGA-BRCA and ROS/MAP, that provide a large set of mediating biomarkers that may be mechanistically involved in gene regulation. However, the two-step regression framework outlined in MeTWAS allows for importing mediator intensity models trained in other cohorts to estimate the germline portion of total gene expression from distal variants.

In conclusion, MOSTWAS provides a user-friendly and intuitive tool that extends transcriptomic imputation and association studies to include distal genetic variants. We demonstrated that the methods in MOSTWAS based on two-step regression and mediation analysis generally out-perform local-only models in both transcriptomic prediction and TWAS power, though at the cost of longer computational times. MOSTWAS enables users to utilize rich reference multi-omic datasets for enhanced gene mapping to better understand the genetic etiology of polygenic traits and diseases with more direct insight into functional follow-up studies.

5. Supplemental Data

Document S1: Supplementary Methods, Supplementary Figures 1–6, Supplementary Tables 1–4 (supplement.pdf)

6. Web Resources

MOSTWAS, `https://github.com/bhattacharya-a-bt/MOSTWAS`
TCGA GDC Legacy Archive, `https://portal.gdc.cancer.gov/legacy-archive/`
GDAC Firehose Browser, `https://gdac.broadinstitute.org/`
ROS/MAP data, `https://www.synapse.org/#!Synapse:syn3219045`
IGAP data, `http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php`
PGC data, `https://www.med.unc.edu/pgc/download-results/mdd/`
UKBB GWAX data, `http://gwas-browser.nygenome.org/downloads/gwas-browser/`

7. Declaration of Interests

The authors have no competing interests.

8. Acknowledgements

We thank Colin Begg, Terry Furey, Michael Gandal, Yun Li, Karen Mohlke, Brandon Pierce, Bogdan Pasaniuc, Hudson Santos, and Jason Stein for engaging conversation and guidance during the research process.

Funding for BCAC and iCOGS came from: Cancer Research UK [grant numbers C1287/A16563, C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), the European Community's Seventh Framework Programme under grant agreement n° 223175 [HEALTHF2-2009-223175] (COGS), the National Institutes of Health [CA128978] and Post-Cancer GWAS initiative [1U19 CA148537, 1U19 CA148065-01 (DRIVE) and 1U19 CA148112 - the GAME-ON initiative], the Department of Defence [W81XWH-10-1-0341], and the Canadian Institutes of Health Research CIHR) for the CIHR Team in Familial Risks of Breast Cancer [grant PSR-SIIRI-701]. All studies and funders as listed in Michailidou K et al (2013 and 2015) and in Guo Q et al (2015) are acknowledged for their contributions.

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i–Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n°) 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.

We also thank the Psychiatric Genomics Consortium for their publicly available GWAS summary statistics (`https://www.med.unc.edu/pgc/`) and the Pickrell Lab at the New York Genome Center for their GWAS browser and GWAX summary statistics (`http://gwas-browser.nygenome.org/downloads/gwas-browser/`).

9. Accession Numbers

Birdseed genotype files of 914 subject were downloaded from the Genome Data Commons (GDC) legacy (GRCh37/hg19) archive (`https://portal.gdc.cancer.gov/legacy-archive/`). Processed mRNA/miRNA expression and CpG methylation data was accessed from the Firehose repository from the Broad Institute via FireBrowse (`https://gdac.broadinstitute.org/`) [32]. ROS/MAP multiomics data (genotype, mRNA and miRNA expression, CpG methylation) was accessed through the AMP-AD Knowledge Portal on Synapse (Synapse ID: syn3219045) [36].

References

[1] Y. Hasin, M. Seldin, A. Lusis, Multi-omics approaches to disease (5 2017). doi:10.1186/s13059-017-1215-1.

[2] F. R. Pinu, D. J. Beale, A. M. Paten, K. Kouremenos, S. Swarup, H. J. Schirra, D. Wishart, Systems biology and multi-omics integration: Viewpoints from the metabolomics research community, Metabolites 9 (4). doi:10.3390/metabo9040076.

[3] E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, D. L. Nicolae, N. J. Cox, H. K. Im, H. K. Im, A gene-based association method for mapping traits using reference transcriptome data, Nature Genetics 47 (9) (2015) 1091–1098. doi:10.1038/ng.3367.
URL http://www.nature.com/articles/ng.3367

[4] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. J. H. Penninx, R. Jansen, E. J. C. de Geus, D. I. Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimäki, E. Raitoharju, M. Kähönen, I. Seppälä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, B. Pasaniuc, Integrative approaches for large-scale transcriptome-wide association studies, Nature Genetics 48 (3) (2016) 245–252. doi:10.1038/ng.3506.
URL http://www.ncbi.nlm.nih.gov/pubmed/26854917http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4767558http://www.nature.com/articles/ng.3506

[5] B. Brynedal, J. M. Choi, T. Raj, R. Bjornson, B. E. Stranger, B. M. Neale, B. F. Voight, C. Cotsapas, Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of Transcriptional Co-regulation, American Journal of Human Genetics 100 (4) (2017) 581–591. doi:10.1016/j.ajhg.2017.02.004.

[6] C. Liu, Y. Xiao, Z. Tao, X.-C. Hu, Identification of immune microenvironment subtypes of breast cancer in TCGA set: Implications for immunotherapy., Journal of Clinical Oncology 37 (15_suppl) (2019) e14205–e14205. doi:10.1200/jco.2019.37.15{_}suppl.e14205.

[7] E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic., Cell 169 (7) (2017) 1177–1186. doi:10.1016/j.cell.2017.05.038.
URL http://www.ncbi.nlm.nih.gov/pubmed/28622505http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5536862

[8] C. D. Brown, L. M. Mangravite, B. E. Engelhardt, Integrative Modeling of eQTLs and Cis-Regulatory Elements Suggests Mechanisms Underlying Cell Type Specificity of eQTLs, PLoS Genetics 9 (8) (2013) e1003649. doi:10.1371/journal.pgen.1003649.

URL http://www.ncbi.nlm.nih.gov/pubmed/23935528http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3731231

[9] X. He, C. K. Fuller, Y. Song, Q. Meng, B. Zhang, X. Yang, H. Li, Sherlock: Detecting gene-disease associations by matching patterns of expression QTL and GWAS, American Journal of Human Genetics 92 (5) (2013) 667–680. doi:10.1016/j.ajhg.2013.03.022.

[10] B. L. Pierce, L. Tong, L. S. Chen, R. Rahaman, M. Argos, F. Jasmine, S. Roy, R. Paul-Brutus, H. J. Westra, L. Franke, T. Esko, R. Zaman, T. Islam, M. Rahman, J. A. Baron, M. G. Kibriya, H. Ahsan, Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians, PLoS Genetics 10 (12). doi:10.1371/journal.pgen.1004818.

[11] F. Yang, J. Wang, B. L. Pierce, L. S. Chen, Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis, Genome Research 27 (11) (2017) 1859–1871. doi:10.1101/gr.216754.116.

[12] B. L. Pierce, L. Tong, M. Argos, K. Demanelis, F. Jasmine, M. Rakibuz-Zaman, G. Sarwar, M. T. Islam, H. Shahriar, T. Islam, M. Rahman, M. Yunus, M. G. Kibriya, L. S. Chen, H. Ahsan, Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms, Nature Communications 9 (1) (2018) 1–12. doi:10.1038/s41467-018-03209-9.

[13] N. Shan, Z. Wang, L. Hou, Identification of trans-eQTLs using mediation analysis with multiple mediators, BMC Bioinformatics 20. doi:10.1186/s12859-019-2651-6.

[14] M. van der Wijst, D. de Vries, H. Groot, C. Hon, M. Nawijn, Y. Idaghdour, P. van der Harst, C. Ye, J. Powell, F. Theis, A. Mahfouz, M. Heinig, L. Franke, Single-cell eQTLGen Consortium: a personalized understanding of disease, arXiv.

[15] J. Arloth, G. Eraslan, T. F. M. Andlauer, J. Martins, S. Iurato, B. Kühnel, M. Waldenberger, J. Frank, R. Gold, B. Hemmer, F. Luessi, S. Nischwitz, F. Paul, H. Wiendl, C. Gieger, S. Heilmann-Heimbach, T. Kacprowski, M. Laudes, T. Meitinger, A. Peters, R. Rawal, K. Strauch, S. Lucae, B. Müller-Myhsok, M. Rietschel, F. J. Theis, E. B. Binder, N. S. Mueller, DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning, PLOS Computational Biology 16 (2) (2020) e1007616. doi:10.1371/journal.pcbi.1007616.
URL https://dx.plos.org/10.1371/journal.pcbi.1007616

[16] W. Zhang, G. Voloudakis, V. M. Rajagopal, B. Readhead, J. T. Dudley, E. E. Schadt, J. L. M. Björkegren, Y. Kim, J. F. Fullard, G. E. Hoffman, P. Roussos, Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits, Nature Communications 10 (1) (2019) 3834. doi:10.1038/s41467-019-11874-7.
URL http://www.nature.com/articles/s41467-019-11874-7

[17] R. McLendon, A. Friedman, D. Bigner, E. G. Van Meir, D. J. Brat, G. M. Mastrogianakis, J. J. Olson, T. Mikkelsen, N. Lehman, K. Aldape, W. K. Yung, O. Bogler, J. N. Weinstein, S. VandenBerg, M. Berger, M. Prados, D. Muzny, M. Morgan, S. Scherer, A. Sabo, L. Nazareth, L. Lewis, O. Hall, Y. Zhu, Y. Ren, O. Alvi, J. Yao, A. Hawes, S. Jhangiani, G. Fowler, A. San Lucas, C. Kovar, A. Cree, H. Dinh, J. Santibanez, V. Joshi, M. L. Gonzalez-Garay, C. A. Miller, A. Milosavljevic, L. Donehower, D. A. Wheeler, R. A. Gibbs, K. Cibulskis, C. Sougnez, T. Fennell, S. Mahan, J. Wilkinson, L. Ziaugra, R. Onofrio, T. Bloom, R. Nicol, K. Ardlie, J. Baldwin, S. Gabriel, E. S. Lander, L. Ding, R. S. Fulton, M. D. McLellan, J. Wallis, D. E. Larson, X. Shi, R. Abbott, L. Fulton, K. Chen, D. C. Koboldt, M. C. Wendl, R. Meyer, Y. Tang, L. Lin, J. R. Osborne, B. H. Dunford-Shore, T. L. Miner, K. Delehaunty, C. Markovic, G. Swift, W. Courtney, C. Pohl, S. Abbott, A. Hawkins, S. Leong, C. Haipek, H. Schmidt, M. Wiechert, T. Vickery, S. Scott, D. J. Dooling, A. Chinwalla, G. M. Weinstock, E. R. Mardis, R. K. Wilson, G. Getz, W. Winckler, R. G. Verhaak, M. S. Lawrence, M. O'Kelly, J. Robinson, G. Alexe, R. Beroukhim, S. Carter, D. Chiang, J. Gould, S. Gupta, J. Korn, C. Mermel, J. Mesirov, S. Monti, H. Nguyen, M. Parkin, M. Reich, N. Stransky, B. A. Weir, L. Garraway, T. Golub, M. Meyerson, L. Chin, A. Protopopov, J. Zhang, I. Perna, S. Aronson, N. Sathiamoorthy, G. Ren, J. Yao, W. R. Wiedemeyer, H. Kim, W. K. Sek, Y. Xiao, I. S. Kohane, J. Seidman, P. J. Park, R. Kucherlapati, P. W. Laird, L. Cope, J. G. Herman, D. J. Weisenberger, F. Pan, D. Van Den Berg, L. Van Neste, M. Y. Joo, K. E. Schuebel, S. B. Baylin, D. M. Absher, J. Z. Li, A. Southwick, S. Brady, A. Aggarwal, T. Chung, G. Sherlock, J. D. Brooks, R. M. Myers, P. T. Spellman, E. Purdom, L. R. Jakkula, A. V. Lapuk, H. Marr, S. Dorton, G. C. Yoon, J. Han, A. Ray, V. Wang, S. Durinck, M. Robinson, N. J. Wang, K. Vranizan, V. Peng, E. Van Name, G. V. Fontenay, J. Ngai, J. G. Conboy, B. Parvin, H. S. Feiler, T. P. Speed, J. W. Gray, C. Brennan, N. D. Socci, A. Olshen, B. S. Taylor, A. Lash, N. Schultz, B. Reva, Y. Antipin, A. Stukalov, B. Gross, E. Cerami, Q. W. Wei, L. X. Qin, V. E. Seshan, L. Villafania, M. Cavatore, L. Borsu, A. Viale, W. Gerald, C. Sander, M. Ladanyi, C. M. Perou, D. N. Hayes, M. D. Topal, K. A. Hoadley, Y. Qi, S. Balu, Y. Shi, J. Wu, R. Penny, M. Bittner, T. Shelton, E. Lenkiewicz, S. Morris, D. Beasley, S. Sanders, A. Kahn, R. Sfeir, J. Chen, D. Nassau, L. Feng, E. Hickey, A. Barker, D. S. Gerhard, J. Vockley, C. Compton, J. Vaught, P. Fielding, M. L. Ferguson, C. Schaefer, J. Zhang, S. Madhavan, K. H. Buetow, F. Collins, P. Good, M. Guyer, B. Ozenberger, J. Peterson, E. Thomson, Comprehensive genomic characterization defines human glioblastoma genes and core pathways, Nature 455 (7216) (2008) 1061–1068. doi:10.1038/nature07385.

[18] P. L. De Jager, Y. Ma, C. McCabe, J. Xu, B. N. Vardarajan, D. Felsky, H. U. Klein, C. C. White, M. A. Peters, B. Lodgson, P. Nejad, A. Tang, L. M. Mangravite, L. Yu, C. Gaiteri, S. Mostafavi, J. A. Schneider, D. A. Bennett, Data descriptor: A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research, Scientific Data 5. doi:10.1038/sdata.2018.142.

[19] J. Friedman, T. Hastie, R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordi-

nate Descent, Journal of Statistical Software 33 (1) (2010) 1–22. doi:10.18637/jss.v033.i01.
URL http://www.jstatsoft.org/v33/i01/

[20] J. B. Endelman, Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP, The Plant Genome 4 (3) (2011) 250–255. doi:10.3835/plantgenome2011.08.0024.

[21] J. Yang, A. Bakshi, Z. Zhu, G. Hemani, A. A. E. Vinkhuyzen, S. H. Lee, M. R. Robinson, J. R. B. Perry, I. M. Nolte, J. V. van Vliet-Ostaptchouk, H. Snieder, T. L. C. LifeLines Cohort Study, T. Esko, L. Milani, R. Mägi, A. Metspalu, A. Hamsten, P. K. E. Magnusson, N. L. Pedersen, E. Ingelsson, N. Soranzo, M. C. Keller, N. R. Wray, M. E. Goddard, P. M. Visscher, Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index., Nature genetics 47 (10) (2015) 1114–1120. doi:10.1038/ng.3390.
URL http://www.ncbi.nlm.nih.gov/pubmed/26323059http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4589513

[22] M. E. Sobel, Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models, Sociological Methodology 13 (1982) 290. doi:10.2307/270723.

[23] D. P. Mackinnon, C. M. Lockwood, J. Williams, Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods, Multivariate Behavioral Research 39 (1) (2004) 99–128. doi:10.1207/s15327906mbr3901{_}4.

[24] J. D. Storey, A. J. Bass, A. Dabney, D. Robinson, G. Warnes, qvalue: Q-value estimation for false discovery rate control (2019).
URL http://github.com/jdstorey/qvalue

[25] B. Pasaniuc, N. Zaitlen, H. Shi, G. Bhatia, A. Gusev, J. Pickrell, J. Hirschhorn, D. P. Strachan, N. Patterson, A. L. Price, J. Kelso, Fast and accurate imputation of summary statistics enhances evidence of functional enrichment, Bioinformatics 30 (20) (2014) 2906–2914. doi:10.1093/bioinformatics/btu416.
URL https://academic.oup.com/bioinformatics/article-abstract/30/20/2906/2422225

[26] N. Mancuso, M. K. Freund, R. Johnson, H. Shi, G. Kichaev, A. Gusev, B. Pasaniuc, Probabilistic fine-mapping of transcriptome-wide association studies, Nature Genetics 51 (4) (2019) 675–682. doi:10.1038/s41588-019-0367-1.

[27] J. O'Connell, D. Gurdasani, O. Delaneau, N. Pirastu, S. Ulivi, M. Cocca, M. Traglia, J. Huang, J. E. Huffman, I. Rudan, R. McQuillan, R. M. Fraser, H. Campbell, O. Polasek, G. Asiki, K. Ekoru, C. Hayward, A. F. Wright, V. Vitart, P. Navarro, J.-F. Zagury, J. F. Wilson, D. Toniolo, P. Gasparini, N. Soranzo, M. S. Sandhu, J. Marchini, A General Approach for Haplotype Phasing across the Full Spectrum

of Relatedness, PLoS Genetics 10 (4) (2014) e1004234. doi:10.1371/journal.pgen.1004234.
URL https://dx.plos.org/10.1371/journal.pgen.1004234

[28] O. Delaneau, J. Marchini, J.-F. Zagury, A linear complexity phasing method for thousands of genomes, Nature Methods 9 (2) (2012) 179–181. doi:10.1038/nmeth.1785.
URL http://www.nature.com/articles/nmeth.1785

[29] B. N. Howie, P. Donnelly, J. Marchini, A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies, PLoS Genetics 5 (6) (2009) e1000529. doi:10.1371/journal.pgen.1000529.
URL https://dx.plos.org/10.1371/journal.pgen.1000529

[30] J. E. Wigginton, D. J. Cutler, G. R. Abecasis, A Note on Exact Tests of Hardy-Weinberg Equilibrium, Tech. rep. (2005).
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1199378/pdf/AJHGv76p887.pdf

[31] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. De Bakker, M. J. Daly, P. C. Sham, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses, Am. J. Hum. Genet 81 (2007) 559–575. doi:10.1086/519795.
URL www.ajhg.org

[32] J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, C. Sander, J. M. Stuart, K. Chang, C. J. Creighton, C. Davis, L. Donehower, J. Drummond, D. Wheeler, A. Ally, M. Balasundaram, I. Birol, Y. S. Butterfield, A. Chu, E. Chuah, H. J. E. Chun, N. Dhalla, R. Guin, M. Hirst, C. Hirst, R. A. Holt, S. J. Jones, D. Lee, H. I. Li, M. A. Marra, M. Mayo, R. A. Moore, A. J. Mungall, A. G. Robertson, J. E. Schein, P. Sipahimalani, A. Tam, N. Thiessen, R. J. Varhol, R. Beroukhim, A. S. Bhatt, A. N. Brooks, A. D. Cherniack, S. S. Freeman, S. B. Gabriel, E. Helman, J. Jung, M. Meyerson, A. I. Ojesina, C. S. Pedamallu, G. Saksena, S. E. Schumacher, B. Tabak, T. Zack, E. S. Lander, C. A. Bristow, A. Hadjipanayis, P. Haseley, R. Kucherlapati, S. Lee, E. Lee, L. J. Luquette, H. S. Mahadeshwar, A. Pantazi, M. Parfenov, P. J. Park, A. Protopopov, X. Ren, N. Santoso, J. Seidman, S. Seth, X. Song, J. Tang, R. Xi, A. W. Xu, L. Yang, D. Zeng, J. T. Auman, S. Balu, E. Buda, C. Fan, K. A. Hoadley, C. D. Jones, S. Meng, P. A. Mieczkowski, J. S. Parker, C. M. Perou, J. Roach, Y. Shi, G. O. Silva, D. Tan, U. Veluvolu, S. Waring, M. D. Wilkerson, J. Wu, W. Zhao, T. Bodenheimer, D. N. Hayes, A. P. Hoyle, S. R. Jeffreys, L. E. Mose, J. V. Simons, M. G. Soloway, S. B. Baylin, B. P. Berman, M. S. Bootwalla, L. Danilova, J. G. Herman, T. Hinoue, P. W. Laird, S. K. Rhie, H. Shen, T. Triche, D. J. Weisenberger, S. L. Carter, K. Cibulskis, L. Chin, J. Zhang, C. Sougnez, M. Wang, G. Getz, H. Dinh, H. V. Doddapaneni, R. Gibbs, P. Gunaratne, Y. Han, D. Kalra, C. Kovar, L. Lewis, M. Morgan, D. Morton, D. Muzny, J. Reid, L. Xi, J. Cho, D. Dicara, S. Frazer, N. Gehlenborg, D. I. Heiman, J. Kim, M. S. Lawrence, P. Lin, Y. Liu, M. S. Noble, P. Stojanov, D. Voet, H. Zhang,

L. Zou, C. Stewart, B. Bernard, R. Bressler, A. Eakin, L. Iype, T. Knijnenburg, R. Kramer, R. Kreisberg, K. Leinonen, J. Lin, Y. Liu, M. Miller, S. M. Reynolds, H. Rovira, I. Shmulevich, V. Thorsson, D. Yang, W. Zhang, S. Amin, C. J. Wu, C. C. Wu, R. Akbani, K. Aldape, K. A. Baggerly, B. Broom, T. D. Casasent, J. Cleland, D. Dodda, M. Edgerton, L. Han, S. M. Herbrich, Z. Ju, H. Kim, S. Lerner, J. Li, H. Liang, W. Liu, P. L. Lorenzi, Y. Lu, J. Melott, L. Nguyen, X. Su, R. Verhaak, W. Wang, A. Wong, Y. Yang, J. Yao, R. Yao, K. Yoshihara, Y. Yuan, A. K. Yung, N. Zhang, S. Zheng, M. Ryan, D. W. Kane, B. A. Aksoy, G. Ciriello, G. Dresdner, J. Gao, B. Gross, A. Jacobsen, A. Kahles, M. Ladanyi, W. Lee, K. V. Lehmann, M. L. Miller, R. Ramirez, G. Rätsch, B. Reva, N. Schultz, Y. Senbabaoglu, R. Shen, R. Sinha, S. O. Sumer, Y. Sun, B. S. Taylor, N. Weinhold, S. Fei, P. Spellman, C. Benz, D. Carlin, M. Cline, B. Craft, M. Goldman, D. Haussler, S. Ma, S. Ng, E. Paull, A. Radenbaugh, S. Salama, A. Sokolov, T. Swatloski, V. Uzunangelov, P. Waltman, C. Yau, J. Zhu, S. R. Hamilton, S. Abbott, R. Abbott, N. D. Dees, K. Delehaunty, L. Ding, D. J. Dooling, J. M. Eldred, C. C. Fronick, R. Fulton, L. L. Fulton, J. Kalicki-Veizer, K. L. Kanchi, C. Kandoth, D. C. Koboldt, D. E. Larson, T. J. Ley, L. Lin, C. Lu, V. J. Magrini, E. R. Mardis, M. D. McLellan, J. F. McMichael, C. A. Miller, M. O'Laughlin, C. Pohl, H. Schmidt, S. M. Smith, J. Walker, J. W. Wallis, M. C. Wendl, R. K. Wilson, T. Wylie, Q. Zhang, R. Burton, M. A. Jensen, A. Kahn, T. Pihl, D. Pot, Y. Wan, D. A. Levine, A. D. Black, J. Bowen, J. Frick, J. M. Gastier-Foster, H. A. Harper, C. Helsel, K. M. Leraas, T. M. Lichtenberg, C. McAllister, N. C. Ramirez, S. Sharpe, L. Wise, E. Zmuda, S. J. Chanock, T. Davidsen, J. A. Demchok, G. Eley, I. Felau, M. Sheth, H. Sofia, L. Staudt, R. Tarnuzzer, Z. Wang, L. Yang, J. Zhang, L. Omberg, A. Margolin, B. J. Raphael, F. Vandin, H. T. Wu, M. D. Leiserson, S. C. Benz, C. J. Vaske, H. Noushmehr, D. Wolf, L. V. T. Veer, D. Anastassiou, T. H. O. Yang, N. Lopez-Bigas, A. Gonzalez-Perez, D. Tamborero, Z. Xia, W. Li, D. Y. Cho, T. Przytycka, M. Hamilton, S. McGuire, S. Nelander, P. Johansson, R. Jörnsten, T. Kling, The cancer genome atlas pan-cancer analysis project (10 2013). doi:10.1038/ng.2764.

[33] D. A. Bennett, J. A. Schneider, Z. Arvanitakis, R. S. Wilson, Overview and Findings from the Religious Orders Study, Current Alzheimer Research 9 (6) (2013) 628–645. doi:10.2174/156720512801322573.

[34] D. A. Bennett, J. A. Schneider, A. S. Buchman, L. L. Barnes, P. A. Boyle, R. S. Wilson, Overview and Findings from the Rush Memory and Aging Project, Current Alzheimer Research 9 (6) (2013) 646–663. doi:10.2174/156720512801322663.
URL http://www.ncbi.nlm.nih.gov/pubmed/22471867http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3439198

[35] P. L. De Jager, J. M. Shulman, L. B. Chibnik, B. T. Keenan, T. Raj, R. S. Wilson, L. Yu, S. E. Leurgans, D. Tran, C. Aubin, C. D. Anderson, A. Biffi, J. J. Corneveaux, M. J. Huentelman, J. Rosand, M. J. Daly, A. J. Myers, E. M. Reiman, D. A. Bennett, D. A. Evans, A genome-wide scan for common variants affecting the rate of age-related cognitive decline, Neurobiology of Aging 33 (5) (2012) 1–1017.

doi:10.1016/j.neurobiolaging.2011.09.033.
URL http://www.ncbi.nlm.nih.gov/pubmed/22054870http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3307898

[36] P. L. De Jager, Y. Ma, C. McCabe, J. Xu, B. N. Vardarajan, D. Felsky, H. U. Klein, C. C. White, M. A. Peters, B. Lodgson, P. Nejad, A. Tang, L. M. Mangravite, L. Yu, C. Gaiteri, S. Mostafavi, J. A. Schneider, D. A. Bennett, Data descriptor: A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research, Scientific Data 5. doi:10.1038/sdata.2018.142.

[37] Q. Guo, M. K. Schmidt, P. Kraft, S. Canisius, C. Chen, S. Khan, J. Tyrer, M. K. Bolla, Q. Wang, J. Dennis, K. Michailidou, M. Lush, S. Kar, J. Beesley, A. M. Dunning, M. Shah, K. Czene, H. Darabi, M. Eriksson, D. Lambrechts, C. Weltens, K. Leunen, S. E. Bojesen, B. G. Nordestgaard, S. F. Nielsen, H. Flyger, J. Chang-Claude, A. Rudolph, P. Seibold, D. Flesch-Janys, C. Blomqvist, K. Aittomäki, R. Fagerholm, T. A. Muranen, F. J. Couch, J. E. Olson, C. Vachon, I. L. Andrulis, J. A. Knight, G. Glendon, A. M. Mulligan, A. Broeks, F. B. Hogervorst, C. A. Haiman, B. E. Henderson, F. Schumacher, L. Le Marchand, J. L. Hopper, H. Tsimiklis, C. Apicella, M. C. Southey, A. Cox, S. S. Cross, M. W. R. Reed, G. G. Giles, R. L. Milne, C. McLean, R. Winqvist, K. Pylkäs, A. Jukkola-Vuorinen, M. Grip, M. J. Hooning, A. Hollestelle, J. W. M. Martens, A. M. W. van den Ouweland, F. Marme, A. Schneeweiss, R. Yang, B. Burwinkel, J. Figueroa, S. J. Chanock, J. Lissowska, E. J. Sawyer, I. Tomlinson, M. J. Kerin, N. Miller, H. Brenner, A. K. Dieffenbach, V. Arndt, B. Holleczek, A. Mannermaa, V. Kataja, V.-M. Kosma, J. M. Hartikainen, J. Li, J. S. Brand, K. Humphreys, P. Devilee, R. A. E. M. Tollenaar, C. Seynaeve, P. Radice, P. Peterlongo, B. Bonanni, P. Mariani, P. A. Fasching, M. W. Beckmann, A. Hein, A. B. Ekici, G. Chenevix-Trench, R. Balleine, K.-A. Phillips, J. Benitez, M. P. Zamora, J. I. Arias Perez, P. Menéndez, A. Jakubowska, J. Lubinski, K. Jaworska-Bieniek, K. Durda, U. Hamann, M. Kabisch, H. U. Ulmer, T. Rüdiger, S. Margolin, V. Kristensen, S. Nord, D. G. Evans, J. E. Abraham, H. M. Earl, L. Hiller, J. A. Dunn, S. Bowden, C. Berg, D. Campa, W. R. Diver, S. M. Gapstur, M. M. Gaudet, S. E. Hankinson, R. N. Hoover, A. Hüsing, R. Kaaks, M. J. Machiela, W. Willett, M. Barrdahl, F. Canzian, S.-F. Chin, C. Caldas, D. J. Hunter, S. Lindstrom, M. García-Closas, P. Hall, D. F. Easton, D. M. Eccles, N. Rahman, H. Nevanlinna, P. D. P. Pharoah, P. D. P. Pharoah, Identification of Novel Genetic Markers of Breast Cancer Survival, JNCI: Journal of the National Cancer Institute 107 (5). doi:10.1093/jnci/djv081.
URL http://www.ncbi.nlm.nih.gov/pubmed/25890600http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4555642https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djv081

[38] K. Michailidou, P. Hall, A. Gonzalez-Neira, M. Ghoussaini, J. Dennis, R. L. Milne, M. K. Schmidt, J. Chang-Claude, S. E. Bojesen, M. K. Bolla, Q. Wang, E. Dicks, A. Lee, C. Turnbull, N. Rahman, O. Fletcher, J. Peto, L. Gibson, I. dos Santos Silva, H. Nevanlinna, T. A. Muranen, K. Aittomäki,

C. Blomqvist, K. Czene, A. Irwanto, J. Liu, Q. Waisfisz, H. Meijers-Heijboer, M. Adank, R. B. van der Luijt, R. Hein, N. Dahmen, L. Beckman, A. Meindl, R. K. Schmutzler, B. Müller-Myhsok, P. Lichtner, J. L. Hopper, M. C. Southey, E. Makalic, D. F. Schmidt, A. G. Uitterlinden, A. Hofman, D. J. Hunter, S. J. Chanock, D. Vincent, F. Bacot, D. C. Tessier, S. Canisius, L. F. A. Wessels, C. A. Haiman, M. Shah, R. Luben, J. Brown, C. Luccarini, N. Schoof, K. Humphreys, J. Li, B. G. Nordestgaard, S. F. Nielsen, H. Flyger, F. J. Couch, X. Wang, C. Vachon, K. N. Stevens, D. Lambrechts, M. Moisse, R. Paridaens, M.-R. Christiaens, A. Rudolph, S. Nickels, D. Flesch-Janys, N. Johnson, Z. Aitken, K. Aaltonen, T. Heikkinen, A. Broeks, L. J. V. Veer, C. E. van der Schoot, P. Guénel, T. Truong, P. Laurent-Puig, F. Menegaux, F. Marme, A. Schneeweiss, C. Sohn, B. Burwinkel, M. P. Zamora, J. I. A. Perez, G. Pita, M. R. Alonso, A. Cox, I. W. Brock, S. S. Cross, M. W. R. Reed, E. J. Sawyer, I. Tomlinson, M. J. Kerin, N. Miller, B. E. Henderson, F. Schumacher, L. Le Marchand, I. L. Andrulis, J. A. Knight, G. Glendon, A. M. Mulligan, A. Lindblom, S. Margolin, M. J. Hooning, A. Hollestelle, A. M. W. van den Ouweland, A. Jager, Q. M. Bui, J. Stone, G. S. Dite, C. Apicella, H. Tsimiklis, G. G. Giles, G. Severi, L. Baglietto, P. A. Fasching, L. Haeberle, A. B. Ekici, M. W. Beckmann, H. Brenner, H. Müller, V. Arndt, C. Stegmaier, A. Swerdlow, A. Ashworth, N. Orr, M. Jones, J. Figueroa, J. Lissowska, L. Brinton, M. S. Goldberg, F. Labrèche, M. Dumont, R. Winqvist, K. Pylkäs, A. Jukkola-Vuorinen, M. Grip, H. Brauch, U. Hamann, T. Brüning, P. Radice, P. Peterlongo, S. Manoukian, B. Bonanni, P. Devilee, R. A. E. M. Tollenaar, C. Seynaeve, C. J. van Asperen, A. Jakubowska, J. Lubinski, K. Jaworska, K. Durda, A. Mannermaa, V. Kataja, V.-M. Kosma, J. M. Hartikainen, N. V. Bogdanova, N. N. Antonenkova, T. Dörk, V. N. Kristensen, H. Anton-Culver, S. Slager, A. E. Toland, S. Edge, F. Fostira, D. Kang, K.-Y. Yoo, D.-Y. Noh, K. Matsuo, H. Ito, H. Iwata, A. Sueta, A. H. Wu, C.-C. Tseng, D. Van Den Berg, D. O. Stram, X.-O. Shu, W. Lu, Y.-T. Gao, H. Cai, S. H. Teo, C. H. Yip, S. Y. Phuah, B. K. Cornes, M. Hartman, H. Miao, W. Y. Lim, J.-H. Sng, K. Muir, A. Lophatananon, S. Stewart-Brown, P. Siriwanarangsan, C.-Y. Shen, C.-N. Hsiung, P.-E. Wu, S.-L. Ding, S. Sangrajrang, V. Gaborieau, P. Brennan, J. McKay, W. J. Blot, L. B. Signorello, Q. Cai, W. Zheng, S. Deming-Halverson, M. Shrubsole, J. Long, J. Simard, M. Garcia-Closas, P. D. P. Pharoah, G. Chenevix-Trench, A. M. Dunning, J. Benitez, D. F. Easton, P. D. P. Pharoah, G. Chenevix-Trench, A. M. Dunning, J. Benitez, D. F. Easton, Large-scale genotyping identifies 41 new loci associated with breast cancer risk, Nature Genetics 45 (4) (2013) 353–361. doi:10.1038/ng.2563.
URL http://www.ncbi.nlm.nih.gov/pubmed/23535729http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3771688http://www.nature.com/articles/ng.2563

[39] K. Michailidou, J. Beesley, S. Lindstrom, S. Canisius, J. Dennis, M. J. Lush, M. J. Maranian, M. K. Bolla, Q. Wang, M. Shah, B. J. Perkins, K. Czene, M. Eriksson, H. Darabi, J. S. Brand, S. E. Bojesen, B. G. Nordestgaard, H. Flyger, S. F. Nielsen, N. Rahman, C. Turnbull, O. Fletcher, J. Peto, L. Gibson, I. dos Santos-Silva, J. Chang-Claude, D. Flesch-Janys, A. Rudolph, U. Eilber, S. Behrens, H. Nevanlinna, T. A. Muranen, K. Aittomäki, C. Blomqvist, S. Khan, K. Aaltonen,

H. Ahsan, M. G. Kibriya, A. S. Whittemore, E. M. John, K. E. Malone, M. D. Gammon, R. M. Santella, G. Ursin, E. Makalic, D. F. Schmidt, G. Casey, D. J. Hunter, S. M. Gapstur, M. M. Gaudet, W. R. Diver, C. A. Haiman, F. Schumacher, B. E. Henderson, L. Le Marchand, C. D. Berg, S. J. Chanock, J. Figueroa, R. N. Hoover, D. Lambrechts, P. Neven, H. Wildiers, E. van Limbergen, M. K. Schmidt, A. Broeks, S. Verhoef, S. Cornelissen, F. J. Couch, J. E. Olson, E. Hallberg, C. Vachon, Q. Waisfisz, H. Meijers-Heijboer, M. A. Adank, R. B. van der Luijt, J. Li, J. Liu, K. Humphreys, D. Kang, J.-Y. Choi, S. K. Park, K.-Y. Yoo, K. Matsuo, H. Ito, H. Iwata, K. Tajima, P. Guénel, T. Truong, C. Mulot, M. Sanchez, B. Burwinkel, F. Marme, H. Surowy, C. Sohn, A. H. Wu, C.-c. Tseng, D. Van Den Berg, D. O. Stram, A. González-Neira, J. Benitez, M. P. Zamora, J. I. A. Perez, X.-O. Shu, W. Lu, Y.-T. Gao, H. Cai, A. Cox, S. S. Cross, M. W. R. Reed, I. L. Andrulis, J. A. Knight, G. Glendon, A. M. Mulligan, E. J. Sawyer, I. Tomlinson, M. J. Kerin, N. Miller, A. Lindblom, S. Margolin, S. H. Teo, C. H. Yip, N. A. M. Taib, G.-H. Tan, M. J. Hooning, A. Hollestelle, J. W. M. Martens, J. M. Collée, W. Blot, L. B. Signorello, Q. Cai, J. L. Hopper, M. C. Southey, H. Tsimiklis, C. Apicella, C.-Y. Shen, C.-N. Hsiung, P.-E. Wu, M.-F. Hou, V. N. Kristensen, S. Nord, G. I. G. Alnaes, G. G. Giles, R. L. Milne, C. McLean, F. Canzian, D. Trichopoulos, P. Peeters, E. Lund, M. Sund, K.-T. Khaw, M. J. Gunter, D. Palli, L. M. Mortensen, L. Dossus, J.-M. Huerta, A. Meindl, R. K. Schmutzler, C. Sutter, R. Yang, K. Muir, A. Lophatananon, S. Stewart-Brown, P. Siriwanarangsan, M. Hartman, H. Miao, K. S. Chia, C. W. Chan, P. A. Fasching, A. Hein, M. W. Beckmann, L. Haeberle, H. Brenner, A. K. Dieffenbach, V. Arndt, C. Stegmaier, A. Ashworth, N. Orr, M. J. Schoemaker, A. J. Swerdlow, L. Brinton, M. Garcia-Closas, W. Zheng, S. L. Halverson, M. Shrubsole, J. Long, M. S. Goldberg, F. Labrèche, M. Dumont, R. Winqvist, K. Pylkäs, A. Jukkola-Vuorinen, M. Grip, H. Brauch, U. Hamann, T. Brüning, P. Radice, P. Peterlongo, S. Manoukian, L. Bernard, N. V. Bogdanova, T. Dörk, A. Mannermaa, V. Kataja, V.-M. Kosma, J. M. Hartikainen, P. Devilee, R. A. E. M. Tollenaar, C. Seynaeve, C. J. Van Asperen, A. Jakubowska, J. Lubinski, K. Jaworska, T. Huzarski, S. Sangrajrang, V. Gaborieau, P. Brennan, J. McKay, S. Slager, A. E. Toland, C. B. Ambrosone, D. Yannoukakos, M. Kabisch, D. Torres, S. L. Neuhausen, H. Anton-Culver, C. Luccarini, C. Baynes, S. Ahmed, C. S. Healey, D. C. Tessier, D. Vincent, F. Bacot, G. Pita, M. R. Alonso, N. Álvarez, D. Herrero, J. Simard, P. P. D. P. Pharoah, P. Kraft, A. M. Dunning, G. Chenevix-Trench, P. Hall, D. F. Easton, P. Kraft, A. M. Dunning, G. Chenevix-Trench, P. Hall, D. F. Easton, Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer, Nature Genetics 47 (4) (2015) 373–380. doi:10.1038/ng.3242.
URL http://www.ncbi.nlm.nih.gov/pubmed/25751625http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4549775http://www.nature.com/articles/ng.3242

[40] J. C. Lambert, C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, G. Jun, A. L. DeStefano, J. C. Bis, G. W. Beecham, B. Grenier-Boley, G. Russo, T. A. Thornton-Wells, N. Jones, A. V. Smith, V. Chouraki, C. Thomas, M. A. Ikram, D. Zelenika, B. N. Vardarajan, Y. Kamatani,

C. F. Lin, A. Gerrish, H. Schmidt, B. Kunkle, N. Fiévet, P. Amouyel, F. Pasquier, V. Deramecourt, R. F. De Bruijn, N. Amin, A. Hofman, C. M. Van Duijn, M. L. Dunstan, P. Hollingworth, M. J. Owen, M. C. O'Donovan, L. Jones, P. A. Holmans, V. Moskvina, J. Williams, C. Baldwin, L. A. Farrer, S. H. Choi, K. L. Lunetta, A. L. Fitzpatrick, T. B. Harris, B. M. Psaty, J. R. Gilbert, K. L. Hamilton-Nelson, E. R. Martin, M. A. Pericak-Vance, J. L. Haines, V. Gudnason, P. V. Jonsson, G. Eiriksdottir, M. T. Bihoreau, M. Lathrop, O. Valladares, L. B. Cantwell, L. S. Wang, G. D. Schellenberg, A. Ruiz, M. Boada, C. Reitz, R. Mayeux, A. Ramirez, W. Maier, O. Hanon, W. A. Kukull, J. D. Buxbaum, D. Campion, D. Wallon, D. Hannequin, P. K. Crane, E. B. Larson, T. Becker, C. Cruchaga, A. M. Goate, D. Craig, J. A. Johnston, B. Mc-Guinness, S. Todd, P. Passmore, C. Berr, K. Ritchie, O. L. Lopez, P. L. De Jager, D. Evans, S. Lovestone, P. Proitsi, J. F. Powell, L. Letenneur, P. Barberger-Gateau, C. Dufouil, J. F. Dartigues, F. J. Morón, D. C. Rubinsztein, P. St. George-Hyslop, K. Sleegers, K. Bettens, C. Van Broeckhoven, M. J. Huentelman, M. Gill, K. Brown, K. Morgan, M. I. Kamboh, L. Keller, L. Fratiglioni, R. Green, A. J. Myers, S. Love, E. Rogaeva, J. Gallacher, A. Bayer, J. Clarimon, A. Lleo, D. W. Tsuang, L. Yu, D. A. Bennett, M. Tsolaki, P. Bossù, G. Spalletta, J. Collinge, S. Mead, S. Sorbi, B. Nacmias, F. Sanchez-Garcia, M. C. Deniz Naranjo, N. C. Fox, J. Hardy, P. Bosco, R. Clarke, C. Brayne, D. Galimberti, M. Mancuso, F. Matthews, S. Moebus, P. Mecocci, M. Del Zompo, H. Hampel, A. Pilotto, M. Bullido, F. Panza, P. Caffarra, M. Mayhaus, S. Pichler, W. Gu, M. Riemenschneider, L. Lannfelt, M. Ingelsson, H. Hakonarson, M. M. Carrasquillo, F. Zou, S. G. Younkin, D. Beekly, V. Alvarez, E. Coto, C. Razquin, P. Pastor, I. Mateo, O. Combarros, K. M. Faber, T. M. Foroud, H. Soininen, M. Hiltunen, D. Blacker, T. H. Mosley, C. Graff, C. Holmes, T. J. Montine, J. I. Rotter, A. Brice, M. A. Nalls, J. S. Kauwe, E. Boerwinkle, R. Schmidt, D. Rujescu, C. Tzourio, M. M. Nöthen, L. J. Launer, S. Seshadri, Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease, Nature Genetics 45 (12) (2013) 1452–1458. doi:10.1038/ng.2802.

[41] N. R. Wray, S. Ripke, M. Mattheisen, M. Trzaskowski, E. M. Byrne, A. Abdellaoui, M. J. Adams, E. Agerbo, T. M. Air, T. M. Andlauer, S. A. Bacanu, M. Bækvad-Hansen, A. F. Beekman, T. B. Bigdeli, E. B. Binder, D. R. Blackwood, J. Bryois, H. N. Buttenschøn, J. Bybjerg-Grauholm, N. Cai, E. Castelao, J. H. Christensen, T. K. Clarke, J. I. Coleman, L. Colodro-Conde, B. Couvy-Duchesne, N. Craddock, G. E. Crawford, C. A. Crowley, H. S. Dashti, G. Davies, I. J. Deary, F. Degenhardt, E. M. Derks, N. DIrek, C. V. Dolan, E. C. Dunn, T. C. Eley, N. Eriksson, V. Escott-Price, F. H. F. Kiadeh, H. K. Finucane, A. J. Forstner, J. Frank, H. A. Gaspar, M. Gill, P. Giusti-Rodríguez, F. S. Goes, S. D. Gordon, J. Grove, L. S. Hall, E. Hannon, C. S. Hansen, T. F. Hansen, S. Herms, I. B. Hickie, P. Hoffmann, G. Homuth, C. Horn, J. J. Hottenga, D. M. Hougaard, M. Hu, C. L. Hyde, M. Ising, R. Jansen, F. Jin, E. Jorgenson, J. A. Knowles, I. S. Kohane, J. Kraft, W. W. Kretzschmar, J. Krogh, Z. Kutalik, J. M. Lane, Y. Li, Y. Li, P. A. Lind, X. Liu, L. Lu, D. J. MacIntyre, D. F. MacKinnon, R. M. Maier, W. Maier, J. Marchini, H. Mbarek, P. McGrath, P. McGuffin, S. E. Medland, D. Mehta, C. M. Middeldorp, E. Mihailov, Y. Milaneschi, L. Milani, J. Mill, F. M. Mondimore, G. W. Montgomery, S. Mostafavi,

N. Mullins, M. Nauck, B. Ng, M. G. Nivard, D. R. Nyholt, P. F. O'Reilly, H. Oskarsson, M. J. Owen, J. N. Painter, C. B. Pedersen, M. G. Pedersen, R. E. Peterson, E. Pettersson, W. J. Peyrot, G. Pistis, D. Posthuma, S. M. Purcell, J. A. Quiroz, P. Qvist, J. P. Rice, B. P. Riley, M. Rivera, S. Saeed Mirza, R. Saxena, R. Schoevers, E. C. Schulte, L. Shen, J. Shi, S. I. Shyn, E. Sigurdsson, G. B. Sinnamon, J. H. Smit, D. J. Smith, H. Stefansson, S. Steinberg, C. A. Stockmeier, F. Streit, J. Strohmaier, K. E. Tansey, H. Teismann, A. Teumer, W. Thompson, P. A. Thomson, T. E. Thorgeirsson, C. Tian, M. Traylor, J. Treutlein, V. Trubetskoy, A. G. Uitterlinden, D. Umbricht, S. Van Der Auwera, A. M. Van Hemert, A. Viktorin, P. M. Visscher, Y. Wang, B. T. Webb, S. M. Weinsheimer, J. Wellmann, G. Willemsen, S. H. Witt, Y. Wu, H. S. Xi, J. Yang, F. Zhang, V. Arolt, B. T. Baune, K. Berger, D. I. Boomsma, S. Cichon, U. Dannlowski, E. C. De Geus, J. R. Depaulo, E. Domenici, K. Domschke, T. Esko, H. J. Grabe, S. P. Hamilton, C. Hayward, A. C. Heath, D. A. Hinds, K. S. Kendler, S. Kloiber, G. Lewis, Q. S. Li, S. Lucae, P. F. Madden, P. K. Magnusson, N. G. Martin, A. M. McIntosh, A. Metspalu, O. Mors, P. B. Mortensen, B. Müller-Myhsok, M. Nordentoft, M. M. Nöthen, M. C. O'Donovan, S. A. Paciga, N. L. Pedersen, B. W. Penninx, R. H. Perlis, D. J. Porteous, J. B. Potash, M. Preisig, M. Rietschel, C. Schaefer, T. G. Schulze, J. W. Smoller, K. Stefansson, H. Tiemeier, R. Uher, H. Völzke, M. M. Weissman, T. Werge, A. R. Winslow, C. M. Lewis, D. F. Levinson, G. Breen, A. D. Børglum, P. F. Sullivan, Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression, Nature Genetics 50 (5) (2018) 668–681. doi:10.1038/s41588-018-0090-3.

[42] J. Z. Liu, Y. Erlich, J. K. Pickrell, Case-control association mapping by proxy using family history of disease, Nature Genetics 49 (3) (2017) 325–331. doi:10.1038/ng.3766.
URL http://www.ncbi.nlm.nih.gov/pubmed/28092683

[43] A. A. Shabalin, Gene expression Matrix eQTL: ultra fast eQTL analysis via large matrix operations, Bioinformatics (Oxford, England) 28 (10) (2012) 1353–1358. doi:10.1093/bioinformatics/bts163.
URL http://www.ncbi.nlm.nih.gov/pubmed/22492648http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3348564http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL

[44] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple, Tech. Rep. 1 (1995).
URL https://www.jstor.org/stable/pdf/2346101.pdf?refreqid=excelsior%3A6411207ed4b0feb82c3964cc8b8151cb

[45] C. Reitz, Genetic loci associated with Alzheimer's disease, Future Neurology 9 (2) (2014) 119–122. doi:10.2217/fnl.14.1.
URL http://www.ncbi.nlm.nih.gov/pubmed/25530720http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4266532

[46] R. Sims, S. J. Van Der Lee, A. C. Naj, C. Bellenguez, N. Badarinarayan, J. Jakobsdottir, B. W. Kunkle, A. Boland, R. Raybould, J. C. Bis, E. R. Martin, B. Grenier-Boley, S. Heilmann-Heimbach, V. Chouraki, A. B. Kuzma, K. Sleegers, M. Vronskaya, A. Ruiz, R. R. Graham, R. Olaso, P. Hoffmann, M. L. Grove, B. N. Vardarajan, M. Hiltunen, M. M. Nöthen, C. C. White, K. L. Hamilton-Nelson, J. Epelbaum, W. Maier, S. H. Choi, G. W. Beecham, C. Dulary, S. Herms, A. V. Smith, C. C. Funk, C. Derbois, A. J. Forstner, S. Ahmad, H. Li, D. Bacq, D. Harold, C. L. Satizabal, O. Valladares, A. Squassina, R. Thomas, J. A. Brody, L. Qu, P. Sánchez-Juan, T. Morgan, F. J. Wolters, Y. Zhao, F. S. Garcia, N. Denning, M. Fornage, J. Malamon, M. C. D. Naranjo, E. Majounie, T. H. Mosley, B. Dombroski, D. Wallon, M. K. Lupton, J. Dupuis, P. Whitehead, L. Fratiglioni, C. Medway, X. Jian, S. Mukherjee, L. Keller, K. Brown, H. Lin, L. B. Cantwell, F. Panza, B. McGuinness, S. Moreno-Grau, J. D. Burgess, V. Solfrizzi, P. Proitsi, H. H. Adams, M. Allen, D. Seripa, P. Pastor, L. A. Cupples, N. D. Price, D. Hannequin, A. Frank-García, D. Levy, P. Chakrabarty, P. Caffarra, I. Giegling, A. S. Beiser, V. Giedraitis, H. Hampel, M. E. Garcia, X. Wang, L. Lannfelt, P. Mecocci, G. Eiriksdottir, P. K. Crane, F. Pasquier, V. Boccardi, I. Henández, R. C. Barber, M. Scherer, L. Tarraga, P. M. Adams, M. Leber, Y. Chen, M. S. Albert, S. Riedel-Heller, V. Emilsson, D. Beekly, A. Braae, R. Schmidt, D. Blacker, C. Masullo, H. Schmidt, R. S. Doody, G. Spalletta, W. T. Jr, T. J. Fairchild, P. Bossù, O. L. Lopez, M. P. Frosch, E. Sacchinelli, B. Ghetti, Q. Yang, R. M. Huebinger, F. Jessen, S. Li, M. I. Kamboh, J. C. Morris, O. Sotolongo-Grau, M. J. Katz, C. Corcoran, M. Dunstan, A. Braddel, C. Thomas, A. Meggy, R. Marshall, A. Gerrish, J. Chapman, M. Aguilar, S. Taylor, M. Hill, M. D. Fairén, A. Hodges, B. Vellas, H. Soininen, I. Kloszewska, M. Daniilidou, J. Uphill, Y. Patel, J. T. Hughes, J. Lord, J. Turton, A. M. Hartmann, R. Cecchetti, C. Fenoglio, M. Serpente, M. Arcaro, C. Caltagirone, M. D. Orfei, A. Ciaramella, S. Pichler, M. Mayhaus, W. Gu, A. Lleó, J. Fortea, R. Blesa, I. S. Barber, K. Brookes, C. Cupidi, R. G. Maletta, D. Carrell, S. Sorbi, S. Moebus, M. Urbano, A. Pilotto, J. Kornhuber, P. Bosco, S. Todd, D. Craig, J. Johnston, M. Gill, B. Lawlor, A. Lynch, N. C. Fox, J. Hardy, R. L. Albin, L. G. Apostolova, S. E. Arnold, S. Asthana, C. S. Atwood, C. T. Baldwin, L. L. Barnes, S. Barral, T. G. Beach, J. T. Becker, E. H. Bigio, T. D. Bird, B. F. Boeve, J. D. Bowen, A. Boxer, J. R. Burke, J. M. Burns, J. D. Buxbaum, N. J. Cairns, C. Cao, C. S. Carlson, C. M. Carlsson, R. M. Carney, M. M. Carrasquillo, S. L. Carroll, C. C. Diaz, H. C. Chui, D. G. Clark, D. H. Cribbs, E. A. Crocco, C. Decarli, M. Dick, R. Duara, D. A. Evans, K. M. Faber, K. B. Fallon, D. W. Fardo, M. R. Farlow, S. Ferris, T. M. Foroud, D. R. Galasko, M. Gearing, D. H. Geschwind, J. R. Gilbert, N. R. Graff-Radford, R. C. Green, J. H. Growdon, R. L. Hamilton, L. E. Harrell, L. S. Honig, M. J. Huentelman, C. M. Hulette, B. T. Hyman, G. P. Jarvik, E. Abner, L. W. Jin, G. Jun, A. Karydas, J. A. Kaye, R. Kim, N. W. Kowall, J. H. Kramer, F. M. Laferla, J. J. Lah, J. B. Leverenz, A. I. Levey, G. Li, A. P. Lieberman, K. L. Lunetta, C. G. Lyketsos, D. C. Marson, F. Martiniuk, D. C. Mash, E. Masliah, W. C. McCormick, S. M. McCurry, A. N. McDavid, A. C. McKee, M. Mesulam, B. L. Miller, C. A. Miller, J. W. Miller, J. R. Murrell, A. J. Myers, S. O'Bryant, J. M. Olichney, V. S.

Pankratz, J. E. Parisi, H. L. Paulson, W. Perry, E. Peskind, A. Pierce, W. W. Poon, H. Potter, J. F. Quinn, A. Raj, M. Raskind, B. Reisberg, C. Reitz, J. M. Ringman, E. D. Roberson, E. Rogaeva, H. J. Rosen, R. N. Rosenberg, M. A. Sager, A. J. Saykin, J. A. Schneider, L. S. Schneider, W. W. Seeley, A. G. Smith, J. A. Sonnen, S. Spina, R. A. Stern, R. H. Swerdlow, R. E. Tanzi, T. A. Thornton-Wells, J. Q. Trojanowski, J. C. Troncoso, V. M. Van Deerlin, L. J. Van Eldik, H. V. Vinters, J. P. Vonsattel, S. Weintraub, K. A. Welsh-Bohmer, K. C. Wilhelmsen, J. Williamson, T. S. Wingo, R. L. Woltjer, C. B. Wright, C. E. Yu, L. Yu, F. Garzia, F. Golamaully, G. Septier, S. Engelborghs, R. Vandenberghe, P. P. De Deyn, C. M. Fernadez, Y. A. Benito, H. Thonberg, C. Forsell, L. Lilius, A. Kinhult-Stählbom, L. Kilander, R. Brundin, L. Concari, S. Helisalmi, A. M. Koivisto, A. Haapasalo, V. Dermecourt, N. Fievet, O. Hanon, C. Dufouil, A. Brice, K. Ritchie, B. Dubois, J. J. Himali, C. D. Keene, J. Tschanz, A. L. Fitzpatrick, W. A. Kukull, M. Norton, T. Aspelund, E. B. Larson, R. Munger, J. I. Rotter, R. B. Lipton, M. J. Bullido, A. Hofman, T. J. Montine, E. Coto, E. Boerwinkle, R. C. Petersen, V. Alvarez, F. Rivadeneira, E. M. Reiman, M. Gallo, C. J. O'Donnell, J. S. Reisch, A. C. Bruni, D. R. Royall, M. Dichgans, M. Sano, D. Galimberti, P. St George-Hyslop, E. Scarpini, D. W. Tsuang, M. Mancuso, U. Bonuccelli, A. R. Winslow, A. Daniele, C. K. Wu, O. Peters, B. Nacmias, M. Riemenschneider, R. Heun, C. Brayne, D. C. Rubinsztein, J. Bras, R. Guerreiro, A. Al-Chalabi, C. E. Shaw, J. Collinge, M. Tsolaki, J. Clarimón, R. Sussams, S. Lovestone, M. C. O'Donovan, M. J. Owen, T. W. Behrens, S. Mead, A. G. Uitterlinden, C. Holmes, C. Cruchaga, M. Ingelsson, D. A. Bennett, J. Powell, T. E. Golde, C. Graff, P. L. De Jager, K. Morgan, N. Ertekin-Taner, O. Combarros, B. M. Psaty, P. Passmore, S. G. Younkin, C. Berr, V. Gudnason, D. Rujescu, D. W. Dickson, J. F. Dartigues, A. L. Destefano, S. Ortega-Cubero, H. Hakonarson, D. Campion, M. Boada, J. K. Kauwe, L. A. Farrer, C. Van Broeckhoven, M. A. Ikram, L. Jones, J. L. Haines, C. Tzourio, C. Tzourio, V. Escott-Price, R. Mayeux, J. F. Deleuze, N. Amin, A. M. Goate, M. A. Pericak-Vance, P. Amouyel, C. M. Van Duijn, A. Ramirez, L. S. Wang, J. C. Lambert, S. Seshadri, J. Williams, G. D. Schellenberg, Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease, Nature Genetics 49 (9) (2017) 1373–1384. doi:10.1038/ng.3916.
URL http://www.ncbi.nlm.nih.gov/pubmed/28714976http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5669039

[47] X. Z. Yuan, S. Sun, C. C. Tan, J. T. Yu, L. Tan, The Role of ADAM10 in Alzheimer's Disease (2017). doi:10.3233/JAD-170061.
URL http://www.ncbi.nlm.nih.gov/pubmed/28409746

[48] X. Guo, W. Lin, J. Bao, Q. Cai, X. Pan, M. Bai, Y. Yuan, J. Shi, Y. Sun, M. R. Han, J. Wang, Q. Liu, W. Wen, B. Li, J. Long, J. Chen, W. Zheng, A Comprehensive cis-eQTL Analysis Revealed Target Genes in Breast Cancer Susceptibility Loci Identified in Genome-wide Association Studies, American Journal of Human Genetics 102 (5) (2018) 890–903. doi:10.1016/j.ajhg.2018.03.016.

[49] S. Khan, R. Fagerholm, L. Kadalayil, W. Tapper, K. Aittomäki, J. Liu, C. Blomqvist, D. Eccles, H. Nevanlinna, Meta-analysis of three genome-wide association studies identifies two loci that predict survival and treatment outcome in breast cancer., Oncotarget 9 (3) (2018) 4249–4257. doi:10.18632/oncotarget.22747.
URL http://www.ncbi.nlm.nih.gov/pubmed/29423119http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5790536

[50] A. Bhattacharya, M. García-Closas, A. F. Olshan, C. M. Perou, M. A. Troester, M. I. Love, A framework for transcriptome-wide association studies in breast cancer in diverse study populations, Genome Biology 21 (1) (2020) 42. doi:10.1186/s13059-020-1942-6.
URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-1942-6

[51] A. Quiroz-Zárate, B. J. Harshfield, R. Hu, N. Knoblauch, A. H. Beck, S. E. Hankinson, V. Carey, R. M. Tamimi, D. J. Hunter, J. Quackenbush, A. Hazra, Expression Quantitative Trait loci (QTL) in tumor adjacent normal breast tissue and breast tumor tissue, PLOS ONE 12 (2) (2017) e0170181. doi:10.1371/journal.pone.0170181.
URL https://dx.plos.org/10.1371/journal.pone.0170181

[52] M. J. Dörfel, G. J. Lyon, The biological functions of Naa10 - From amino-terminal acetylation to human disease (8 2015). doi:10.1016/j.gene.2015.04.085.

[53] I. Lambertz, C. Kumps, S. Claeys, S. Lindner, A. Beckers, E. Janssens, D. R. Carter, A. Cazes, B. B. Cheung, M. De Mariano, A. De Bondt, S. De Brouwer, O. Delattre, J. Gibbons, I. Janoueix-Lerosey, G. Eve Laureys, C. Liang, G. M. Marchall, M. Porcu, J. Takita, D. Camacho Trujillo, I. Van, D. Wyngaert, N. V. Roy, A. V. Goethem, T. Van Maerken, P. Zabrocki, J. Cools, J. H. Schulte, J. Vialard, F. Speleman, K. De Preter, Biology of Human Tumors Upregulation of MAPK Negative Feedback Regulators and RET in Mutant ALK Neuroblastoma: Implications for Targeted Treatmentdoi:10.1158/1078-0432.CCR-14-2024.
URL http://clincancerres.aacrjournals.org/

[54] B. Whitton, H. Okamoto, G. Packham, S. J. Crabb, Vacuolar ATPase as a potential therapeutic target and mediator of treatment resistance in cancer (8 2018). doi:10.1002/cam4.1594.

[55] M. Matsubara, M. J. Bissell, Inhibitors of Rho kinase (ROCK) signaling revert the malignant phenotype of breast cancer cells in 3D context, Oncotarget 7 (22) (2016) 31602–31622. doi:10.18632/oncotarget.9395.

[56] F. Chang, Y. Zhang, J. Mi, Q. Zhou, F. Bai, X. Xu, D. E. Fisher, Q. Sun, X. Wu, ROCK inhibitor enhances the growth and migration of BRAF-mutant skin melanoma cells, Cancer Science 109 (11) (2018) 3428–3437. doi:10.1111/cas.13786.
URL http://doi.wiley.com/10.1111/cas.13786

[57] Y. Ni, S. Seballos, B. Fletcher, T. Romigh, L. Yehia, J. Mester, L. Senter, F. Niazi, M. Saji, M. D. Ringel, T. Laframboise, C. Eng, G. H. Risk, Germline compound heterozygous poly-glutamine deletion in USF3 may be involved in predisposition to heritable and sporadic epithelial thyroid carcinomadoi:10.1093/hmg/ddw382.

[58] A. Gusev, N. Mancuso, H. Won, M. Kousi, H. K. Finucane, Y. Reshef, L. Song, A. Safi, S. McCarroll, B. M. Neale, R. A. Ophoff, M. C. O'Donovan, G. E. Crawford, D. H. Geschwind, N. Katsanis, P. F. Sullivan, B. Pasaniuc, A. L. Price, M. C. O'Donovan, G. E. Crawford, D. H. Geschwind, N. Katsanis, P. F. Sullivan, B. Pasaniuc, A. L. Price, Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights, Nature Genetics 50 (4) (2018) 538–548. doi:10.1038/s41588-018-0092-1.
URL http://www.nature.com/articles/s41588-018-0092-1

[59] T. Raj, Y. I. Li, G. Wong, J. Humphrey, M. Wang, S. Ramdhani, Y. C. Wang, B. Ng, I. Gupta, V. Haroutunian, E. E. Schadt, T. Young-Pearse, S. Mostafavi, B. Zhang, P. Sklar, D. A. Bennett, P. L. De Jager, Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility, Nature Genetics 50 (11) (2018) 1584–1592. doi:10.1038/s41588-018-0238-1.

[60] C. Blauwendraat, M. Francescatto, J. R. Gibbs, I. E. Jansen, J. Simón-Sánchez, D. G. Hernandez, A. A. Dillman, A. B. Singleton, M. R. Cookson, P. Rizzu, P. Heutink, Comprehensive promoter level expression quantitative trait loci analysis of the human frontal lobe, Genome Medicine 8 (1). doi:10.1186/s13073-016-0320-1.

[61] N. Y. A. Sey, B. Hu, W. Mah, H. Fauni, J. C. McAfee, P. Rajarajan, K. J. Brennand, S. Akbarian, H. Won, A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles, Nature Neuroscience (2020) 1–11doi:10.1038/s41593-020-0603-0.
URL http://www.nature.com/articles/s41593-020-0603-0

[62] L. M. Sng, P. C. Thomson, D. Trabzuni, Genome-wide human brain eQTLs: In-depth analysis and insights using the UKBEC dataset, Scientific Reports 9 (1) (2019) 1–21. doi:10.1038/s41598-019-55590-0.

[63] S. Nagpal, X. Meng, M. P. Epstein, L. C. Tsoi, M. Patrick, G. Gibson, P. L. De Jager, D. A. Bennett, A. P. Wingo, T. S. Wingo, J. Yang, TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances Gene Mapping of Complex Traits, American Journal of Human Genetics 105 (2) (2019) 258–266. doi:10.1016/j.ajhg.2019.05.018.

[64] B. Bischl, M. Lang, O. Mersmann, J. Rahnenfuhrer, C. Weihs, BatchJobs and BatchExperiments: Abstraction Mechanism for Using R in Batch Environments, The Journal of Statistical Software 64 (11) (2015) 1–25.

[65] J. Kö Ster, S. Rahmann, Genome analysis Snakemake -a scalable bioinformatics workflow engine, Bioinformatics 28 (2012) 2520–2522. doi:10.1093/bioinformatics/bts480.
URL http://snakemake.googlecode.com.

[66] F. Bao, P. R. LoVerso, J. N. Fisk, V. B. Zhurkin, F. Cui, p53 binding sites in normal and cancer cells are characterized by distinct chromatin context, Cell Cycle 16 (21) (2017) 2073–2085. doi:10.1080/15384101.2017.1361064.

[67] J. X. Zhou, Z. Isik, C. Xiao, I. Rubin, S. A. Kauffman, M. Schroeder, S. Huang, Systematic drug perturbations on cancer cells reveal diverse exit paths from proliferative state, Oncotarget 7 (7) (2016) 7415–7425. doi:10.18632/oncotarget.7294.

[68] D. A. J. Ahmad, O. H. Negm, M. L. Alabdullah, S. Mirza, M. R. Hamed, V. Band, A. R. Green, I. O. Ellis, E. A. Rakha, Clinicopathological and prognostic significance of mitogen-activated protein kinases (MAPK) in breast cancers, Breast Cancer Research and Treatment 159 (3) (2016) 457–467. doi:10.1007/s10549-016-3967-9.
URL http://link.springer.com/10.1007/s10549-016-3967-9

[69] M. M. Ryan, H. E. Lockstone, S. J. Huffaker, M. T. Wayland, M. J. Webster, S. Bahn, Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes, Molecular Psychiatry 11 (10) (2006) 965–978. doi:10.1038/sj.mp.4001875.

[70] K. Baker, S. L. Gordon, D. Grozeva, M. Van Kogelenberg, N. Y. Roberts, M. Pike, E. Blair, M. E. Hurles, W. K. Chong, T. Baldeweg, M. A. Kurian, S. G. Boyd, M. A. Cousin, F. L. Raymond, Identification of a human synaptotagmin-1 mutation that perturbs synaptic vesicle cycling, Journal of Clinical Investigation 125 (4) (2015) 1670–1678. doi:10.1172/JCI79765.

[71] S. Heyes, W. S. Pratt, E. Rees, S. Dahimene, L. Ferron, M. J. Owen, A. C. Dolphin, Genetic disruption of voltage-gated calcium channels in psychiatric and neurological disorders (11 2015). doi:10.1016/j.pneurobio.2015.09.002.

[72] Y. A. Savva, L. E. Rieder, R. A. Reenan, The ADAR protein family, Genome Biology 13 (12) (2012) 259. doi:10.1186/gb-2012-13-12-252.

[73] W. Slotkin, K. Nishikura, Adenosine-to-inosine RNA editing and human disease (11 2013). doi:10.1186/gm508.

[74] K. J. Preacher, J. P. Selig, Advantages of Monte Carlo Confidence Intervals for Indirect Effects, Communication Methods and Measures 6 (2012) 77–98. doi:10.1080/19312458.2012.679848.

[75] A. Saha, A. Battle, False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors [version 2; peer review: 3 approved], F1000Research 7 (2019) 1860. doi:10.12688/f1000research.17145.2.

List of Figures

1 *Modeling schemes implemented in MOSTWAS.* (A) Two-step least squares for mediator-enriched TWAS (MeTWAS). (B) Distal-eQTL prioritization by via mediation analysis (DePMA). 35

2 *Comparison of TWAS power via simulations using MOSTWAS and local-only models.* (A) Proportion of gene-trait associations at $P < 2.5 \times 10^{-6}$ using local-only (red) and the most predictive MOSTWAS (blue) models across various local and distal expression heritabilities, trait heritability, and causal proportions. (B) Proportion of significant gene-trait associations across the same simulation parameters with no distal effect on the trait in the simulated external GWAS panel. 36

3 *Predictive adjusted R^2 from cross-validation across local-only, MeTWAS, and DePMA models.* If a given gene does not have $h^2 > 0$ with $P < 0.05$, we set the predictive adjusted R^2 to 0 here for comparison. We compare local-only and MeTWAS in TCGA-BRCA (A) and ROS/MAP (D), local-only and DePMA in TCGA-BRCA (B) and ROS/MAP (E), and MeTWAS and DePMA in TCGA-BRCA (C) and ROS/MAP (F), all axes indicating the CV adjusted R^2 for different models. 37

4 *External validation of MOSTWAS and gene-trait associations using MOSTWAS models.* (A) Median predictive adjusted R^2 in held-out cohorts from TCGA-BRCA and ROS/MAP in local-only, MeTWAS, and DePMA models that have in-sample significant heritability. The interval shows the 25% and 75% quantiles for external cohort predictive R^2. (B) Associations with 11 known Alzheimer's risk lock, as identified in literature, using MOSTWAS, local-only, and TIGAR Dirichlet process regression (DPR). (C) TWAS associations for breast cancer-specific survival using GWAS summary statistics from iCOGs. Loci are colored and labelled if the overall association achieves FDR-adjusted $P < 0.05$. Loci are labelled with 'P' if the permutation test achieves FDR-adjusted $P < 0.05$ and 'D' if the added-last test achieves FDR-adjusted $P < 0.05$. (D) TWAS associations for major depressive disorder risk using GWAS summary statistics from PGC. Loci are colored red if the overall association achieves FDR-adjusted $P < 0.05$ and the permutation test also achieves FDR-adjusted $P < 0.05$. We label the 7 loci that were independently validated with UK Biobank GWAX summary statistics at FDR-adjusted $P < 0.05$ for both the overall association test and permutation test. 38

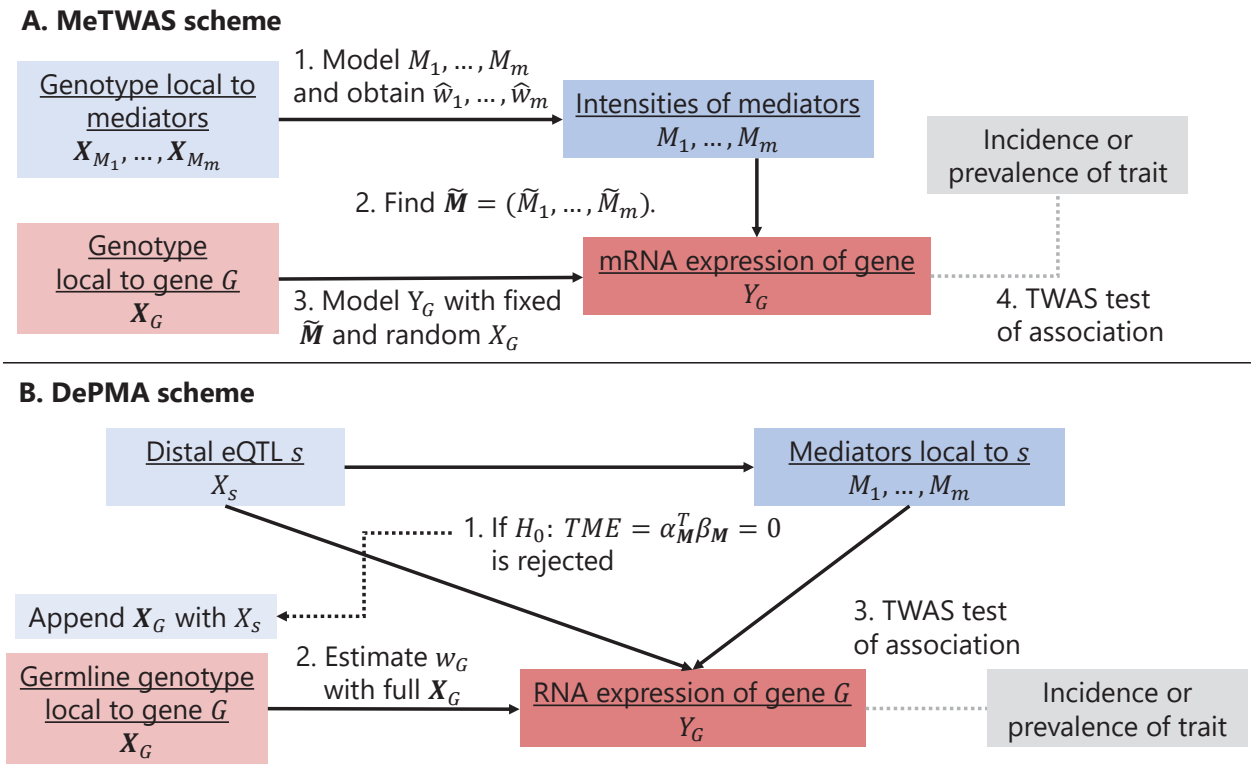

Figure 1: *Modeling schemes implemented in MOSTWAS.* (A) Two-step least squares for mediator-enriched TWAS (MeTWAS). (B) Distal-eQTL prioritization by via mediation analysis (DePMA).

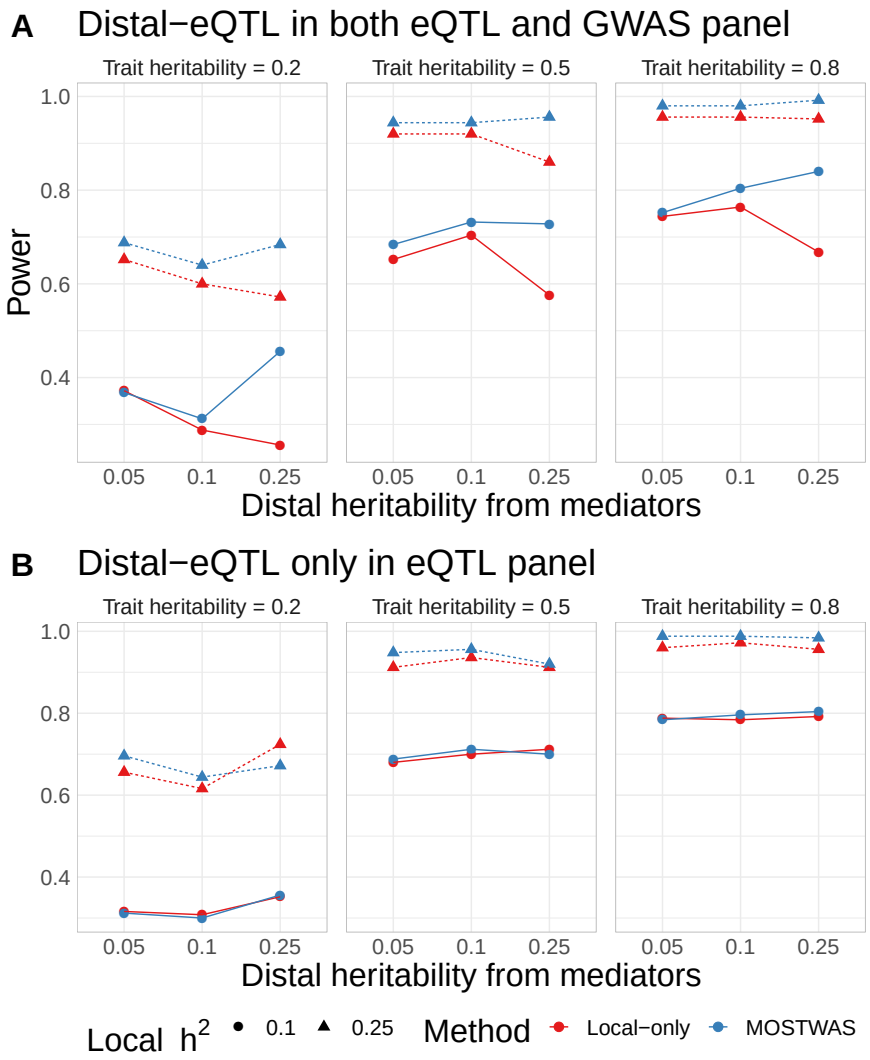

Figure 2: *Comparison of TWAS power via simulations using MOSTWAS and local-only models.* (A) Proportion of gene-trait associations at $P < 2.5 \times 10^{-6}$ using local-only (red) and the most predictive MOSTWAS (blue) models across various local and distal expression heritabilities, trait heritability, and causal proportions. (B) Proportion of significant gene-trait associations across the same simulation parameters with no distal effect on the trait in the simulated external GWAS panel.

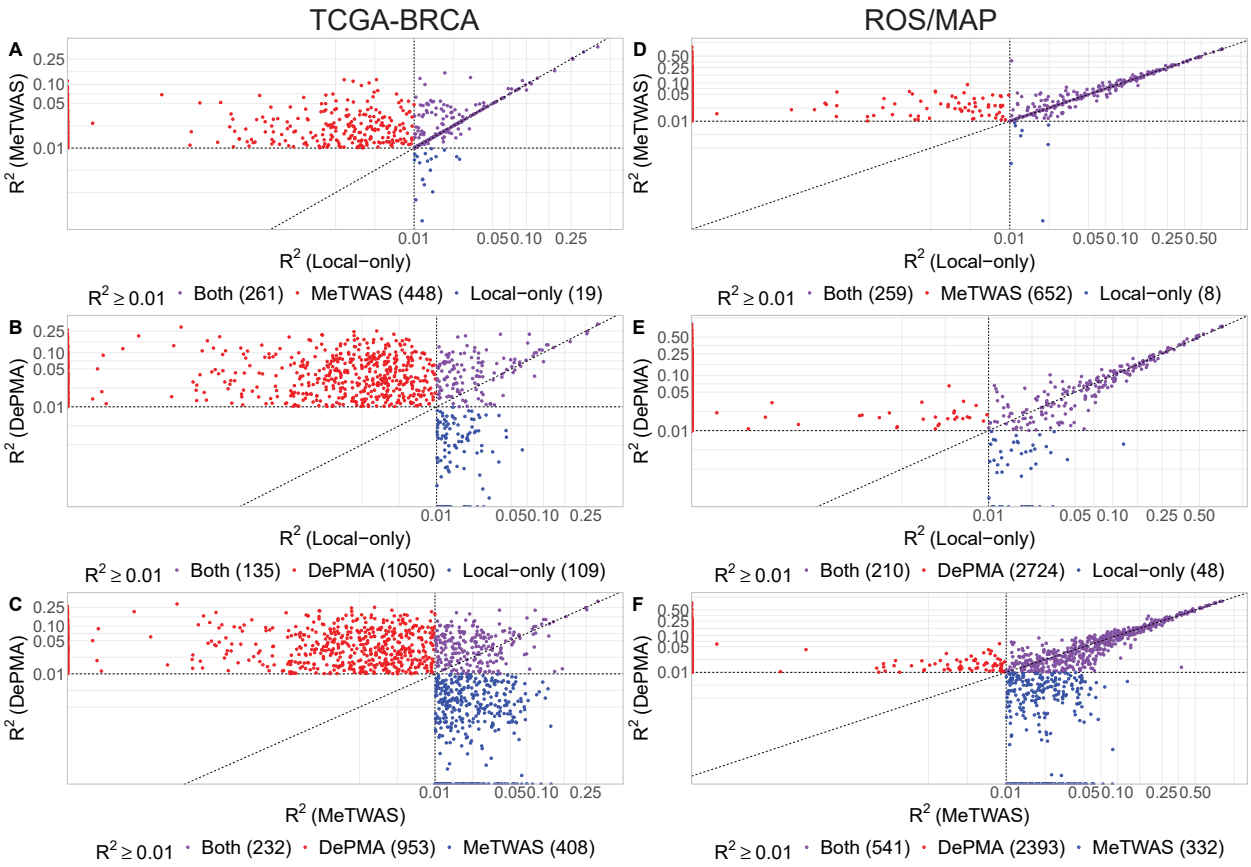

Figure 3: *Predictive adjusted R^2 from cross-validation across local-only, MeTWAS, and DePMA models.* If a given gene does not have $h^2 > 0$ with $P < 0.05$, we set the predictive adjusted R^2 to 0 here for comparison. We compare local-only and MeTWAS in TCGA-BRCA (A) and ROS/MAP (D), local-only and DePMA in TCGA-BRCA (B) and ROS/MAP (E), and MeTWAS and DePMA in TCGA-BRCA (C) and ROS/MAP (F), all axes indicating the CV adjusted R^2 for different models.

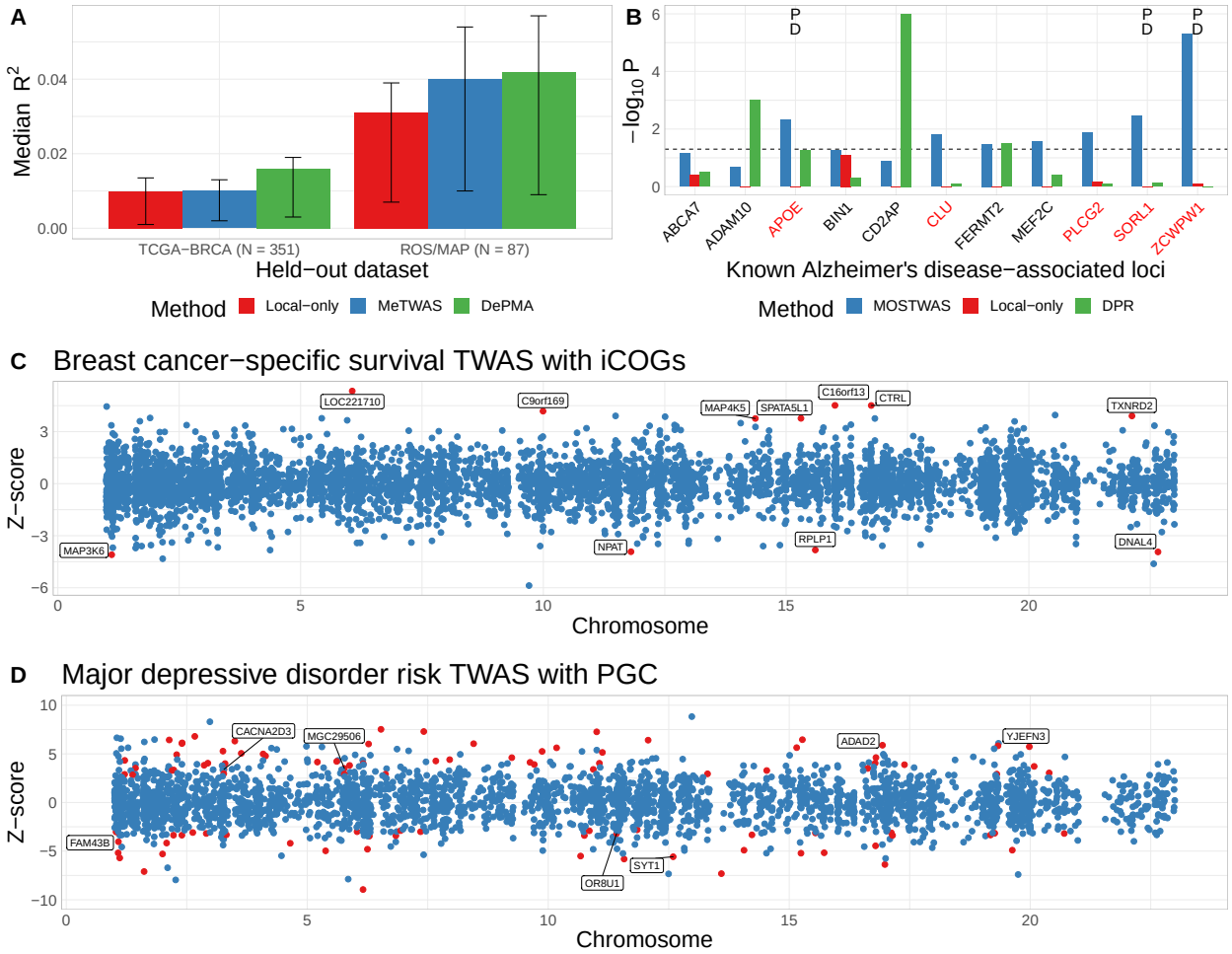

Figure 4: *External validation of MOSTWAS and gene-trait associations using MOSTWAS models.* (A) Median predictive adjusted R^2 in held-out cohorts from TCGA-BRCA and ROS/MAP in local-only, MeTWAS, and DePMA models that have in-sample significant heritability. The interval shows the 25% and 75% quantiles for external cohort predictive R^2. (B) Associations with 11 known Alzheimer's risk lock, as identified in literature, using MOSTWAS, local-only, and TIGAR Dirichlet process regression (DPR). (C) TWAS associations for breast cancer-specific survival using GWAS summary statistics from iCOGs. Loci are colored and labelled if the overall association achieves FDR-adjusted $P < 0.05$. Loci are labelled with 'P' if the permutation test achieves FDR-adjusted $P < 0.05$ and 'D' if the added-last test achieves FDR-adjusted $P < 0.05$. (D) TWAS associations for major depressive disorder risk using GWAS summary statistics from PGC. Loci are colored red if the overall association achieves FDR-adjusted $P < 0.05$ and the permutation test also achieves FDR-adjusted $P < 0.05$. We label the 7 loci that were independently validated with UK Biobank GWAX summary statistics at FDR-adjusted $P < 0.05$ for both the overall association test and permutation test.