

# UC San Diego

## UC San Diego Previously Published Works

### Title

A twin analysis to estimate genetic and environmental factors contributing to variation in weighted gene co-expression network module eigengenes.

### Permalink

<https://escholarship.org/uc/item/7351w6cp>

### Journal

American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 198(1)

### Authors

Gillespie, Nathan

Bell, Tyler

Hearn, Gentry

et al.

### Publication Date

2025

### DOI

10.1002/ajmg.b.33003

Peer reviewed



# HHS Public Access

Author manuscript

*Am J Med Genet B Neuropsychiatr Genet.* Author manuscript; available in PMC 2025 January 29.

Published in final edited form as:

*Am J Med Genet B Neuropsychiatr Genet.* 2025 January ; 198(1): e33003. doi:10.1002/ajmg.b.33003.

## A twin analysis to estimate genetic and environmental factors contributing to variation in weighted gene co-expression network module eigengenes

Nathan A. Gillespie<sup>1,2</sup>, Tyler R. Bell<sup>3,4</sup>, Gentry C. Hearn<sup>5</sup>, Jonathan L. Hess<sup>5</sup>, Ming T. Tsuang<sup>3</sup>, Michael J. Lyons<sup>5</sup>, Carol E. Franz<sup>3,4</sup>, William S. Kremen<sup>3,4</sup>, Stephen J. Glatt<sup>6</sup>

<sup>1</sup>Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Virginia, USA

<sup>2</sup>QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia

<sup>3</sup>Department of Psychiatry, University of California San Diego, La Jolla, California, USA

<sup>4</sup>Center for Behavior Genetics of Aging, University of California San Diego, La Jolla, California, USA

<sup>5</sup>Department of Psychological and Brain Sciences, Boston University, Boston, Massachusetts, USA

<sup>6</sup>Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, New York, USA

### Abstract

Multivariate network-based analytic methods such as weighted gene co-expression network analysis are frequently applied to human and animal gene-expression data to estimate the first principal component of a module, or module eigengene (ME). MEs are interpreted as multivariate summaries of correlated gene-expression patterns and network connectivity across genes within a module. As such, they have the potential to elucidate the mechanisms by which molecular genomic variation contributes to individual differences in complex traits. Although increasingly used to test for associations between modules and complex traits, the genetic and environmental etiology of MEs has not been empirically established. It is unclear if, and to what degree, individual differences in blood-derived MEs reflect random variation versus familial aggregation arising from heritable or shared environmental influences. We used biometrical genetic analyses

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

**Correspondence:** Nathan A. Gillespie, Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, VA, USA. [nathan.gillespie@vcuhealth.org](mailto:nathan.gillespie@vcuhealth.org). William S. Kremen and Stephen J. Glatt are joint senior authors.

#### AUTHOR CONTRIBUTIONS

Data collection: **Stephen J. Glatt, William S. Kremen**. Conceptualization: All authors. Funding acquisition: **Stephen J. Glatt, William S. Kremen**. Data analyses: **Nathan A. Gillespie, Tyler R. Bell, and Gentry C. Hearn**. Manuscript preparation: All authors.

#### CONFLICT OF INTEREST STATEMENT

All authors declare no conflict of interest.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

to estimate the contribution of genetic and environmental influences on MEs derived from blood lymphocytes collected on a sample of  $N = 661$  older male twins from the Vietnam Era Twin Study of Aging (VETSA) whose mean age at assessment was 67.7 years ( $SD = 2.6$  years, range = 62–74 years). Of the 26 detected MEs, 14 (56%) had statistically significant additive genetic variation with an average heritability of 44% ( $SD = 0.08$ , range = 35%–64%). Despite the relatively small sample size, this demonstration of significant family aggregation including estimates of heritability in 14 of the 26 MEs suggests that blood-based MEs are reliable and merit further exploration in terms of their associations with complex traits and diseases.

## Keywords

heritability; module eigengenes; weighted gene co-expression network analysis

## 1 | BACKGROUND

Gene expression (GE) is the first link in the chain connecting genetic polymorphisms to complex outcomes. GE is understood to play a critical role in many human diseases (Carroll, 2005; Emilsson et al., 2008; Kleinjan & Lettice, 2008; Kudaravalli et al., 2009; Schadt et al., 2005; Wray et al., 2007). Expression quantitative trait loci (eQTLs), which are single nucleotide polymorphisms (SNPs) regulating GE, provide a direct link between genome-wide association and GE studies (Franke & Jansen, 2009; Lehrmann & Freed, 2008). eQTL analysis can be useful for discerning transcriptome adaptations underlying complex behaviors (Bhattacharya & Mariani, 2009; Coppola & Geschwind, 2006; Gu et al., 2002). The identification of eQTLs located in transcription-factor-binding sites, splice sites, or other regulatory regions is likely to reveal mechanisms by which genetic variants contribute to individual differences in complex behaviors (Hindorff et al., 2009; Liu, 2011; Shastri, 2009). Yet, GE is also influenced by the environment, via epigenetic modifications. Thus, GE can be conceptualized as the primordial phenotype, and although likely influenced by genetic (G) and environmental (E) main effects as well as gene-by-environment interactions, the relative influence of genes and environment on GE is largely unknown.

A drawback of transcriptomic analysis is that functional genes with smaller effects on phenotypes may be undetectable at the single-gene level, particularly when relying on underpowered univariate statistics with high Type-I error rates (Langfelder et al., 2013; Langfelder & Horvath, 2008). Such analyses are also unlikely to capture biological interactions between genes, which seem likely to occur in complex traits (Horvath et al., 2006; Zhi et al., 2013). Indeed, notwithstanding questions regarding the direction of causation, dysregulation of disease-related gene networks is believed to be associated with many neurodevelopmental and psychiatric disorders (Oldham et al., 2008).

These limitations can be addressed with multivariate network-based analytic methods such as weighted gene co-expression network analysis (WGCNA) (Zhang & Horvath, 2005). WGCNA is a well-validated method for identifying genes associated with traits (DiLeo et al., 2011; Fuller et al., 2007; Langfelder & Horvath, 2008; MacLennan et al., 2009; Mason et al., 2009; Plaisier et al., 2009; Winden et al., 2011; Zhang & Horvath, 2005). It works by

classifying a set of genes with correlated expression and connectivity patterns into ‘network modules’. Each of these networks is summarized via a single quantitative metric called a ‘module eigengene’ (ME). Each ME is essentially the first principal component of a module, and as such, it summarizes some of the correlated expression across all the genes in the module. Yet, despite the increasing application of MEs in human and animal studies for exploring associations with complex behaviors and disorders (Chen et al., 2021; Liu et al., 2017; Wang et al., 2021; Zeng et al., 2020), the genetic and environmental etiology of MEs has never been empirically established. For instance, it is unclear to what extent individual differences in MEs derived from blood lymphocyte transcriptomes are simply a reflection of random variation (including measurement error) versus familial aggregation attributable to either genetic or shared environmental influences. To address this gap, we used biometrical genetic analyses that rely on human twin data to explore the genetic and environmental variation in MEs derived from blood lymphocytes collected on a sample of older male twins.

## 2 | MATERIALS AND METHODS

### 2.1 | Subjects

Participants comprise middle-aged male twins from the Vietnam Era Twin Study of Aging (VETSA) (Kremen et al., 2013). The VETSA ascertained monozygotic and dizygotic twin-pairs who were concordant for US military service at some time between 1965 and 1975, although nearly 80% reported no combat experience. Based on data from the US National Center for Health Statistics, the sample is very similar to American men in their age cohort with respect to health and lifestyle characteristics (Schoeneborn & Heyman, 2009). The sample is 88.3% non-Hispanic white, 5.3% African-American, 3.4% Hispanic, and 3.0% “other” participants.

To date, VETSA twins have been assessed on three occasions, each 5–6 years apart. Briefly, Wave 1 took place between 2001 and 2007 (Kremen et al., 2006) (mean age = 55.9, SD = 2.4, range = 51.1–60.7). Wave 2 occurred approximately 5.7 years later (mean age = 61.7, SD = 2.5, range = 56.0–67.0). Wave 3 occurred a further 5.9 years later (mean age = 67.6, SD = 2.5, range = 61.4–73.3). Data for this study came from a subset of Wave 3 subjects ( $N = 695$ ) who were approached to participate in a gene expression (GE) study. Here, the average age of subjects at the time of blood draw was 67.69 years (SD = 2.6 years, range = 62–74 years). Subjects who participated in the GE study were marginally older than non-participants (67.4 vs. 67.7,  $t = -2.03$ ,  $df = 1157.20$ ,  $p$ -value = 0.04).

### 2.2 | Ethics approval

Written informed consent was obtained from all participants. The University of California, San Diego (UCSD) and Boston University institutional review boards approved the proposal to collect these data (Projects 4013E, 150572, 150572, 150537, 140361, 071446, 031639, and 151333). Data are publicly available through requests at the VETSA website (<http://www.vetsatwins.org>).

### 2.3 | Blood collection, RNA extraction, and quality control

PAXgene tubes were used to collect 9 mL of blood in total required for the RNA isolation. Total RNA was extracted from PAXgene tubes using whole blood gene RNA purification kits (QIAGEN). RNA from all samples was run on an Agilent Bioanalyzer to assess RNA quality and estimate RNA concentrations. Post-RNA quality assessment included total RNA conversion to cDNA, amplification, and purification using the Ambion Illumina TotalPrep RNA Amplification Kit (Ambion).

### 2.4 | RNA-seq analysis

RNA sequencing was performed on Illumina NovaSeq 6000 Sequencing System at the University of California, San Francisco and the UCSD IGM Genomics Center. Raw sequencing data in *Fastq* format (Babraham Bioinformatics, 2011) were transferred to SUNY Upstate Medical University and processed as follows.

Samples were sequenced on eight 96-well plates. An average 23 million paired-end 151 base pair reads per sample were obtained from each run. For quality control, three samples were duplicated on every plate. Sequencing quality was accessed by *fastQC* v0.11.8, with multiple quality statistics collected and evaluated at each processing step. One plate failed quality checks for 49 out of 96 samples and was discarded; all samples from this plate were re-sequenced as the eighth plate.

Low-quality bases and reads were filtered out of *Fastq* files by *trimmomatic* v0.39 (Bolger et al., 2014) with a quality score cutoff of 30. Trimmed reads were then mapped to the *Gencode* GRCh38.p12 release 30 human reference genome using *STAR aligner* v2.7.2b (Dobin et al., 2013). Mapping statistics were collected from alignment results using *PicardTools* v2.20.8 (Broad Institute, 2018). Reads were then summarized using *featureCounts* in *subread* v1.6.4 (Liao et al., 2014), and raw count data were combined and annotated using customized *R* scripts.

Transcripts, including protein-coding genes, miRNAs, and lncRNAs, were filtered to include only those with greater than or equal to one count per million in at least 50% of samples. Counts of zero were replaced with 0.5 and data were normalized to effective library size via the *TMM* method in *edgeR* v3.28.1 (Robinson et al., 2010). Batch effects were reduced using *ComBat* (Wang et al., 2017).

### 2.5 | Weighted gene co-expression network analysis

We performed weighted gene co-expression network analysis (WGCNA) (Zhang & Horvath, 2005) on the GE data to identify the GE networks using the WGCNA v1.36 package in *R* (Langfelder et al., 2013; Langfelder & Horvath, 2008; Zhang & Horvath, 2005). This approach identifies higher-order interactions between genes by assembling the GE data into 'network modules'. Each network module in WGCNA was then represented using a 'module eigengene' (ME), which is a quantitative summary of the correlated expression and connectedness across all the genes within each module.

The *goodsamplesgenes()* function identifies genes and samples with low variance or too much missingness, and excludes these from network construction. This function was run to

ensure suitability of the features and subjects, detecting no samples or genes to be removed. Both functions relied on default parameters. The *blockwiseModules()* function was then used to create a signed co-expression network. Soft threshold power was set at 12 following the developer's recommendation (Langfelder & Horvath, 2008) *deepSplit* was set to 2, with a minimum module size of 30. Using these parameters, expression data for 4571 transcripts were organized into 26 modules with sizes ranging from 114 to 1200 transcripts. The 26 modules include a 'gray' or unclassified bin containing genes not reliably connected (ME0).

## 2.6 | Residualization and descriptive statistics

MEs were generated for 692 VETSA subjects. However, to improve the distribution, all ME means and variances were residualized for the effects of testing site, time between blood draw and storage, and storage times at  $-20^{\circ}\text{C}$  and  $-80^{\circ}\text{C}$  using the *umx\_residualize* function (version 4.9.0) in R (version 4.1.1). A total of 31 cases were lost during residualization due to missing covariates. Another two subjects were removed due to missing zygosity. Descriptive statistics for each ME following residualization are shown in Table 1. Among the 661 individuals with complete MEs there were 160 monozygotic twin pairs, 67 monozygotic singletons (incomplete pairs), 100 dizygotic twin pairs, and 74 dizygotic singletons.

## 2.7 | Statistical analyses

The *OpenMx*<sub>2.9.9.1</sub> software package (Boker et al., 2011) in *R*<sub>3.4.1</sub> (R Development Core Team, 2018) was used to fit univariate biometric genetic twin models (Neale & Cardon, 1992) to estimate the relative contribution of genetic and environmental influences in each ME. Given the number of incomplete twin pairs, methods such as weighted least squares will result in significant listwise deletion thereby altering the accuracy of each ME's mean and variance. Fortunately, the raw data full information maximum likelihood (FIML) option in *OpenMx*<sub>2.9.9.1</sub> (Boker et al., 2011) is robust to violations of non-normality while enabling analysis of missing or incomplete data as well as the direct estimation of covariate effects. We included age at blood draw as a covariate. This enables accurate estimates of means thereby improving estimation of the variances and covariance structure used to test our competing hypotheses regarding the genetic and environmental etiology of each ME.

## 2.8 | Univariate analyses

In univariate analyses, the total variation in each of the 26 MEs was decomposed into additive (A) genetic (heritability), shared or common environmental (C), and non-shared or unique (E) environmental variance components (see Figure 1). This approach is referred to as an 'ACE' variance components model. The decomposition is achieved by exploiting the expected genetic and environmental correlations between monozygotic (MZ) and dizygotic (DZ) twin pairs. MZ twin pairs are genetically identical, whereas DZ twin pairs share, on average, half of their genes. Therefore, the MZ and DZ twin pair correlations for the additive genetic (A) effects are fixed to  $r_A = 1.0$  and  $r_A = 0.5$ , respectively. The model also assumes that shared or common (C) environmental effects are equal for MZ and DZ twin pairs (also known as the equal environment assumption), that there are no genotype-by-environmental interactions or correlations, and that mating is random. Therefore, the MZ and DZ twin pair correlations for these 'C' effects are both fixed to one, that is,  $r_C = 1$ . Finally, since all non-

shared environmental influences (E), which include measurement error, are by definition uncorrelated, the MZ and DZ twin pair correlation ( $r_E$ ) for these influences is fixed to zero (Neale & Cardon, 1992). Shared environmental effects are defined as those that make siblings similar. Non-shared environmental factors are defined as those that make siblings different; this term also includes measurement error. Note that these variance components are derived from the association between phenotypic and genotypic similarities of the twin pairs.

## 2.9 | Model fit

For each univariate analysis, we fitted an A + C + E model before determining the most likely sources of variance by fitting three additional sub-models in which the (i) A, (ii) C, and (iii) A and C influences were fixed to zero. In other words, the significance of each nested sub-model (i.e., 'AE', 'CE', and 'E') was determined using the change in minus twice the log-likelihood ( $-2LL$ ) when compared to the full ACE model. Under certain regularity conditions, this  $-2LL$  is asymptotically distributed as chi-squared with the degrees of freedom equal to the difference in the number of free parameters in the two models (Steiger et al., 1985). Our determination of the best-fitting model was also based on the optimal balance of complexity and explanatory power by selecting the model with the lowest Akaike's information criterion (AIC) value (Akaike, 1987).

## 3 | RESULTS

### 3.1 | Twin pair correlations

Table 1 shows the twin pair correlations by zygosity for all 26 MEs. If familial aggregation is entirely attributable to shared environmental influences, then the MZ and DZ twin pair correlations will be statistically equal. If, however, familial aggregation is attributable to shared additive (or non-additive) genetic factors, then DZ correlations will be less than  $\frac{1}{2}$  the MZ twin pair correlations. Finally, if the hypothesis is that blood-based MEs are unreliable or that variability arises entirely from either random influences that are unshared between siblings or from measurement error, then both MZ and DZ twin pair correlations will not statistically differ from zero. For ME4, ME6, ME8, MEs10–11, MEs13–14, ME17 and ME0, the DZ twin pair correlations were greater than  $\frac{1}{2}$  of their MZ twin pair counterparts, which is consistent with familial aggregation attributable to shared or common environmental (C) influences. In contrast, DZ twin pair correlations for MEs ME3, ME7, ME12, MEs18–19, ME21, and ME23 were approximately  $\frac{1}{2}$  ( $\pm 0.05$ ) of their MZ twin pair counterparts, which is consistent with familial aggregation driven by additive genetic influences. For the remaining MEs (ME1–2, ME9, MEs15–ME16, ME20, ME22, and MEs24–25), the DZ twin pair correlations were well below their MZ twin pair counterparts, which is consistent with non-additivity and dominance (D) influencing familial aggregation. In the classical twin design, 'C' and 'D' influences are negatively confounded, and therefore, cannot be modeled simultaneously (Martin et al., 1978). Since the sample sizes required to detect 'D' as a source of variation are very large, even for variables measured on a continuous liability scale, we chose to model 'C' influences in all subsequent univariate and multivariate models below.

### 3.2 | Twin modeling

Table S1 shows the full univariate model fitting results, including comparisons between the ‘ACE’ and the ‘AE’, ‘CE’, and ‘E’ sub-models. Table S1 also includes 95% confidence intervals for all A, C, and E variance components along with 95% confidence intervals. Among the 26 MEs examined, there were 15 MEs with evidence of significant familial aggregation attributable to either additive (A) genetic or common (C) environmental influences (see Table 2).

### 3.3 | MEs with significant additive genetic familial aggregation

There were 14 MEs (ME1, ME2, ME5, ME7, ME9, ME16, and MEs18–25) in which the 95% confidence intervals surrounding the additive genetic ‘A’ point estimates in the saturated ACE model did not include zero (Table S1). For these MEs, dropping the ‘A’ influences resulted in a significant change in chi-square, whereas removing the common ‘C’ environmental influences (i.e., the AE model) improved the model fit in terms of a non-significant change in chi-square and producing the lowest AIC value. As illustrated in Figure 2, additive genetic influences explained 35%–64% of the total variance across these 14 MEs (mean = 45%, median = 43%, and standard deviation = 9%).

**3.3.1 | MEs with significant common environment familial aggregation—**There were two MEs (ME10 and ME13) in which the 95% confidence intervals surrounding the ‘C’ point estimates in the saturated ACE univariate model did not include zero. In both cases, dropping the ‘C’ influences resulted in a significant change in chi-square, whereas removing the ‘A’ influences (i.e., the CE model) improved the model fit in terms of a non-significant change in chi-square. Common environmental influences explained 42%–54% of the total variances in ME10 and ME13, respectively.

In small twin studies, estimates of ‘C’ should always be interpreted in the context of statistical power. We, therefore, performed post-hoc power calculations using the *umx* package *power.ACE.test* (Bates et al., 2019) to determine the minimum amount of ‘C’ that could be detected at 80% power (see Appendix S1). Given the current sample size, we estimated that there was 80% power to detect estimates of ‘C’ as low as 44%. In terms of the power to detect the ‘C’ estimate of 42% observed in ME10, the power was 73%.

**3.3.2 | MEs without significant familial aggregation—**For the remaining 11 MEs (ME3, ME4, ME6, ME8, ME11–12, ME14–15, ME17, and ME0) there was insufficient power to choose between the competing ‘AE’ and ‘CE’ univariate models. Consequently, we retained the saturated ‘ACE’ model as the best fitting in each instance. Note that in each of these MEs, the 95% confidence intervals surrounding both the ‘A’ and ‘C’ point estimates spanned zero.

## 4 | DISCUSSION

To our knowledge, this is the first study to explore the genetic and environmental contributions to weighted gene co-expression network analysis module eigengenes (MEs). Among the 26 derived MEs, 50% had statistically significant additive genetic variation

with an average heritability of 46% (SD = 0.09, range = 35%–64%). Additionally, there were two MEs with evidence of significant shared environmental influences accounting for 42%–54% of the total variance. For the remaining 11 MEs, estimates of additive genetic *and* shared environmental influences were non-significantly different from zero. Despite the small sample size for a classical twin study, the demonstration of significant family aggregation including estimates of significant heritability suggests that blood-based gene expression networks are reliable and merit further exploration in terms of their associations with complex traits and diseases.

Our average estimate of ME heritability is higher than the heritability estimates of individual differentially expressed genes reported previously. For example, Monks et al.'s (2004) analyses of human lymphoblastoid cell lines (LCLs) taken from 167 individuals found that among  $N = 2340$  differentially expressed genes, 31% ( $N = 762$ ) were significantly heritable ( $h^2$ ), with a median  $h^2$  of 34%. McRae et al. (2007) estimated a mean heritability of 31% for normalized gene expression levels based on  $N = 15,887$  transcripts in LCLs taken from a small sample of monozygotic twin pairs (MZ). Powell et al. (2012) analyzed 9555 genes in LCLs and whole blood (WB) taken from 50 and 47 MZ twin pairs, and estimated average heritabilities for differential gene expression of 38% for LCLs and 32% for WB. We note that the methods used by these three reports captured broad sense heritability and assumed an absence of common environmental influences; our results have shown these 'C' influences to be present in at least two module eigengenes. Wright et al. (2014) used the classical twin design, which unconfounds additive genetic and common environmental influences, to analyze gene expression levels based on peripheral venous blood taken from 2752 twins. Among differentially expressed genes, they reported an average heritability of 14% ( $\pm 15\%$ ). While these previous reports demonstrate a clear pattern of heritability for QTLs, our findings suggest that the average heritability of MEs is potentially higher than differentially expressed genes. Complex gene networks and interactions are the hallmark of complex traits (Horvath et al., 2006; Zhi et al., 2013), and since MEs are multivariate summaries of correlated gene expression patterns and network connectivity across all genes in a module, the signal detection power when based on the heritable MEs is likely to be greater compared to analyses based on univariate GE loci (Horvath, 2011).

Of interest are the two MEs with evidence of significant shared environmental (C) influences. As with additive genetics, 'C' influences also drive twin pair similarity. The advantage of the twin methodology applied here is that the use of age-matched siblings reared together provides a natural control for environmental influences that were either shared during infancy and youth, continue to be shared, or are persistent and continue to exert an impact. Naturally, as siblings age and spend less time together, the number of directly shared environmental influences, relative to the time in their lives when reared together, is expected to diminish. Despite the expected waning of environmental influences shared between siblings, familial aggregation in ME10 and ME13 was unequivocally explained by shared environmental influences ranging from 42% to 54% of the total variance. Given that batch effects can also drive twin pair similarity when correlated within twin pairs, all 26 MEs were corrected for the effects of testing location, the time from blood draw to placement in  $-20^\circ\text{C}$  storage, total time spent in  $-20^\circ\text{C}$  storage and  $-80^\circ\text{C}$  storage, the time between blood draw and RNA extraction, and finally, the time between blood draw

and sequencing. Although caution should normally be applied when interpreting ‘C’ in small twin studies, the power to detect the ‘C’ influence of 41% for ME10 in this sample was 73%. If, after future efforts at validation and replication based on larger samples, estimates of C persist, then further analyses should elucidate which aspects of environments shared between siblings, broadly indexed by ‘C’ here, are responsible for these observed individual differences for example, allostatic stressors, diet, geographical location, and so forth.

Our results must be interpreted in the context of at least two limitations. First, between-group differences are likely to exist and may be complex (Royse et al., 2021). Therefore, our results may not generalize to women or other age or ancestral groups. It is unclear to what extent there are sex differences in the genetic and environmental etiology of MEs in similarly aged women. It is plausible that the genetic and environmental etiologies of these MEs may differ between ancestral groups. Only by ascertaining larger and ancestrally varied samples can we begin to test hypotheses regarding sex and ancestral group differences.

Second, while estimating GE patterns from whole blood has become a common alternative to *postmortem* brain studies (GTEx Consortium, 2013, 2015), our analyses did not account for cell heterogeneity, which can bias results. For instance, cell heterogeneity can result in GE differences due to different proportions of cells between samples, or mask important disease-relevant differences. It is therefore important to control for such differences in order to extract accurate and meaningful GE information. One solution would be to account for cell-specific contributions is assessing GE patterns in certain cell populations derived from fluorescence-activated cell sorting (FACS) approaches. An alternative, emerging solution would be to rely on ‘*computational deconvolution*’ methods to infer cell type-specific information from data generated from heterogeneous samples (Novershtern et al., 2011; Shay & Kang, 2013; Shen-Orr & Gaujoux, 2013).

## 5 | CONCLUSION

Weighted gene co-expression network analysis MEs are interpreted as multivariate summaries of correlated gene-expression patterns and network connectivity across genes within a module. Because they have the potential to elucidate the mechanisms by which molecular genomic variation contributes to individual differences in complex traits, they have been increasingly used to test for associations between modules and complex traits. This is the first empirical study to explore the genetic and environmental contributions to MEs. Among 14 MEs with statistically significant additive genetic variation, the average heritability was 46%. Because network modules may not be well preserved between blood and brain (Cai et al., 2010), blood-derived GE may not recapitulate neurobiological mechanisms of complex behaviors and disorders. Therefore, while blood GE and MEs can be used as proxy biomarkers for brain-based phenotypes and complex traits, blood tissue should only be considered a useful proxy for gene expression in the brain provided that the relevant gene(s) are expressed in both tissues (Joehanes et al., 2012; Sullivan et al., 2006). To address this problem, we have recently developed a novel transcriptome-imputation method called Brain Gene Expression and Network Imputation Engine (BrainGENIE) (Hess et al., 2023) that is capable of imputing brain-region-specific gene expression levels from peripheral blood gene expression. In addition to applying BrainGENIE to the current data, it

remains an empirical question as to how well blood-based MEs assessed in early old age can be leveraged as useful biomarkers for complex traits and disorders that emerge later in life.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We kindly thank the twins from the Vietnam Era Twin Study of Aging (VETSA) for their participation.

## FUNDING INFORMATION

This study was principally supported by grant R01AG054002 from the U.S. National Institutes of Health (NIH) National Institute on Aging to Stephen J. Glatt and Ming T. Tsuang. Dr. Hess was supported by grants from the NIH National Institute of Mental Health (NIMH) (R21MH126494), the NIH National Institute of Neurological Disorders and Stroke (R01NS128535), and The Brain & Behavior Research Foundation. This project was also funded by the NIH National Institute of Mental Health (R01MH125902).

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

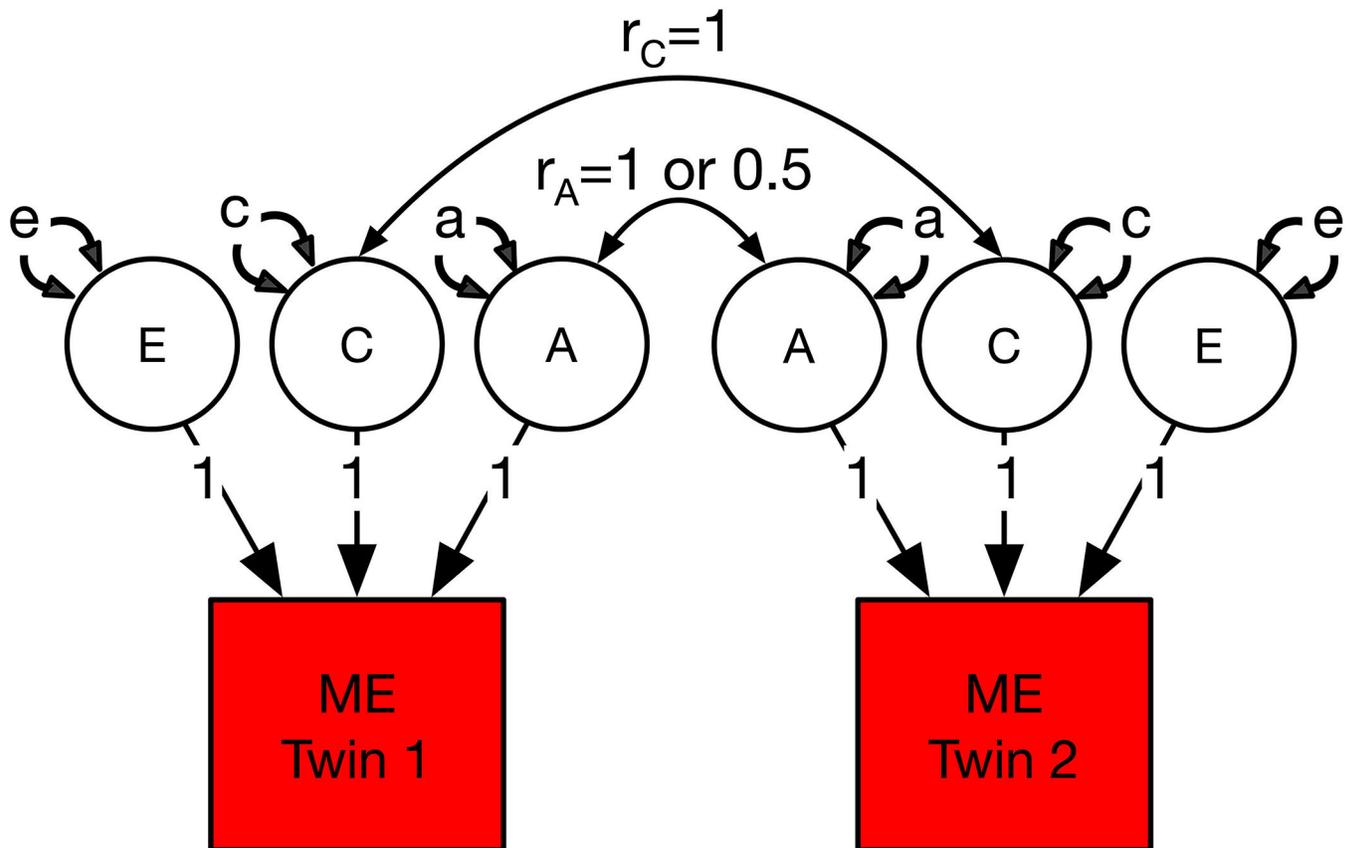
## REFERENCES

- Akaike H (1987). Factor analysis and AIC. *Psychometrika*, 52, 317–332.
- Babraham Bioinformatics. (2011). Fastqc: A quality control tool for high throughput sequence. Data. In (version v0.11.9). <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bates TC, Neale M, & Maes HH (2019). Umx: A library for structural equation and twin modelling in R. *Twin Research and Human Genetics*, 22, 27–41. 10.1017/thg.2019.2 [PubMed: 30944056]
- Bhattacharya S, & Mariani TJ (2009). Array of hope: Expression profiling identifies disease biomarkers and mechanism. *Biochemical Society Transactions*, 37(Pt 4), 855–862. 10.1042/BST0370855 [PubMed: 19614607]
- Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, & Fox J (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, 76(2), 306–317. 10.1007/s11336-010-9200-6 [PubMed: 23258944]
- Bolger AM, Lohse M, & Usadel B (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. 10.1093/bioinformatics/btu170 [PubMed: 24695404]
- Institute Broad. (2018). GitHub repository <http://broadinstitute.github.io/picard/>
- Cai C, Langfelder P, Fuller TF, Oldham MC, Luo R, van den Berg LH, Ophoff RA, & Horvath S (2010). Is human blood a good surrogate for brain tissue in transcriptional studies? *BMC Genomics*, 11, 589. 10.1186/1471-2164-11-589 [PubMed: 20961428]
- Carroll SB (2005). Evolution at two levels: On genes and form. *PLoS Biology*, 3(7), e245 (1545x–7885 (electronic)). [PubMed: 1600021]
- Chen WY, Chen H, Hamada K, Gatta E, Chen Y, Zhang H, Drnevich J, Krishnan HR, Maienschein-Cline M, Grayson DR, Pandey SC, & Lasek AW (2021). Transcriptomics identifies STAT3 as a key regulator of hippocampal gene expression and anhedonia during withdrawal from chronic alcohol exposure. *Translational Psychiatry*, 11(1), 298. 10.1038/s41398-021-01421-8 [PubMed: 34016951]
- Coppola G, & Geschwind DH (2006). Technology insight: Querying the genome with microarrays—Progress and hope for neurological disease. *Nature Clinical Practice. Neurology*, 2(3), 147–158. 10.1038/ncpneuro0133 [PubMed: 16932541]

- DiLeo MV, Strahan GD, den Bakker M, & Hoekenga OA (2011). Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. *PLoS One*, 6(10), e26683. 10.1371/journal.pone.0026683 [PubMed: 22039529]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, & Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. 10.1093/bioinformatics/bts635 [PubMed: 23104886]
- Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M, Steinthorsdottir V, Eiriksdottir GH, Bjornsdottir G, Reynisdottir I, Gudbjartsson D, Helgadóttir A, Jonasdottir A, Jonasdottir A, ... Stefansson K (2008). Genetics of gene expression and its effect on disease. *Nature*, 452, 423–428. 1476–4687 (Electronic). [PubMed: 18344981]
- Franke L, & Jansen RC (2009). eQTL analysis in humans. *Methods in Molecular Biology*, 573, 311–328. 10.1007/978-1-60761-247-6\_17 [PubMed: 19763935]
- Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusk AJ, & Horvath S (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. *Mammalian Genome*, 18(6–7), 463–472. 10.1007/s00335-007-9043-3 [PubMed: 17668265]
- GTE Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585. 10.1038/ng.2653 [PubMed: 23715323]
- GTE Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 348(6235), 648–660. 10.1126/science.1262110 [PubMed: 25954001]
- Gu CC, Rao DC, Stormo G, Hicks C, & Province MA (2002). Role of gene expression microarray analysis in finding complex disease genes. *Genetic Epidemiology*, 23(1), 37–56. 10.1002/gepi.220 [PubMed: 12112247]
- Hess JL, Quinn TP, Zhang C, Hearn GC, Chen S, Neuropsychiatric Consortium for Analysis and Sharing of Transcriptomes, Kong SW, Cairns M, Tsuang MT, Faraone SV, & Glatt SJ (2023). BrainGENIE: The brain gene expression and network imputation engine. *Translational Psychiatry*, 13(1), 98. 10.1038/s41398-023-02390-w [PubMed: 36949060]
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, & Manolio TA (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*, 106(23), 9362–9367. 10.1073/pnas.0903103106 [PubMed: 19474294]
- Horvath S (2011). *Weighted network analysis: Applications in genomics and systems biology*. Springer.
- Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z, Lee Y, Scheck AC, Liao LM, Wu H, Geschwind DH, Febbo PG, Kornblum HI, Cloughesy TF, Nelson SF, & Mischel PS (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), 17402–17407. 10.1073/pnas.0608396103 [PubMed: 17090670]
- Joehanes R, Johnson AD, Barb JJ, Raghavachari N, Liu P, Woodhouse KA, O'Donnell CJ, Munson PJ, & Levy D (2012). Gene expression analysis of whole blood, peripheral blood mononuclear cells, and lymphoblastoid cell lines from the Framingham heart study. *Physiological Genomics*, 44(1), 59–75. 10.1152/physiolgenomics.00130.2011 [PubMed: 22045913]
- Kleinjan DA, & Lettice LA (2008). Long-range gene control and genetic disease. *Advances in Genetics*, 61, 339–388. [PubMed: 18282513]
- Kremen WS, Thompson-Brenner H, Leung YM, Grant MD, Franz CE, Eisen SA, Jacobson KC, Boake C, & Lyons MJ (2006). Genes, environment, and time: The Vietnam Era Twin Study of Aging (VETSA). *Twin Research and Human Genetics*, 9(6), 1009–1022. 10.1375/183242706779462750 [PubMed: 17254445]
- Kremen WS, Franz CE, & Lyons MJ (2013). VETSA: The Vietnam Era Twin Study of Aging. *Twin Research and Human Genetics*, 16(1), 399–402. 10.1017/thg.2012.86 [PubMed: 23110957]

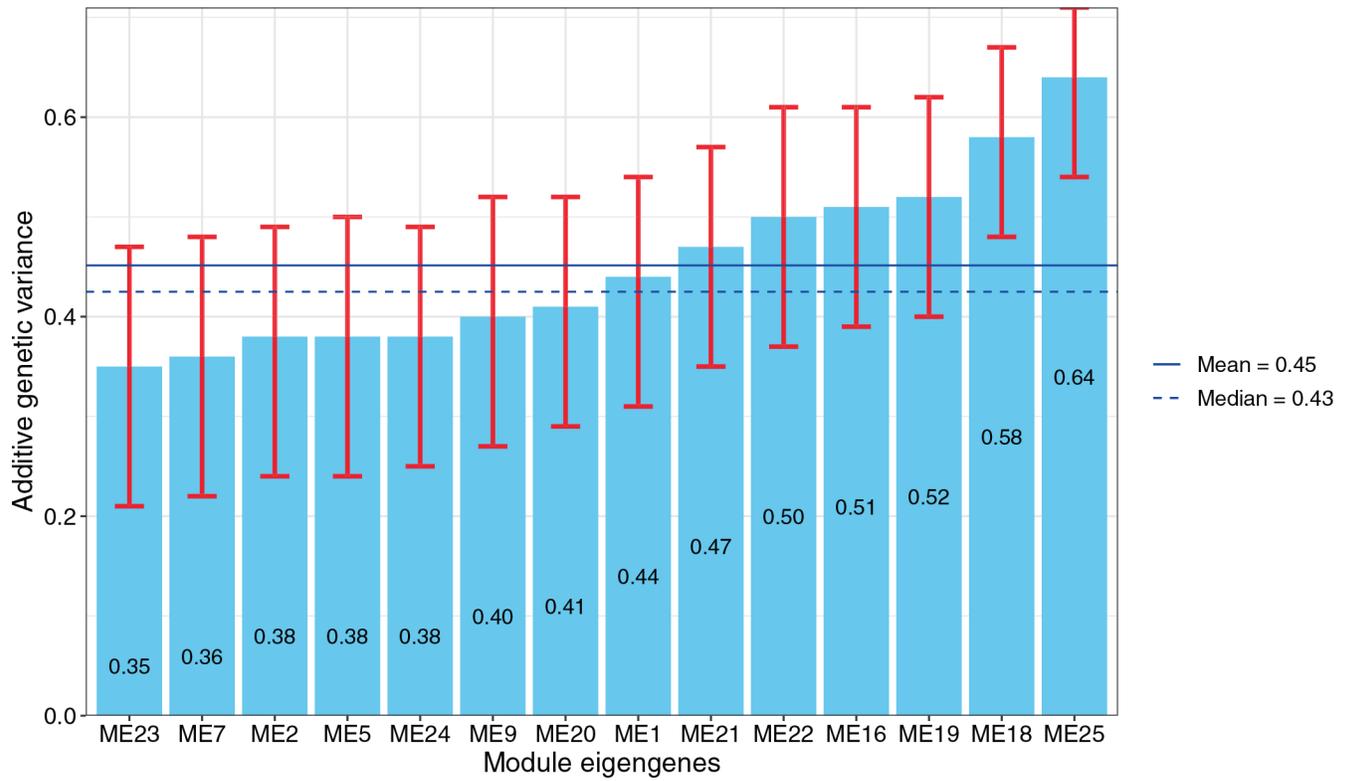
- Kudaravalli S, Veyrieras JB, Stranger BE, Dermitzakis ET, & Pritchard JK (2009). Gene expression levels are a target of recent natural selection in the human genome. *Molecular Biology and Evolution*, 26, 649–658. [PubMed: 19091723]
- Langfelder P, & Horvath S (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559. 10.1186/1471-2105-9-559 [PubMed: 19114008]
- Langfelder P, Mischel PS, & Horvath S (2013). When is hub gene selection better than standard meta-analysis? *PLoS One*, 8(4), e61505. [PubMed: 23613865]
- Lehrmann E, & Freed WJ (2008). Transcriptional correlates of human substance use. *Annals of the New York Academy of Sciences*, 1139, 34–42. 10.1196/annals.1432.027 [PubMed: 18991846]
- Liao Y, Smyth GK, & Shi W (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923–930. [PubMed: 24227677]
- Liu C (2011). Brain expression quantitative trait locus mapping informs genetic studies of psychiatric diseases. *Neuroscience Bulletin*, 27(2), 123–133. 10.1007/s12264-011-1203-5 [PubMed: 21441974]
- Liu W, Li L, Ye H, & Tu W (2017). Weighted gene co-expression network analysis in biomedicine research. *Sheng Wu Gong Cheng Xue Bao*, 33(11), 1791–1801. 10.13345/j.cjb.170006 [PubMed: 29202516]
- MacLennan NK, Dong J, Aten JE, Horvath S, Rahib L, Ornelas L, Dipple KM, & McCabe ER (2009). Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. *Molecular Genetics and Metabolism*, 98(1–2), 203–214. 10.1016/j.ymgme.2009.05.004 [PubMed: 19546021]
- Martin NG, Eaves LJ, Kearsley MJ, & Davies P (1978). The power of the classical twin study. *Heredity (Edinb)*, 40(1), 97–116. (MEDLINE). [PubMed: 272366]
- Mason MJ, Fan G, Plath K, Zhou Q, & Horvath S (2009). Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*, 10, 327. 10.1186/1471-2164-10-327 [PubMed: 19619308]
- McRae AF, Matigian NA, Vadlamudi L, Mulley JC, Mowry B, Martin NG, Berkovic SF, Hayward NK, & Visscher PM (2007). Replicated effects of sex and genotype on gene expression in human lymphoblastoid cell lines. *Human Molecular Genetics*, 16(4), 364–373. 10.1093/hmg/ddl456 [PubMed: 17164263]
- Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, Phillips JW, Sachs A, & Schadt EE (2004). Genetic inheritance of gene expression in human cell lines. *American Journal of Human Genetics*, 75(6), 1094–1105. 10.1086/426461 [PubMed: 15514893]
- Neale MC, & Cardon LR (1992). *Methodology for genetic studies of twins and families* (1st ed.). Kluwer Academic Publishers.
- Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N, Chang CY, Shay T, Frampton GM, Drake AC, Leskov I, Nilsson B, Preffer F, Dombkowski D, Evans JW, Liefeld T, Smutko JS, ... Ebert BL (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*, 144(2), 296–309. 10.1016/j.cell.2011.01.004 [PubMed: 21241896]
- Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, & Geschwind DH (2008). Functional organization of the transcriptome in human brain. *Nature Neuroscience*, 11(11), 1271–1282. 10.1038/nn.2207 [PubMed: 18849986]
- Plaisier CL, Horvath S, Huertas-Vazquez A, Cruz-Bautista I, Herrera MF, Tusie-Luna T, Aguilar-Salinas C, & Pajukanta P (2009). A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genetics*, 5(9), e1000642. 10.1371/journal.pgen.1000642 [PubMed: 19750004]
- Powell JE, Henders AK, McRae AF, Wright MJ, Martin NG, Dermitzakis ET, Montgomery GW, & Visscher PM (2012). Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Research*, 22(3), 456–466. 10.1101/gr.126540.111 [PubMed: 22183966]
- R Development Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

- Robinson MD, McCarthy DJ, & Smyth GK (2010). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. 10.1093/bioinformatics/btp616 [PubMed: 19910308]
- Royse SK, Cohen AD, Snitz BE, & Rosano C (2021). Differences in Alzheimer's disease and related dementias pathology among African American and Hispanic women: A qualitative literature review of biomarker studies. *Frontiers in Systems Neuroscience*, 15, 685957. 10.3389/fnsys.2021.685957 [PubMed: 34366799]
- Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, ... Lusis AJ (2005). An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics*, 37(7), 710–717. 10.1038/ng1589 [PubMed: 15965475]
- Schoeneborn CA, & Heyman KM (2009). Health characteristics of adults aged 55 years and over: United States, 2004–2007. Hyattsville, MD.
- Shastri BS (2009). SNPs: Impact on gene function and phenotype. *Methods in Molecular Biology*, 578, 3–22. 10.1007/978-1-60327-411-1\_1 [PubMed: 19768584]
- Shay T, & Kang J (2013). Immunological genome project and systems immunology. *Trends in Immunology*, 34(12), 602–609. 10.1016/j.it.2013.03.004 [PubMed: 23631936]
- Shen-Orr SS, & Gaujoux R (2013). Computational deconvolution: Extracting cell type-specific information from heterogeneous samples. *Current Opinion in Immunology*, 25(5), 571–578. 10.1016/j.coi.2013.09.015 [PubMed: 24148234]
- Steiger JH, Shapiro A, & Browne MW (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50, 253–264.
- Sullivan PF, Fan C, & Perou CM (2006). Evaluating the comparability of gene expression in blood and brain. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 141B(3), 261–268. 10.1002/ajmg.b.30272 [PubMed: 16526044]
- Wang M, Huang J, Liu Y, Ma L, Potash JB, & Han S (2017). COMBAT: A combined association test for genes using summary statistics. *Genetics*, 207(3), 883–891. 10.1534/genetics.117.300257 [PubMed: 28878002]
- Wang W, Li W, Wu Y, Tian X, Duan H, Li S, Tan Q, & Zhang D (2021). Genome-wide DNA methylation and gene expression analyses in monozygotic twins identify potential biomarkers of depression. *Translational Psychiatry*, 11(1), 416. 10.1038/s41398-021-01536-y [PubMed: 34341332]
- Windén KD, Karsten SL, Bragin A, Kudo LC, Gehman L, Ruidera J, Geschwind DH, & Engel J Jr. (2011). A systems level, functional genomics analysis of chronic epilepsy. *PLoS One*, 6(6), e20763. 10.1371/journal.pone.0020763 [PubMed: 21695113]
- Wray NR, James MR, Mah SP, Nelson M, Andrews G, Sullivan PF, Montgomery GW, Birley AJ, Braun A, & Martin NG (2007). Anxiety and comorbid measures associated with PLXNA2. *Archives of General Psychiatry*, 64(3), 318–326. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=17339520](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17339520) [PubMed: 17339520]
- Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, Madar V, Jansen R, Chung W, Zhou YH, Abdellaoui A, Batista S, Butler C, Chen G, Chen TH, D'Ambrosio D, Gallins P, Ha MJ, Hottenga JJ, ... Boomsma DI (2014). Heritability and genomics of gene expression in peripheral blood. *Nature Genetics*, 46(5), 430–437. 10.1038/ng.2951 [PubMed: 24728292]
- Zeng D, He S, Ma C, Wen Y, Song W, Xu Q, Zhao N, Wang Q, Yu Y, Shen Y, Huang J, & Li H (2020). Network-based approach to identify molecular signatures in the brains of depressed suicides. *Psychiatry Research*, 294, 113513. 10.1016/j.psychres.2020.113513 [PubMed: 33137553]
- Zhang B, & Horvath S (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, 4, Article17. 10.2202/1544-6115.1128 [PubMed: 16646834]
- Zhi W, Minturn J, Rappaport E, Brodeur G, & Li H (2013). Network-based analysis of multivariate gene expression data. *Methods in Molecular Biology*, 972, 121–139. 10.1007/978-1-60327-337-4\_8 [PubMed: 23385535]



**FIGURE 1.**

Univariate structural equation model used to decompose and estimate the relative contribution of additive genetic (A), common or shared environments (C), and non-shared environmental (E) influences to individual differences or variation in each module eigengene (ME). Individual differences in each ME are decomposed into additive genetic (A), common or shared environmental influences (C), and unshared or random environmental influences as well as measurement error (E). Double-headed arrows represent the freely estimated within-person genetic and environmental latent factor variances (a, c and e) and the fixed twin pair correlations ( $r_C$ ,  $r_A$ ).  $r_C$  = correlation of 1 for MZ and DZ twin pairs.  $r_A = 1$  or 0.5 for MZ and DZ twin-pairs, respectively. Single-headed arrows represent the causal contributions of the A, C, and E effects for a given ME. Phenotypic means not shown.



**FIGURE 2.**

Standardized estimates (including 95% confidence intervals) of additive genetic variance in ascending order for the 14 MEs with significant heritability.

Descriptive statistics along with monozygotic (MZ) and dizygotic (DZ) twin pair polychoric correlations and their 95% confidence intervals for each module eigengene (ME).

**TABLE 1**

	N	Mean	SD	Minimum	Maximum	MZ (95% CI)	DZ (95% CI)
ME1	661	0.0	0.0279	-0.1118	0.1279	0.49 (0.35, 0.60)	0.09 (-0.13, 0.30)
ME2	661	0.0	0.0281	-0.0967	0.1246	0.43 (0.28, 0.55)	0.05 (-0.17, 0.26)
ME3	661	0.0	0.0346	-0.1549	0.1206	0.38 (0.22, 0.51)	0.14 (-0.06, 0.33)
ME4	661	0.0	0.0270	-0.1050	0.0775	0.46 (0.33, 0.58)	0.36 (0.15, 0.53)
ME5	661	0.0	0.0337	-0.1263	0.1239	0.38 (0.24, 0.51)	0.13 (-0.07, 0.32)
ME6	661	0.0	0.0338	-0.1600	0.1050	0.38 (0.23, 0.51)	0.29 (0.09, 0.47)
ME7	661	0.0	0.0338	-0.1080	0.1448	0.36 (0.22, 0.49)	0.13 (-0.09, 0.33)
ME8	661	0.0	0.0374	-0.0940	0.1497	0.42 (0.28, 0.54)	0.30 (0.10, 0.47)
ME9	661	0.0	0.0349	-0.1151	0.1405	0.42 (0.28, 0.55)	0.15 (-0.06, 0.35)
ME10	661	0.0	0.0326	-0.1080	0.1130	0.45 (0.31, 0.56)	0.36 (0.18, 0.52)
ME11	661	0.0	0.0368	-0.1403	0.1001	0.40 (0.25, 0.52)	0.35 (0.17, 0.51)
ME12	661	0.0	0.0291	-0.1460	0.1178	0.51 (0.38, 0.61)	0.27 (0.07, 0.44)
ME13	661	0.0	0.0300	-0.1281	0.0767	0.58 (0.46, 0.67)	0.47 (0.30, 0.60)
ME14	661	0.0	0.0357	-0.1452	0.1035	0.40 (0.25, 0.53)	0.30 (0.11, 0.47)
ME15	661	0.0	0.0364	-0.1144	0.1291	0.42 (0.27, 0.54)	0.14 (-0.07, 0.33)
ME16	661	0.0	0.0365	-0.1186	0.1139	0.50 (0.38, 0.60)	0.19 (-0.01, 0.37)
ME17	661	0.0	0.0371	-0.1369	0.1247	0.27 (0.11, 0.41)	0.26 (0.04, 0.45)
ME18	661	0.0	0.0370	-0.1045	0.1114	0.59 (0.48, 0.68)	0.26 (0.06, 0.43)
ME19	661	0.0	0.0380	-0.1307	0.1418	0.50 (0.37, 0.61)	0.27 (0.07, 0.45)
ME20	661	0.0	0.0342	-0.1052	0.1267	0.44 (0.31, 0.56)	0.15 (-0.06, 0.34)
ME21	661	0.0	0.0359	-0.1214	0.1208	0.48 (0.35, 0.59)	0.22 (0.02, 0.40)
ME22	661	0.0	0.0353	-0.1135	0.1380	0.52 (0.40, 0.62)	0.03 (-0.17, 0.23)
ME23	661	0.0	0.0365	-0.0642	0.1998	0.34 (0.20, 0.46)	0.12 (-0.05, 0.29)
ME24	661	0.0	0.0365	-0.0933	0.1368	0.39 (0.26, 0.51)	0.10 (-0.16, 0.34)
ME25	661	0.0	0.0367	-0.0733	0.1187	0.65 (0.55, 0.73)	0.25 (0.06, 0.41)
ME0	661	0.0	0.0351	-0.1283	0.1258	0.38 (0.24, 0.51)	0.33 (0.13, 0.50)

*Note:* Residualization was performed using the `umx_residualize` function (version 4.9.0) in R (version 4.1.1). All MEs were residualized for (i) testing site, (ii) time in hours from blood draw to storage in  $-20^{\circ}\text{C}$ , (iii) time in days spent in  $-20^{\circ}\text{C}$  storage, (iv) time in days spent in  $-80^{\circ}\text{C}$  storage, (v) time in days between blood draw and RNA extraction, and (vi) time in days between blood draw and sequencing. A total of 32 cases were lost due to missing covariates.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Standardized estimates (including 95% confidence intervals [95% CI]) of additive genetic (A), common environment (C), and non-shared environment (E) influences under the best fitting univariate models for each module eigengene (ME).

**TABLE 2**

	A	95%CI	C	95%CI	E	95%CI
MEs with significant additive genetic variation						
ME1	0.44	(0.31, 0.54)	-	-	0.56	(0.46, 0.69)
ME2	0.38	(0.24, 0.49)	-	-	0.62	(0.51, 0.76)
ME5	0.38	(0.24, 0.50)	-	-	0.62	(0.50, 0.76)
ME7	0.36	(0.22, 0.48)	-	-	0.64	(0.52, 0.78)
ME9	0.40	(0.27, 0.52)	-	-	0.60	(0.48, 0.73)
ME16	0.51	(0.39, 0.61)	-	-	0.49	(0.39, 0.61)
ME18	0.58	(0.48, 0.67)	-	-	0.42	(0.33, 0.52)
ME19	0.52	(0.40, 0.62)	-	-	0.48	(0.38, 0.60)
ME20	0.41	(0.29, 0.52)	-	-	0.59	(0.53, 0.71)
ME21	0.47	(0.35, 0.57)	-	-	0.53	(0.43, 0.65)
ME22	0.50	(0.37, 0.61)	-	-	0.50	(0.39, 0.63)
ME23	0.35	(0.21, 0.47)	-	-	0.65	(0.53, 0.79)
ME24	0.38	(0.25, 0.49)	-	-	0.62	(0.51, 0.75)
ME25	0.64	(0.54, 0.71)	-	-	0.36	(0.29, 0.46)
MEs with significant common environment variation						
ME10	-	0.42	(0.31, 0.51)	0.58	(0.49, 0.69)	
ME13	-	0.54	(0.45, 0.62)	0.46	(0.38, 0.55)	
MEs with no significant familial aggregation						
ME3	0.39	(-0.09, 0.91)	-0.04	(-0.50, 0.37)	0.65	(0.52, 0.79)
ME4	0.09	(-0.31, 1.00)	0.35	(-0.10, 0.69)	0.56	(0.45, 0.68)
ME6	0.18	(-0.26, 0.66)	0.20	(-0.23, 0.56)	0.62	(0.50, 0.77)
ME8	0.27	(-0.14, 0.72)	0.15	(-0.25, 0.50)	0.57	(0.46, 0.71)
ME11	0.14	(-0.26, 0.57)	0.27	(-0.11, 0.59)	0.59	(0.47, 0.73)
ME12	0.33	(-0.08, 0.81)	0.15	(-0.31, 0.50)	0.52	(0.42, 0.64)
ME14	0.15	(-0.29, 0.62)	0.24	(-0.18, 0.59)	0.61	(0.49, 0.76)
ME15	0.46	(-0.03, 0.99)	-0.07	(-0.56, 0.36)	0.61	(0.50, 0.74)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

		A	95%CI	C	95%CI	E	95%CI
ME17	0.15	(-0.33, 0.63)	0.15	(-0.26, 0.52)	0.70	(0.56, 0.87)	
ME0	-0.02	(-0.44, 0.46)	0.38	(-0.06, 0.73)	0.64	(0.52, 0.77)	