A systematic analysis of genetically regulated differences in gene expression and the role of co-expression networks across 16 psychiatric disorders and substance use phenotypes

Zachary F Gerring[a] PhD, Jackson G Thorp[a] BSc (Hons), Eric R Gamazon[b,c,d], PhD, Eske M Derks[a] PhD

[a]Translational Neurogenomics Laboratory; QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
[b]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
[c]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA
[d]Clare Hall & MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

Running title: Transcriptomic analysis of mental health phenotypes

Keywords: Genetic correlation, transcriptomic correlation, gene expression, translational neuroscience, network analysis, systems biology

Correspondence

Zachary Gerring, Ph.D
Translational Neurogenomics Laboratory
300 Herston Road
Brisbane City QLD 4006
zachary.gerring@qirmberghofer.edu.au

ABSTRACT

Genome-wide association studies (GWASs) have identified thousands of risk loci for many psychiatric and substance use phenotypes, however the biological consequences of these loci remain largely unknown. We performed a transcriptome-wide association study of 10 psychiatric disorders and 6 substance use phenotypes (collectively termed "mental health phenotypes") using expression quantitative trait loci data from 532 prefrontal cortex samples. We estimated the correlation due to predicted genetically regulated expression between pairs of mental health phenotypes, and compared the results with the genetic correlations. We identified 1,645 genes with at least one significant trait association, comprising 2,176 significant associations across the 16 mental health phenotypes of which 572 (26%) are novel. Overall, the transcriptomic correlations for phenotype pairs were significantly higher than the respective genetic correlations. For example, attention deficit hyperactivity disorder and autism spectrum disorder, both childhood developmental disorders, showed a much higher transcriptomic correlation (r=0.84) than genetic correlation (r=0.35). Finally, we tested the enrichment of phenotype-associated genes in gene co-expression networks built from prefrontal cortex. Phenotype-associated genes were enriched in multiple gene co-expression modules and the implicated modules contained genes involved in mRNA splicing and glutamatergic receptors, among others. Together, our results highlight the utility of gene expression data in the understanding of functional gene mechanisms underlying psychiatric disorders and substance use phenotypes.

INTRODUCTION

Psychiatric and substance use disorders are a leading cause of disease burden and account for 28.5% of global years lived with disability (1). Genome-Wide Association Studies (GWAS) have identified hundreds of genomic regions that are linked to the risk to develop a psychiatric disorder (2–4) and provide novel insights into the genetic architecture of psychiatric disorders and into the sharing of genetic risk factors across mental health phenotypes (5). The Brainstorm Consortium used GWAS data to estimate genetic correlations (r_g) across ten psychiatric disorders, revealing considerable sharing of common genetic risk (5). However, relatively few studies have explored sharing of pathological or molecular mechanisms which contribute to these disorders. The majority (~93%) of disease-associated genetic variants are located in non-protein coding regions of the genome (6) suggesting that genetic mutations act through the regulation of gene expression rather than by directly altering the protein product. In the present study, we will integrate genetic and transcriptomic information from the brain to explore sharing of transcriptomic mechanisms across 16 mental health phenotypes.

Previous studies have integrated genetic and gene expression data to gain pathophysiological insights into a more limited subset of psychiatric disorders (7,8). Gandal *et al.* compared levels of differential gene expression in postmortem brain samples from patients with autism (ASD), schizophrenia (SCZ), bipolar disorder (BD), depression (DEP), and matched healthy controls and revealed significant overlap of disease-related signatures between ASD, SCZ, BD, and DEP (9). Transcriptomic changes were most severe in ASD and least severe in DEP, with SCZ and BD showing intermediate levels of severity. Although this study provided important new insights into the sharing of molecular mechanisms across mental health disorders, comparison of observed levels of gene expression is susceptible to reverse causation, where traits may affect gene expression levels (10,11). Imputation of gene expression levels

based on whole-genome and RNA sequence reference data from healthy participants provides a unique way to investigate how the genetically regulated component of gene expression is shared across disorders (10,11).

We and others have shown that biologically relevant functional networks are critical for understanding pathway convergence of manifold genetic risk variants in neuropsychiatric diseases (12). Genetic co-expression networks model correlated levels of gene expression and provide a way to explore how the activity of multiple biologically related genes within the same co-expression network influence disease risk. We have previously generated co-expression networks in 13 brain tissues from healthy GTEx donors and report an association between four co-expression networks and Major Depressive Disorder, suggesting a role for synaptic signalling and neuronal development pathways (12). In their study of post-mortem gene expression in patients vs. healthy controls, Gandal et al. (13) explored module-level differential expression and showed that a module strongly enriched for microglial markers was upregulated specifically in ASD while several other modules were downregulated across ASD, SCZ, and BD.

In the present study, we conduct a comprehensive exploration of differences in genetically regulated levels of gene expression across 16 mental health phenotypes, including 11 psychiatric disorders and 6 substance use phenotypes. First, we integrate GWAS summary statistics with gene expression data from the prefrontal cortex of 533 healthy PsychENCODE donors. Second, we perform a systematic exploration of differences in genetically regulated levels of gene expression for the 16 individual phenotypes and delineate genetic and transcriptomic overlap across phenotypes. Third, we generate co-expression networks and explore enrichment of GWAS signal within network modules. Finally, we will use LD score

regression (14) to partition GWAS heritability and determine the contribution from SNPs included in phenotype-associated networks before and after accounting for baseline functional annotations.

METHODS

Description of the GWAS Summary statistics

We included 10 psychiatric disorders and 6 substance use phenotypes (which we collectively refer to as "mental health" phenotypes) in our analyses. We selected only mental health phenotypes with significant SNP-based heritability (Z-score>2). Details on the individual GWAS samples, including sample sizes and SNP-based heritability estimates, are provided in Table 1 and Supplementary Table 1.

PsychENCODE RNAseq data

We obtained a gene expression matrix derived from the prefrontal cortex in 532 healthy control subjects from the PsychECODE project (http://resource.psychencode.org/) (15). The gene expression data were normalised from the full (healthy subjects and diseased cases) Fragments Per Kilobase of transcript per Million (FPKM) count matrix expression matrix as described by Gandal et al. (16), and filtered so that only genes with FPKM >= 0.1 in at least 10 samples are retained.

TWAS FUSION

We used TWAS FUSION (11) to integrate eQTL information from the PsychENCODE project with GWAS summary statistics for 16 mental health phenotypes (Table 1) to identify genes whose genetically predicted expression levels are associated with each phenotype. We used expression weights generated by the PsychENCODE consortium (16), and Linkage Disequilibrium information from the 1000 Genomes Project Phase 3 (17). These data were processed with the beta coefficients or odds ratios from each GWAS to estimate the expression-GWAS association statistic. For each phenotype, we corrected for

multiple testing using the false discovery rate (18) (FDR<0.05). We performed empirical Brown's test (19) to combine TWAS FUSION *P* values to rank order genes based on their strength of association across the 16 phenotypes. We restricted this analysis to genes for which TWAS FUSION association statistics results were available for all 16 phenotypes.

MAGMA

We performed gene-based analyses using MAGMA v1.07 (20), which assigns SNPs to their nearest gene using a pre-defined genomic window. We defined the genomic window as 35 kb upstream or 10 kb downstream of a gene body. The gene-based test statistic based was calculated using the default *snp-wise=mean* model, which uses the weighted sum of the SNP –log(10) P values while accounting for the correlation (i.e. linkage disequilibrium) between nearby SNPs. Linkage disequilbrium information was obtained from the 1000 Genomes Project Phase 3 (17). Multiple testing correction was performed across all phenotypes using FDR<0.05. We identified novel significant TWAS FUSION genes for each mental health phenotype by intersecting significant TWAS FUSION results with those identified using conventional proximity-based methods in MAGMA, as well as the FUMA SNP2GENE function (see Web resources). For the latter, we used positional gene mapping for all genes (i.e., we included non-protein coding genes as these were also included in the Fusion analysis), and lead SNPs were identified using default settings (R^2 threshold to define independent significant SNPs=0.6).

Transcriptome-wide correlation analysis

We estimated the genome-wide genetic correlation between each pair of mental health phenotypes as a function of the predicted gene expression effect from TWAS FUSION using RhoGE (21), after excluding the MHD region. Briefly, RhoGE estimates the mediating effect of genetically regulated gene expression (estimated from TWAS FUSION) before calculating the correlation of effect sizes between pairs of traits.

149

150 *Genetic correlations and estimates of h^2_{SNP}*

151 LD Score Regression was used to estimate SNP-based heritability (h^2_{SNP}) and genetic correlations
152 between each pair of the 16 traits, after exclusion of the MHC region. SNP-based heritability for case-
153 control phenotypes was estimated on the liability scale (see Supplementary Table 1 for the sample and
154 population prevalence of each trait). Multiple testing was corrected for by adjusting *P* values based on
155 false discovery rate (FDR) across all tests.

156

157 *Hierarchical cluster analysis*

158 We performed hierarchical clustering analysis for both transcriptomic and genetic correlations in order
159 to examine the underlying genetic and transcriptomic structure between the 16 traits. Complete-linkage
160 clustering was implemented using the hclust function in R (22), where dissimilarity between trait pairs
161 was defined as one minus the (genetic or transcriptomic) correlation.

162

163 *Gene co-expression network analysis*

164 Gene co-expression networks were individually constructed from 532 prefrontal cortex control samples,
165 generated by the PsychENCODE project, using the weighted gene co-expression network analysis
166 (WGCNA) package in R (23). A signed pairwise correlation matrix using Pearson's product moment
167 correlation coefficient was calculated. A "soft-thresholding" value of 14 was selected by plotting the
168 strength of correlation against a series (range 2 to 20) of soft threshold powers. The correlation matrix
169 was subsequently transformed into an adjacency matrix, where nodes correspond to genes and edges to
170 the connection strength between genes. The adjacency matrix was normalised using a topological
171 overlap function. Hierarchical clustering was performed using average linkage, with one minus the
172 topological overlap matrix as the distance measure. The hierarchical cluster tree was cut into gene

modules using the dynamic tree cut algorithm (24), with a minimum module size of 30 genes. We amalgamated modules if the correlation between their eigengenes – defined as the first principal component of their genes' expression values – was greater or equal to 0.8.

Gene-set analysis to explore enrichment of heritability in gene co-expression networks

To identify gene co-expression networks enriched with candidate risk genes for each mental health trait, we performed gene-set analysis of TWAS FUSION gene-level results in tissue-specific gene co-expression networks using the gene-set analysis function in MAGMA v1.07 (20). For each mental health trait, we generated MAGMA-format annotation (.annot) files using the default --annot function. For the gene-based analysis, we used the *--snp-wise=mean* function, which calculates an association statistic for each gene using the weighted sum of *P* values for a predefined genomic window (5 kilobases upstream and 1.5 kilobases downstream). The 1000 Genomes European reference panel (Phase 3) (17) was used to account for Linkage Disequilibrium between SNPs. Finally, we tested for the enrichment of gene-based association signals within gene co-expression networks from the prefrontal cortex. First, we modified the intermediary .raw files generated in the gene-based test by substituting each MAGMA gene z-score with the corresponding TWAS FUSION gene z-score. The module enrichment analyses were re-performed after excluding genes in the MHC region.

Characterisation of gene expression modules

Gene expression modules enriched with neuropsychiatric GWAS association signals were assessed for biological pathways using g:Profiler (https://biit.cs.ut.ee/gprofiler/) (25). Ensembl gene identifiers within enriched gene modules were used as input; we tested for the over-representation of module genes in Gene Ontology (GO) biological process terms, as well as KEGG (26) and Reactome (27) gene pathways. The g:Profiler algorithm uses a Fisher's one-tailed test for gene pathway enrichment; the smaller the P value, the lower the probability a gene belongs to both a co-expression module and a biological term or

196 pathway purely by chance. Multiple testing correction was done using g:SCS; this approach accounts for
197 the correlated structure of GO terms and biological pathways, and corresponds to an experiment-wide
198 threshold of α=0.05.

199 *Partitioned heritability analysis*

200 We used stratified LD score regression (S-LDSC) to estimate the enrichment and the standardized effect
201 size of the six associated gene co-expression modules (28). S-LDSC assumes the association statistic for
202 an associated SNP captures the effects of all nearby tagged SNPs. If a phenotype has a polygenic
203 architecture, SNPs with a high LD score will have larger association statistics on average than SNPs with
204 a low LD score. As such, LD within a functional category that is enriched for heritability will increase the
205 association statistic relative to that of a category that does not contribute to heritability. Thus, S-LDSC
206 will identify functional categories if SNPs with high LD to that category have higher association statistics
207 than SNPs with low LD to that category. We generated customized annotation-specific LD scores based
208 on the gene sets from the genetic co-expression modules using the python script provided by the
209 developers of LDSC. LD scores were calculated using a default window size of 100kb and 1KG genotype
210 data as reference data (see Web resources). We calculated the enrichment of heritability including LD
211 scores of the co-expression modules before correcting for baseline functional annotations (28) (see Web
212 resources). The baseline-LD model contains 52 functional annotations, including coding, conserved, and
213 regulatory annotations (e.g., promoter, enhancer, histone marks, transcription factor [TF] binding sites).

214 **RESULTS**

215 ***Gene-based results***

216 We calculated the association between imputed genetically regulated gene expression from prefrontal
217 cortex and 16 neuropsychiatric phenotypes using TWAS FUSION (Supplementary Table 2). We identified
218 1,645 genes with at least one significant phenotype association (after correction using the false

discovery rate [FDR] <0.05), comprising 2,176 significant associations across the 16 phenotypes. Of these, 1,236 were related to psychiatric disorders and 940 to substance use phenotypes. Within psychiatric phenotypes, the largest number of TWAS FUSION associations was observed for schizophrenia (N=597) followed by depression (N=185), while smoking initiation (N=312) and drinks per week (N=260) accounted for the largest number of substance use associations. When compared with two commonly-used gene mapping tools, conventional MAGMA and FUMA SNP2gene, a total of 572 genes (26%) of the TWAS FUSION genes were novel (Supplementary Table 3). We conducted empirical Brown's tests to rank-order genes of which the imputed gene expression levels are most strongly associated across the 16 phenotypes. Figure 1A illustrates the strength of association for the top 20 genes from the Brown's analysis across the 16 phenotypes. Interestingly, 8 of the top 20 most strongly associated genes across all phenotypes showed a significant association with concordant effects in depression and schizophrenia. It should be noted, however, that most of the associations were linked to the MHC region, highlighting its importance in mental health phenotypes. Full results are presented in Supplementary Table 4 (including MHC) and Supplementary Table 5 (excluding MHC). We also selected the top gene for each phenotype and visualize effect sizes across all 16 phenotypes (Figure 1B).

Genetic and transcriptomic correlations across 16 phenotypes

All phenotypes exhibited significant SNP-based heritability (Table 1 and Supplementary Table 6). We estimated correlations across the 16 pairs of phenotypes based on genetic variation (ρ_g) using LDSC (Figure 2A; below diagonal) and predicted expression (ρ_t) using RhoGE (Figure 2A, above diagonal). Tabulated data for genetic and transcriptomic correlations are shown in Supplementary Table 7 and Supplementary Table 8, and Supplementary Figure 1 shows the pairwise correlation differences between phenotypes. The genetic correlations (mean absolute ρ_g = 0.23; SD = 0.25) were significantly lower than the average transcriptomic correlations (mean absolute ρ_t = 0.30; SD = 0.31) (paired sample

t-test; t-statistic=-3.48; $P < 0.001$). The genetic correlations explained a large proportion of the variance in transcriptomic correlations (Figure 2B; $R^2 = 0.7808$; $P < 2.2 \times 10^{-16}$), with the most pronounced difference between ADHD and ASD, with $\rho_g = 0.35$ (SE = 0.05) and $\rho_t = 0.84$ (SE = 0.05). A hierarchical cluster analyses of genetic and transcriptomic correlations showed similar groupings between genetic and transcriptomic correlations (Supplementary Figure 2). Both analyses, for example, were suggestive of strong sharing of genetic risk factors between anxiety and depression, and between bipolar disorder and schizophrenia. However, despite these similarities, some interesting differences were also revealed; for example, ADHD and ASD were grouped together in the transcriptomic cluster analysis but not the genetic cluster analysis in line with the results from the genetic and transcriptomic correlation analysis.

Co-expression network analysis

We identified 25 gene co-expression modules which ranged between 85 and 3,042 genes in size. Biological pathway enrichment analysis showed each module contained genes involved in the same or similar biological pathways (for example, the immune response [module M9] or trans-synaptic signaling [M25]; Supplementary Table 9). We tested for the enrichment of TWAS FUSION gene-based association signals within each module, while adjusting for gene size, gene density, and correlated expression. Six modules were associated with at least one psychiatric disorder (FDR<0.05) (Figure 3). The strongest association was found between module M19, enriched with genes involved in mRNA splicing, and anxiety (FDR = 0.0063). Full results are provided in Supplementary Table 10 (including the MHC region) and Supplementary Table 11 (excluding MHC). We partitioned the heritability explained by the six modules (Figure 4A) and showed 7 significant associations after Bonferroni correction for number of modules and traits (P<0.000054). After including baseline functional annotations, a single module (module M10), enriched with nucleic acid and RNA metabolism pathways, remained significant in bipolar disorder and schizophrenia.

DISCUSSION

We performed a systematic, network-based, analysis of genetic and transcriptomic risk factors underlying neuropsychiatric and substance use phenotypes. By integrating GWAS summary statistics for 10 neuropsychiatric and 6 substance use phenotypes with gene expression data from the prefrontal cortex, we identified 2,176 significant (FDR<0.05) gene-trait associations (representing unique 1,645 genes). After the removal of known gene-based associations, schizophrenia had the largest number of novel gene-based associations, followed by the substance use traits smoking initiation and drinks per week. We found evidence of widespread pleiotropic effects underlying phenotype-associated genetically regulated gene expression. This was most noticeable with depression and schizophrenia, where 8 of the top 20 most strongly associated genes across all phenotypes, including genes (N=6) in the MHC region, showed a significant association with concordant effects. We estimated the correlation between genetically regulated gene expression levels underlying neuropsychiatric and substance use traits. The transcriptomic correlations were significantly larger than the genetic correlations, and several phenotype pairs—for example, ASD and ADHD—showed a large difference in the magnitude of correlation between each method. Gene co-expression modules built from control (i.e. healthy) prefrontal cortex tissue samples were enriched with neuropsychiatric and substance use association signals and implicated multiple biologically meaningful pathways in disease/trait susceptibility. Collectively, these data suggest genetic regulation of gene expression measured from healthy subjects contains highly relevant biological information for the interpretation of disease susceptibility.

To prioritise genes whose expression is most strongly associated with multiple traits, we combined and ranked association signals for the investigated traits using Brown's method. Increased expression of the most strongly associated gene, *PSMA4*, was significantly associated with schizophrenia, cigarettes per

day, and smoking cessation. The gene *PSMA4*, located within the 15q25.1 gene cluster, has previously been associated with nicotine dependence and lung cancer (29), and we recently linked its expression in multiple GTEx brain tissues to cigarettes per day and smoking cessation (30). *PSMA4* has also been identified as one of six "high confidence" genes in schizophrenia, based on probabilistic fine mapping approaches and observed expression profiles (31). Interestingly, using observed expression data, these authors found *decreased PSMA4* expression in prefrontal cortex and hippocampus was associated with schizophrenia, while we reported the opposite effect direction with imputed (genetically regulated) gene expression in prefrontal cortex. Our association is consistent with previously reported *PSMA4* associations for schizophrenia in brain using transcriptome imputation methods TWAS FUSION (21) and S-PrediXcan (32). It is possible the observed expression data (GSE21138) were confounded by a hidden or surrogate variable, such as current smoking status, which may explain the association with *PSMA4* in schizophrenia cases compared to controls, rather than a causal disease process. This highlights a major advantage of transcriptome imputation methods, which remove environmental noise by focussing on the genetically regulated component of gene expression.

We estimated genome-wide genetic correlations at the level of predicted expression and show that 56 of the 112 trait pairs are significantly correlated at FDR<0.05. In line with the large genetic overlap between these disorders (5), predicted expression levels were strongly correlated between bipolar disorder and schizophrenia (ρ_t=0.84) and between anxiety and depression (ρ_t=0.93). A systematic comparison of the transcriptomic and genetic correlations revealed a strong relationship (R^2 = 0.78, $P < 2.2 \times 10^{-16}$), although on average the predicted expression levels were found to be more strongly correlated than genetic variation. For example, ASD and ADHD, two common childhood onset neurodevelopmental disorders, are more strongly correlated at the transcriptomic level (ρ_t=0.84) than the genetic level (ρ_g=0.35). The strong transcriptomic correlation between ASD and ADHD is not only

315 supported by the genetic correlation between the disorders but also their phenotypic similarity, where a
316 large proportion of children (37-85%) with ASD have comorbid symptoms of ADHD (33). Furthermore,
317 exome sequencing of children with ASD and ADHD indicated that they have a similar burden of rare
318 protein-truncating variants (34). While clinical guidelines dictate ASD cannot be diagnosed in the
319 presence of ADHD, our data suggests the high co-occurrence of these disorders is due to a shared
320 genetic regulation (35).
321
322 We can only speculate as to why the genetic correlations are generally lower than their respective
323 transcriptomic correlations. One possible explanation is the assumptions of LDSC, such as a highly
324 polygenic genetic architecture underlying the investigated phenotypes, may be violated in our study. For
325 example, it is possible LDSC yields an underestimate of shared genetic regulation by incorrectly
326 modelling the contribution of genomic regions more strongly enriched for heritability for some mental
327 health traits, while the transcriptomic correlation captures a truly high genetic overlap. However, it is
328 also possible the transcriptomic correlations are inflated due to the local correlation structure of gene
329 expression at a locus associated with two or more phenotypes. These scenarios may be investigated
330 using recently developed computational tools for casual inference, such as FOCUS (36) or MR-JTI (37), to
331 identify a reliable set of independent causal genes underlying each phenotype.
332
333 Our gene co-expression network analysis of prefrontal cortex identified modules of genes enriched with
334 gene-based associations for four neuropsychiatric disorders (anxiety, bipolar disorder, obsessive
335 compulsive disorder, autism spectrum disorder), and three substance use phenotypes (cigarettes per
336 day, cannabis initiation, and age of smoking initiation). The most strongly associated module was
337 associated with anxiety and strongly enriched in biological pathways associated with mRNA splicing.
338 Splicing is genetically regulated (38) and can influence gene expression in particular tissues, giving rise to

different functional effects such as altered neuronal connectivity and synaptic firing properties in the brain (39). Alternative mRNA splicing events are associated with diverse neuropsychiatric disorders, including schizophrenia (40), autism spectrum disorder (41), bipolar disorder (42), and major depression (43), highlighting the importance of alternative splicing in neuropsychiatric disease susceptibility. Current genomic resources, such as the latest release (version 8) of the Genotype-Tissue Expression study (GTEx) (38), will help researchers better understand how genetic variants affect gene expression through alternative splicing events. Other trait-associated modules were enriched with biologically meaningful pathways. For example, the module M1 was associated with bipolar disorder and enriched with genes involved in the regulation of metabotropic glutamate receptors. Glutamatergic receptors are the primary effectors of glutamate, a critical excitatory neurotransmitter, and their dysregulation is implicated in many neuropsychiatric disorders (44), including bipolar disorder (45). Collectively, these data suggest gene co-expression networks may be used as a molecular substrate for the biological characterisation of genetic risk factors underlying neuropsychiatric and substance use traits.

Stratified heritability analyses revealed significant enrichment of network module co-expression with mental health traits. Annotations for a single module, enriched with genes involved in nucleic acid and RNA processing, was significantly associated with bipolar disorder and schizophrenia after adjusting for baseline annotations. The loss of most modular enrichments after baseline annotation adjustment is in line with the findings of Kim *et al.*, who explored the association between genes with network connectivity and 42 traits and showed that significant enrichments of genetic networks were fully explained by excess overlap between network annotations and regulatory annotations from the baseline LD-models (46). The loss of module enrichments following baseline annotation can be expected, and most likely show that observed modular enrichment is explained by current knowledge on functional and regulatory elements in the human genome, rather than some unexplained biological process.

364 The findings of this study should be interpreted in view of the following limitations. First, the TWAS
365 FUSION expression imputation approach is only valid if disease risk is mediated through expression and
366 the expression weights were generated in a disease-relevant or appropriate proxy tissue or cell type. For
367 example, expression changes associated with depression are most strongly associated with microglial
368 cells (47), while altered expression underlying schizophrenia is enriched in neurons (48,49). Expression
369 weights from PyschENCODE were not available at single cell resolution. Therefore, the imputed
370 expression effects may reflect a mosaic of expression effects from multiple cell types rather than a single
371 causal cell type, or the sharing of genetic regulation of gene expression (37). The generation of large
372 single-cell eQTL datasets from the human brain will provide a valuable resource to disentangle cell-
373 specific effects (50,51). Second, the TWAS FUSION approach does not test whether gene expression and
374 a phenotype are affected by the same causal SNP in a *cis*-eQTL region. As such, the approach does not
375 provide direct evidence of causal relationship between expression and disease risk. Mendelian
376 randomisation-based approaches, such as SMR (52) and MR-JTI (37), may refine our list of gene
377 candidates by selecting genes most likely associated through pleiotropy, where gene expression and a
378 phenotype are affected by the same causal variant. Finally, our gene co-expression analyses rely on the
379 stability (i.e. robustness) of gene co-expression networks in prefrontal cortex. We built signed networks
380 using similar parameters described by Gandal *et al.* (16). Using a permutation procedure, these authors
381 compared each module's density (that is, the average strength of association, or connectivity, between
382 genes in a module) to the density of modules of equivalent size. These authors concluded psychENCODE
383 prefrontal cortex modules were robust to the influence of outlier samples on network architecture,
384 providing confidence in the stability of our co-expression network.

385

386 Our study highlights the benefits of integrating GWAS studies from mental health phenotypes with large
387 scale transcriptomic information to identify the functional impact of disease-causing variants. By

integrating transcriptomic data from prefrontal cortex with GWAS data, we identified hundreds of candidate risk genes not previously identified using commonly-used proximity-based and eQTL gene mapping methods. We found a significant difference between transcriptomic and genetic correlations across all phenotype pairs, and the magnitude of the difference was particularly large for ADHD and ASD. These data suggest transcriptomic correlations, which take correlations across genes into account, may provide additional insight into the functional relationship between mental health phenotypes. Finally, we observed some enrichment and convergence of candidate risk genes for mental health traits within co-expression networks from prefrontal cortex, suggesting our approach will prove useful in characterising the functional impact of trait-associated genetic variation. Future analyses could extend our approach by incorporating additional sources of genomic (for example, epigenetic marks) and statistical (e.g. SNP priors) information within co-expression networks.

REFERENCES

1. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T (2015): The Global Burden of Mental, Neurological and Substance Use Disorders: An Analysis from the Global Burden of Disease Study 2010. *PLoS One* 10: e0116820.

2. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, *et al.* (2019): Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 51: 237–244.

3. Lee PH, Anttila V, Won H, Feng Y-CA, Rosenthal J, Zhu Z, *et al.* (2019): Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* 179: 1469-1482.e11.

4. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, *et al.* (2018): Psychiatric Genomics: An Update and an Agenda, 2017/10/03. *Am J Psychiatry* 175: 15–27.

5. Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, *et al.* (2018): Analysis of shared heritability in common disorders of the brain. *Science* 360. https://doi.org/10.1126/science.aap8757

6. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, *et al.* (2012): Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science (80-)* 337: 1190–1195.

7. Gamazon ER, Zwinderman AH, Cox NJ, Denys D, Derks EM (2019): Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat Genet* 51: 933–940.

8. Huckins LM, Dobbyn A, Ruderfer DM, Hoffman G, Wang W, Pardiñas AF, *et al.* (2019): Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. *Nat Genet* 51: 659–674.

9. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, *et al.* (2018): Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science (80-)* 359: 693 LP – 697.

10. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V, Aquino-Michaels K, Carroll RJ, *et al.* (2015): A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* advance on.

11. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, *et al.* (2016): Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 48: 245–252.

12. Gerring ZF, Gamazon ER, Derks EM (2019): A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. *PLOS Genet* 15: e1008245.

13. Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH (2016): The road to precision psychiatry: translating genetics into disease mechanisms. *Nat Neurosci* 19: 1397–1407.

14. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Consortium SWG of the PG, *et al.* (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47: 291–295.

15. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, *et al.* (2018): Comprehensive functional genomic

439 resource and integrative model for the human brain. *Science (80-)* 362: eaat8464.

440 16. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, *et al.* (2018): Transcriptome-wide
441 isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science (80-)* 362:
442 eaat8127.

443 17. Delaneau O, Marchini J, Consortium T 1000 GP (2014): Integrating sequence and array data to create
444 an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* 5: 3934.

445 18. Storey JD, Bass AJ, Dabney A, Robinson D (2015): qvalue: Q-value estimation for false discovery rate
446 control.

447 19. Poole W, Gibbs DL, Shmulevich I, Bernard B, Knijnenburg TA (2016): Combining dependent P-values
448 with an empirical adaptation of Brown's method. *Bioinformatics* 32: i430–i436.

449 20. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: Generalized Gene-Set Analysis of
450 GWAS Data. *PLOS Comput Biol* 11: e1004219.

451 21. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B (2017): Integrating Gene Expression
452 with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am J*
453 *Hum Genet* 100: 473–487.

454 22. Müllner D (2013): Fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *J*
455 *Stat Softw* 53: 1–18.

456 23. Langfelder P, Horvath S (2008): WGCNA: an R package for weighted correlation network analysis.
457 *BMC Bioinformatics* 9: 1–13.

458 24. Langfelder P, Zhang B, Horvath S (2008): Defining clusters from a hierarchical cluster tree: The
459 Dynamic Tree Cut package for R. *Bioinformatics* 24: 719–720.

460 25. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J (2016): g:Profiler—a web server
461 for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* 44: W83–W89.

462 26. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, *et al.* (2008): KEGG for linking genomes
463 to life and the environment. *Nucleic Acids Res* 36: D480–D484.

464 27. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, *et al.* (2005): Reactome: A
465 knowledgebase of biological pathways. *Nucleic Acids Res* 33: D428–D432.

466 28. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, *et al.* (2015): Partitioning
467 heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*
468 47: 1228–1235.

469 29. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, *et al.* (2008): Genome-wide association scan of
470 tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 40: 616–622.

471 30. Marees AT, Gamazon ER, Gerring Z, Vorspan F, Fingal J, den Brink W van, *et al.* (2019): Post-GWAS
472 analysis of six substance use traits improves the identification and functional interpretation of
473 genetic risk loci: Full list of International Cannabis Consortium members. *Drug Alcohol Depend*
474 107703.

475 31. Ma C, Gu C, Huo Y, Li X, Luo X-J (2018): The integrated landscape of causal genes and pathways in
476 schizophrenia. *Transl Psychiatry* 8: 67.

32. Pividori M, Rajagopal PS, Barbeira A, Liang Y, Melia O, Bastarache L, *et al.* (2020): PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *Sci Adv* 6: eaba2083.

33. Leitner Y (2014): The co-occurrence of autism and attention deficit hyperactivity disorder in children - what do we know? *Front Hum Neurosci* 8: 268.

34. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, *et al.* (2019): Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat Neurosci* 22: 1961–1965.

35. Rommelse NNJ, Geurts HM, Franke B, Buitelaar JK, Hartman CA (2011): A review on cognitive and brain endophenotypes that may be common in autism spectrum disorder and attention-deficit/hyperactivity disorder and facilitate the search for pleiotropic genes. *Neurosci Biobehav Rev* 35: 1363–1396.

36. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, Pasaniuc B (2019): Probabilistic fine-mapping of transcriptome-wide association studies. *Nat Genet* 51: 675–682.

37. Zhou D, Jiang Y, Zhong X, Cox NJ, Liu C, Gamazon ER (2020): A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat Genet*. https://doi.org/10.1038/s41588-020-0706-2

38. The GTEx Consortium atlas of genetic regulatory effects across human tissues (2020): *Science (80-)* 369: 1318 LP – 1330.

39. Vuong CK, Black DL, Zheng S (2016): The neurogenetics of alternative splicing. *Nat Rev Neurosci* 17: 265–281.

40. Flaherty E, Zhu S, Barretto N, Cheng E, Deans PJM, Fernando MB, *et al.* (2019): Neuronal impact of patient-specific aberrant NRXN1α splicing. *Nat Genet* 51: 1679–1690.

41. Farini D, Cesari E, Weatheritt RJ, La Sala G, Naro C, Pagliarini V, *et al.* (2020): A Dynamic Splicing Program Ensures Proper Synaptic Connections in the Developing Cerebellum. *Cell Rep* 31: 107703.

42. Harrison PJ, Geddes JR, Tunbridge EM (2018): The Emerging Neurobiology of Bipolar Disorder, 2017/11/20. *Trends Neurosci* 41: 18–30.

43. Le François B, Zhang L, Mahajan GJ, Stockmeier CA, Friedman E, Albert PR (2018): A Novel Alternative Splicing Mechanism That Enhances Human 5-HT1A Receptor RNA Stability Is Altered in Major Depression. *J Neurosci* 38: 8200 LP – 8210.

44. Crupi R, Impellizzeri D, Cuzzocrea S (2019): Role of Metabotropic Glutamate Receptors in Neurological Disorders. *Front Mol Neurosci* 12: 20.

45. Martí SB, Cichon S, Propping P, Nöthen M (2002): Metabotropic glutamate receptor 3 (GRM3) gene variation is not associated with schizophrenia or bipolar affective disorder in the German population. *Am J Med Genet* 114: 46–50.

46. Kim SS, Dai C, Hormozdiari F, van de Geijn B, Gazal S, Park Y, *et al.* (2019): Genes with High Network Connectivity Are Enriched for Disease Heritability. *Am J Hum Genet* 104: 896–913.

47. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, *et al.* (2018): Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 50: 668–681.

516 48. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, *et al.* (2018): Heritability
517 enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat*
518 *Genet* 50: 621–629.

519 49. Dong X, Liao Z, Gritsch D, Hadzhiev Y, Bai Y, Locascio JJ, *et al.* (2018): Enhancers active in dopamine
520 neurons are a primary link between genetic variation and neuropsychiatric disease. *Nat Neurosci*
521 21: 1482–1492.

522 50. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, *et al.* (2015): A survey of human brain
523 transcriptome diversity at the single cell level. *Proc Natl Acad Sci* 112: 7285 LP – 7290.

524 51. Agarwal D, Sandor C, Volpato V, Caffrey TM, Monzón-Sandoval J, Bowden R, *et al.* (2020): A single-
525 cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological
526 disorders. *Nat Commun* 11: 4183.

527 52. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, *et al.* (2016): Integration of summary data
528 from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 48: 481–487.

529

530

ACKNOWLEDMENTS

E.R.G. is supported by the National Human Genome Research Institute of the National Institutes of Health under Award Numbers R35HG010718 and R01HG011138.

DISCLOSURES

The authors have no conflicts of interest to disclose.

Table 1: Sample descriptions and SNP-based heritabilities of 16 mental health traits

Phenotype (abbreviation)	Sample size (N)			SNP-based heritability		TWAS FUSION (N)	
	Case	Control	Total	h^2_{snp} (se)	Z	Genes	Novel
Anorexia Nervosa (AN)	16992	55525	72517	0.429 (0.0282)	15.21	74	20
Attention Deficit Hyperactivity Disorder (ADHD)	19099	34194	53293	0.217 (0.0141)	15.41	53	12
Autism Spectrum Disorder (ASD)	18382	27969	46351	0.112 (0.0097)	11.54	54	13
Anxiety Disorders (ANX)	31977	82114	114019	0.125 (0.0090)	13.92	42	41
Bipolar Disorder (BIP)	20352	31358	51710	0.200 (0.0101)	19.83	171	58
Depression (DEP)	246363	561190	807553	0.073 (0.0025)	29.00	185	37
Obsessive Compulsive Disorder (OCD)	2688	7037	9725	0.280 (0.0432)	6.48	15	15
Post Traumatic Stress Disorder (PTSD)	23212	151447	174659	0.053 (0.0095)	5.53	18	13
Schizophrenia (SCZ)	40675	64643	105318	0.234 (0.0083)	28.17	597	112
Tourette's Syndrome (TS)	4819	9488	14307	0.213 (0.0248)	8.59	27	15
Drinks Per Week (DrnkWk)	-	-	537349	0.049 (0.0021)	23.19	260	62
Smoking Initiation (SmkInit)	311629	321173	632802	0.104 (0.0033)	31.64	312	72
Cigarettes Per Day (CigDay)	-	NA	263954	0.073 (0.0069)	10.52	163	36
Smoking Cessation (SmkCes)	92573	220248	312821	0.060 (0.0039)	15.36	61	19
Age of Smoking Initiation (AgeSmk)	-	-	262990	0.047 (0.0028)	16.93	70	23
Cannabis Use Initiation (CanInit)	43380	118702	162082	0.118 (0.0075)	15.69	74	24

ADHD: (Demontis *et al.*, 2019); AN: (Duncan *et al.*, 2017); ASD: (Grove *et al.*, 2019); ANX: (Otowa *et al.*, 2016); BIP, (Stahl *et al.*, 2019); DEP, (Howard *et al.*, 2019); OCD: (Arnold *et al.*, 2018); PTSD: (Nievergelt *et al.*, 2019); TS: (Yu *et al.*, 2019); CanInit: (Pasman *et al.*, 2018); SCZ: (Pardiñas *et al.*, 2018); DrnkWk, SmkInit, CigDay, SmkCes, AgeSmk: (Liu *et al.*, 2019).

537

FIGURE LEGENDS

Figure 1A: Brown's method Z scores for the top 20 genes across 16 mental health phenotypes.

Figure 1B: Brown's method Z scores for the top gene for each mental health phenotype.

Figure 2A: Correlations across the 16 pairs of mental health phenotypes (excluding MHC region) based on genetic variation (below diagonal) and genetically regulated gene expression (above diagonal).

Figure 2B: Scatter plot of genetic and transcriptomic correlations (excluding MHC region) across 16 mental health phenotype pairs.

Figure 3: Circos plot of TWAS FUSION Z scores, modular enrichments, and significant transcriptomic correlations across 16 mental health phenotypes. Notes: The outermost circle highlights significant (FDR<0.05) TWAS FUSION associations; second middle layer shows the distribution of TWAS FUSION Z scores for each phenotype; the inner most layer shows the enrichment Z scores for each of the six significant co-expression modules in prefrontal cortex, with darker shading signifying greater enrichment; the inner ribbons represent significant transcriptomic correlations across phenotype pairs.

Figure 4A. Heritability Enrichment of six co-expression network annotations. The figure illustrates heritability enrichment of network annotations for Alzheimer's Disease GWAS. Coloured squares represent significant enrichment after Bonferroni correction for 16*6 tests ($P<5.2 \times 10^{-4}$). Notes: Enrichment Z scores outside the bounds -20 to 20 have been truncated. See Supplementary Table 12 for full list of heritability enrichment Z scores.

Figure 4B. Heritability Enrichment of six co-expression network annotations when taking baseline functional annotations into account. The figure illustrates heritability enrichment of network and baseline annotations for Alzheimer's Disease GWAS. Coloured squares represent significant enrichment after Bonferroni correction for 16*58 tests ($P<5.3 \times 10^{-5}$). Notes: Enrichment Z scores outside the bounds -20 to 20 have been truncated. See Supplementary Table 13 for full list of heritability enrichment with baseline annotation Z scores.

Supplementary Figure 1: Pairwise differences in genetic and transcriptomic correlations across 16 mental health phenotypes.

Supplementary Figure 2: Hierarchical cluster analyses of genetic and transcriptomic correlations for 16 mental health phenotypes.

WEB RESOURCES

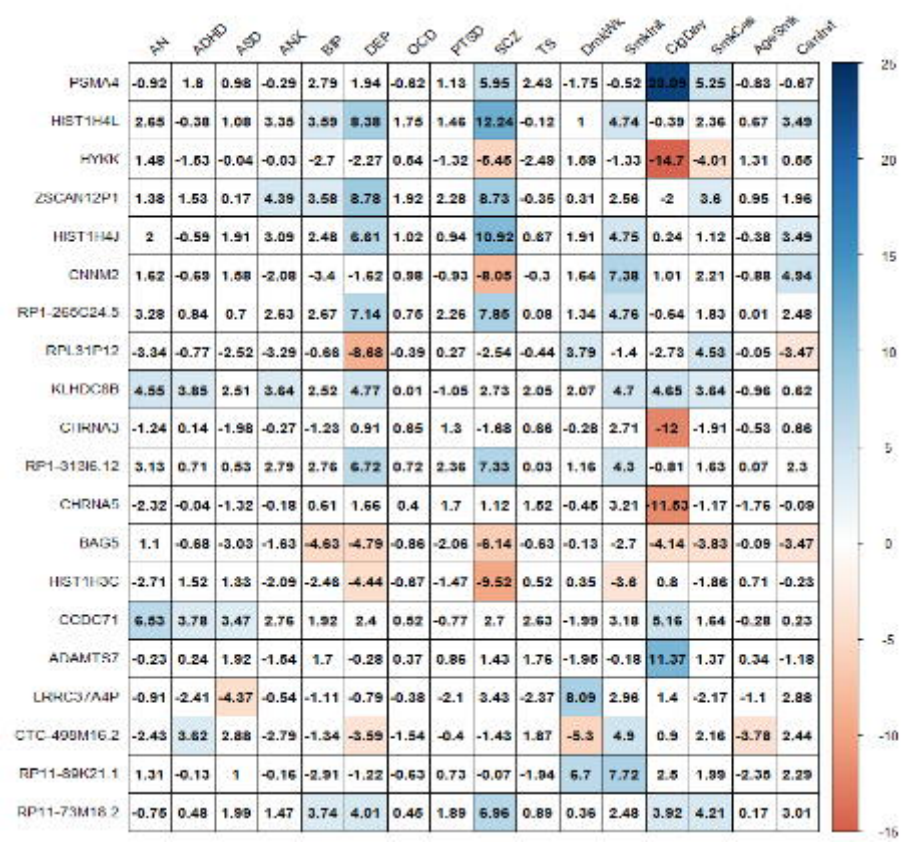
TWAS FUSION http://gusevlab.org/projects/fusion/

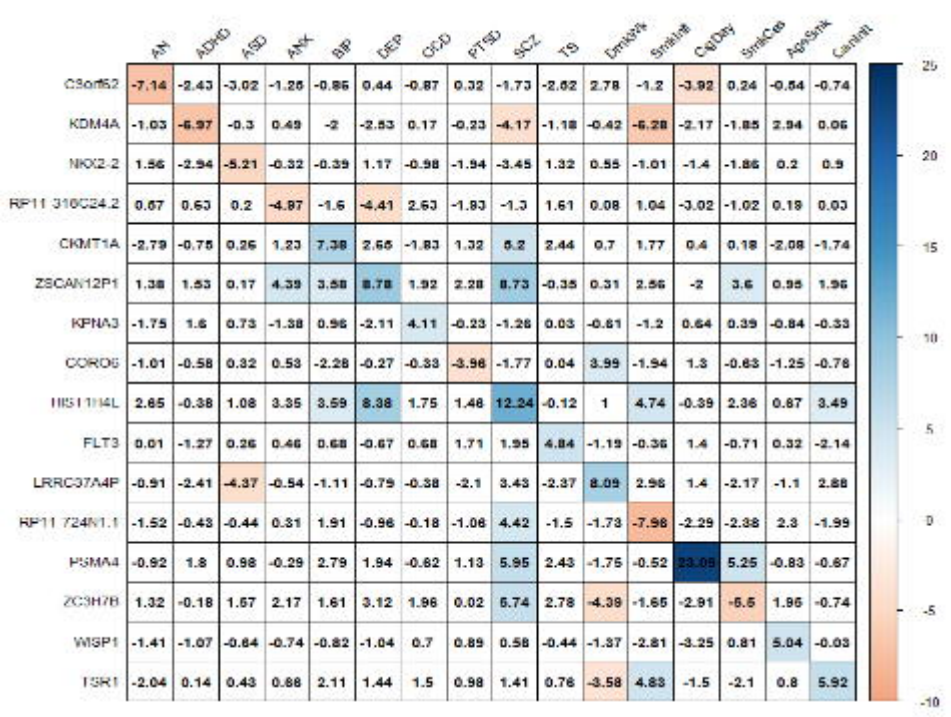
PsychENCODE constortium http://resource.psychencode.org/

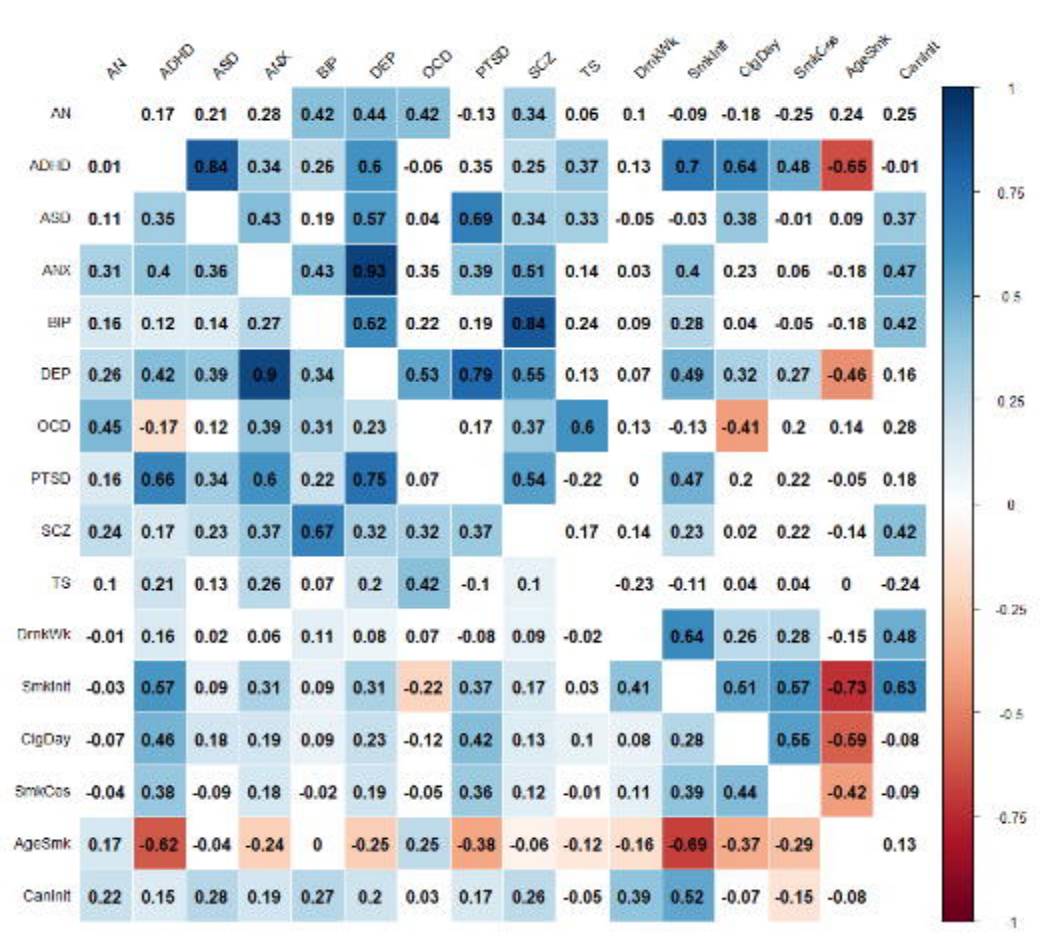
MAGMA https://ctg.cncr.nl/software/magma

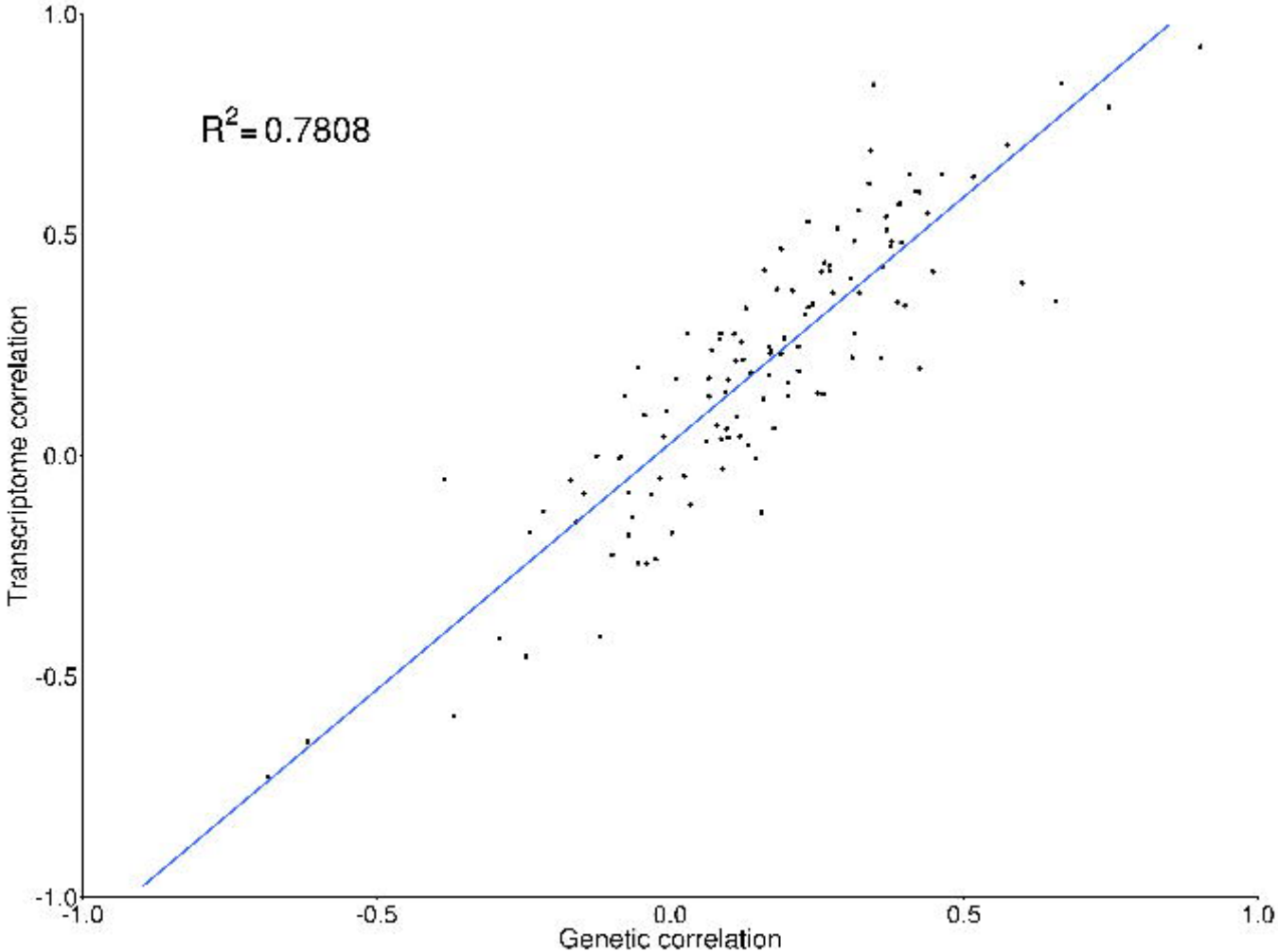
RhoGE https://github.com/bogdanlab/RHOGE

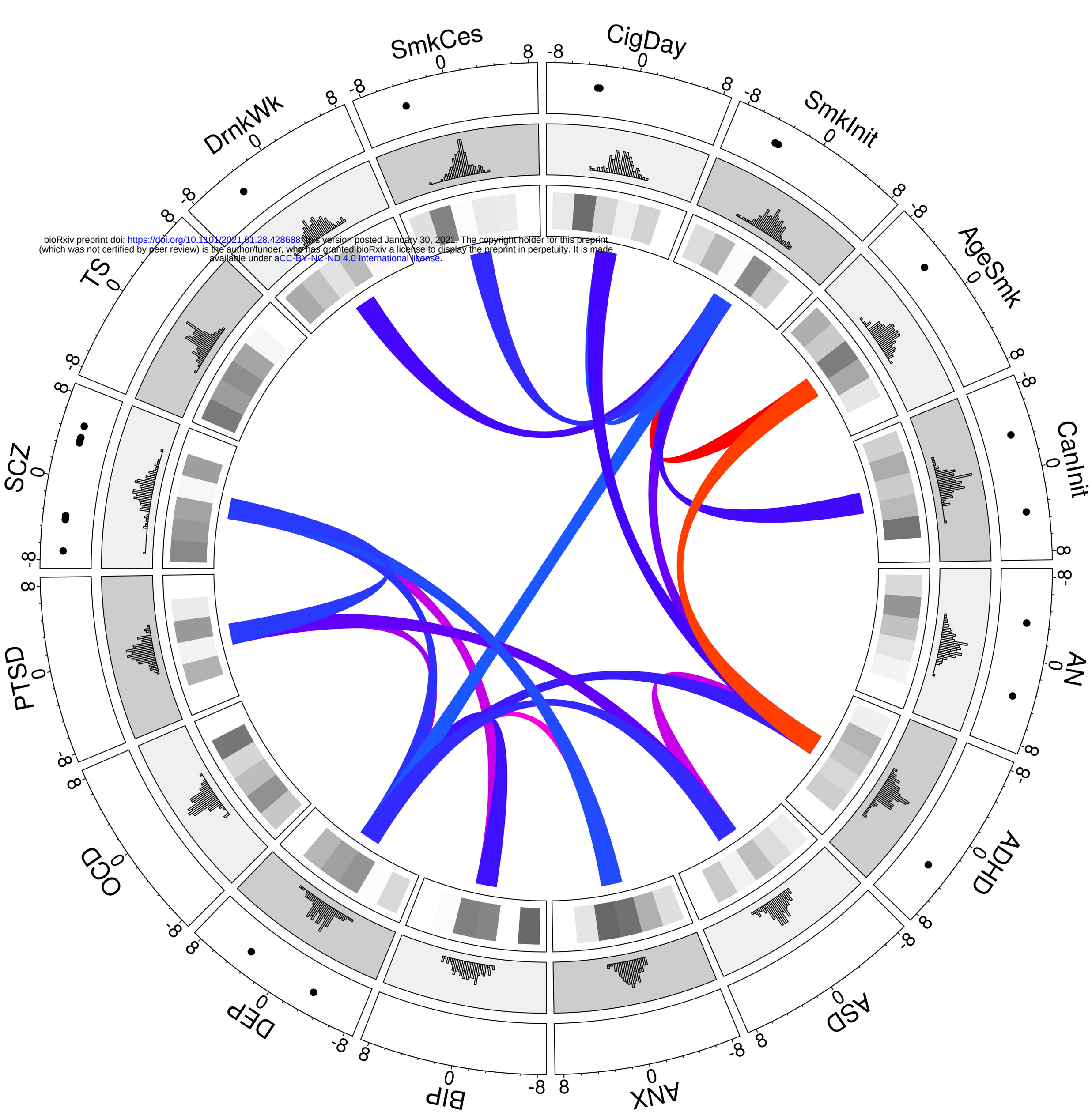
WGCNA https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/

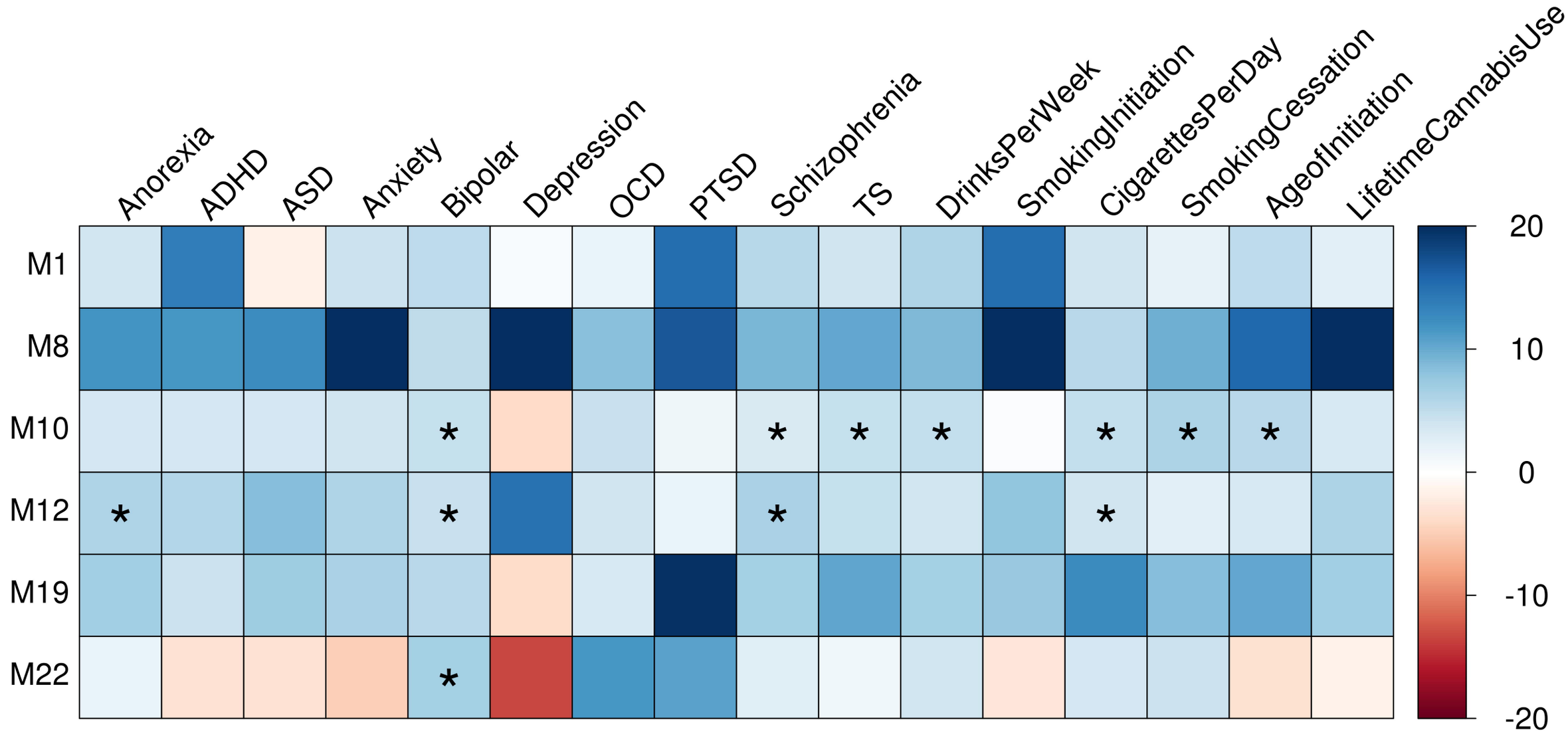

AN
ADHD
ASD
ANX
BIP
DEP
OCD
PTSD
SCZ
TS
DrugWk
SmkHist
CigDay
SmkCess
AgeSmk
Cannbt
PSMA4 -0.92 1.8 0.98 -0.29 2.79 1.94 -0.82 1.13 5.95 2.43 -1.75 -0.52 20.09 5.25 -0.83 -0.87
HIST1H4L 2.65 -0.38 1.08 3.35 3.59 8.38 1.75 1.46 12.24 -0.12 1 4.74 -0.39 2.36 0.67 3.49
HYKK 1.48 -1.63 -0.04 -0.03 -2.7 -2.27 0.64 -1.32 -5.45 -2.48 1.69 -1.33 -14.7 -4.01 1.31 0.65
ZSCAN12P1 1.38 1.53 0.17 4.39 3.58 8.78 1.92 2.28 8.73 -0.35 0.31 2.56 -2 3.8 0.95 1.96
HIST1H4J 2 -0.59 1.91 3.09 2.48 6.81 1.02 0.94 10.92 0.67 1.91 4.75 0.24 1.12 -0.38 3.49
CNNM2 1.62 -0.69 1.68 -2.08 -3.4 -1.62 0.98 -0.93 -8.05 -0.3 1.64 7.38 1.01 2.21 -0.88 4.94
RP1-265C24.5 3.28 0.84 0.7 2.63 2.67 7.14 0.76 2.26 7.85 0.08 1.34 4.76 -0.64 1.83 0.01 2.48
RPL31P12 -3.34 -0.77 -2.52 -3.29 -0.66 -8.68 -0.39 0.27 -2.54 -0.44 3.79 -1.4 -2.73 4.53 -0.05 -3.47
KLHDC8B 4.55 3.85 2.51 3.64 2.52 4.77 0.01 -1.05 2.73 2.05 2.07 4.7 4.65 3.64 -0.96 0.62
CTRNA1 -1.24 0.14 -1.98 -0.27 -1.23 0.91 0.85 1.3 -1.88 0.88 -0.28 2.71 -12 -1.91 -0.53 0.88
RP1-313I6.12 3.13 0.71 0.53 2.79 2.76 6.72 0.72 2.36 7.33 0.03 1.16 4.3 -0.81 1.63 0.07 2.3
CHRNA5 -2.32 -0.04 -1.32 -0.18 0.61 1.66 0.4 1.7 1.12 1.52 -0.45 3.21 -11.53 -1.17 -1.76 -0.09
BAG5 1.1 -0.68 -3.03 -1.63 -4.63 -4.79 -0.86 -2.06 -6.14 -0.63 -0.13 -2.7 -4.14 -3.83 -0.09 -3.47
HIST1H3C -2.71 1.52 1.33 -2.09 -2.48 -4.44 -0.87 -1.47 -9.52 0.52 0.35 -3.6 0.8 -1.86 0.71 -0.23
CCDC71 6.53 3.78 3.47 2.76 1.92 2.4 0.52 -0.77 2.7 2.63 -1.99 3.18 5.16 1.64 -0.28 0.23
ADAMTS7 -0.23 0.24 1.92 -1.64 1.7 -0.28 0.37 0.86 1.43 1.76 -1.95 -0.18 11.37 1.37 0.34 -1.18
LRRC37A4P -0.91 -2.41 -4.37 -0.54 -1.11 -0.79 -0.38 -2.1 3.43 -2.37 8.09 2.96 1.4 -2.17 -1.1 2.88
CTC-499M16.2 -2.43 3.62 2.88 -2.79 -1.34 -3.59 -1.54 -0.4 -1.43 1.87 -5.3 4.9 0.9 2.16 -3.78 2.44
RP11-89K21.1 1.31 -0.13 1 -0.16 -2.91 -1.22 -0.63 0.73 -0.07 -1.94 6.7 7.72 2.5 1.99 -2.35 2.29
RP11-73M18.2 -0.75 0.48 1.99 1.47 3.74 4.01 0.46 1.89 6.96 0.89 0.36 2.48 3.92 4.21 0.17 3.01
25
20
15
10
5
0
-5
-10
-15


AN
ADHD
ASD
ANX
BP
DEP
OCD
PTSD
SCZ
TS
Drink/Wk
Smk/Hlf
Cig/Day
Smk/Cess
Age Smk
Current
C3orf62 -7.14 -2.43 -3.02 -1.26 -0.86 0.44 -0.87 0.32 -1.73 -2.62 2.78 -1.2 -3.92 0.24 -0.64 -0.74
KDM4A -1.03 -6.97 -0.3 0.49 -2 -2.53 0.17 -0.23 -4.17 -1.18 -0.42 -6.28 -2.17 -1.85 2.94 0.06
NKX2-2 1.56 -2.94 -5.21 -0.32 -0.39 1.17 -0.98 -1.94 -3.45 1.32 0.55 -1.01 -1.4 -1.86 0.2 0.9
RP11 316C24.2 0.67 0.63 0.2 -4.97 -1.6 -4.41 2.63 -1.83 -1.3 1.61 0.08 1.04 -3.02 -1.02 0.19 0.03
CKMT1A -2.79 -0.75 0.26 1.23 7.38 2.65 -1.83 1.32 6.2 2.44 0.7 1.77 0.4 0.18 -2.08 -1.74
ZSCAN12P1 1.38 1.53 0.17 4.39 3.58 8.78 1.92 2.28 8.73 -0.35 0.31 2.56 -2 3.6 0.95 1.96
KPNA3 -1.75 1.6 0.73 -1.38 0.96 -2.11 4.11 -0.23 -1.26 0.03 -0.61 -1.2 0.64 0.39 -0.84 -0.33
CORO6 -1.01 -0.58 0.32 0.53 -2.28 -0.27 -0.33 -3.96 -1.77 0.04 3.99 -1.94 1.3 -0.63 -1.25 -0.78
HIST1H4L 2.65 -0.38 1.08 3.35 3.59 8.38 1.75 1.48 12.24 -0.12 1 4.74 -0.39 2.36 0.87 3.49
FLT3 0.01 -1.27 0.26 0.46 0.68 -0.67 0.68 1.71 1.95 4.84 -1.19 -0.36 1.4 -0.71 0.32 -2.14
LRRC37A4P -0.91 -2.41 -4.37 -0.54 -1.11 -0.79 -0.38 -2.1 3.43 -2.37 8.09 2.96 1.4 -2.17 -1.1 2.88
RP11 724N1.1 -1.52 -0.43 -0.44 0.31 1.91 -0.96 -0.18 -1.06 4.42 -1.5 -1.73 -7.98 -2.29 -2.38 2.3 -1.99
PSMA4 -0.92 1.8 0.98 -0.29 2.79 1.94 -0.62 1.13 5.95 2.43 -1.75 -0.52 23.05 5.25 -0.83 -0.67
ZC3H7B 1.32 -0.18 1.67 2.17 1.61 3.12 1.96 0.02 5.74 2.78 -4.39 -1.65 -2.91 -5.5 1.96 -0.74
WISP1 -1.41 -1.07 -0.64 -0.74 -0.82 -1.04 0.7 0.89 0.58 -0.44 -1.37 -2.81 -3.25 0.81 5.04 -0.03
TSR1 -2.04 0.14 0.43 0.88 2.11 1.44 1.5 0.98 1.41 0.76 -3.58 4.83 -1.5 -2.1 0.8 5.92
25
20
15
10
5
0
-5
-10


AN ADHD ASD ANX BIP DEP OCD PTSD SCZ TS DrnkWk SmkInit CigDay SmkCes AgeSmk CanInit
AN 0.17 0.21 0.28 0.42 0.44 0.42 -0.13 0.34 0.06 0.1 -0.09 -0.18 -0.25 0.24 0.25
ADHD 0.01 0.84 0.34 0.26 0.6 -0.06 0.35 0.25 0.37 0.13 0.7 0.64 0.48 -0.65 -0.01
ASD 0.11 0.35 0.43 0.19 0.57 0.04 0.69 0.34 0.33 -0.05 -0.03 0.38 -0.01 0.09 0.37
ANX 0.31 0.4 0.36 0.43 0.93 0.35 0.39 0.51 0.14 0.03 0.4 0.23 0.06 -0.18 0.47
BIP 0.16 0.12 0.14 0.27 0.62 0.22 0.19 0.84 0.24 0.09 0.28 0.04 -0.05 -0.18 0.42
DEP 0.26 0.42 0.39 0.9 0.34 0.53 0.79 0.55 0.13 0.07 0.49 0.32 0.27 -0.46 0.16
OCD 0.45 -0.17 0.12 0.39 0.31 0.23 0.17 0.37 0.6 0.13 -0.13 -0.41 0.2 0.14 0.28
PTSD 0.16 0.66 0.34 0.6 0.22 0.75 0.07 0.54 -0.22 0 0.47 0.2 0.22 -0.05 0.18
SCZ 0.24 0.17 0.23 0.37 0.67 0.32 0.32 0.37 0.17 0.14 0.23 0.02 0.22 -0.14 0.42
TS 0.1 0.21 0.13 0.26 0.07 0.2 0.42 -0.1 0.1 -0.23 -0.11 0.04 0.04 0 -0.24
DrnkWk -0.01 0.16 0.02 0.06 0.11 0.08 0.07 -0.08 0.09 -0.02 0.64 0.26 0.28 -0.15 0.48
SmkInit -0.03 0.57 0.09 0.31 0.09 0.31 -0.22 0.37 0.17 0.03 0.41 0.51 0.57 -0.73 0.63
CigDay -0.07 0.46 0.18 0.19 0.09 0.23 -0.12 0.42 0.13 0.1 0.08 0.28 0.55 -0.59 -0.08
SmkCes -0.04 0.38 -0.09 0.18 -0.02 0.19 -0.05 0.36 0.12 -0.01 0.11 0.39 0.44 -0.42 -0.09
AgeSmk 0.17 -0.62 -0.04 -0.24 0 -0.25 0.25 -0.38 -0.06 -0.12 -0.16 -0.69 -0.37 -0.29 0.13
CanInit 0.22 0.15 0.28 0.19 0.27 0.2 0.03 0.17 0.26 -0.05 0.39 0.52 -0.07 -0.15 -0.08
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1


$R^2 = 0.7808$
Transcriptome correlation
Genetic correlation
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0




Anorexia
ADHD
ASD
Anxiety
Bipolar
Depression
OCD
PTSD
Schizophrenia
TS
DrinksPerWeek
SmokingInitiation
CigarettesPerDay
SmokingCessation
AgeofInitiation
LifetimeCannabisUse
M1
M8
M10
M12
M19
M22
20
10
0
-10
-20