Complex Psychiatry

Research Article

Complex Psychiatry 2022;8:35–46
DOI: 10.1159/000523748

Received: September 10, 2021
Accepted: February 16, 2022
Published online: February 28, 2022

Association of Predicted Expression and Multimodel Association Analysis of Substance Abuse Traits

Darius M. Bost[a, b] Chris Bizon[a, b] Jeffrey L. Tilson[a, b] Dayne L. Filer[a, b]
Ian R. Gizer[c] Kirk C. Wilhelmsen[a, b, d, e]

[a]Department of Genetics, School of Medicine, UNC-Chapel Hill, Chapel Hill, NC, USA; [b]Renaissance Computing Institute, Chapel Hill, NC, USA; [c]Department of Psychological Sciences, University of Missouri, Columbia, MO, USA; [d]Department of Neurology, School of Medicine, UNC-Chapel Hill, Chapel Hill, NC, USA; [e]Department of Neurology, West Virginia University Rockefeller Neuroscience Institute, Morgantown, WV, USA

Keywords
Predictive genomics · Meta-analysis · Burden test · Addiction

Abstract
Introduction: Genome-wide association studies (GWAS) have played a critical role in identifying many thousands of loci associated with complex phenotypes and diseases. This has led to several translations of novel disease susceptibility genes into drug targets and care. This however has not been the case for analyses where sample sizes are small, which suffer from multiple comparisons testing. The present study examined the statistical impact of combining a burden test methodology, PrediXcan, with a multimodel meta-analysis, cross phenotype association (CPASSOC). ***Methods:*** The analysis was conducted on 5 addiction traits: family alcoholism, cannabis craving, alcohol, nicotine, and cannabis dependence and 10 brain tissues: anterior cingulate cortex BA24, cerebellar hemisphere, cortex, hippocampus, nucleus accumbens basal ganglia, caudate basal ganglia, cerebellum, frontal cortex BA9, hypothalamus, and putamen basal ganglia. Our sample consisted of 1,640 participants from the University of California, San Francisco (UCSF) Family Alcoholism Study. Genotypes were obtained through low pass whole genome sequencing and the use of Thunder, a linkage disequilibrium variant caller. ***Results:*** The post-PrediXcan, gene-phenotype association without aggregation resulted in 2 significant results, *HCG27* and *SPPL2B*. Aggregating across phenotypes resulted no significant findings. Aggregating across tissues resulted in 15 significant and 5 suggestive associations: *PPIE, RPL36AL, FOXN2, MTERF4, SEPTIN2, CIAO3, RPL36AL, ZNF304, CCDC66, SSPOP, SLC7A9, LY75, MTRF1L, COA5,* and *RRP7A*; *RPS23, GNMT, ERV3-1, APIP*, and *HLA-B,* respectively. ***Discussion:*** Given the relatively small size of the cohort, this multimodel approach was able to find over a dozen significant associations between predicted gene expression and addiction traits. Of our findings, 8 had prior associations with similar phenotypes through investigation of the GWAS Atlas. With the onset of improved transcriptome data, this approach should increase in efficacy.

Introduction

The medical community has embraced genetic analysis to understand the pathogenesis of clinically important phenotypes. Most common phenotypes, such as addiction, have complex modes of inheritance. Genome-wide

Karger@karger.com
www.karger.com/cxp

Correspondence to:
Darius M. Bost, dariusmb@ad.unc.edu

Table 1. Proportion endorsed and heritability

	Alcohol Dep.	Family alcoholism	Cannabis craving	Nicotine Dep.	Cannabis Dep.
Proportion endorsed	60.49	85.93	17.38	49.76	13.2
Heritability	0	0.63	0.57	0.4	0.23

Displays proportions of each phenotype endorsed by the cohort as well as the estimated heritability (by EPACTS). Dep. is shortened from Dependence.

association studies (GWAS) have been instrumental in detecting thousands of correlations between genetic variants and complex phenotypes [1]. GWAS limited by small sample size (thousands) typically detect few, if any, loci with modest effect sizes and explain little variation due to poor statistical power [2]. As sample sizes for some phenotypes have increased to hundreds of thousands, GWAS have detected many additional loci with diminishing effect sizes, affirming most complex phenotypes are highly polygenic. More efficient approaches are needed to elucidate genetic pathogenicity when sample sizes are limiting.

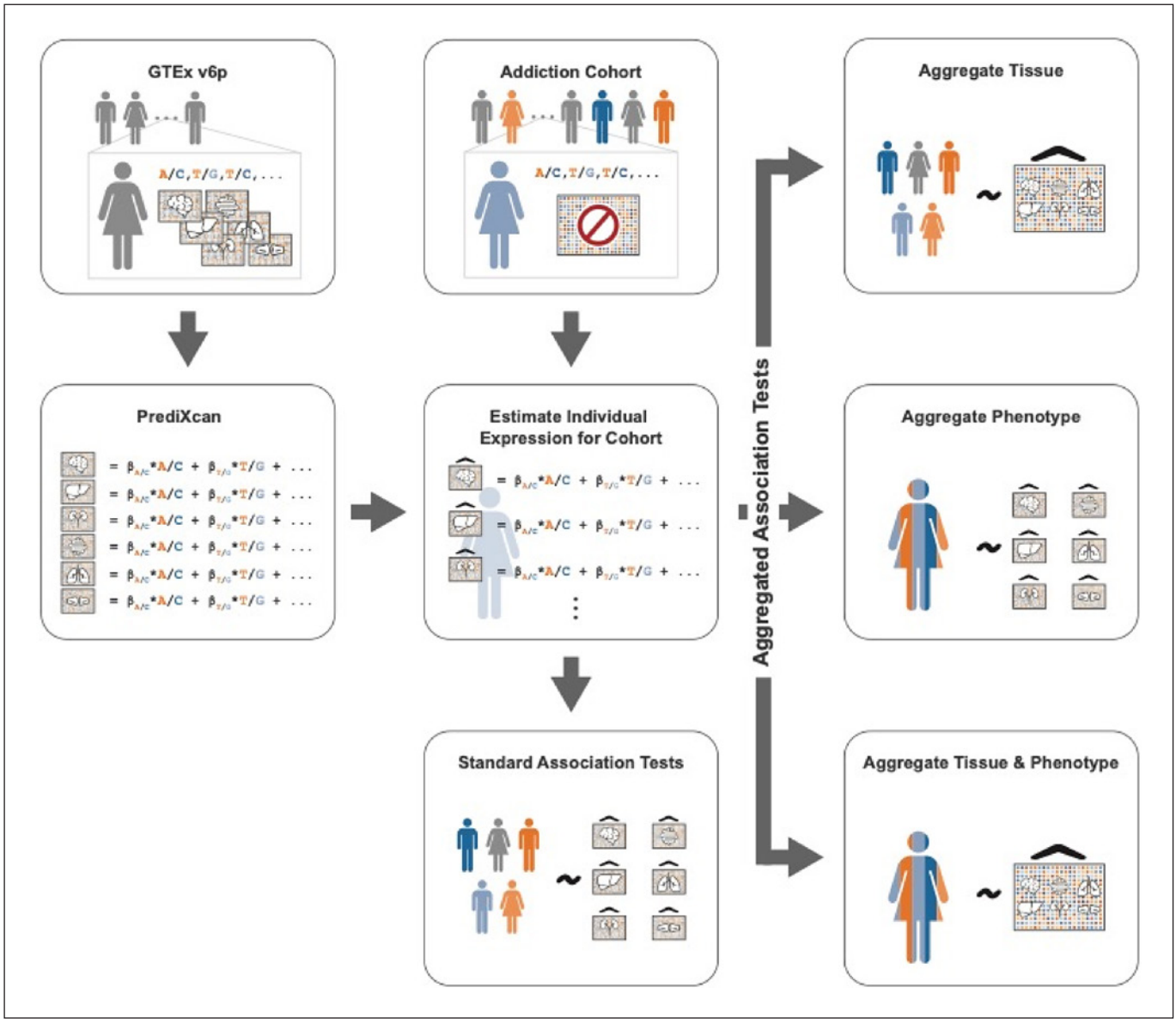
GWAS have identified several significant loci that affect addiction phenotypes. For example, findings suggest signals within the alcohol dehydrogenase (*ADH*) gene cluster (*ADH4, ADH1B*, and *ADH1C*), *ANKS1, AGBL4, GCKR,* and *MTIF2* influence risk for alcohol dependence and alcohol use disorders [3–6]. For nicotine dependence, the most statistically significant signals are in the nicotinic acetylcholine receptor gene clusters (*CHRNA5, CHRNA3, CHRNB4, CHRNB3, CHRNA6, CHRNA4*). Other implicated genes include *DBH, DNMT3B, FAM163B, HYKK, CTNNA3, NOL4L* [7–9]. Some significant associations mapped for cannabis dependence are located in or near *FOXP2* and *CHRNA2* as well as *AKFN1, CHST11, PIM3, UCHL5,* and *AL157388.1* [10–12]. These variants account for a small proportion of the expected genetic variation in the described traits. With the observed estimated heritability of these phenotypes, we expect to detect many more loci as sample sizes increase.

An organizing principle for understanding disease pathogenesis and identifying possible treatment modalities is to understand which genes are implicated. Identifying the mechanism of action of variants detected by GWAS is often arduous [13]. Most variants detected by GWAS fall in intergenic regions with indeterminate function. Functional intergenic variants are expected to mediate phenotypes by changing transcription of both proximal and very distant genes. A burden test that aggregates variants by their effect on gene-specific transcription has the potential to both increase power and our understanding of pathogenesis.

PrediXcan provides a framework for a gene burden test using models trained with genotype and specific expression data from several data consortiums [14]. As an alternative, transcript-wide association studies (TWAS) use a similar approach to predict gene expression, but rely on eQTL reference summary statistics [15]. Both approaches predict tissue-specific gene expression based on genotypes of study individuals for whom expression was not measured. Predicted expression values in study individuals are then tested for phenotypic association. Direct gene-phenotype associations bypass the arduous mechanistic interpretation of intergenic variants, though such variants may be excluded form analysis if their relation to gene expression cannot be quantified. Nonetheless, a further advantage of these methods, which is shared with other burden tests, is that it raises the significance threshold for declaring an association test result significant. In PrediXcan, the divisor is reduced from all variants to the number of tested genes, greatly increasing power.

In the current study, we applied PrediXcan to the University of California, San Francisco (UCSF) Family Alcoholism study to interrogate addiction phenotypes. The UCSF Family Alcoholism study seeks to identify genetic variation that increases or decreases susceptibility to addiction and other related phenotypes [16]. Previous analyses have revealed several loci for alcohol and nicotine dependence through linkage analysis [17, 18]. However, association analysis of candidate genes located within the identified linkage support intervals has not been successful.

In addition to the described difficulties related to single variant testing, association studies can have reporting bias if appropriate corrections for significance thresholds are not made when several related phenotypic models are tested. A more rigorous threshold for association testing is required when multiple, related phenotypes are analyzed. Furthermore, when multiple tissue-specific expression models (e.g., frontal cortex, striatum, parietal cortex, etc.) are tested for association with a phenotype, a more stringent threshold is also required for all association tests using the PrediXcan approach. One way to increase power is to use a multimodel association analysis of associa-



Color version available online

Fig. 1. The workflow starts with PrediXcan models trained using eqtl data from the GTEx consortium. The weights of these models are utilized to estimate expression for the UCSF addiction cohort. With our predicted expressions obtained, a standard association test is performed for each transcript per trait and tissue. To increase power by raising the Bonferroni threshold, aggregate multimodel associations were performed using CPASSOC. Aggregated tissue by phenotype, aggregated phenotype by tissue, aggregated tissue and phenotype.

tion statistics for multiple phenotypes and tissue-specific expression models [19, 20]. Multi-phenotype association analysis takes advantage of pleiotropy. Multi-tissue expression association analysis takes advantage of estimates of gene expression across neural tissues. We predicted that combining the test statistics across tissues would be a particularly effective approach, given that many brain regions have correlated gene expression.

Methods

UCSF Family Alcoholism Study

The cohort was obtained through a nationwide recruitment effort to understand the genetics of substance use disorders, including, but not limited to, alcoholism. Further objectives and designs for this study have previously been detailed [16, 21]. The cohort includes small nuclear families and unrelated subjects, with 92% reporting as White [11, 16]. The remaining racial distribution was

3% each for African American and Hispanic and 1% each Native American and other. In regards to the sex distribution of the cohort, 58% were female. Diagnosis of lifetime alcohol dependence and availability of both parents or one sibling for participation was required for proband study invitation. Relatives of probands were consented after probands obtained permission to contact their relatives. Out of 1,218 family pedigrees, 2,524 individuals were enrolled over the course of the study. Probands were excluded if they reported the use of opiates, cocaine, or stimulants weekly for 6 months or daily for more than 3 months. Probands with current or past diagnoses of bipolar disorder, schizophrenia, other psychiatric illnesses involving psychotic symptoms, the inability to read or speak English, or a threatening illness were also excluded [22].

Given quality control measures mentioned in prior work, our cohort was refined to 1,640. This final sample consisted of 761 families and an average family size of 2.17. The majority of the sample self-identified as White (98.1%), and female (63%). The mean age of the sample was 49.7 years old with a standard deviation of 12.9 years [23].

To collect data on alcohol, nicotine, drug-use history, medical history, and demographics, a modified version of the Semi-Structured Assessment for Genetics of Alcoholism (SSAGA) was administered via a telephone interview. This allowed for Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) diagnosis of alcohol and other substance misuse. Recruitment details of all participants have been previously published [22]. The phenotypes used were DSM-IV alcohol, nicotine, and cannabis dependence. In addition, we used the DSM-IV qualifier, cannabis craving ("In situations where you couldn't use cannabis, did you ever have such a strong desire for it that you couldn't think of anything else?"). Lastly, family alcoholism (phenotype designated in the SSAGA through inquiry of family's history with alcohol) was also included. A single-question screening approach was used within the modified SSAGA to acquire this trait ("Does alcoholism run in your family?"). It is reported that this method correlates well with diagnosis obtained by in-person interviews and the family history research diagnostic criteria for parental alcoholism [16]. Table 1 displays a data summary of the proportion endorsed and estimated heritability for each phenotype. Heritability estimates were computed using Efficient Mixed Model Association expedited (EMMAX) within the Efficient and Parallelizable Association Container Toolbox software (EPACTS) pipeline. These are pseudo-heritability estimates, which resemble estimates derived from pedigree, that are fractions of phenotypic variance explained by the empirically estimated relatedness matrix. The selection of these phenotypes was based on their estimated heritability and the assumption that some associated genes would have pleiotropic effects. Alcohol dependence, being an original focus of the UCSF study, was also included even though its heritability is approximately 0 [16].

Workflow

PrediXcan models of tissue-specific gene expression were applied to the genotypes in the UCSF family cohort to predict each individual's expression. Linear mixed models (GEMMA) were then used to simultaneously correct for family structure and test for associations between predicted expression and addiction phenotypes for the cohort. Finally, association test results were combined using a multimodel analysis. The workflow is illustrated in Figure 1.

Predicting Gene Expression

Briefly, PrediXcan GTEx models [14] (http://predictdb.hakyimlab.org) are based on the measured tissue-specific gene expression and genotypes for the samples collected by the GTEx Consortium. PrediXcan expression models are trained using an Elastic Net variable selection method on a restricted search space of both typed and imputed cis-single nucleotide polymorphisms (SNPs) within 1 Mb of gene start or end. Our analyses utilized PrediXcan models based on GTEX v8. The genotypes used to predict expression for the UCSF cohort were collected with low-pass whole genome sequencing (WGS) and linkage disequilibrium variant caller Thunder [11]. Low-pass WGS obtains increases in the genomic landscape interrogated relative to GWAS and WES, thus allowing for a greater chance of discovering novel variants and associations. While low pass WGS is not as good as deep WGS for detecting, with confidence, rare alleles or alleles that are specifically typed with fixed content arrays, LD-based variant calling increases calling accuracy in low-coverage samples to the point at which low-coverage approaches become viable [24]. We utilized 10 tissue models for analysis: anterior cingulate cortex BA24, cerebellar hemisphere, cortex, hippocampus, nucleus accumbens basal ganglia, caudate basal ganglia, cerebellum, frontal cortex BA9, hypothalamus, and putamen basal ganglia.

GEMMA

To account for non-independence due to family and population structure we first constructed a centered relatedness matrix based on genotypes using the Genome-wide Efficient Mixed Model Association (GEMMA) application [24]. We further utilized GEMMA to apply a univariate linear mixed model between predicted expression levels and phenotypes while controlling for family/population structure with the above relatedness matrix and including as covariates age, sex, and the first two ancestry principal components. These principal components were calculated within the EIGENSOFT package [25]. For each association we evaluated quantile-quantile plots (qq-plots) to detect evidence of bias.

Multimodel Analysis

In our initial analysis, there were 50 tissue-phenotype combinations (10 tissues × 5 phenotypes). To reduce the number of comparisons we employed Cross Phenotype Association (CPASSOC) version 1.01 to perform multimodel analysis [26]. This method delivers two statistics, S_{hom} and S_{het}, that result from the integration of association evidence from multiple GWASs. Though comparable to METAL [25], a fixed meta-analysis, S_{hom} alternatively accounts for correlated summary statistics resulting from related traits, relatedness of participants within a cohort, and overlapping participants across cohorts. It operates more efficiently when the effect sizes are homogeneous. When those effect sizes are heterogeneous, S_{het} is more powerful [26]. CPASSOC allows for relatedness or overlapping among samples within and between different studies and cohorts being analyzed. This is achieved by building a correlation matrix among all summary statistics from each GWAS being analyzed. It also removes targets with very large effect sizes as they may represent true association and inflate those correlations. Multimodel association was performed by aggregating effect sizes across all phenotypes within a given tissue and across all tissues within a given phenotype. CPASSOC excludes genes from further analysis when there are missing values for the different models (phenotype-tissue combinations). For example, if gene *X*

Table 2. CPASSOC results for the "phenotype across all tissues" analysis, annotated with false discovery rate and Bonferroni-adjusted *p* value calculated

Tissue	Phet	FDRphet	Bonfphet†	Phom	FDRphom	Bonfphom	chr	Gene name
Hypothalamus	4.80E – 06	1.07E – 01	1.07E – 01	3.75E – 04	3.27E – 01	1.00E + 00	6	*HCG27*
Hippocampus	3.60E – 05	4.02E – 01	8.04E – 01	2.87E – 05	2.23E – 01	6.41E – 01	19	*SPPL2B*
Hypothalamus	1.39E – 04	6.80E – 01	1.00E + 00	6.52E – 06	1.45E – 01	1.45E – 01	17	*PRPSAP2*
Cortex	1.64E – 03	1.00E + 00	1.00E + 00	2.99E – 05	2.23E – 01	6.68E – 01	1	*MIER1*
Cerebellar hemisphere	2.70E – 04	8.61E – 01	1.00E + 00	4.27E – 05	2.38E – 01	9.54E – 01	2	*SLC4A5*
Caudate basal ganglia	4.18E – 04	9.23E – 01	1.00E + 00	7.11E – 05	3.17E – 01	1.00E + 00	2	*COX5B*
Cerebellar hemisphere	3.78E – 04	9.23E – 01	1.00E + 00	1.07E – 04	3.27E – 01	1.00E + 00	11	*SSRP1*
Cerebellar hemisphere	1.02E – 02	1.00E + 00	1.00E + 00	1.67E – 04	3.27E – 01	1.00E + 00	11	*SIPA1*
Cortex	1.03E – 02	1.00E + 00	1.00E + 00	1.71E – 04	3.27E – 01	1.00E + 00	12	*CAMKK2*
Cortex	3.27E – 03	1.00E + 00	1.00E + 00	1.79E – 04	3.27E – 01	1.00E + 00	1	*RBBP5*
Hypothalamus	6.95E – 05	3.66E – 01	1.00E + 00	1.30E – 05	3.13E – 01	3.13E – 01	17	*PRPSAP2*
Hippocampus	1.80E – 05	2.16E – 01	4.33E – 01	5.74E – 05	4.79E – 01	1.00E + 00	19	*SPPL2B*
Cortex	8.21E – 04	7.29E – 01	1.00E + 00	5.99E – 05	4.79E – 01	1.00E + 00	1	*MIER1*
Cerebellar hemisphere	1.35E – 04	4.63E – 01	1.00E + 00	8.55E – 05	5.13E – 01	1.00E + 00	2	*SLC4A5*
Caudate basal ganglia	2.09E – 04	5.02E – 01	1.00E + 00	1.42E – 04	6.83E – 01	1.00E + 00	2	*COX5B*
Cerebellar hemisphere	1.89E – 04	5.02E – 01	1.00E + 00	2.14E – 04	7.04E – 01	1.00E + 00	11	*SSRP1*
Cerebellar hemisphere	5.08E – 03	7.29E – 01	1.00E + 00	3.34E – 04	7.04E – 01	1.00E + 00	11	*SIPA1*
Cortex	5.17E – 03	7.29E – 01	1.00E + 00	3.41E – 04	7.04E – 01	1.00E + 00	12	*CAMKK2*
Cortex	1.63E – 03	7.29E – 01	1.00E + 00	3.59E – 04	7.04E – 01	1.00E + 00	1	*RBBP5*
Cerebellar hemisphere	1.01E – 03	7.29E – 01	1.00E + 00	3.86E – 04	7.04E – 01	1.00E + 00	12	*ARHGAP9*

Top portion is based on phom and bottom is based on phet. Values shown: *Shom* homogeneous *p* value (phom), *Shet* heterogeneous *p* value (phet), FDR of phom and phet (FDRphom, FDRphet), chromosome number (chr), Bonferroni-adjusted *p* value (Bonfphet, Bonfphom), and annotated gene name (gene name). No results listed are significant by the Bonferroni threshold 1.08×10^{-6}. Table is sorted by column notated with †.

is only included in one of two models being aggregated, gene *X* will not be used in the multimodel analysis. Because there are not validated models for every gene in every tissue, many genes would be excluded from multimodel analysis. To remedy this, we replaced missing gene *p* values with values sampled from the gene-specific distribution of observed *p* values across tissue. To account for false-positives we permuted and ranked the "across all tissues" analysis 10 times.

"Gene Dropping" Single Marker Association Analysis

To demonstrate the utility of gene-based burden tests and multimodel analysis, we calculated single-marker association statistics to serve as a baseline comparison. This was accomplished by implementing the EMMAX software package through the EPACTS pipeline. EMMAX accounts for population stratification due to relatedness and ancestry through a variance component mixed model approach. Covariates included were sex, age, and the first two ancestry principal components. Utilizing a simulation scheme detailed in prior work [27], we determined empirical *p* values for single-marker association analysis. EPACTS was used to calculate association statistics for each simulated genotype, correcting for family structure [28]. To generate a single permutation, an initial allele frequency was chosen, and for each founder in the pedigree, genotypes were assigned based on this frequency. Subsequent generations were then assigned genotypes by randomly assigning one allele from each parent (gene-dropping). Once genotypes were assigned for all successive generations, the minor allele frequency was calculated for the replicate. Only simulated replicates with a similar minor allele frequency to the test marker were retained for comparison. The simulations were limited to 10^9 due to computational costs [27].

Results

Single Marker Association (GWAS)

As a baseline, we performed a GWAS analysis with EMMAX. This analysis was conducted over each phenotype, with approximately 33.7 million markers being tested in each. Taken individually, the single marker analyses identified no genome-wide significant results after Bonferroni correction (1.48×10^{-8}) for all markers. The Bonferroni threshold however is derived from the combination of all analyses, our 5 phenotypes, so our new threshold was more stringent (2.96×10^{-9}) with 168.7 million tests. Thresholding was obtained by a conservative Bonferroni approach to illustrate the scope of the problem of multiple testing. QQ-plots of each trait association test shows *p* values stratified by minor allele frequency. These figures illustrate inflation caused by extremely rare mark-

Table 3. CPASSOC results for the "tissue across all phenotypes" analysis, annotated with false discovery rate and Bonferroni-adjusted *p* value calculated

Phenotype	Phet	FDRphet	Bonfphet†	Phom	FDRphom	Bonfphom	chr	Gene name
Nic Dep	3.49E − 09	2.50E − 05	2.50E − 05	7.16E − 06	1.43E − 03	5.14E − 02	2	*COA5*
Alc Dep	1.76E − 06	6.32E − 03	1.26E − 02	4.75E − 01	9.17E − 01	1.00E + 00	6	*HLA-B*
Alc Dep	1.42E − 05	2.42E − 02	1.02E − 01	1.42E − 02	1.13E − 01	1.00E + 00	7	*GPC2*
Alc Dep	1.55E − 05	2.42E − 02	1.11E − 01	1.78E − 04	1.04E − 02	1.00E + 00	6	*HCG27*
Cannabis craving	1.69E − 05	2.42E − 02	1.21E − 01	6.55E − 09	1.98E − 05	4.70E − 05	19	*ZNF304*
Can Dep	2.43E − 05	2.80E − 02	1.75E − 01	1.10E − 08	1.98E − 05	7.93E − 05	2	*LY75*
Family alcoholism	2.73E − 05	2.80E − 02	1.96E − 01	5.51E − 09	1.98E − 05	3.95E − 05	2	*SEPTIN2*
Alc Dep	5.31E − 05	4.41E − 02	3.81E − 01	9.09E − 01	9.90E − 01	1.00E + 00	17	*TRPV3*
Can Dep	5.53E − 05	4.41E − 02	3.97E − 01	1.87E − 08	2.43E − 05	1.34E − 04	6	*MTRF1L*
Nic Dep	6.21E − 05	4.46E − 02	4.46E − 01	1.08E − 08	1.98E − 05	7.76E − 05	7	*SSPOP*
Family alcoholism	1.36E − 05	7.86E − 02	5.50E − 01	1.10E − 08	2.23E − 04	4.44E − 04	2	*SEPTIN2*
Cannabis craving	8.44E − 06	6.80E − 02	3.40E − 01	1.31E − 08	2.23E − 04	5.28E − 04	19	*ZNF304*
Nic Dep	3.11E − 05	1.37E − 01	1.00E + 00	2.16E − 08	2.23E − 04	8.72E − 04	7	*SSPOP*
Can Dep	1.22E − 05	7.86E − 02	4.90E − 01	2.21E − 08	2.23E − 04	8.90E − 04	2	*LY75*
Can Dep	2.77E − 05	1.37E − 01	1.00E + 00	3.74E − 08	2.73E − 04	1.51E − 03	6	*MTRF1L*
Cannabis craving	4.58E − 05	1.37E − 01	1.00E + 00	4.06E − 08	2.73E − 04	1.64E − 03	3	*CCDC66*
Family alcoholism	4.76E − 05	1.37E − 01	1.00E + 00	6.75E − 08	3.88E − 04	2.72E − 03	1	*PPIE*
Alc Dep	1.42E − 04	2.71E − 01	1.00E + 00	1.77E − 07	8.91E − 04	7.13E − 03	16	*CIAO3*
Can Dep	1.73E − 04	2.75E − 01	1.00E + 00	2.09E − 07	9.21E − 04	8.41E − 03	22	*RRP7A*
Nic Dep	1.85E − 04	2.75E − 01	1.00E + 00	2.33E − 07	9.21E − 04	9.38E − 03	19	*SLC7A9*
Family alcoholism	2.68E − 04	3.33E − 01	1.00E + 00	2.51E − 07	9.21E − 04	1.01E − 02	14	*RPL36AL*
Family alcoholism	2.90E − 04	3.37E − 01	1.00E + 00	4.02E − 07	1.27E − 03	1.62E − 02	2	*FOXN2*
Alc Dep	7.52E − 05	2.02E − 01	1.00E + 00	4.11E − 07	1.27E − 03	1.65E − 02	14	*RPL36AL*
Family alcoholism	2.67E − 04	3.33E − 01	1.00E + 00	5.56E − 07	1.60E − 03	2.24E − 02	2	*MTERF4*
Nic Dep	7.21E − 04	5.36E − 01	1.00E + 00	8.17E − 07	2.06E − 03	3.29E − 02	5	*RPS23*
Family alcoholism	3.53E − 04	3.56E − 01	1.00E + 00	8.18E − 07	2.06E − 03	3.30E − 02	6	*GNMT*
Family alcoholism	6.96E − 04	5.36E − 01	1.00E + 00	9.02E − 07	2.14E − 03	3.63E − 02	7	*ERV3-1*
Alc Dep	8.85E − 04	5.67E − 01	1.00E + 00	1.04E − 06	2.32E − 03	4.17E − 02	11	*APIP*

Top portion is based on phom and bottom is based on phet. Values shown: *Shom* homogeneous *p* value (phom), *Shet* heterogeneous *p* value (phet), FDR of phom and phet (FDRphom, FDRphet), chromosome number (chr), Bonferroni-adjusted *p* value (Bonfphet, Bonfphom), and annotated gene name (gene name). Fifteen genes listed are significant by the Bonferroni threshold 7.66×10^{-7}. Four genes are suggestive by the Bonferroni-adjusted *p* value. Table is sorted by column notated with †.

ers which drove our decision to conduct permutations (online suppl. Section 4; for all online suppl. material, see www.karger.com/doi/10.1159/000523748).

Gene-Phenotype Associations per Tissue (No Aggregation)

Application of PrediXcan for each of the 5 addiction phenotypes and 10 brain tissues requires substantial correction for multiple comparisons. A total of 230,690 phenotype-tissue-gene tests were conducted from the 50 phenotype-tissue combinations. Two genes reached Bonferroni-adjusted significance at the 0.05 threshold (2.17×10^{-7}). *HCG27* expression within the frontal cortex BA9 and alcohol dependence had the strongest association (3.80×10^{-9}) and *SPPL2B* within hippocampus, also with alcohol dependence, was second (3.89×10^{-8}).

Gene-Tissue Associations (Aggregated by Phenotype)

By implementing multimodel association analysis and conducting tests after aggregating across all phenotypes for 10 tissues using the meta-analytic approaches implemented in CPASSOC, we are able to reduce the tests to 46,128. The strongest associations for the aggregated phenotype analysis are listed in Table 2. No genes reached the Bonferroni threshold for significance (1.08×10^{-6}).

Gene-Phenotype Associations (Aggregated by Tissue)

Alternatively, aggregating across all tissues reduces the number of test comparisons to 65,270. The most significant results of the aggregated tissue analysis are summarized in Table 3. Fourteen associations passed the new significance threshold (7.66×10^{-7}) when filtered by the *p* value of the S_{hom} statistic. Familial alcoholism produced

the strongest overall association, *SEPTIN2*, and the most associations, 5 (*PPIE, RPL36AL, FOXN2, MTERF4, SEPTIN2*). Alcohol dependence yielded 2 significant results, *CIAO3* and *RPL36AL*. Cannabis craving yielded 2 significant results, *ZNF304* and *CCDC66* with the former being the second most significant result overall. Nicotine dependence showed a significant relation with *SSPOP, COA5*, and *SLC7A9*. Finally, Cannabis dependence identified 3 significant associations, *LY75, MTR1L,* and *RRP7A*. *RPS23, GNMT, ERV3-1, HLA-B,* and *APIP* did not reach Bonferroni significance but were suggestive based on *p* values adjusted by Bonferroni using the p.adjust function in R [28].

Discussion

The field of addiction genetics aims to identify genes involved in addiction pathogenesis. Addiction researchers typically have collected rich phenotypic data with the expectation that there are many important sub-phenotypes. In most cases, cohorts with rich phenotype data are small and lack power for conventional GWAS. Progress has been made using GWAS with large sample sizes with limited phenotype data to identify associated variants. Most variants found in these studies are noncoding and are believed to affect gene expression. More work needs to be done to establish how these variants affect genes. Because a vast majority of detected variants have small effect sizes, it is unlikely these marker associations will advance the field unless the genes they affect are proven to be key players in the pathogenesis of addiction. Currently, only a small fraction of the genetic risk of addiction has been accounted for because multiple comparison testing limits the power of GWAS. We can circumvent some of the issues of identifying genes and limited power by utilizing gene-based burden and multi-phenotype tests.

PrediXcan, is a gene-burden test approach that relies on models that weigh SNPs within 1 Mb of a translation start site to predict gene transcription levels. The models are then used to predict gene expression levels for cohorts without direct RNA analysis. The predicted expression levels can then be associated with phenotypes. The PrediXcan approach reduces the number of association tests by 3 orders of magnitude (from 10–20 million SNPs in the genome to 10–20 thousand genes). In addition to improving power, PrediXcan directly implicates genes and their expression to phenotypes which allow for a better understanding of biological mechanisms at work. Despite the potential advantages of the PrediXcan approach, it is limited by the precision of the models used to predict expression. There is not a model for every gene in each tissue which hampered part of our multimodel analysis. We were able to test 14,219 transcripts in total but only with an average of 4,613 genes analyzed per tissue. The GTEx consortium, upon which these models are derived, only used, on average, 203 samples per brain tissue. It is expected that the models will improve as more data is collected [29–31].

The multimodel analysis approach, CPASSOC, used to combine phenotypic and tissue-specific results, reduced the number of tests from 230,690 in our base univariate association to 46,128, when aggregating across phenotype, and 65,270, when aggregating across tissues. The success of the aggregated tissue-phenotype analysis is likely due to two factors: (1) the need to correct for fewer comparisons and (2) the correlation between gene expression in brain tissues allows for a better estimation of predicted gene expression. In contrast, the aggregated phenotype-tissue analysis detected fewer significant findings compared to aggregated tissue-phenotype analysis. This could be the result of low correlations between phenotypes as they are focused on varying substances. There is also the potential to lose specificity in understanding which phenotype or tissue is driving an association. Forest plots were utilized to display multimodel analysis results with the phenotype by tissue association, GWAS association and GTEx model associations, to elucidate which predictor was driving a result (Fig. 2).

Single marker association tests without cross-phenotype or cross-tissue aggregation resulted in no significant findings. This was in part due to severe corrections for multiple comparisons. Low precision of predicted transcript levels limited power and resulted in the initial PrediXcan tests, without aggregation, to result in just two significant findings. Nonetheless, this does reinforce the benefit in the burden test methodology in relation to conventional methods. Using multimodel analyses however 1 suggestive transcript was detected and 15 significant results were found when aggregating across tissues. Aggregating by tissue performed better because a better gene specific estimate was obtained likely due to the correlations in gene expression across tissues (online suppl. Section 5.1).

The mechanism of action of how the genes associated with addiction in this study exert their effect is speculative. Many of the genes detected by aggregated tissue-phenotype association analysis are implicated in cognitive and behavioral phenotypes (See Table 4). Many of the

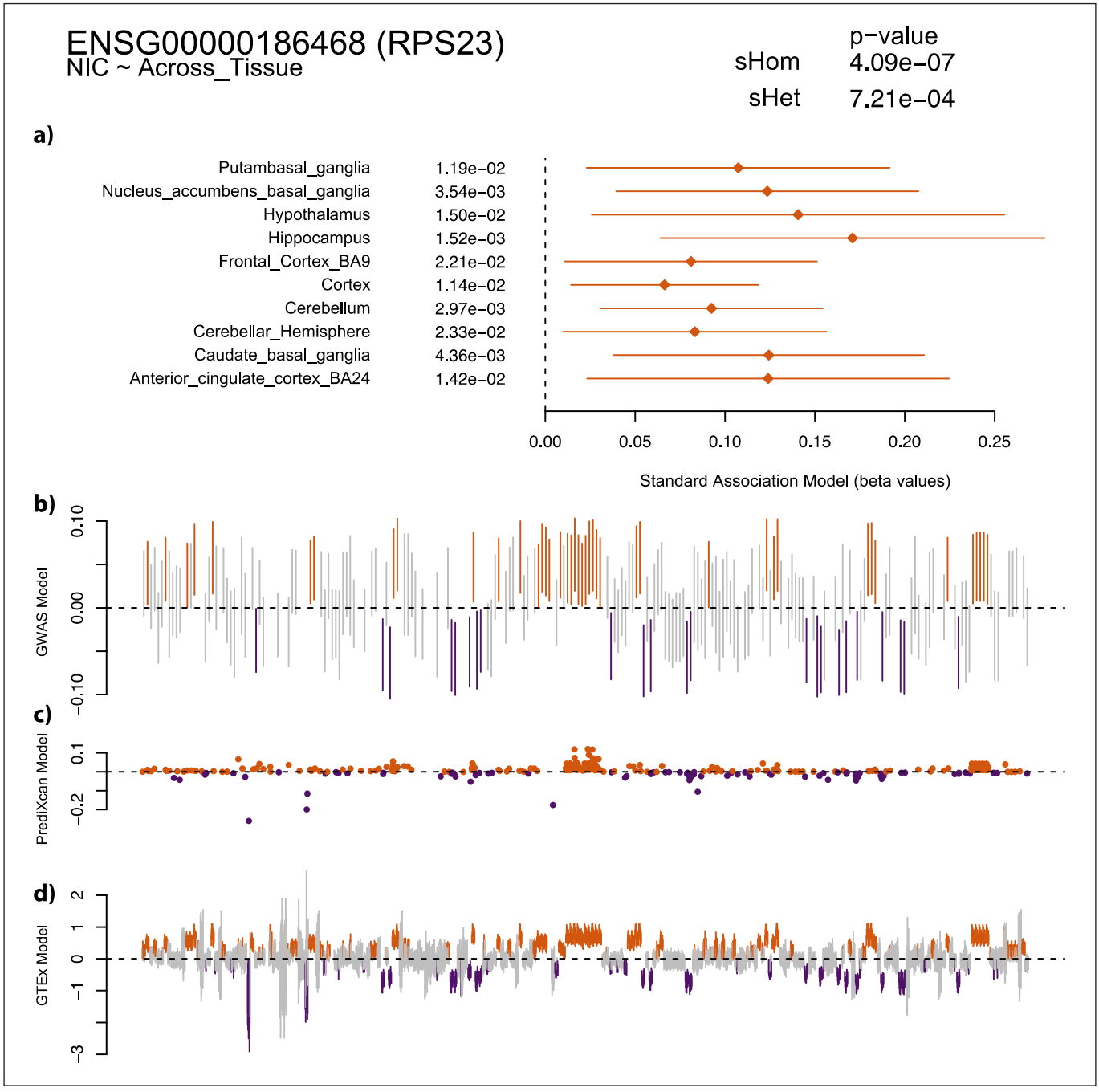


Fig. 2. Forest plot of nicotine dependence with RPS23. **a** Shows predicted expression in all tissues exhibit a positive (orange) correlation (βs) with Nicotine dependence. The hippocampus has the strongest association. The trend for all tissue-RPS23 correlations is in the same direction. **b** Shows the GWAS nicotine association (βs) for SNPs that appear in any of the PrediXcan models. The SNPs are all within 1 MB of RPS23 but are not spaced by their physical separations. **c** PrediXcan model displays the predicted weights for all SNPs in every tissue; for every SNP there is a value for each of the 10 tissues. GTEx (**d**) models displays (βs) of the same set of SNPs in (**b**, **c**) among the tissues. All SNPs in every model are ordered the same as so all SNPs are aligned and allow for comparisons of the correlation pattern across the 3 model sets (**b**–**d**).

Table 4. Prior evidence for genes implicated by aggregated tissue-phenotype association analysis

Locus	Product	Addiction phenotype	Prior associations
SEPTIN2	Septin 2	Family alcoholism	Neurofibrillary tangle properties, CSF IL-8 level, diffusion barrier for cilium (1–3)
ZNF304	Zinc finger protein 304	Cannabis craving	Transcription factor(4)
SSPOP	SCO-spondin pseudogene	Nicotine dependence	Pseudogene, transcript affects neurite outgrowth (5, 6)
LY75	CD205	Cannabis dependence	Unipolar depression, mathematical ability, lymphocyte antigen processing (7–9)
MTRF1L	Mitochondrial translational release factor 1-like	Cannabis dependence	Mitochondrial electron transport and amygdala volume (10, 11)
CCDC66	Coiled-coil domain containing 66	Cannabis craving	Interacts with microtubules, cognitive function measurement, mathematical ability (9, 12)
PPIE	Peptidylprolyl isomerase E	Family alcoholism	Peptidyl-prolyl cis-trans isomerase which affects protein folding (13)
CIAO3	Cytosolic iron-sulfur assembly component 3	Alcohol dependence	Iron binding (14)
RRP7A	Ribosomal RNA processing 7 homolog A	Cannabis dependence	Microcephaly and neurogenesis, rRNA processing (15)
SLC7A9	Solute carrier family 7 member 9	Nicotine dependence	Cystinuria, amino acid transporter (16, 17)
RPL36AL	Ribosomal protein L36a Like	Family alcoholism and alcohol dependence	Behavioral disinhibition measurement (negative urgency) (18, 19)
FOXN2	Forkhead box N2	Family alcoholism	Autism spectrum disorder, ADHD, bipolar disorder, major depressive disorder, and schizophrenia (combined), transcription factor, Alzheimer's age of onset, math ability, handedness (9, 20–23)
MTERF4	Mitochondrial transcription termination factor 4	Family alcoholism	Mitochondrial ribosomal biogenesis (24)
RPS23	Ribosomal protein S23	Nicotine dependence	Ribosomal protein, brachycephaly, trichomegaly, and developmental delay (25)
GMNT	Glycine methyltransferase	Family alcoholism	Enzyme to regulate the ratio of S-adenosylmethionine to S-adenosylhomocysteine, detoxification pathway in liver cells (26)
ERV3-1	Endogenous retrovirus group 3 member 1, envelope	Family alcoholism	Down regulated in choriocarcinoma (27)
APIP	APAF1 interacting protein	Alcohol dependence	Suppresses mitochondrial-mediated apoptosis (28)
COA5	Cytochrome C oxidase assembly factor 5	Nicotine dependence	Mitochondrial complex IV assembly factor (29)
HLA-B	Major histocompatibility complex, class I, B	Alcohol dependence	Immune response, depression, schizophrenia, smoking behavior, autism spectrum, anxiety (30–34)

(For Footnote see next page.)

Table 4 (Footnote)

1. Hu Q, Milenkovic L, Jin H, Scott MP, Nachury M V., Spiliotis ET, et al. A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. Science [Internet]. 2010 Jul 23 [cited 2021 May 23];329(5990):436–9. Available from: https://pubmed.ncbi.nlm.nih.gov/20558667/. 2. Zhang R, Song J, Isgren A, Jakobsson J, Blennow K, Sellgren CM, et al. Genome-wide study of immune biomarkers in cerebrospinal fluid and serum from patients with bipolar disorder and controls. Transl Psychiatry [Internet]. 2020 Dec 1 [cited 2021 May 23];10(1). Available from: https://pubmed.ncbi.nlm.nih.gov/32066700/. 3. Wang H, Yang J, Schneider JA, De Jager PL, Bennett DA, Zhang HY. Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. Neurobiol Aging [Internet]. 2020 Sep 1 [cited 2021 May 28];93:61–8. Available from: https://pubmed.ncbi.nlm.nih.gov/32450446/. 4. Sabater L, Ashhab Y, Caro P, Kolkowski EC, Pujol-Borrell R, Domínguez O. Identification of a KRAB-containing zinc finger protein, ZNF304, by AU-motif-directed display method and initial characterization in lymphocyte activation. Biochem Biophys Res Commun [Internet]. 2002 [cited 2021 May 23];293(3):1066–72. Available from: https://pubmed.ncbi.nlm.nih.gov/12051768/. 5. Gobron S, Creveaux I, Meiniel R, Didier R, Herbet A, Bamdad M, et al. Subcommissural organ/reissner's fiber complex: characterization of SCO spondin, a glycoprotein with potent activity on neurite outgrowth. Glia [Internet]. 2000 Nov [cited 2021 May 28];32(2):177-91. Available from: https://onlinelibrary.wiley.com/doi/10.1002/1098–1136(200011)32:2%3C177. 6. Han X, Ong JS, An J, Hewitt AW, Gharahkhani P, MacGregor S. Using Mendelian randomization to evaluate the causal relationship between serum C-reactive protein levels and age-related macular degeneration. Eur J Epidemiol [Internet]. 2020 Feb 1 [cited 2021 May 23];35(2):139–46. Available from: http://europepmc.org/article/MED/31900758. 7. Jiang W, Swiggard WJ, Heufler C, Peng M, Mirza A, Steinman RM, et al. The receptor DEC-205 expressed by dendritic cells and thymic epithelial cells is involved in antigen processing [Internet]. Nature. 1995 [cited 2021 May 23];375:151–5. Available from: https://pubmed.ncbi.nlm.nih.gov/7753172/. 8. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet [Internet]. 2016 Sep 1 [cited 2021 May 23];48(9):1031–6. Available from: https://pubmed.ncbi.nlm.nih.gov/27479909/. 9. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet [Internet]. 2018 Aug 1 [cited 2021 May 23];50(8):1112–21. Available from: https://pubmed.ncbi.nlm.nih.gov/30038396/. 10. Soleimanpour-Lichaei HR, Kühl I, Gaisne M, Passos JF, Wydro M, Rorbach J, et al. mtRF1a Is a human mitochondrial translation release factor decoding the major termination codons UAA and UAG. Mol Cell [Internet]. 2007 Sep 7 [cited 2021 May 23];27(5):745–57. Available from: https://pubmed.ncbi.nlm.nih.gov/17803939/. 11. Alliey-Rodriguez N, Grey TA, Shafee R, Asif H, Lutz O, Bolo NR, et al. NRXN1 is associated with enlargement of the temporal horns of the lateral ventricles in psychosis. Transl Psychiatry [Internet]. 2019 Dec 1 [cited 2021 May 28];9(1):230. Available from: https://pubmed.ncbi.nlm.nih.gov/31530798/. 12. Conkar D, Culfa E, Odabasi E, Rauniyar N, Yates JR, Firat-Karalar EN. The centriolar satellite protein CCDC66 interacts with CEP290 and functions in cilium formation and trafficking. J Cell Sci [Internet]. 2017 Apr 15 [cited 2021 May 23];130(8):1450–62. Available from: https://pubmed.ncbi.nlm.nih.gov/28235840/. 13. Colgan J, Asmal M, Yu B, Luban J. Cyclophilin A-deficient mice are resistant to immunosuppression by cyclosporine. J Immunol [Internet]. 2005 May 15 [cited 2021 May 23];174(10):6030–8. Available from: https://pubmed.ncbi.nlm.nih.gov/15879096/. 14. Huang J, Song D, Flores A, Zhao Q, Mooney SM, Shaw LM, et al. IOP1, a novel hydrogenase-like protein that modulates hypoxia-inducible factor-1α activity. Biochem J [Internet]. 2007 Jan 1 [cited 2021 May 23];401(1):341–52. Available from: https://pubmed.ncbi.nlm.nih.gov/16956324/. 15. Farooq M, Lindbæk L, Krogh N, Doganli C, Keller C, Mönnich M, et al. RRP7A links primary microcephaly to dysfunction of ribosome biogenesis, resorption of primary cilia, and neurogenesis. Nat Commun [Internet]. 2020 Dec 1 [cited 2021 May 23];11(1). Available from: https://pubmed.ncbi.nlm.nih.gov/33199730/. 16. Strologo L Dello, Pras E, Pontesilli C, Beccia E, Ricci-Barbini V, De Sanctis L, et al. Comparison between SLC3A1 and SLC7A9 cystinuria patients and carriers: a need for a new classification. J Am Soc Nephrol [Internet]. 2002 Oct 1 [cited 2021 May 23];13(10):2547–53. Available from: https://pubmed.ncbi.nlm.nih.gov/12239244/. 17. Colombo R. Dating the origin of the V170M mutation causing non-type I cystinuria in Libyan Jews by linkage disequilibrium and physical mapping of the SLC7A9 gene. Genomics [Internet]. 2000 Oct 1 [cited 2021 May 23];69(1):131–4. Available from: https://pubmed.ncbi.nlm.nih.gov/11013083/. 18. Sanchez-Roige S, Fontanillas P, Elson SL, Gray JC, De Wit H, MacKillop J, et al. Genome-wide association studies of impulsive personality traits (BIS-11 and UPPS-P) and drug experimentation in up to 22,861 adult research participants identify loci in the CACNA1I and CADM2 genes. J Neurosci [Internet]. 2019 Mar 27 [cited 2021 May 23];39(13):2562–72. Available from: https://pubmed.ncbi.nlm.nih.gov/30718321/. 19. Feo S, Davies B, Fried M. The mapping of seven intron-containing ribosomal protein genes shows they are unlinked in the human genome. Genomics [Internet]. 1992 [cited 2021 May 23];13(1):201–7. Available from: https://pubmed.ncbi.nlm.nih.gov/1577483/. 20. Li C, Lusis AJ, Sparkes R, Tran SM, Gaynor R. Characterization and chromosomal mapping of the gene encoding the cellular DNA binding protein HTLF. Genomics [Internet]. 1992 [cited 2021 May 23];13(3):658–64. Available from: https://pubmed.ncbi.nlm.nih.gov/1639393/. 21. Herold C, Hooli BV., Mullin K, Liu T, Roehr JT, Mattheisen M, et al. Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with OSBPL6, PTPRG, and PDCL3. Mol Psychiatry [Internet]. 2016 Nov 1 [cited 2021 May 23];21(11):1608–12. Available from: https://pubmed.ncbi.nlm.nih.gov/26830138/. 22. Smoller JW, Kendler K, Craddock N, Lee PH, Neale BM, Nurnberger JN, et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genom-wide analysis. Lancet [Internet]. 2013 Apr 1 [cited 2021 May 23];381(9875):1371–9. Available from:/pmc/articles/PMC3714010/. 23. Cuellar-Partida G, Tung JY, Eriksson N, Albrecht E, Aliev F, Andreassen OA, et al. Genome-wide association study identifies 48 common genetic variants associated with handedness. Nat Hum Behav [Internet]. 2021 Jan 1 [cited 2021 May 28];5(1):59–70. Available from: https://pubmed.ncbi.nlm.nih.gov/32989287/. 24. Cámara Y, Asin-Cayuela J, Park CB, Metodiev MD, Shi Y, Ruzzenente B, et al. MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. Cell Metab [Internet]. 2011 May 4 [cited 2021 May 23];13(5):527–39. Available from: https://pubmed.ncbi.nlm.nih.gov/21531335/. 25. Paolini NA, Attwood M, Sondalle SB, Vieira CM dos S, van Adrichem AM, di Summa FM, et al. A ribosomopathy reveals decoding defective ribosomes driving human dysmorphism. Am J Hum Genet [Internet]. 2017 Mar 2 [cited 2021 May 23];100(3):506–22. Available from: https://pubmed.ncbi.nlm.nih.gov/28257692/. 26. Chen YMA, Chen LY, Wong FH, Lee CM, Chang TJ, Yang-Feng TL. Genomic structure, expression, and chromosomal localization of the human glycine N-methyltransferase gene. Genomics [Internet]. 2000 May 15 [cited 2021 May 23];66(1):43–7. Available from: https://pubmed.ncbi.nlm.nih.gov/10843803/. 27. Kato N, Shimotohno K, VanLeeuwen D, Cohen M. Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger protein related to Krüppel. Mol Cell Biol [Internet]. 1990 Aug [cited 2021 May 23];10(8):4401–5. Available from: https://pubmed.ncbi.nlm.nih.gov/2115127/. 28. Cho DH, Hong YM, Lee HJ, Woo HN, Pyo JO, Mak TW, et al. Induced inhibition of ischemic/hypoxic injury by APIP, a novel Apaf-1-interacting protein. J Biol Chem [Internet]. 2004 Sep 17 [cited 2021 May 23];279(38):39942–50. Available from: https://pubmed.ncbi.nlm.nih.gov/15262985/. 29. Huigsloot M, Nijtmans LG, Szklarczyk R, Baars MJH, Van Den Brand MAM, Hendriksfranssen MGM, et al. A mutation in C2orf64 causes impaired cytochrome c oxidase assembly and mitochondrial cardiomyopathy. Am J Hum Genet [Internet]. 2011 Apr 8 [cited 2021 May 23];88(4):488–93. Available from: https://pubmed.ncbi.nlm.nih.gov/21457908/. 30. Howard DM, Adams MJ, Shirali M, Clarke TK, Marioni RE, Davies G, et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nat Commun [Internet]. 2018 Dec 1 [cited 2021 May 23];9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29662059/. 31. Ikeda M, Takahashi A, Kamatani Y, Okahisa Y, Kunugi H, Mori N, et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. Mol Psychiatry. 2018 Mar 1;23(3):639–47. 32. Hällfors J, Palviainen T, Surakka I, Gupta R, Buchwald J, Raevuori A, et al. Genome-Wide Association Study in Finnish twins highlights the connection between nicotine addiction and neurotrophin signaling pathway. Addict Biol [Internet]. 2019 May 1 [cited 2021 May 23];24(3):549–61. Available from: https://pubmed.ncbi.nlm.nih.gov/29532581/. 33. Nagel M, Watanabe K, Stringer S, Posthuma D, Van Der Sluis S. Item-level analyses reveal genetic heterogeneity in neuroticism. Nat Commun [Internet]. 2018 Dec 1 [cited 2021 May 23];9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29500382/. 34. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism [Internet]. 2017 [cited 2021 May 23];8:21. Available from: https://pubmed.ncbi.nlm.nih.gov/28540026/.

genes implicated by aggregated tissue-phenotype association analysis are known to have biosynthetic, regulatory, and signal transductions roles (*SEPTIN2* – organelle biogenesis; *ZNF304* – functions as a transcriptional repressor; *MTRF1L* – plays a role in mitochondrial translation termination; *PPIE* – peptidyl-prolyl cis-trans isomerase; *RRP7A* – ribosomal RNA processing; *SLC7A9* – involved in the transport of many compounds; *RPL36AL* – rRNA processing in the nucleus and cytosol; *FOXN2* – forkhead domain-binding protein; *MTERF4* – mitochondrial transcription termination factor; and *GNMT* – an enzyme that catalyzes the conversion of S-adenosyl-L-methionine to S-adenosyl-L-homocysteine and sarcosine). Changes in expression of *SSPOP*, a Spondin pseudogene, may not produce a functional translation product. The *SSPOP* transcript may act as a regulatory noncoding RNA. *RPS23* – component of the large 40S ribonucleoprotein complex, associated with nicotine dependence in our study has pri-

or association with smokeless tobacco [32]. Of all significant genes mentioned, 7 had prior associations with relatively similar traits obtained from interrogation of GWAS ATLAS [33]. *PPIE* – alcohol intake frequency (6.64×10^{-3}); *SLC7A9* – cigarettes per day (8.55×10^{-3}); *RPS23* – maternal smoking around birth (4.67×10^{-3}); *GMNT* – drinks per day (1.25×10^{-2}); *ERV3-1* – frequency of needing morning drink of alcohol after heavy drinking session in last year (7.68×10^{-3}); *APIP* – alcohol consumption (dichotomous, female) 5 (8.02×10^{-3}); drinks per day (8.60×10^{-4}) [33–38]. *COA5* – with cigarettes per day (7.67×10^{-4}), smoking status (1.07×10^{-3}), current tobacco smoking (5.71×10^{-3}). The one suggestive gene *HLA-B* also had prior associations with average weekly beer (2.92×10^{-4}), alcohol intake frequency (2.94×10^{-4}) [33, 34]. All other implicated genes are novel for addiction/substance use phenotypes.

In summary, this study detected associations between predicted gene expression and addiction traits, in a relatively small, well-characterized cohort where the more conventional analysis was unsuccessful. Multi-trait and multi-tissue analysis is a promising approach to identify risk genes because it reduces the need to correct for multiple comparisons. The approach should benefit from improved models that will be made possible when additional reference transcriptional data becomes available. Because of the extremely polygenic nature of most traits with complex modes of inheritance, we expect that this study detected only a small fraction of the genes where the level of transcription is associated with addiction.

Acknowledgments

We thank Cindy Ehlers, Qian Peng, and Clark Jefferies for participation in writing and/or technical editing of manuscript.

Statement of Ethics

As all of the data used in this manuscript are based on previously published literature, a statement of ethics is not applicable.

Conflict of Interest Statement

The authors report no conflicts of interest.

Funding Sources

This research was supported by the National Institute on Drug Abuse Grant R01 DA030976. Additional funding was provided by the State of California and the Ernest Gallo Clinic and the Research Center for Medical Research on Alcohol and Substance Abuse through the University of California at San Francisco.

Author Contributions

D.M.B. prepared the text and figures, produced R code, and performed analysis. D.M.B. and K.C.W. conceived and designed the experiment. C.B. helped write and test software in R and Python to perform analysis. D.L.F. produced Figure 1 as well as code for producing forest plots (Fig. 2 and online supplemental). J.L.T. assisted in data management and procedural testing. K.C.W. and I.R.G. are responsible for data usage. All authors participated in reviews, revisions, and approval of manuscript.

Data Availability Statement

Nicotine and cannabis UCSF data used in this study are available through dbGaP accession phs001458.v1.p1. Alcohol phenotype data were not included in that dbGaP data submission but is available upon contacting corresponding author.

References

1 Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of Genome-Wide Association Studies. Nat Rev Genet. 2019;20(8):467–84.

2 Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101:5–22.

3 Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, et al. Genome-Wide Association Study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. Mol Psychiatry. 2014;19(1):41–9.

4 Luo X, Guo X, Luo X, Tan Y, Zhang P, Yang K, et al. Significant, replicable, and functional associations between KTN1 variants and alcohol and drug codependence. Addict Biol. 2021 Mar 1 [cited 2021 May 12];26(2):e12888. http: //dx.doi.org/10.1111/adb.12888.

5 Kranzler HR, Zhou H, Kember RL, Vickers Smith R, Justice AC, Damrauer S, et al. Genome-Wide Association Study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun [Internet]. 2019 Dec 1 [cited 2021 May 12];10(1):1499. Available from:

6 Sun Y, Chang S, Wang F, Sun H, Ni Z, Yue W, et al. Genome-Wide Association Study of alcohol dependence in male Han Chinese and cross-ethnic polygenic risk score comparison. Transl Psychiatry [Internet]. 2019 Dec 1 [cited 2021 May28];9(1):249. Available from:

7 Hancock DB, Markunas CA, Bierut LJ, Johnson EO. Human genetics of addiction: new insights and future directions. Curr Psychiatry Rep. 2018 Feb 1 [cited 2021 May 12];20(2):8. [

8 Zuo L, Garcia-Milian R, Guo X, Zhong C, Tan Y, Wang Z, et al. Replicated risk nicotinic cholinergic receptor genes for nicotine dependence. Genes [Internet]. 2016;7:95. Available from: www.mdpi.com/journal/genes-Genes2016.
9 Yin X, Bizon C, Tilson J, Lin Y, Gizer IR, Ehlers CL, et al. Genome- wide meta-analysis identifies a novel susceptibility signal at CACNA2D3 for nicotine dependence. Am J Med Genet Part B Neuropsychiatr Genet. 2017;174(5):557–67.
10 Johnson EC, Demontis D, Thorgeirsson TE, Walters RK, Polimanti R, Hatoum AS, et al. A Large-Scale Genome-Wide Association Study meta-analysis of cannabis use disorder. Lancet Psychiatry [Internet]. 2020 Dec 1 [cited 2021 May 13];7(12):1032–45. Available from:
11 Gizer IR, Bizon C, Gilder DA, Ehlers CL, Wilhelmsen KC. Whole Genome Sequence Study of cannabis dependence in two independent cohorts. Addict Biol. 2018;23(1):461–73.
12 Agrawal A, Lynskey MT, Hinrichs A, Grucza R, Saccone SF, Krueger R, et al. A Genome-Wide Association Study of DSM-IV: cannabis dependence. Addict Biol [Internet]. 2011 Jul [cited 2021 May 28];16(3):514–8. Available from:
13 Moutsianas L, Agarwala V, Fuchsberger C, Flannick J, Rivas MA, Gaulton KJ, et al. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. PLoS Genet. 2015;11(4.
14 Gamazon ER, Wheeler HE, Shah K, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. PrediXcan: trait mapping using human transcriptome regulation. bioRxiv [Internet]. 2015. Available from: http://biorxiv.org/lookup/doi/10.1101/020164.
15 Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative Approaches for Large-Scale Ttranscriptome-Wide Association Studies. Nat Genet. 2016;48(3):245–52.
16 Vieten C, Seaton KL, Feiler HS, Wilhelmsen KC. The University of California, San Francisco Family Alcoholism Study. I. Design, methods, and demographics. Alcohol Clin Exp Res. 2004;28(10):1509–16.
17 Gizer IR, Ehlers CL, Vieten C, Seaton-Smith KL, Feiler HS, Lee JV, et al. Linkage scan of alcohol dependence in the UCSF Family Alcoholism Study. Drug Alcohol Depend. 2011;113(2–3):125–32.
18 Wilhelmsen KC, Swan GE, Cheng LSC, Lessov-Schlaggar CN, Amos CI, Feiler HS, et al. Support for previously identified alcoholism susceptibility loci in a cohort selected for smoking behavior. Alcohol Clin Exp Res. 2005;29(12):2108–15.
19 Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet. 2018 Feb 1 [cited 2021 May 13];50(2):229–37.
20 O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin MR, et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS One [Internet]. 2012 May 2 [cited 2021 May 13];7(5):34861. Available from: www.plosone.org.
21 Seaton KL, Cornell JL, Wilhelmsen KC, Vieten C. Effective strategies for recruiting families ascertained through alcoholic probands. Alcohol Clin Exp Res. 2004;28(1):78–84.
22 Gizer IR, Seaton-Smith KL, Ehlers CL, Vieten C, Wilhelmsen KC. Heritability of MMPI-2 scales in the UCSF Family Alcoholism Study. J Addict Dis. 2010;29(1):84–97.
23 Otto JM, Gizer IR, Ellingson JM, Wilhelmsen KC. Genetic variation in the exome: associations with alcohol and tobacco co-use. Psychol Addict Behav. 2017 May 1 [cited 2021 Sep 6];31(3):354–66.
24 Bizon C, Spiegel M, Chasse SA, Gizer IR, Li Y, Malc EP, et al. Variant calling in low-coverage whole genome sequencing of a native American population sample. BMC Genomics. 2014 Jan 30 [cited 2021 Dec 22];15(1):85.
25 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in Genome-Wide Association Studies. Nat Genet. 2006 [cited 2022 Feb 6];38(8):904–9. Available from: http://www.nature.com/naturegenetics.
26 Zhou X, Stephens M. Genome-Wide Efficient Mixed-Model Analysis for Association Studies. Nat Genet. 2012;44(7):821–4.
27 Li X, Zhu X. Cross-phenotype association analysis using summary statistics from GWAS. Methods Mol Biol. 2017;1666:455–67.
28 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190–1.
29 Ehlers CL, Gizer IR, Bizon C, Slutske W, Peng Q, Schork NJ, et al. Single nucleotide polymorphisms in the REG-CTNNA2 region of chromosome 2 and NEIL3 associated with impulsivity in a native American sample. Genes, Brain Behav. 2016;15(6):568–77.
30 Kang HM. Efficient and Parallelizable Association Container Toolbox. (EPACT) University of Michigan Center for Statistical Genetics; 2014.
31 Jafari M, Ansari-Pour N. Why, when and how to adjust your p values? Cell J. 2019 [cited 2021 Dec 29];20(4):604.
32 Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, et al. Genetic effects on gene expression across human tissues. Nature [Internet]. 2017;550(7675):204–13. Available from: http: //www.nature.com/doifinder/10.1038/nature24277.
33 Rohatgi N, Matta A, Kaur J, Srivastava A, Ralhan R. Novel molecular targets of smokeless tobacco (khaini) in cell culture from oral hyperplasia. Toxicology. 2006;224(1–2):1–13.
34 Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 2018;9(1):1825.
35 Watanabe K, Stringer S, Frei O, Mirkov MU, de Leeuw C, Polderman TJC, et al. Author correction: a global overview of pleiotropy and genetic architecture in complex traits. Nat Genet. 2020 Mar 1;52(3):353.
36 Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019 Feb 1 [cited 2021 Dec 17];51(2):237.
37 Schumann G, Liu C, O'Reilly P, Gao H, Song P, Xu B, et al. KLB is associated with alcohol drinking, and its gene product β-Klotho is necessary for FGF21 regulation of alcohol preference. Proc Natl Acad Sci U S A. 2016 Dec 13 [cited 2021 Dec 17];113(50):14372–7.
38 Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019 Jun 27 [cited 2021 Dec 17];570(7762):514.