

UC San Diego

UC San Diego Previously Published Works

Title

Post-GWAS analysis of six substance use traits improves the identification and functional interpretation of genetic risk loci

Permalink

<https://escholarship.org/uc/item/5zj1229p>

Authors

Marees, Andries T
Gamazon, Eric R
Gerring, Zachary
[et al.](#)

Publication Date

2020

DOI

10.1016/j.drugalcdep.2019.107703

Peer reviewed



HHS Public Access

Author manuscript

Drug Alcohol Depend. Author manuscript; available in PMC 2022 June 02.

Published in final edited form as:

Drug Alcohol Depend. 2020 January 01; 206: 107703. doi:10.1016/j.drugalcdep.2019.107703.

Post-GWAS analysis of six substance use traits improves the identification and functional interpretation of genetic risk loci

Andries T. Mares^{1,2,*}, Eric R. Gamazon^{3,4}, Zachary Gerring², Florence Vorspan^{5,6}, Josh Fingal¹, Wim van den Brink¹, Dirk J.A. Smit¹, Karin J.H. Verweij^{1,7}, Henry R. Kranzler⁸, Richard Sherva⁹, Lindsay Farrer⁹, International Cannabis Consortium, Joel Gelernter¹⁰, Eske M. Derks^{1,2}

¹Department of Psychiatry, Amsterdam UMC, Amsterdam Neuroscience, University of Amsterdam, Meibergdreef 9, Amsterdam, the Netherlands

²QIMR Berghofer, Translational Neurogenomics group, Brisbane, Australia

³Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN

⁴Clare Hall, University of Cambridge, Cambridge, CB3 9AL, United Kingdom

⁵Assistance Publique – Hôpitaux de Paris, Hôpital Fernand Widal, Département de Psychiatrie et de Médecine Addictologique, 200 rue du Faubourg Saint Denis, 75010 Paris, France

⁶Inserm umr-s 1144, Université Paris Descartes, Université Paris Diderot, 4 avenue de l'Observatoire, 75006 Paris, France

⁷Behavioural Science Institute, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, the Netherlands

⁸Center for Studies of Addiction, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine and Crescenz VAMC, Philadelphia, PA 19104, USA

⁹Section of Biomedical Genetics, Department of Medicine, Boston University School of Medicine, Boston, MA, United States

¹⁰Department of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine, New Haven, Connecticut

Abstract

Background: Little is known about the functional mechanisms through which genetic loci associated with substance use traits ascertain their effect. This study aims to identify and functionally annotate loci associated with substance use traits based on their role in genetic regulation of gene expression.

Methods: We evaluated expression Quantitative Trait Loci (eQTLs) from 13 brain regions and whole blood of the Genotype-Tissue Expression (*GTEX*) database, and from whole blood of the Depression Genes and Networks (DGN) database. The role of single eQTLs was examined for six substance use traits: alcohol consumption (N=537,349), cigarettes per day (CPD; N=263,954),

*Correspondence: andriestm@hotmail.com; Tel.: +31 6 21626999.

former vs. current smoker (N=312,821), age of smoking initiation (N=262,990), ever smoker (N=632,802), and cocaine dependence (N=4,769). Subsequently we conducted a gene level analysis of gene expression on these substance use traits using S-PrediXcan.

Results: Using an FDR threshold of $q=0.05$ we found 2,976 novel candidate genetic loci for substance use traits, and identified genes and tissues through which these loci potentially exert their effects. Using S-PrediXcan we identified significantly associated genes for all substance traits.

Discussion: Annotating genes based on transcriptomic regulation improves the identification and functional characterization of candidate loci and genes for substance use traits.

Keywords

addiction; eQTLs; functional annotation; GTEx; substance use; S-PrediXcan

1. Introduction

In recent years, large-scale genome-wide association studies (GWAS) of substance use traits (i.e., substance use disorders and quantitative measures of substance use) have been conducted (Gelernter et al., 2014; Liu et al., 2019). These GWAS revealed multiple genome-wide significant loci ($p < 5.0 \cdot 10^{-8}$). However, the GWAS approach faces two major challenges. First, the effects of individual single nucleotide polymorphisms (SNPs) are generally small, and a typical GWAS in which single-variant tests of association are performed is generally underpowered to detect trait-associated SNPs with small effect sizes. One possible solution to this problem is to increase sample sizes by combining samples in GWAS meta-analyses (Evangelou and Ioannidis, 2013), but even in these studies, the identified loci explain a modest proportion of the trait variance in substance use traits (Liu et al., 2019; Pasman et al., 2018). A second limitation of the GWAS approach is that the functional relevance of identified SNPs remains unclear. The majority (~93%) of trait-associated SNPs are located in non-coding regions of the genome (Maurano et al., 2012) suggesting that these SNPs act through the regulation of gene expression rather than by altering the protein product. Furthermore, due to extensive linkage disequilibrium (LD) in the genome, GWAS alone are unable to distinguish causal variants from correlated non-functional variants within an LD block. Constrained by these challenges, GWAS have been largely unsuccessful in elucidating the biological mechanisms involved in most substance use traits.

To facilitate the identification of biological mechanisms underlying substance use traits, it is essential to study the genetic regulation of gene expression in relevant tissues, i.e., human brain tissue from specific brain regions. The relevance of the brain for understanding the etiology of substance use (and disorders) is supported by numerous gene expression studies, which aimed to explore the role of epigenetic regulation and the transcriptional machinery in the addicted brain. In their review of these studies, Zhou et al. (Zhou et al., 2014) summarized the findings of post-mortem human studies, including those that report differences in gene expression in various brain regions between cases with substance use disorder vs. non-addicted controls. Overall, the reviewed studies suggest a prominent role

for mechanisms involved in transcriptomic regulation (Albertson et al., 2004; Albertson et al., 2006; Celentano et al., 2009; Zhou et al., 2011).

However, in general differential expression analysis between cases and controls does not reveal the direction of causality, i.e., it does not answer the question whether altered gene expression in a brain region represents susceptibility to substance dependence or whether excessive substance use is responsible for the altered gene expression. Because a large proportion of disease-associated variants exert their effects by regulating gene expression (Maurano et al., 2012), we applied an integrative approach that combines transcriptome data with summary statistics from recent GWAS to explore the role of regulatory genetic variants in substance use traits. Genetic variants that are associated with messenger RNA (mRNA) expression levels of one or more genes, in one or more tissues, are known as “expression quantitative trait loci” (eQTLs). The genes that are under the influence of at least one eQTL, in one or more tissues are called eGenes. Two types of eQTLs have been identified: cis-eQTLs, influencing expression levels of genes on the same locus (located at ± 1 MB from the gene), and trans-eQTLs that have their effect on genes at a different locus (e.g., those on a different chromosome). Because the statistical power to detect trans-eQTLs is low at current sample sizes, we focused only on cis-eQTLs. We determined the eQTL status of 13 human brain tissues and whole blood using the Genotype-Tissue Expression (GTEx) project, the most comprehensive eQTL database available to date, in terms of the diversity of tissues included (Consortium, 2018). GTEx (V7) provides samples from 53 different tissues, including 13 from the brain, obtained from post-mortem adult subjects. In addition, we performed the same procedure for whole blood eQTLs from the Depression Genes and Networks (DGN) database (Battle et al., 2014).

In the current study, we investigated associations between cis-eQTLs and substance use traits, by filtering GWAS variants on eQTL status for specific brain regions and whole blood. This approach has several benefits. First, it reduces the multiple-testing burden by focusing specifically on variants that are involved in genetic regulation, thereby facilitating the identification of novel candidate variants which might be especially beneficial for relatively small GWAS studies. Second, functional interpretation of eQTLs is relatively straightforward because an eQTL variant tags a causal regulatory variant of an eGene. Therefore, the eGene targeted by the eQTL is a sensible target for future follow-up studies. Third, by investigating whether eQTLs in particular tissues (e.g., the brain or specific regions within the brain) are significant, we obtain information on which (brain) tissues are involved in substance use traits at a genetic level. Fourth, this method provides causal associations between gene-expression and substance use, because eQTL- gene expression associations have been evaluated in healthy individuals, and are thus not influenced by substance use. In addition to this single variant approach, we conducted an integrative analysis to investigate expression on a gene level using S-Predixcan (Barbeira et al., 2018), which can provide further mechanistic insights into the substance use traits under investigation.

We analyzed eQTLs in 13 different brain regions (i.e., amygdala, anterior cingulate cortex, caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, putamen, nucleus accumbens, substantia nigra, and cervical spine) and in

whole blood as a comparison tissue. The role of eQTLs was examined for six substance use traits. We used summary statistics of the discovery samples obtained from major GWAS for these traits (Gelernter et al., 2014; Liu et al., 2019). The substance use traits examined in the current study were included if sufficiently large GWAS (i.e., > 4000 subjects) are available. The aims of this study were: i) to identify novel genetic loci associated with these substance use traits; ii) to improve the functional characterization of novel and known genetic loci; and iii) to obtain information about the mediating effects of gene expression levels on substance use traits by using a gene-level association approach.

2. Materials and Methods

2.1 GWAS summary statistics

We obtained summary statistics from previously conducted GWAS meta-analyses, all of which were based on large samples; alcohol consumption (N=537,349), CPD (N=263,954), former vs. current smoker (N=312,821), age of smoking initiation (N=262,990), ever smoker (N=632,802), and cocaine dependence (N=4,769). Information on subjects, sample preparation, and analytic methods can be found in the original articles of the corresponding GWAS (Gelernter et al., 2014; Liu et al., 2019). For some samples, genome-wide data was only provided for the discovery samples, explaining the difference in the number of subjects between the current study and the samples described in the original papers. Detailed information about the samples is provided in Supplementary Table 1.

2.2 eQTL data

Details concerning the GTEx data used for this study are described elsewhere (Aguet et al., 2016; Ardlie et al., 2015). To filter and annotate SNPs of the GWAS summary statistics with eQTL information, we downloaded significant SNP-gene associations (FDR-adjusted p-value < 0.05; i.e., eQTL-eGene associations) for 13 brain tissues: amygdala, anterior cingulate cortex, caudate, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen, spine (cervical), substantia nigra, and whole blood from the GTEx Portal V7 (http://www.gtportal.org/home/GTEx_Analysis_v7_eQTL.tar). The sample sizes varied across the 13 brain tissues (N ranges from 80 to 154); for whole blood the SNP-gene associations were obtained from 369 individuals (Aguet et al., 2016). Detailed information about the GTEx sample size per tissue and the number of significant eQTLs and eGenes per tissue can be found in Supplementary Table 2.

DGN (N=922)(Battle et al., 2014) data consists of a single tissue: whole blood. This data was used as an independent sample to conduct additional eQTL informed analyses. For the DGN database we used the same method as for GTEx database described above.

2.3 Statistical analyses

For the single variant analyses we extracted those SNPs that are eQTLs according to the summary statistics of all GWAS SNPs. This extraction was performed for each tissue. We applied the Benjamini-Hochberg False Discovery Rate (FDR) approach (Benjamini and Hochberg, 1995), which controls the expected proportion of false positives among all signals

with a FDR value below a fixed threshold, to determine significance. Using the GTEx database, we applied FDR on all 14 tissues combined for each of the sex substance use traits, which provided us with six sets of eQTLs with an FDR value. The single tissue DGN data set was analysed separately, also using FDR. Subsequently, we determined significant trait associations within those eQTL-sets using a FDR threshold of $q=0.05$.

Significant eGenes were identified by applying the FDR 0.05 threshold on the eQTLs, and linking the eQTLs with significant trait associations to the targeted eGenes. Since a single genetic locus can include multiple eQTLs that are in high LD, we clumped eQTLs with significant trait associations per tissue using PLINK2 (Chang et al., 2015). As clumping cut-offs, we used an R^2 of >0.1 and a physical distance of 1000kb, which generated a list of “index” eQTLs (i.e., independent eQTLs), which we considered as distinct loci. Due to the limited power all identified index eQTL, eGenes, and tissues should be considered as candidate findings.

Statistical analyses for the single variants were performed using the open-source programming language R (<https://www.r-project.org/>). For visualization, we used the R-library “ggplot2”. To generate a heatmap, a matrix was generated based on the R squared of the number of index eQTLs that met the trait-association FDR threshold of $q=0.05$. To investigate gene expression levels we used S-PrediXcan (Barbeira et al., 2018), which integrated eQTL information from summary statistics of the substance use GWAS in an aggregated manner. S-PrediXcan estimates gene expression weights by training a linear prediction model in samples with both gene expression and SNP genotype data. Subsequently these weights are used to predict gene expression from GWAS summary statistics, while incorporating the variance and co-variance of SNPs from an LD reference panel. The current study used expression weights for the 14 tissues central in this study from the GTEx Project (V7) and whole blood from the DGN cohort (Battle et al., 2014; Gamazon et al., 2018), and LD information from the 1000 Genome Project Phase 3 (Delaneau et al., 2014). These data were processed with beta values and standard errors from substance use summary statistics to estimate the expression-GWAS association statistic. A Bonferroni correction was used to determine the transcriptome-wide significant threshold (adjusting for all tissues and genes per trait).

3. Results

The GWAS data sets of the six substance use traits used in this study showed different levels of genetic signal when plotting all SNPs (Supplementary Figure 1). Alcohol consumption, CPD, ever smoker, age of smoking initiation, and former vs. current smoker showed significant trait associations (FDR-adjusted p -value <0.05). In contrast, cocaine dependence did not reveal any trait associated SNPs in the full GWAS. These results are largely in line with the original GWAS reports, and serve to compare the full (i.e., ‘uninformed’) GWAS against the eQTL informed GWAS analyses central in the current study.

After extracting GTEx eQTLs from the full GWAS summary statistics, significant trait associations (FDR-adjusted p -value <0.05) were observed, in all tissues, for the following substance use traits: alcohol consumption, CPD, ever smoker, age of smoking initiation, and

former vs. current smoker (Figure 1, Table 1, and Supplementary Table 3), but no significant trait association were found for any tissue for cocaine dependence. Using this methodology, we identified 2,976 (GTEx) and 811 (DGN) novel loci (i.e., loci not identified in the uninformed GWAS of the six substance use traits; FDR-adjusted p-value <0.05; Table 2). Moreover, these eQTLs show evidence of exerting their effects in the various tissues through numerous eGenes (Table 1, Supplementary Tables 3 and 4). Consistent over all substance use traits, we observed that eGenes targeted by index eQTLs were generally different from the nearest gene (Table 1 and Supplementary Tables 3). In fact, in 66.3% of the cases the nearest gene was not the eGene in the GTEx informed analyses. This suggests that proximity is a limited measure in determining the functional relevance of a gene.

For alcohol consumption, we found eQTLs with significant trait associations in all 14 GTEx tissues (Figure 1), and identified considerably more significant eGenes than for the other substance use traits (Supplementary Table 5). In total, we identified 949 novel index eQTLs for alcohol consumption, i.e. these SNPs were not significant in the uninformed GWAS.

For CPD, we found significant trait associations in all 14 GTEx tissues. We identified 509 index eQTLs which were not identified in the uninformed GWAS for CPD (Table 2). Many of the strongest associations were found in brain, and were located within the gene cluster: *IREB2-CHRNA3-CHRNA5-CHRNA4-HYKK-PSMA4* (Table 1, Supplementary Table 3). For the other smoking traits (i.e., former vs. current smoker, age of smoking initiation, ever vs. never smoker) we identified 117, 66, and 1,335 novel loci, respectively (Table 2). For these traits the *IREB2-CHRNA3-CHRNA5-CHRNA4-HYKK-PSMA4* gene cluster showed considerably weaker associations (Table 1, Supplementary Table 3). Furthermore, the GTEx eQTL informed GWAS for cocaine dependence did not reveal any significant trait associations, similar to the uninformed GWAS.

The single variant eQTL analyses informed by the DGN data, showed 224 significant index eQTLs for alcohol consumption, 136 for CPD, 401 for ever smoker, 25 for age of smoking initiation, and 35 for former vs. current smoker, which were not found by the uninformed GWAS (Table 2 and Supplementary Table 5). While the genes most proximal to significant index eQTL based on DGN were occasionally similar as found in the GTEx analyses, they often targeted different eGenes. Comparing DGN to GTEx whole blood, different eGenes were targeted by the same eQTL in 65.6 % of the cases.

The S-PrediXcan analyses, informed by the GTEx eQTLs, revealed large numbers of significantly associated genes (corrected for the number of genes and tissues) for alcohol consumption, CPD, age of smoking initiation, ever smoker, and former vs. current smoker (Tables 3,4 and Supplementary Table 6). For alcohol consumption, CPD, and ever smoker, we found differentially expressed genes in all tissues under investigation (Table 3). Noteworthy, the identified genes of the S-PrediXcan analyses showed evidence of tissue specificity, as many of our findings were either unique to brain or to whole blood (Table 4). Moreover, comparing the S-PrediXcan results informed by GTEx and DGN, we observed many findings to be unique to a specific reference panel (Supplementary Table 7).

4. Discussion

The overarching aim of this study was to explore whether functionally annotating SNPs using information on their role in the regulation of gene expression facilitates the identification and functional interpretation of candidate loci involved in substance use traits. Furthermore, we aimed to identify genes with differentially genetically regulated levels of gene expression associated with substance use traits. Using the GTEx database we explored the role of regulatory genetic variants in 14 different tissues and tested association with six substance use traits. In addition, we conducted independent eQTL analyses using the DGN database.

We identified index eQTLs with significant trait associations (FDR-adjusted p-value <0.05) for five of the six traits examined: alcohol consumption, CPD ever smoker, age of onset of smoking, former vs. current smoker. For cocaine dependence, no GTEx eQTLs with significant trait associations were found. Compared to the full (i.e., eQTL-uninformed) GWAS results, functional annotation of eQTLs improved the power to detect significant trait associations. Overall, informed by GTEx, this method allowed us to identify 2,976 novel index eQTLs, we interpret these index eQTLs as candidate loci for substance use traits. These candidate loci were not previously detected by the uninformed GWAS of the six substance use traits.

We will discuss a few highlights of our findings. For alcohol consumption, significant index eQTLs were observed in all brain regions and in whole blood. Most of these brain regions have previously been demonstrated to play a role in the susceptibility to alcohol use disorders (Acheson et al., 2009; Cheetham et al., 2014; Hanson et al., 2010; Herting et al., 2011; Herting et al., 2010; Sjoerds et al., 2013). Interesting findings were index eQTLs in the putamen, caudate and cervical spine targeting alcohol dehydrogenase 1C (*ADH1C*) to be significantly associated with alcohol consumption. *ADH1C* was previously reported to be associated with alcohol dependence and consumption (Clarke et al., 2017; Frank et al., 2012; Park et al., 2013; Treutlein et al., 2009). Moreover, we identified multiple index eQTLs targeting many different eGenes in various tissues on chromosome 17q21.31 for alcohol consumption. Chromosome 17q21.31 has been described as one of the genome's most structurally complex and evolutionary dynamic regions, and genes in this region have – among other traits – been implicated in alcohol use (Liu et al., 2019; Nelson et al., 2010; Pennisi, 2008). This underscores the complexity of this region and of the potential importance of gene regulatory mechanisms (Louro et al., 2009). Our results, in combination with the emerging literature, suggest that this region includes multiple functional genetic variants that contribute to individual differences in alcohol consumption. It should be noted, however, that previous research indicates that these regulation hotspots should be interpreted with caution due to the complex correlation structure of gene expression, which could lead to false positive associations (de Koning and Haley, 2005; Peng et al., 2007).

The associations found for CPD on chromosome 15q25.1 were magnitudes stronger than the ones found for the other substance use traits examined in the current study (i.e., 10^{-40} vs 10^{-211}). The index eQTLs with the strongest associations, all located within the *IREB2-CHRNA3-CHRNA5-CHRNA4-HYKK-PSMA4* gene cluster, predominantly

targeted *CHRNA5*. This gene cluster has been associated many times in the literature with smoking severity (Barrie et al., 2016; Furberg et al., 2010). Our findings extend the previously observed genetic association by showing that the CPD-associated eQTLs regulate gene expression of *CHRNA3* and *CHRNA5* in striatal brain areas. A study by Barrie et al., in which GTEx tissue-specific eQTLs from the gene cluster *CHRNA5/CHRNA3/CHRNB4* were explored to investigate their role in nicotine dependence, found significant striatal eQTLs targeting the same eGenes (*CHRNA3*, *CHRNA5*, *RP11-650L12.2*) as we found here for CPD (Barrie et al., 2016). The striatal eQTLs identified in the current study and the findings of Barrie et al. highlight the importance of genetic regulation in the striatum for smoking behaviors. For the other smoking related traits no associations, comparable to the strength of those for CPD, were found for the *IREB2-CHRNA3-CHRNA5-CHRNB4-HYKK-PSMA4* gene cluster. This suggests that vulnerability for these traits, in part, goes through other biological pathways. This observation is in line with the literature which shows genetic correlations between CPD and other smoking related traits to be below 0.5 (Liu et al., 2019).

For cocaine dependence, we found no index eQTLs with significant trait associations in the GTEx analyses. The absence of significant results is most likely due to the limited sample size of this GWAS. However, the DGN informed analyses identified one index eQTL, which targets the eGene *C11orf9*. Previously this gene has been shown to be the target of miRNA's which are increased in patients with alcohol use (Miguel-Hidalgo, 2018).

We observed modest overlap between proximal genes and eGenes: 66.3% of GTEx's significant index eQTLs targeted eGenes other than their proximal gene. This implies that the search for functionally relevant genes using GWAS results should not merely focus on physical proximity but should instead take genetic regulation into account as an important biological mechanism. The DGN informed single variant analyses revealed significant eQTLs in the same traits as the GTEx whole blood sample and also in cocaine dependence. While significant index eQTLs were found in many of the same loci in both databases for alcohol consumption, CPD, age of smoking initiation, ever smoker, and former vs. current smoker, different eGenes were targeted by these eQTLs. In fact, in 65.6 percent another eGene was targeted. This is probably due to the fact that both GTEx and DGN are underpowered, and thus providing incomplete, not overlapping, eQTL information.

The S-PrediXcan analyses partly confirmed the results of the single variant eQTL analyses. For example, similar to the single variant analysis for CPD, the GTEx informed S-PrediXcan analysis identified the genes *CHRNA5*, *CHRNA3* and *PSMA4* to be associated with CPD for various tissues. Furthermore, for alcohol consumption, the strongest association was found for the gene *RFC1* in both the single variant and S-PrediXcan analyses. Previous research also shown a role for *RFC1* in alcohol consumption and alcohol use disorder (Liu et al., 2019; Sanchez-Roige et al., 2019). Noteworthy, 64.9 percent of our finding were detected only in brain-tissues, highlighting the importance of transcriptomic annotation, to assess the role of difficult-to-acquire tissues in substance use traits. Previous research made similar observations for psychiatric disorders (Gamazon et al., 2019).

Our results suggest that whole blood may be an interesting biomarker for substance use traits since whole blood consistently, both in the single variant and the S-PrediXcan analyses, showed significant results also found other tissues. In fact, whole blood showed more significant results than other tissues. However, this is, in part, explained by the larger sample size of whole blood. The significant associations in whole blood may be due to whole blood-brain tissue eQTL overlap as a reflection of causative brain tissue specific eQTLs, which are detectable in whole blood, rather than pointing to a causative role for whole blood in substance use traits (McKenzie et al., 2014; Wainberg et al., 2019).

The findings and conclusions of this study should be interpreted in view of some key limitations. Despite the comparatively large sample sizes of the GWAS, our study may still be underpowered to detect small genetic effects, which is especially true for the cocaine dependence sample. In addition, although GTEx and DGN belong to the most comprehensive genetic expression databases, the statistical power for eQTL discovery is still modest for some tissues (Ardlie et al., 2015). The GTEx brain sample sizes are smaller than those for whole blood, which is reflected in fewer identified index eQTLs in the separate brain tissues. It is therefore likely that we tested only a subset of the total number of true eQTLs for the various tissues. Our analyses focus on the role of eQTLs in a wide range of tissues, however, recently it has been shown that eQTL effects may differ between cell types within a specific tissue (van der Wijst et al., 2018). Therefore, to better understand the role of gene expression in substance use traits future studies should focus on cell type specific analyses (van der Wijst et al., 2018). Furthermore, the genes identified with the single variant and S-PrediXcan analyses should be seen as ‘candidates’ as correlated levels of gene expression may be observed in high LD genomic regions which makes it challenging to identify the true causal genes (Wainberg et al., 2019). Moreover, the GTEx data was composed of subjects of European ancestry, while the GWAS results used in the current study for and cocaine dependence are based on European and African American subjects. Since eQTLs may not completely overlap across ethnic populations, this may have reduced our ability to detect novel loci. However, with sample size always being a limiting factor in complex trait GWAS, we included the combined sample of European and African American subjects for cocaine dependence to improve statistical power. By no means we claim to present the full set of eQTLs, eGenes, and tissues involved in the substance use traits under investigation. Therefore, the index eQTLs, eGenes, and tissues identified by this study to be involved in substance use traits, should be seen as ‘candidates’. Moreover, independent replication of these novel candidate loci is necessary before strong conclusions can be drawn regarding the role of these loci in substance use traits.

5. Conclusions

Our findings suggest that there is great value in utilizing brain and whole blood eQTL annotations for enhancing the discovery of novel genetic susceptibility loci for substance use traits. The tissue- focused GTEx eQTL analyses revealed 2,976 index eQTLs which were not identified in the discovery samples using the same threshold (FDR adjusted $p < 0.05$), implying that these candidate loci might be interesting for further research. In addition, the functional annotation of GWAS data revealed that most of the identified candidate eGenes targeted by the trait-associated index eQTLs are not the nearest genes,

underscoring the importance of studying genetic regulation of gene expression for functional annotation of genetic loci. Finally, the S-PrediXcan results validated some of the candidate findings of the single variant analyses. Both the single variant and gene-level analyses confirm the conclusions of Gamazon et al. regarding the importance of multiple tissue eQTL investigation (Gamazon et al., 2018), as our analyses identified many interesting (novel) candidate eGenes and tissues for substance use traits. In conclusion, annotating genes based on transcriptomic regulation in brain and non-brain tissues improves both the identification of novel candidate genes and the functional characterization of genetic risk factors for substance use traits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Full list of International Cannabis Consortium members

Sven Stringer^{1,11}, Camelia C. Minic¹², Hamdi Mbarek¹², Manon Bernard¹³, Jaime Derringer¹⁴, Kristel R. van Eijk¹⁵, Joshua D. Isen¹⁶, Anu Loukola¹⁷, Dominique F. Maciejewski¹⁸, Evelin Mihailov¹⁹, Peter J. van der Most²⁰, Cristina Sánchez-Mora^{21,22,23}, Leonie Roos²⁴, Raymond Walters^{25,26,27}, Jennifer J. Ware^{28,29}, Abdel Abdellaoui¹², Timothy B. Bigdeli³⁰, Susan J.T. Branje³¹, Sandra A. Brown³², Marcel Bruinenberg³³, Miguel Casas^{22,23,34}, Tõnu Esko¹⁹, Iris Garcia-Martinez^{21,22}, Scott D. Gordon³⁵, Juliette M. Harris²⁴, Catharina A. Hartman³⁶, Anjali K. Henders³⁵, Andrew C. Heath³⁷, Ian B. Hickie³⁸, Matthew Hickman²⁸, Christian J. Hopfer³⁹, Jouke Jan Hottenga¹², Anja C. Huizink¹⁸, Daniel E. Irons¹⁶, René S. Kahn¹⁵, Tellervo Korhonen^{17,40,41}, Ken Krauter⁴², Pol A.C. van Lier¹⁸, Gitta H. Lubke^{12,43}, Pamela A.F. Madden³⁷, Reedik Mägi¹⁹, Matt K. McGue¹⁶, Sarah E. Medland³⁵, Wim H.J. Meeus^{31,44}, Michael B. Miller¹⁶, Grant W. Montgomery³⁵, Michel G. Nivard¹², Ilja M. Nolte²⁰, Albertine J. Oldehinkel⁴⁵, Zdenka Pausova^{13,46}, Beenish Qaiser¹⁷, Lydia Quayle²⁴, Josep A. Ramos-Quiroga^{22,23,34}, Vanesa Richarte²², Richard J. Rose⁴⁷, Jean Shin¹³, Michael C. Stallings⁴⁸, Alex I. Stiby²⁸, Tamara L. Wall⁴⁹, Margaret J. Wright³⁵, Hans M. Koot¹⁸, Tomas Paus^{50,51,52}, John K. Hewitt⁴⁸, Marta Ribasés^{21,22,23}, Jaakko Kaprio^{17,41,53}, Marco P. Boks¹⁵, Harold Snieder²⁰, Tim Spector²⁴, Marcus R. Munafò^{28,54}, Andres Metspalu¹⁹, Dorret I. Boomsma^{12,55}, William G. Iacono¹⁶, Nicholas G. Martin³⁵, Nathan A. Gillespie^{30,35}, & Jacqueline M. Vink^{7,12, **}

** Corresponding author International Cannabis Consortium: Jacqueline M. Vink, j.vink@bsi.ru.nl; Tel.: +31 24 3616097

Affiliations:

¹¹Department of Complex Trait Genetics, VU Amsterdam, Center for Neurogenomics and Cognitive Research, Amsterdam, The Netherlands

¹²Department of Biological Psychology/Netherlands Twin Register, VU University, Amsterdam, The Netherlands

¹³The Hospital for Sick Children Research Institute, Toronto, Canada

¹⁴Department of Psychology, University of Illinois Urbana-Champaign, Champaign, Illinois, USA

¹⁵Department of Human Neurogenetics, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

¹⁶Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA

¹⁷Department of Public Health, University of Helsinki, Hjelt Institute, Helsinki, Finland

¹⁸VU University, Department of Developmental Psychology and EMGO Institute for Health and Care Research, Amsterdam, The Netherlands

¹⁹Estonian Genome Center, University of Tartu, Tartu, Estonia

²⁰Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

²¹Psychiatric Genetics Unit, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain

²²Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain

²³Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain

²⁴Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom.

²⁵Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

²⁶Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

²⁷Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

²⁸School of Social and Community Medicine, University of Bristol, Bristol, UK

²⁹MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK

³⁰Department of Psychiatry, Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, Virginia, USA

³¹Research Centre Adolescent Development, Utrecht University, Utrecht, the Netherlands

³²Department of Psychology and Psychiatry, University of California San Diego, La Jolla, California, USA

³³The LifeLines Cohort Study, University of Groningen, Groningen, The Netherlands

- ³⁴Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain
- ³⁵Genetic Epidemiology, Molecular Epidemiology and Neurogenetics Laboratories, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
- ³⁶Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
- ³⁷Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri, USA
- ³⁸Brain & Mind Research Institute, University of Sydney, Sydney, NSW, Australia
- ³⁹Department of Psychiatry, University of Colorado Denver, Aurora, Colorado, USA
- ⁴⁰University of Eastern Finland, Institute of Public Health & Clinical Nutrition, Kuopio, Finland
- ⁴¹Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland
- ⁴²Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, Colorado, USA
- ⁴³Department of Psychology, University of Notre Dame, Notre Dame, Indiana, USA
- ⁴⁴Developmental Psychology, Tilburg University, Tilburg, The Netherlands
- ⁴⁵Interdisciplinary Center for Pathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
- ⁴⁶Physiology and Nutritional Sciences, University of Toronto, Toronto, Canada
- ⁴⁷Department of Psychological & Brain Sciences, Indiana University Bloomington, Bloomington, Indiana, USA
- ⁴⁸Institute for Behavioral Genetics, Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, Colorado, USA
- ⁴⁹Department of Psychiatry, University of California San Diego, La Jolla, California, USA
- ⁵⁰Rotman Research Institute, Baycrest, Toronto, Canada
- ⁵¹Psychology and Psychiatry, University of Toronto, Toronto, Canada
- ⁵²Center for the Developing Brain, Child Mind Institute, New York, USA
- ⁵³Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

⁵⁴UK Centre for Tobacco and Alcohol Studies and School of Experimental Psychology, University of Bristol, Bristol, UK

⁵⁵Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

References

- Acheson A, Robinson JL, Glahn DC, Lovallo WR, Fox PT, 2009. Differential activation of the anterior cingulate cortex and caudate nucleus during a gambling simulation in persons with a family history of alcoholism: studies from the Oklahoma Family Health Patterns Project. *Drug Alcohol Depend* 100(1–2), 17–23. [PubMed: 19026496]
- Aguet F, Brown AA, Castell SE, Davis JR, Pejman M, 2016. Local genetic effects on gene expression across 44 human tissues. *bioRxiv*
- Albertson DN, Pruetz B, Schmidt CJ, Kuhn DM, Kapatos G, Bannon MJ, 2004. Gene expression profile of the nucleus accumbens of human cocaine abusers: evidence for dysregulation of myelin. *J Neurochem* 88(5), 1211–1219. [PubMed: 15009677]
- Albertson DN, Schmidt CJ, Kapatos G, Bannon MJ, 2006. Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse. *Neuropsychopharmacol* 31(10), 2304–2312.
- Ardlie KG, DeLuca DS, Segre AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M, Ward LD, Kheradpour P, Iriarte B, Meng Y, Palmer CD, Esko T, Winckler W, Hirschhorn JN, Kellis M, MacArthur DG, Getz G, Shabalin AA, Li G, Zhou YH, Nobel AB, Rusyn I, Wright FA, Lappalainen T, Ferreira PG, Ongen H, Rivas MA, Battle A, Mostafavi S, Monlong J, Sammeth M, Mele M, Reverter F, Goldmann JM, Koller D, Guigo R, McCarthy MI, Dermitzakis ET, Gamazon ER, Im HK, Konkashbaev A, Nicolae DL, Cox NJ, Flutre T, Wen XQ, Stephens M, Pritchard JK, Tu ZD, Zhang B, Huang T, Long Q, Lin L, Yang JL, Zhu J, Liu J, Brown A, Mestichelli B, Tidwell D, Lo E, Salvatore M, Shad S, Thomas JA, Lonsdale JT, Moser MT, Gillard BM, Karasik E, Ramsey K, Choi C, Foster BA, Syron J, Fleming J, Magazine H, Hasz R, Walters GD, Bridge JP, Miklos M, Sullivan S, Barker LK, Traino HM, Mosavel M, Siminoff LA, Valley DR, Rohrer DC, Jewell SD, Branton PA, Sobin LH, Barcus M, Qi LQ, McLean J, Hariharan P, Um KS, Wu SP, Tabor D, Shive C, Smith AM, Buia SA, Undale AH, Robinson KL, Roche N, Valentino KM, Britton A, Burges R, Bradbury D, Hambright KW, Seleski J, Korzeniewski GE, Erickson K, Marcus Y, Tejada J, Taherian M, Lu CR, Basile M, Mash DC, Volpi S, Struwing JP, Temple GF, Boyer J, Colantuoni D, Little R, Koester S, Carithers LJ, Moore HM, Guan P, Compton C, Sawyer SJ, Demchok JP, Vaught JB, Rabiner CA, Lockhart NC, Ardlie KG, Getz G, Wright FA, Kellis M, Volpi S, Dermitzakis ET, Consortium G, 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 348(6235), 648–660. [PubMed: 25954001]
- Barbeira AN, Dickinson SP, Bonazzola R, Zheng JM, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, Stahl EA, Huckins LM, Nicolae DL, Cox NJ, Im HK, Consortium G, 2018. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 9.
- Barrie ES, Hartmann K, Lee SH, Frater JT, Seweryn M, Wang D, Sadee W, 2016. The CHRNA5/CHRNA3/CHRNA4 Nicotinic Receptor Regulome: Genomic Architecture, Regulatory Variants, and Clinical Associations. *Hum Mutat*
- Battle A, Mostafavi S, Zhu XW, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi JX, Mei R, Urban AE, Montgomery SB, Levinson DF, Koller D, 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome research* 24(1), 14–24. [PubMed: 24092820]
- Benjamini Y, Hochberg Y, 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57(1), 289–300.
- Celentano M, Caprioli D, Di Pasquale P, Cardillo V, Nencini P, Gaetani S, Badiani A, 2009. Drug context differently regulates cocaine versus heroin self-administration and cocaine- versus heroin-induced Fos mRNA expression in the rat. *Psychopharmacology* 204(2), 349–360. [PubMed: 19169671]

- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ, 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. [PubMed: 25722852]
- Cheetham A, Allen NB, Whittle S, Simmons J, Yucel M, Lubman DI, 2014. Volumetric differences in the anterior cingulate cortex prospectively predict alcohol-related problems in adolescence. *Psychopharmacology (Berl)* 231(8), 1731–1742. [PubMed: 24553579]
- Clarke T-K, Adams MJ, Davies J, Howard DM, Hall LS, Padmanabhan S, Murray AD, Smith BH, Campbell A, Hayward C, Porteous DJ, Deary IJ, McIntosh AM, 2017. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol Psychiatr Advance online publication*.
- Consortium G, 2018. Genetic effects on gene expression across human tissues (vol 550, pg 204, 2017). *Nature* 553(7689), 530–530.
- de Koning DJ, Haley CS, 2005. Genetical genomics in humans and model organisms. *Trends Genet* 21(7), 377–381. [PubMed: 15908034]
- Delaneau O, Marchini J, Consortium GP, 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* 5.
- Evangelou E, Ioannidis JPA, 2013. Meta-analysis methods for genome-wide association studies and beyond. *Nat Rev Genet* 14(6), 379–389. [PubMed: 23657481]
- Frank J, Cichon S, Treutlein J, Ridinger M, Mattheisen M, Hoffmann P, Herms S, Wodarz N, Soyka M, Zill P, Maier W, Mossner R, Gaebel W, Dahmen N, Scherbaum N, Schmal C, Steffens M, Lucae S, Ising M, Muller-Myhsok B, Nothen MM, Mann K, Kiefer F, Rietschel M, 2012. Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict Biol* 17(1), 171–180. [PubMed: 22004471]
- Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, Bernardinelli L, Mannucci PM, Mauri F, Merlini PA, Absher D, Assimes TL, Fortmann SP, Iribarren C, Knowles JW, Quertermous T, Ferrucci L, Tanaka T, Bis JC, Furberg CD, Haritunians T, McKnight B, Psaty BM, Taylor KD, Thacker EL, Almgren P, Groop L, Ladenvall C, Boehnke M, Jackson AU, Mohlke KL, Stringham HM, Tuomilehto J, Benjamin EJ, Hwang SJ, Levy D, Preis SR, Vasan RS, Duan J, Gejman PV, Levinson DF, Sanders AR, Shi JX, Lips EH, McKay JD, Agudo A, Barzan L, Bencko V, Benhamou S, Castellsague X, Canova C, Conway DI, Fabianova E, Foretova L, Janout V, Healy CM, Holcatova I, Kjaerheim K, Lagiou P, Lissowska J, Lowry R, Macfarlane TV, Mates D, Richiardi L, Rudnai P, Szeszenia-Dabrowska N, Zaridze D, Znaor A, Lathrop M, Brennan P, Bandinelli S, Frayling TM, Guralnik JM, Milanesechi Y, Perry JRB, Altshuler D, Elosua R, Kathiresan S, Lucas G, Melander O, O'Donnell CJ, Salomaa V, Schwartz SM, Voight BF, Penninx BW, Smit JH, Vogelzangs N, Boomsma DI, de Geus EJC, Vink JM, Willemsen G, Chanock SJ, Gu FY, Hankinson SE, Hunter DJ, Hofman A, Tiemeier H, Uitterlinden AG, van Duijn CM, Walter S, Chasman DI, Everett BM, Pare G, Ridker PM, Li MD, Maes HH, Audrain-McGovern J, Posthuma D, Thornton LM, Lerman C, Kaprio J, Rose JE, Ioannidis JPA, Kraft P, Lin DY, Sullivan PF, 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 42(5), 441–U134. [PubMed: 20418890]
- Gamazon ER, Segre AV, van de Bunt M, Wen XQ, Xi HS, Hormozdiari F, Ongen H, Konkashbaev A, Derks EM, Aguet F, Quan J, Nicolae L, Eskin E, Kellis M, Getz G, McCarthy MI, Dermitzakis ET, Cox NJ, Ardlie KG, Consortium, G., 2018. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* 50(7), 956–+. [PubMed: 29955180]
- Gamazon ER, Zwinderman AH, Cox NJ, Denys D, Derks EM, 2019. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat Genet* 51(6), 933–+. [PubMed: 31086352]
- Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, Kranzler HR, Farrer L, 2014. Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol Psychiatr* 19(6), 717–723.
- Hanson KL, Medina KL, Nagel BJ, Spadoni AD, Gorlick A, Tapert SF, 2010. Hippocampal volumes in adolescents with and without a family history of alcoholism. *The American journal of drug and alcohol abuse* 36(3), 161–167. [PubMed: 20465374]
- Herting MM, Fair D, Nagel BJ, 2011. Altered fronto-cerebellar connectivity in alcohol-naive youth with a family history of alcoholism. *Neuroimage* 54(4), 2582–2589. [PubMed: 20970506]

- Herting MM, Schwartz D, Mitchell SH, Nagel BJ, 2010. Delay discounting behavior and white matter microstructure abnormalities in youth with a family history of alcoholism. *Alcohol Clin Exp Res* 34(9), 1590–1602. [PubMed: 20586754]
- Liu MZ, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, Datta G, Davila-Velderrain J, McGuire D, Tian C, Zhan XW, Choquet H, Docherty AR, Faul JD, Foerster JR, Fritsche LG, Gabrielsen ME, Gordon SD, Haessler J, Hottenga JJ, Huang HY, Jang SK, Jansen PR, Ling Y, Magi R, Matoba N, McMahon G, Mulas A, Orru V, Palviainen T, Pandit A, Reginsson GW, Skogholt AH, Smith JA, Taylor AE, Turman C, Willemsen G, Young H, Young KA, Zajac GJM, Zhao W, Zhou W, Bjornsdottir G, Boardman JD, Boehnke M, Boomsma DI, Chen C, Cucca F, Davies GE, Eaton CB, Ehringer MA, Esko T, Fiorillo E, Gillespie NA, Gudbjartsson DF, Haller T, Harris KM, Heath AC, Hewitt JK, Hickie IB, Hokanson JE, Hopfer CJ, Hunter DJ, Iacono WG, Johnson EO, Kamatani Y, Kardina SLR, Keller MC, Kellis M, Kooperberg C, Kraft P, Krauter KS, Laakso M, Lind PA, Loukola A, Lutz SM, Madden PAF, Martin NG, McGue M, McQueen MB, Medland SE, Metspalu A, Mohlke KL, Nielsen JB, Okada Y, Peters U, Polderman TJC, Posthuma D, Reiner AP, Rice JP, Rimm E, Rose RJ, Runarsdottir V, Stallings MC, Stancakova A, Stefansson H, Thai KK, Tindler HA, Tyringsson T, Wall TL, Weir DR, Weisner C, Whitfield JB, Winsvold BS, Yin J, Zuccolo L, Bierut LJ, Hveem K, Lee JJ, Munafo MR, Saccone NL, Willer CJ, Cornelis MC, David SP, Hinds DA, Jorgenson E, Kaprio J, Stitzel JA, Stefansson K, Thorgeirsson TE, Abecasis G, Liu DJ, Vrieze S, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Furlotte NA, Hinds DA, Hromatka BS, Huber KE, Kleinman A, Litterman NK, McIntyre MH, Mountain JL, Northover CAM, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Tung JY, Vacic V, Wilson CH, Pitts SJ, Mitchell A, Winsvold BS, Sivertsen B, Stordal E, Morken G, Kallestad H, Heuch I, Zwart JA, Fjukstad KK, Pedersen LM, Johnsen MB, Skrove M, Indredavik MS, Drange OK, Bjerkeset O, Borte S, Stensland SO, Team MR, Psychiat HAI, 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(2), 237–+. [PubMed: 30643251]
- Louro R, Smirnova AS, Verjovski-Almeida S, 2009. Long intronic noncoding RNA transcription: Expression noise or expression choice? *Genomics* 93(4), 291–298. [PubMed: 19071207]
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu HZ, Brody J, Shafer A, Neri F, Lee K, Kutayavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA, 2012. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* 337(6099), 1190–1195. [PubMed: 22955828]
- McKenzie M, Henders AK, Caracella A, Wray NR, Powell JE, 2014. Overlap of expression quantitative trait loci (eQTL) in human brain and blood. *Bmc Med Genomics* 7, 31. [PubMed: 24894490]
- Miguel-Hidalgo JJ, 2018. Molecular Neuropathology of Astrocytes and Oligodendrocytes in Alcohol Use Disorders. *Front Mol Neurosci* 11.
- Nelson EC, Agrawal A, Pergadia ML, Wang JC, Whitfield JB, Saccone FS, Kern J, Grant JD, Schrage AJ, Rice JP, Montgomery GW, Heath AC, Goate AM, Martin NG, Madden PA, 2010. H2 haplotype at chromosome 17q21.31 protects against childhood sexual abuse-associated risk for alcohol consumption and dependence. *Addict Biol* 15(1), 1–11. [PubMed: 19878140]
- Park BL, Kim JW, Cheong HS, Kim LH, Lee BC, Seo CH, Kang TC, Nam YW, Kim GB, Shin HD, Choi IG, 2013. Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Human genetics* 132(6), 657–668. [PubMed: 23456092]
- Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, Abdellaoui A, Nivard MG, Baselmans BML, Ong JS, Ip HF, van der Zee MD, Bartels M, Day FR, Fontanillas P, Elson SL, de Wit H, Davis LK, MacKillop J, Derringer JL, Branje SJT, Hartman CA, Heath AC, van Lier PAC, Madden PAF, Magi R, Meeus W, Montgomery GW, Oldehinkel AJ, Pausova Z, Ramos-Quiroga JA, Paus T, Ribases M, Kaprio J, Boks MPM, Bell JT, Spector TD, Gelernter J, Boomsma DI, Martin NG, MacGregor S, Perry JRB, Palmer AA, Posthuma D, Munafo MR, Gillespie NA, Derks EM, Vink JM, Team, a.R., Ps, S.U.D.W.G., Consortium, I.C., 2018. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci* 21(9), 1161–+. [PubMed: 30150663]

- Peng J, Wang P, Tang H, 2007. Controlling for false positive findings of trans-hubs in expression quantitative trait loci mapping. *BMC Proc* 1 Suppl 1, S157.
- Pennisi E, 2008. Genetics. 17q21.31: not your average genomic address. *Science* 322(5903), 842–845. [PubMed: 18988819]
- Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, Adams MJ, Howard DM, Edenberg HJ, Davies G, Crist RC, Deary IJ, McIntosh AM, Clarke TK, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Furlotte NA, Hinds DA, Huber KE, Kleinman A, Litterman NK, McCreight JC, McIntyre MH, Mountain JL, Noblin ES, Northover CAM, Pitts SJ, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Tung JY, Vacic V, Wilson CH, Team, a.R., 2019. Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am J Psychiat* 176(2), 107–118. [PubMed: 30336701]
- Sjoerds Z, Van Tol MJ, Van den Brink W, Van der Wee NJ, Van Buchem MA, Aleman A, Penninx BW, Veltman DJ, 2013. Family history of alcohol dependence and gray matter abnormalities in non-alcoholic adults. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry* 14(8), 565–573.
- Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, Ludwig KU, Frank J, Wichmann HE, Schreiber S, Dragano N, Sommer WH, Leonardi-Essmann F, Lourdasamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nothen MM, Kiefer F, Spanagel R, Mann K, Rietschel M, 2009. Genome-wide association study of alcohol dependence. *Archives of general psychiatry* 66(7), 773–784. [PubMed: 19581569]
- van der Wijst MGP, Brugge H, de Vries DH, Deelen P, Swertz MA, LifeLines Cohort, S., Consortium B, Franke L, 2018. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet* 50(4), 493–497. [PubMed: 29610479]
- Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K, Bjorkegren JLM, Im HK, Pasaniuc B, Rivas MA, Kundaje A, 2019. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* 51(4), 592–599. [PubMed: 30926968]
- Zhou ZF, Enoch MA, Goldman D, 2014. Gene Expression in the Addicted Brain. *Int Rev Neurobiol* 116, 251–273. [PubMed: 25172478]
- Zhou ZF, Yuan QP, Mash DC, Goldman D, 2011. Substance-specific and shared transcription and epigenetic changes in the human hippocampus chronically exposed to cocaine and alcohol. *P Natl Acad Sci USA* 108(16), 6626–6631.

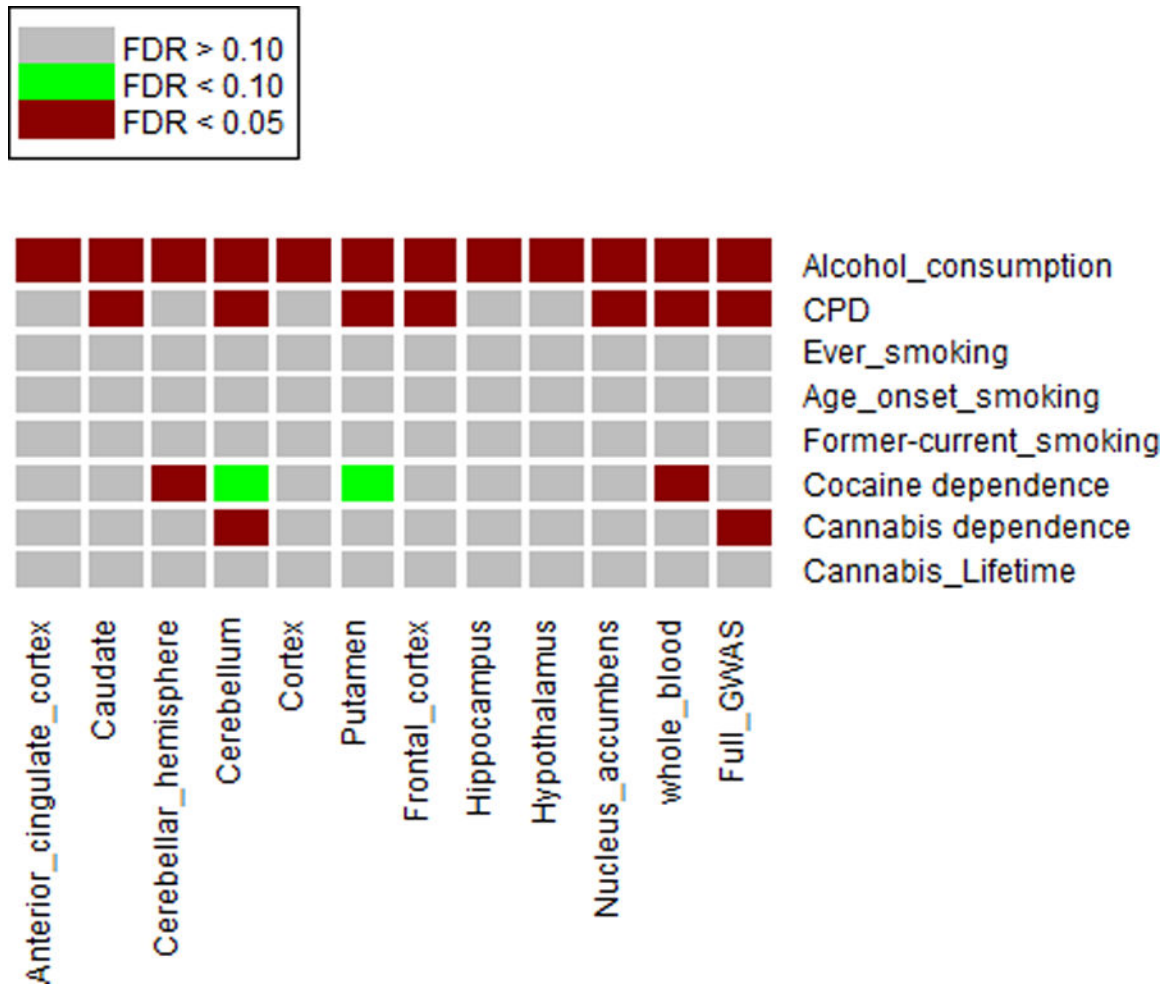


Figure 1. Heatmap of number of GTEx (sqrt transformed) index eQTLs with significant trait association (FDR-adjusted $p < 0.05$) per tissue and per trait

Table 1.

Overview of the five strongest eQTL-eGene relations per substance use trait, identified by the current study.

Tissue	Index eQTL	PVALUE	FDR	Alcohol consumption					eGene	Distance to TSS (kb)*
				GWAS effect	GTEX slope	Proximal gene				
Cerebellar Hemisphere	rs28712821	1.10E-46	3.59E-40	-2.84E-02	-3.58E-01	ENSG00000134962	ENSG00000035928	46		
Whole Blood	rs35538052	3.23E-44	1.42E-38	2.80E-02	1.75E-01	ENSG00000134962	ENSG00000109814	-111		
Cerebellum	rs1260326	3.33E-33	4.18E-28	-2.33E-02	5.54E-01	ENSG00000084734	ENSG00000234072	152		
Cerebellar Hemisphere	rs1260326	3.33E-33	4.18E-28	-2.33E-02	5.00E-01	ENSG00000084734	ENSG00000234072	152		
Whole blood	rs1260326	3.33E-33	4.18E-28	-2.33E-02	-1.39E-01	ENSG00000084734	ENSG00000115216	80		
CPD										
Caudate	rs8034191	4.80E-211	1.95E-205	-9.06E-02	-5.68E-01	ENSG00000188266	ENSG00000261762	-78		
Caudate	rs8034191	4.80E-211	1.95E-205	-9.06E-02	-5.90E-01	ENSG00000188266	ENSG00000169684	-52		
Cortex	rs8034191	4.80E-211	1.95E-205	-9.06E-02	-6.68E-01	ENSG00000188266	ENSG00000169684	-52		
Frontal Cortex	rs8034191	4.80E-211	1.95E-205	-9.06E-02	-6.44E-01	ENSG00000188266	ENSG00000169684	-52		
Nucleus accumbens	rs8034191	4.80E-211	1.95E-205	-9.06E-02	-6.11E-01	ENSG00000188266	ENSG00000169684	-52		
Ever smoker										
Brain Cerebellum	rs1565735	3.42E-17	5.01E-12	1.76E-02	6.80E-01	ENSG00000234770	ENSG00000120903	89		
Caudate	rs1004787	5.27E-17	5.01E-12	-1.24E-02	-4.16E-01	ENSG00000259439	ENSG00000236502	-10		
Anterior Cingulate Cortex	rs240957	5.36E-17	5.01E-12	1.75E-02	-5.60E-01	ENSG00000271789	ENSG00000255389	-303		
Cerebellum	rs240957	5.36E-17	5.01E-12	1.75E-02	-4.12E-01	ENSG00000271789	ENSG00000009413	-187		
Cortex	rs6937734	5.56E-17	5.01E-12	1.59E-02	-4.35E-01	ENSG00000009413	ENSG00000009413	-68		
Age of initiation										
Whole Blood	rs62177761	1.85E-10	4.55E-06	-1.91E-02	1.84E-01	ENSG00000143951	ENSG00000143951	-629		
Nucleus accumbens	rs1607204	2.02E-10	4.55E-06	1.89E-02	-3.15E-01	ENSG00000143951	ENSG00000143951	-554		
Cerebellum	rs4513466	2.39E-10	4.55E-06	-2.03E-02	4.24E-01	ENSG00000175161	ENSG00000239519	-382		
Cerebellar Hemisphere	rs4513466	2.39E-10	4.55E-06	-2.03E-02	5.28E-01	ENSG00000175161	ENSG00000239519	-382		
Nucleus accumbens	rs4513466	2.39E-10	4.55E-06	-2.03E-02	5.25E-01	ENSG00000175161	ENSG00000239519	-382		
Former vs. current smoker										
Whole Blood	rs56113850	2.52E-26	7.40E-20	2.06E-02	9.08E-02	ENSG00000255974	ENSG00000269858	48		

Tissue	Alcohol consumption									
	Index eQTL	PVALUE	FDR	GWAS effect	GTEX slope	Proximal gene	eGene	Distance to TSS (Kb) *		
Brain Cerebellum	rs12459249	3.89E-16	1.63E-10	1.67E-02	-4.61E-01	ENSG00000269843	ENSG00000130612	-57		
Brain Cerebellum	rs111697662	9.82E-15	3.60E-09	2.18E-02	5.54E-01	ENSG00000203900	ENSG00000101204	-18		
Whole Blood	rs7937	5.29E-11	1.55E-05	1.38E-02	-2.55E-01	ENSG00000171570	ENSG00000233622	-15		
Whole Blood	rs7937	5.29E-11	1.55E-05	1.38E-02	-2.16E-01	ENSG00000171570	ENSG00000188493	45		

* Distance of index eQTL to transcription start site (TTS) of target gene (eGene).

Overview of novel significant index eQTLs which were not identified by the uninformed GWAS (FDR adj. p-value<0.05).

Table 2.

Trait	N unique novel loci GTEx informed	N unique novel loci DGN informed
CPD	509	214
Alcohol consumption	949	136
Ever smoker	1335	401
Age of initiation	66	25
Former vs. current smoker	117	35
Cocaine dependence	0	0
Total	2,976	811

Note: the number of unique novel loci identified by GTEx informed analyses is based on 14 tissues, for DGN it is based on a single tissue (Whole blood)

Overview of the five strongest GTEx informed S-PrediXcan gene associations per substance use trait, genes printed in bold indicate significance.

Table 3.

Alcohol consumption						
Gene	Gene_name	Zscore	Effect size	Pvalue	Tissue	
ENSG00000035928.10	RFC1	-8.96	-3.37E-02	3.21E-19	Cerebellar Hemisphere	
ENSG00000238083.3	LRR37A2	-8.61	-2.79E-02	7.47E-18	Cerebellum	
ENSG00000262539.1	RP11-259G18.3	-8.56	-1.87E-02	1.08E-17	Cerebellar Hemisphere	
ENSG00000214425.2	LRR37A4P	8.54	1.89E-02	1.36E-17	Cerebellar Hemisphere	
ENSG00000263503.1	RP11-707O23.5	-8.51	-2.35E-02	1.75E-17	Cerebellar Hemisphere	
CPD						
ENSG00000041357.11	PSMA4	30.86	3.39E-01	4.04E-209	Whole Blood	
ENSG00000041357.11	PSMA4	19.65	2.10E-01	6.10E-86	Substantia nigra	
ENSG00000041357.11	PSMA4	13.88	1.09E-01	7.86E-44	Putamen	
ENSG00000169684.9	CHRNA5	-13.46	-4.46E-02	2.55E-41	Substantia nigra	
ENSG00000169684.9	CHRNA5	-11.22	-3.46E-02	3.31E-29	Spine	
Ever smoker						
ENSG00000076685.14	NT5C2	-7.99	-3.50E-02	1.35E-15	Cerebellar Hemisphere	
ENSG00000166275.11	C10orf32	-7.90	-1.81E-02	2.90E-15	Cerebellum	
ENSG00000166275.11	C10orf32	-7.75	-3.02E-02	9.42E-15	Whole Blood	
ENSG00000235266.1	RP11-753C18.8	-7.73	-1.97E-02	1.09E-14	Cerebellar Hemisphere	
ENSG00000117385.11	LEPRE1	7.68	4.13E-02	1.63E-14	Frontal Cortex	
Age of initiation						
ENSG00000120903.6	CHRNA2	5.71	1.61E-02	1.10E-08	Cerebellum	
ENSG00000120903.6	CHRNA2	5.44	5.12E-02	5.42E-08	Caudate	
ENSG00000271643.1	RP11-10C24.3	-4.89	-6.77E-02	1.01E-06	Substantia nigra	
ENSG00000197386.6	HTT	-4.88	-8.73E-02	1.08E-06	Whole Blood	
ENSG00000120903.6	CHRNA2	4.84	1.57E-02	1.28E-06	Cerebellum	
Former vs. current smoker						
ENSG00000041357.11	PSMA4	6.16	5.10E-02	7.29E-10	Whole Blood	
ENSG00000233622.1	CYP2T2P	5.35	3.97E-02	8.98E-08	Whole Blood	

Alcohol consumption						
Gene	Gene_name	Zscore	Effect size	Pvalue	Tissue	
ENSG00000232630.1	PRPS1P2	-5.10	-3.44E-02	3.46E-07	Cerebellum	
ENSG00000269858.1	EGLN2	-5.02	-1.51E-02	4.96E-07	Whole Blood	
ENSG00000126215.9	XRCC3	5.03	2.31E-02	5.10E-07	Whole Blood	
Cocaine dependence						
ENSG00000257941.1	RP11-290L1.4	3.77	1.73	1.65E-04	Frontal Cortex	
ENSG00000152056.12	AP1S3	3.63	2.44	2.81E-04	Whole_Blood	
ENSG00000271179.1	RP11-629P16.1	-3.62	-1.29	2.95E-04	Nucleus accumbens	
ENSG00000164344.11	KLKB1	-3.60	-4.01E-04	3.14E-04	Cerebellum	
ENSG00000169291.5	SHE	-3.57	-1.46	3.626E-02	Whole Blood	

The Bonferroni corrected threshold for transcriptome-wide significance (adjusted for all tissues and genes per trait) is 9.7×10^{-7} for the GTEx analyses. Genquant gene name (Gene); Gene name (Gene name); S-PrediXcan association result for the gene (Z-score); S-PrediXcan association effect size for the gene (Effect size); p-value for the association statistic (P-value); number of SNPs from GWAS that were used (Number of SNPs); tissue in which the S-PrediXcan association result was found (Tissue).

Number of gene discoveries from PrediXcan analyses and an overview of the number of significant associations in brain tissues, and whole blood tissues.

Table 4.

Tissue ^a	Alcohol consumption	CPD	Ever vs. never smoker	Age of smoking onset	Former vs. current smoker	Cocaine dependence
	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
Amygdala	18	4	6	0	0	0
Anterior cingulate cortex	21	6	11	0	0	0
Caudate	25	8	11	1	0	0
Cerebellar hemisphere	29	7	18	0	0	0
Cerebellum	34	8	27	1	1	0
Cortex	39	6	29	0	0	0
Frontal cortex	18	3	17	0	0	0
Hippocampus	15	7	10	0	0	0
Hypothalamus	17	4	10	0	0	0
Nucleus accumbens	21	7	4	0	0	0
Putamen	17	6	10	0	0	0
Spine	12	4	7	0	0	0
Substantia nigra	9	6	5	0	0	0
Whole blood	27	14	21	0	5	0
<i>n</i> unique genes 'total'	71	35	78	1	6	0
<i>n</i> unique genes brain 'total'	61	27	67	1	1	0
<i>n</i> unique genes whole blood 'only'	10	8	11	0	5	0
<i>n</i> unique genes brain 'only'	44	21	57	1	1	0

^a *n* unique genes 'total' = the number of unique genes that are significant in one or more tissues; *n* unique genes; *n* genes 'brain' = the number of genes that are significant in one or more brain tissues; *n* genes 'brain only' = the number of genes that are significant in one or more brain tissues and not significant whole blood; *n* unique genes 'whole blood' only, the number of genes that are significant in whole blood, but not significant in any of the brain.