

Multi-omics integration analysis identifies novel genes for alcoholism with potential link to neurodegenerative diseases

Manav Kapoor^{1}, Michael Chao¹, Emma C. Johnson², Gloriia Novikova¹, Dongbing Lai³, Jacquelyn Meyers⁵, Jessica Schulman¹, John I Nurnberger Jr⁴, Bernice Porjesz⁵, Yunlong Liu³, COGA collaborators, Tatiana Foroud^{3,4}, Howard J. Edenberg^{3,6}, Edoardo Marcora¹, Arpana Agrawal² and Alison Goate¹*

¹Dept. of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY

²Dept. of Psychiatry, Washington University School of Medicine, St. Louis, MO

³Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN

⁴Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN

⁵Department of Psychiatry, State University of New York, Downstate Medical Center, Brooklyn, NY, USA

⁶Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN

***Corresponding Author**

Manav Kapoor, PhD
Assistant Professor
Dept. of Neuroscience
Dept. of Genetics and Genomic Sciences
Icahn School of Medicine at Mount Sinai
1425 Madison Avenue, Room 10-75B
Box #1065
New York, NY 10029

Tel: (212) 659-5967
email: manav.kapoor@mssm.edu

Abstract

Significance: Identification of causal variants and genes underlying genome-wide association study (GWAS) loci is essential to understanding the biology of alcohol use disorder (AUD).

Methods: Integration of “multi-omics” data is often necessary to nominate candidate causal variants and genes and prioritize them for follow up studies. Here, we used Mendelian randomization to integrate AUD and drinks per week (DPW) GWAS summary statistics with the gene expression and methylation quantitative trait loci (eQTLs and mQTLs) in the largest brain and myeloid datasets. We also used AUD-related single cell epigenetic data to nominate candidate causal variants and genes associated with DPW and AUD.

Results: Our multi-omics integration analyses prioritized unique as well as shared genes and pathways among AUD and DPW. The GWAS variants associated with both AUD and DPW showed significant enrichment in the promoter regions of fetal and adult brains. The integration of GWAS SNPs with mQTLs from fetal brain prioritized variants on chromosome 11 in both AUD and DPW GWASs. The co-localized variants were found to be overlapping with promoter marks for *SPII*, specifically in human microglia, the myeloid cells of the brain. The co-localized SNPs were also strongly associated with *SPII* mRNA expression in myeloid cells from peripheral blood. The prioritized variant at this locus is predicted to alter the binding site for a transcription factor, RXRA, a key player in the regulation of myeloid cell function. Our analysis also identified *MAPT* as a candidate causal gene specifically associated with DPW. mRNA expression of *MAPT* was also correlated with daily amounts of alcohol intake in post-mortem brains (frontal cortex) from alcoholics and controls (N = 92). Results may be queried and visualized in an online public resource of these integrative analysis (https://lcad.shinyapps.io/alc_multomics/). These results highlight overlap between causal genes for neurodegenerative diseases, alcohol use disorder and alcohol consumption.

In conclusion, integrating GWAS summary statistics with multi-omics datasets from multiple sources identified biological similarities and differences between typical alcohol intake and disordered drinking highlighting molecular heterogeneity that might inform future targeted functional and cross-species studies. Interestingly, overlap was also observed with causal genes for neurodegenerative diseases.

Introduction:

Alcohol use disorders (AUD) are complex, moderately heritable (50-60%) ¹⁻⁴, psychiatric disorders associated with heightened morbidity and mortality⁵. An AUD diagnosis includes aspects of physiological dependence, loss of control over drinking, as well as persistent alcohol intake despite physiological, psychological and interpersonal consequences^{6,7}. In contrast, typical alcohol intake, as assessed using measures such as drinks per week, represents the distribution of alcohol use from casual or social drinking to excessive drinking demarcating risk for AUD ^{8,9}. While heritable, measures such as drinks per week (DPW) are more likely to be influenced by environmental and socio-cultural factors and have complex and variable associations with morbidity and mortality ^{8,9}.

Genome-wide association studies (GWASs) of AUD and DPW have identified multiple risk loci. The largest GWAS of problematic alcohol use (PAU; N=435,563) which meta-analyzed AUD with a GWAS of the problem-subscale of the Alcohol Use Disorders Identification Test (AUDIT-P) reported genome-wide associations at 29 loci encompassing 66 genes, the largest tranche of signals for any addictive disorder to date ¹⁰. Alongside, the largest GWAS of typical alcohol intake (N=941,280) identified more than 200 independent genome-wide significant variants within or near more than 150 genes at 81 independent loci ¹¹. Despite a genetic correlation (SNP-rg) of 0.77 between PAU and DPW (less so for AUD and DPW, SNP-rg=0.67), genetic correlations between these aspects of alcohol involvement and other anthropometric, cardio-metabolic and psychiatric disorders revealed marked distinctions ¹⁰⁻¹⁴. For instance, while AUD and PAU appear to be consistently associated with increased genetic liability for other psychiatric disorders and negatively with liability to educational achievement, DPW is genetically uncorrelated with most psychiatric disorders (except ADHD and tobacco use disorder) but correlated negatively with educational achievement and cardio-metabolic disease (which remains uncorrelated with PAU or AUD) ¹⁰⁻¹⁴. These findings strongly hint at some common pathological underpinnings to AUD, PAU and other mental illnesses while genetic liability to DPW appears to be confounded with socio-economic correlates of alcohol use ^{10,11,13,14}.

Few studies have examined the intersection between the loci and genes associated with AUD and DPW, especially with respect to their functional and regulatory significance. As observed in other large GWAS, most genome-wide significant variants associated with AUD and DPW are intergenic and thus not directly mappable to a specific gene ^{15,16}. Furthermore, positionally mapping a non-coding variant to the nearest gene often does not identify the causal gene(s) ¹⁵⁻¹⁷. Indeed, most variants identified by GWAS reside within and affect the activity of regulatory elements (e.g., enhancers and promoters) that regulate the expression of target causal genes in specific cell types; the affected genes are often located at quite a distance from the risk variant/regulatory element ^{13,15,16,18}. Several recent studies have integrated GWAS data with expression

QTLs (eQTLs) using colocalization or integration methods to identify causal variants and genes associated with schizophrenia, Alzheimer's disease and many other complex disorders^{13,15,16,18}. While similar efforts have been targeted at AUD and DPW, they have predominantly relied on bulk mRNA expression data from the small number of brain tissue samples in GTEx¹⁰⁻¹⁴.

In the current study, we used a multi-omics systems approach to identify causal variants and genes associated with AUD and DPW. Using Mendelian Randomization-based methods on the largest available transcriptomic and epigenomic data for brain tissues and myeloid cells, we prioritized regulatory variants that influence AUD and DPW, specifically as well as simultaneously. We also used mRNA expression data from the brains of individuals diagnosed with AUD and controls (N=138) to validate the differential expression of genes prioritized in the GWAS integration analyses. To our knowledge, this is the first and the largest systematic multi-omics integration analysis to identify the functional impact of variants and genes associated with two correlated but etiologically distinct aspects of alcohol involvement.

Results:

AUD meta-analysis

The large meta-analysis of AUD GWAS summary statistics (N=48,545 AUD cases and 187,065 controls) from the Million Veterans Program (MVP)¹⁹, the Psychiatric Genetics consortium (PGC-SUD)¹² and the Collaborative Studies on Genetics of Alcoholism (COGA)²⁰ identified 1157 SNPs (31 independent lead SNPs) within or near 79 genes at 10 independent loci associated with AUD. We didn't include UKB-AUDIT-P in this meta-analysis to specifically focus on AUD. Many of these loci were shared between the AUD GWAS and the drinks per week (DPW) GWAS by Liu et al¹¹ who identified 81 independent loci represented by 5197 (> 200 independent lead) SNPs. A total of 360 SNPs associated with AUD and DPW were in common (i.e., $p < 5 \times 10^{-8}$ in both GWAS). A large proportion (45%) of AUD and DPW associated SNPs were within intronic, UTR and non-coding regions of the genome.

LDSC analysis using tissue specific epigenetic annotations

We used the stratified linkage disequilibrium score (LDSC) regression²¹ to test whether the heritability of AUD and DPW is enriched in regulatory regions surrounding genes in a specific tissue. Using multi-tissue chromatin (ROADMAP and ENTEX) data²², we observed a significant enrichment of promoter-specific epigenetic markers (H3K4me1/me3) in the fetal and the adult (germinal matrix, frontal-cortex) brain ($P < 5 \times 10^{-8}$) (Figure 2; Supplementary table 1) for the SNPs associated with AUD and DPW respectively.

Integration of GWAS and eQTL/ mQTL data from fetal and adult brain

Summary based Mendelian Randomization (SMR) analysis of genome-wide AUD summary statistics with eQTLs and mQTLs in the adult and fetal brain identified 21 genes at 18 loci across the genome (P-SMR FDR < 20%; Heidi > 0.05) [Supplementary Table 3a]. Among these 18 loci, SMR analysis nominated a single candidate causal gene at 16 loci, while more than 1 causal gene was nominated at 3p21.31 (*GPX1*, *AMT*) and 11p11.2 (*SPII*, *MTCH2*, *NUP160*). To avoid the occurrence of false positive co-localizations that might be exclusively driven by stronger eQTL/mQTL signals, we focused on the loci where the strongest SNP was at least suggestively significant in the GWAS (P-GWAS < 5×10^{-5}) [Table 1a; Figure 1]. Because of the much larger sample size of the DPW GWAS, 61 genes at nearly 31 loci passed the threshold for significance (P-SMR FDR < 20%; Heidi > 0.05; GWAS P < 5×10^{-5}) [Table 1b; Figure 1]. On chromosome 11p11.2, our SMR based integration analysis co-localized a fetal brain specific mQTL (*SPII*) and an adult brain specific eQTL (*NUP160*) with both traits (AUD and DPW). On chromosome 17q.21.31, the integration analysis prioritized different candidate genes for AUD (*MAP3K14*) and DPW (*MAPT*, *CRHR1* and *LRRC37A*). Indeed, the AUD and DPW associations at 11p11.2 are likely to be two distinct loci, because the lead co-localized SNPs for each phenotype were not in LD ($r^2 = 0.2$). The DPW association tagged the H2 haplotype at 17q.21.31, while AUD's association with *MAP3K14* was outside the inversion area (Supplementary figure), defined by the H1/H2 haplotypes in this region.

Fine mapping of 17q.21.31. At 17q.21.31, eQTL and/ or mQTL from both fetal and adult brains co-localized with DPW signals. We observed stronger evidence of co-localization (SMR P < 5×10^{-15}) for DPW with *MAPT* and *LRRC37A*, than at any other locus. These genes are within a large inversion polymorphism (approximately 900 kb) that arose about 3 million years ago²³. Since that time, these haplotypes have been recombinationally suppressed and have accumulated many haplotype-specific variants. As a result, there is extended LD within more than 1 Mb, which makes it difficult to fine map the causal variants and genes at this locus. Integration analysis of adult brain eQTL data with the DPW GWAS predicted that increased *MAPT* expression (SMR Beta = 0.01) is associated with increased number of alcoholic drinks per week, while decreased expression of *LRRC37A* (SMR beta = -0.02) was associated with a decrease in drinks per week. The predicted gene expression results from AUD and DPW GWAS were compared with observed expression differences in the brains of AUD subjects and controls to validate the results. Our differential expression analysis of alcohol consumption in the human brain indeed showed that the mRNA expression of *MAPT* was associated with increased alcohol consumption (Figure 2c). We did not observe any association between expression of *LRRC37A* and level of alcohol intake. The co-localized SNPs within the 17q.21.31 locus were also compared with the promoter (H3K27ac, H3K4me3), enhancer (ATAC-Seq) and promoter-enhancer interactome (PLAC-Seq) data from 4 specific brain cell- types (microglia, neuron, astrocytes and oligodendrocytes) to elucidate the functional significance of these variants (Supplementary table). The co-localized mQTLs (rs3785884 and rs17651887) overlapped with the chromatin interaction region specifically in oligodendrocytes and these interactions looped at the *MAPT* promoter (Figure 2b). This observation combined with differential expression data in

the human brain provides strong supporting evidence that *MAPT* is likely to be the causal gene at this locus associated with increased alcohol consumption.

Fine mapping at 11p11.2

SMR analysis with mQTLs from fetal brain and eQTLs from adult brain prioritized *SPII* and *NUP160* respectively at 11p11.2 for association with DPW and AUD (Figure 3A). Both *SPII* and *NUP160* are primarily expressed in myeloid cells. The predicted causal *SPII* mQTL (rs56030824) and *NUP160* eQTL (rs10838753) were in fact in low LD ($R^2 = 0.31$) with each other. rs56030824 (mQTL) showed a stronger association with both AUD ($P = 8.91 \times 10^{-6}$) and DPW (4.90×10^{-12}) than rs10838753 (eQTL) (DPW $P = 1.28 \times 10^{-10}$; AUD $p = 4.85 \times 10^{-5}$). Adding rs56030824 as a covariate in conditional analyses had a larger effect on the association between rs10838753 and AUD ($P_{\text{orig}} = 4.85 \times 10^{-5}$; $P_{\text{cond}} = 0.09$) than with DPW ($P_{\text{orig}} = 1.28 \times 10^{-10}$; $P_{\text{cond}} = 8.1 \times 10^{-3}$). rs56030824 remained significantly associated with both AUD ($P_{\text{orig}} = 8.91 \times 10^{-6}$; $P_{\text{cond}} = 2.0 \times 10^{-2}$) and DPW ($P_{\text{orig}} = 4.90 \times 10^{-12}$; $P_{\text{cond}} = 7.0 \times 10^{-4}$) even after adding rs10838753 as a covariate. Rs56030824 overlapped the promoter marks (H3K27ac, H3K4me3) for *SPII* specifically in microglia (Figure 3A). This SNP also alters the binding site regulatory motif for RXRA, a transcription factor which is involved in promotion of myelin debris phagocytosis and remyelination by macrophages ²⁴. Since *SPII* is expressed in myeloid lineage cells, its mRNA expression in the bulk brain was too low to perform differential expression or integration analysis. Therefore, we chose eQTLs from a large sample of peripheral blood monocytes to examine if rs56030824 is associated with expression of *SPII* in these cells. The effect sizes of eQTLs at 11p11.2 locus were linearly correlated with effect sizes from the DPW GWAS at this locus (Figure 3B and 3C). In fact, rs56030824 had the strongest effect size for *SPII* expression and DPW in the common variant category (Figure 3D). These observations together established rs56030824 as a stronger candidate to be considered as a causal variant and *SPII* as a potential candidate gene associated with AUD and DPW.

This study identified several other genes for DPW with multiple lines of evidence (eQTL, mQTL, differential expression; FDR <20%; HEIDI $P > 0.05$; GWAS $P < 5 \times 10^{-5}$). For example, at locus 16p11.2, *SULT1A1* and *SULT1A2* were the strongest candidates with co-localization evidence emerging from mQTL and eQTLs from adult brain tissue (Table 1b). On chromosome 19, *FUT2* was the strongest candidate; mRNA expression of *FUT2* was also associated with increased alcohol consumption ($\text{Beta} = 0.09$, $P = 4.6 \times 10^{-2}$) when comparing the DLPFC of individuals with AUD and control subjects (Table 1b).

Pathway and network analysis

Ingenuity Pathway Analysis of the prioritized genes associated with DPW showed significant enrichment for pathways related to TR (Thyroid hormone receptor)/RXR (Retinoic X receptor) activation, Lipoate biosynthesis, Estrogen biosynthesis and Sirutuin signaling (Figure 6, supplementary table 4). The DPW associated genes were also part of networks associated with

immune cell trafficking and cellular movements (cell migration). Due to insufficient power and a smaller number of genes passing threshold of significance, we were not able to perform the pathway enrichment analysis for AUD.

Discussion:

In this study, we used a multi-omics integration approach to detect genes relevant to typical drinking (DPW) and alcohol use disorder (AUD). The AUD GWAS used in the current analysis specifically focused on the diagnosis rather than the disordered drinking. Importantly, our work highlights that GWAS variants for alcohol use disorder and drinks per week are enriched in promoter regions of the fetal and adult brain. Using large-scale transcriptomic and epigenomic data from these tissues, we successfully fine mapped complex loci (*17q21.31*, *11p11.2*, *16p11.2*) and identified likely functional variants and candidate causal genes associated with alcoholism. Prior transcriptomic data from human and animal brains highlighted the contribution of immune networks in drinking behaviors ²⁵⁻³¹. But these observations were never consistent with results from GWAS of AUD and DPW, most likely due to lack of power in these genomic studies. Transcriptomic changes can be either a cause or a consequence of chronic excessive alcohol consumption. The identification of genes and/ or pathways involved in immune signaling (*SPII*, *RXR activation*), lipid metabolism (RXR and sulfotransferases) and regulation of alcohol metabolism (Sirutuin signaling) are therefore important as an attempt to fill gaps in our understanding of disease predisposition and underlying biological mechanisms in a genomic context.

For example, our LDSC based enrichment analysis shows that GWAS variants for both AUD and DPW are enriched in the genes expressed during early brain development. Drinking in later years might interact with this genetic predisposition making individuals more susceptible not only to AUD but also to other neuropsychiatric disorders³². Identification of *SPII* and *MAPT* as genes for AUD are good examples of pleiotropy and/ or causal links between the alcohol intake and susceptibility to AUD, other psychiatric disorders (e.g., depression) and even Alzheimer's disease^{15,33} and other neurodegenerative diseases. We found that increased *SPII* expression in myeloid lineage cells was associated with a higher DPW and higher risk for AUD. Recently, Zhang and colleagues³³ observed that protein expression levels of *SPII* in the cerebellum and spleen from subjects with Major depressive disorder and schizophrenia were significantly higher than in controls. In the past, we have also demonstrated that functional variants related to *SPII* expression are associated with risk of Alzheimer's disease ¹⁵. Similarly, higher levels of expression are associated with increased risk for Alzheimer's disease.

SPII (Spi-1 Proto-Oncogene) encodes an ETS-domain transcription factor (PU.1) that regulates gene expression during myeloid and B-lymphoid cell development and homeostasis. This nuclear protein binds to a purine-rich sequence known as the PU-box found near the promoters of target genes and, in coordination with other transcription factors and cofactors, regulates their

expression; among the genes are LXR/RXR nuclear receptors³⁴. In the brain, *SPII* is specifically expressed in microglia¹⁵. Given *SPII*'s control over expression of several downstream genes, this gene may be a major reason enrichment of immune pathways is observed in transcriptomic analysis of human and animal brains. Because of the small fraction of microglia in bulk brain tissue, it is difficult to study expression of this transcription factor in transcriptomic datasets from whole brains. Some studies have reported that chronic alcohol consumption can influence expression of PU.1 in isolated microglia³⁵ and peripheral lung macrophages^{36,37}. However, these studies report the consequences of drinking on PU.1 expression whereas our study uses genomic evidence to demonstrate that regulation of innate immune response likely underlies, at least in part, susceptibility to increased drinking and eventual risk for AUD¹.

MAPT is another example of a pleiotropic relationship between AUD and other neuropsychiatric and neurodegenerative disorders. Located on chromosome 17, *MAPT*, encodes the tau proteins best known medically for their role in central nervous system disorders such as Alzheimer's disease³⁸, frontotemporal dementia³⁹, Parkinson's disease³⁸, and the primary tauopathies progressive supranuclear palsy and corticobasal degeneration⁴⁰. Recently, Hoffman and colleagues⁴¹ showed that alcohol use can upregulate the expression of pTau (Ser199/Ser202) in the hippocampus of C57BL/6J mice. Another study in humans observed differences in CSF-Tau levels in demented alcoholics vs Alzheimer disease patients⁴². *CRHR1* (corticotropin-releasing hormone type I receptor) is another gene on 17q.21.31, that has been reported to be associated with alcoholism⁴³. However, in our analysis, we did not observe an association between *CRHR1* expression and alcohol consumption

We also identified other genes that might be involved in increased alcohol consumption through a variety of biological mechanisms. For example, *VPS4A* at 16q23.1 has been implicated in dopamine regulation, reward anticipation, and hyperactivity in an fMRI study⁴⁴. We also identified functional variants for *SULT1A1* and *SULT1A2* genes that encode for Sulfotransferase Family 1A enzymes catalyzing the sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds⁴⁵. In IPA disease enrichment analysis we also observed a significant overlap between genes implicated in DPW with other neurological, behavioral and immune related disorders (Supplementary Fig 1). The genes associated with DPW also showed significant enrichment for pathways related to TR/ RXR activation, Lipoate biosynthesis, Estrogen biosynthesis and Sirtuin signaling (Figure 6, Supplementary Table). TRs (Thyroid hormone receptor) control the expression of target genes involved in diverse physiological processes and diseases, such as metabolic syndrome, obesity, and cancer, and, therefore, are considered as important targets for therapeutic drug development⁴⁶. RXRs (Retinoic X Receptor) are also known to potentially regulate the ethanol metabolizing enzymes after chronic alcohol consumption⁴⁷. It has been reported that the human aldehyde dehydrogenase-2 (*ALDH2*) promoter contains a retinoid response element, which might be contributing to the regulation of the gene⁴⁷. Sirtuins signaling has been shown to play an

important role in cocaine and morphine Action in the Nucleus Accumbens. Ferguson and colleagues⁴⁸ demonstrated that systemic administration of a nonselective pharmacological activator of all sirtuins can increase the cocaine reward.

In conclusion, our study prioritizes risk variants and genes for subsequent experimental follow-up, which will help interrogate the molecular mechanisms underlying the link between alcohol consumption and AUD. Our database of multi-omics analysis in the fetal and adult brain is also made available with this study (see URLs) and provides a starting point to elucidate the biological mechanisms underlying AUD. We have demonstrated that individuals susceptible to AUD may have altered expression of disease-causing genes at earlier stages of life. Moreover, our results show the pleiotropic role of AUD-related variants in a variety of other brain disorders including Alzheimer's disease. We expect results of multi-omic integration analysis will help researchers to design genetically informed experiments to identify biological mechanisms and drug targets related to AUD.

Materials and methods:

Samples

Alcohol Use Disorder

We meta-analyzed three published GWAS: the Million Veteran Program (MVP)¹⁹ GWAS of AUD (EUR N = 202,004; Ncases = 34,658), with case status derived from International Classification of Diseases (ICD) codes of alcohol-related diagnoses from electronic health records (EHR) data, the Psychiatric Genomics Consortium (PGC) GWAS of alcohol dependence¹² (cases based on DSM-IV diagnoses; EUR unrelated genotyped N = 28,757; Ncases = 8,485; AFR N = 5,799; Ncases = 2,991) and the Collaborative Studies on Genetics of Alcoholism (COGA) GWAS of alcohol dependence (cases based on DSM-IV diagnoses; EUR unrelated genotyped N = 4,849; Ncases = 2,411)²⁰.

Drinks per week

We used genome-wide summary statistics for drinks per week (DPW; EUR N = 537,349 without the 23andMe samples) from GSCAN¹¹ to contrast with AUD.

eQTLs from adult brain

We meta-analyzed three eQTL datasets with data from the dorsolateral prefrontal cortex (DLFPC): PsychEncode (N = 1387)⁴⁹, ROSMAP (N = 461)⁵⁰ and COGA-INIA (N = 138). We genotyped brain samples acquired from the New South Wales Brain Bank (NSWBB) using the UK Biobank Axiom array as part of COGA-INIA collaboration.

mQTLs from adult brain

Brain-mMeta mQTL summary data (Qi et al. 2018 Nat Commun)⁵¹ in SMR binary (BESD) format were obtained from the SMR data resource. This is a set of mQTL data from a meta-analysis of ROSMAP⁵², Hannon et al.⁵³ and Jaffe et al.⁵⁴.

eQTLs from fetal brain

Summary data for eQTLs from developing human brains were obtained from an online repository shared by Heath O'Brien and Nicholas J. Bray⁵⁵. The analyses were performed on 120 human fetal brains from the second trimester of gestation (12 to 19 post-conception weeks).

mQTLs from fetal brain

Summary data for mQTLs from developing human brains were obtained from an online repository shared by Ellis Hannon and Jonathan Mill⁵³. The mQTLs were characterized in a large collection (n=166) of human fetal brain samples spanning 56–166 days post-conception, identifying >16,000 fetal brain mQTLs.

eQTL data from CD14+ monocytes

We used the gene expression and genotype data generated on primary monocytes from 432 healthy Europeans to quantify the relationship between the co-localized SNPs and expression of myeloid lineage gene⁵⁶.

Whole genome transcriptomic data in brain of alcoholics

mRNA expression data in the DLFPC region of the human brain was generated in 138 brains obtained from the New South Wales Brain Bank (NSWBB). We also had access to alcohol consumption (gm/day) data in a subset of 92 brains. Alcohol-dependence diagnoses and consumption data were collected by physician interviews, review of hospital medical records, questionnaires to next-of-kin, and from pathology, radiology, and neuropsychology reports.

Brain cell type specific enhancer and promoter data

We used the promoter (H3K27ac, H3K4me3), enhancer (ATAC-Seq) and promoter-enhancer interactome (PLAC-Seq) data for 4 specific cell types of brain (microglia, neuron, astrocytes and oligodendrocytes) to elucidate the functional significance of colocalized SNPs⁵⁷.

Analysis

eQTL meta-analysis in adult brain

RNA Sequencing data on DLFPC region of the brain for 138 samples was generated as part of COGA-INIA collaboration²⁸. We also genotyped the brain samples using the UK Biobank Axiom array. All NSWBB samples were imputed to 1000 Genomes using the cosmopolitan reference panel (Phase 3, version 5, NCBI GRCh37) using SHAPEIT then Impute2⁵⁸ within each array. Only variants with non-A/T or C/G alleles, missing rates <5%, MAF >5%, and HWE P values >.0001 were used for imputation. Imputed variants with R2 < .30 were excluded, and genotype probabilities were converted to genotypes if probabilities $\geq .90$. All genotyped and imputed variants (4,615,871 SNPs) with missing rates <10%, MAF $\geq 5\%$ and HWE P values $> 1 \times 10^{-6}$ were included in the downstream analyses using MatrixQTL program. The gene expression was corrected for the batch, age, sex, rin, PMI and alcohol status using the “removeBatchEffect” option from limma package. The eQTL summary statistics from all three datasets were processed and munged together at single bp and allele level to remove ambiguity due to dbSNP rsids. The gene labels in all three datasets were also matched to Ensembl ids. The summary statistics were saved in binary format files (BESD) using the SMR. SMR “--meta” option was used to perform the meta-analysis in all three datasets.

LDSC analysis

We performed the partition heritability analyses for functional annotation using LDSC program. We obtained the weights for the multi-cell and tissue chromatin marks and performed the LDSC partition heritability analyses on munged summary statistics of AUD and DPW GWAS²¹.

SMR analysis

To examine whether the GWAS variants associated with both AUD and DPW are mediated by changes in methylation and gene expression patterns, we conducted a summary data-based Mendelian randomization (SMR) analysis⁵⁹ on a set of mQTLs and eQTLs from fetal and adult brains. SMR is a Mendelian randomization-based analysis which integrates GWAS summary statistics with eQTL data in order to test whether the effect size of a SNP on the phenotype of interest is mediated by gene expression. We used this method as it does not require raw eQTL data to build the weights, so we were able to use the meta-analysis of eQTL data for the integration analysis. The gene and SNP positions for the summary statistics of the eQTL and mQTL datasets were standardized and aligned using an in-house summary statistics munging pipeline. The summary statistics were then converted into binary format (BESD) to perform the SMR analysis. The European subset of ADGC GWAS (phs000372.v1.p1) was used as the LD reference panel to perform the SMR analyses. The genes below FDR threshold (FDR < 20%) and with heterogeneity P value > 0.05 were considered to be causal within each combination of analysis.

Conditional and joint (COJO) analysis

We used a summary data based conditional analysis approach to identify the independent lead SNP associated with AUD and DPW. This conditional and Joint analysis (COJO)⁶⁰ approach is implemented in Genome-wide Complex Trait Analysis (GCTA) software⁶¹ package and is valuable when the individual level genotype data is not available for the conditional analysis. To perform the COJO analysis we used the summary statistics of AUD and DPW GWAS along with the European subset of COGA samples as the LD reference panel.

Differential Expression analysis

We first performed a linear regression with alcohol intake as a dependent variable to identify possible covariates (e.g. sex, age, PMI). Gene-level analyses started with the featureCounts-derived sample-by-gene read count matrix. The basic normalization and adjustment pipeline for the expression data matrix consisted of: (i) removal of low expression genes (<1 CPM in > 50% of the individuals); (ii) differential gene expression analysis based upon adjustment for the chosen covariates. We filtered out all genes with lower expression in a substantial fraction of the cohort, with 18,463 genes with at least 1 CPM in at least 50% of the individuals; note that only these genes were carried forward in all subsequent analyses. The log10 normalized alcohol consumption (from NSWBB brains) was used for differential expression analysis using the DeSeq2 program. The analysis was controlled for sex, age, PMI, BMI, RIN, batch and severity of alcoholism (AUDIT scores).

Pathway analysis

The results of integration analysis for the DPW and AUD GWAS (FDR < 20%, Heidi P > 0.05) were used to perform gene ontology and pathway enrichment analyses using the EnrichR and Ingenuity Pathway Analysis (IPA).

Database for query and visualization

We used ShinyApp to create a database for query and visualization of the results of integration analyses. Users can create volcano, Manhattan plots and heatmaps to visualize the results of eQTL, mQTL and epigenetic integration analyses with summary statistics of AUD and DPW GWAS. Users will also be able to see whether the genes of interest are differentially expressed in the brains of alcoholics and controls.

Data availability

Results of the SMR analysis for all the conditions can be found in Supplementary Tables 5a -5h. Additionally all the results can be visualized at our Shiny web app (https://lcad.shinyapps.io/alc_multiomics/).

Code availability

Standard tools (LDSC, SMR, GCTA-COJO, UCSC browser) were used to perform all the integrative analysis reported in this manuscript. Pipelines used in the analysis can be accessed at the GitHub repository (https://github.com/kapoormanav/alc_multiomics).

Table 1a: SMR analysis results with summary statistics of AUD GWAS

Chr	BP	Gene	Adult brain		Fetal brain	
			eQTL p value	mQTL p value	eQTL p value	mQTL p value
11	46399942	<i>SPI1</i>	x	x	x	1.91E-04
	46664086	<i>MTCH2</i>	1.89E-05	x	x	x
	46843734	<i>NUP160</i>	3.88E-04	x	x	x
3	48395716	<i>GPX1</i>	4.93E-05	x	x	x
	48459884	<i>AMT</i>	2.07E-04	x	4.39E-04	4.47E-01
17	43361331	<i>MAP3K14</i>	x	x	x	2.99E-05

Column names: Chr (Chromosome), BP (Start position of the gene), Gene (Candidate causal gene), p value (SMR P values for integration of AUD summary statistics with respective eQTL/ mQTL annotation from adult or fetal brain). X denotes the missing values either due to non-significant results or sub-threshold expression or methylation values in the respective tissue.

Table 1b: SMR analysis results with summary statistics of DPW GWAS

Chr	BP	Gene	Adult brain		Fetal brain		DLFPC
			eQTL P value	mQTL P value	eQTL P value	mQTL P value	
17	42971748	<i>MAPT</i>	4.84E-16	1.77E-08	x	1.80E-12	7.64E-03
	43370099	<i>LRRC37A</i>	8.35E-16	3.89E-13	2.18E-10	x	4.72E-01

	42699267	<i>CRHR1</i>	1.07E-08	8.67E-09	x	x	2.98E-01
	42568094	<i>PLEKHM1</i>	x	3.89E-06	x	x	1.01E-01
11	46399942	<i>SPII</i>	x	x	x	1.11E-06	6.96E-01
	46843734	<i>NUP160</i>	1.43E-06	x	x	x	2.47E-01
	46616211	<i>C1QTNF4</i>	5.26E-06	x	x	x	6.77E-01
16	27620241	<i>SULT1A1</i>	2.09E-09	6.75E-06	x	x	5.55E-01
	27607801	<i>SULT1A2</i>	2.90E-09	1.56E-02	x	x	9.65E-01
	29102570	<i>TBX6</i>	1.70E-06	x	x	x	5.59E-01
	27996147	<i>LAT</i>	1.77E-06	9.09E-02	x	x	3.73E-01
	30022548	<i>DOC2A</i>	x	x	x	6.13E-07	1.48E-01
	28915196	<i>ATP2A1</i>	x	x	x	2.57E-06	9.00E-01
	29662188	<i>PRR14</i>	x	1.62E-04	x	x	7.06E-01
22	40956767	<i>CSDC2</i>	4.17E-06	x	x	x	7.17E-01
	40940449	<i>POLR3H</i>	8.53E-06	4.77E-05	x	4.35E-06	2.57E-01
	40697526	<i>ZC3H7B</i>	1.67E-04	2.66E-05	x	x	1.46E-01
19	48199232	<i>FUT2</i>	9.13E-06	x	x	x	4.61E-02
	48401990	<i>TULP2</i>	2.78E-04	x	x	3.75E-04	1.74E-01

	49342101	<i>PLEKHA4</i>	x	6.10E-05	x	1.05E-03	8.14E-02
3	182967438	<i>ECE2</i>	1.27E-05	x	x	x	7.79E-02
8	37034192	<i>LSM1</i>	1.38E-05	1.35E-05	x	x	5.15E-01
	37088861	<i>DDHD2</i>	9.76E-05	x	x	x	5.06E-02
9	129213652	<i>RPL12</i>	1.65E-05	x	x	x	7.02E-01
	127244764	<i>NR5A1</i>	x	x	x	6.94E-06	8.23E-01
	129186661	<i>ZNF79</i>	x	1.04E-04	x	x	9.07E-02
1	155698234	<i>RRNAD1</i>	x	3.07E-05	x	x	4.19E-01
	156722410	<i>HDGF</i>	x	3.78E-05	x	x	7.11E-01
16	68221032	<i>SNTB2</i>	3.79E-05	x	x	x	4.40E-01
	68345259	<i>VPS4A</i>	4.49E-05	x	x	x	6.89E-02
	70611033	<i>TAT</i>	7.38E-05	x	x	x	2.45E-01
	68760409	<i>NQO1</i>	1.95E-04	x	x	x	2.85E-01
	68373333	<i>NIP7</i>	2.40E-04	x	x	x	5.94E-01
	68796209	<i>WWP2</i>	3.11E-04	x	x	x	2.52E-01
	71879545	<i>ATXN1L</i>	x	x	x	2.77E-04	8.57E-01
12	50785101	<i>SLC4A8</i>	3.95E-05	x	x	x	5.62E-01

2	72612886	<i>ALMS1</i>	4.61E-05	x	x	x	6.54E-01
17	6589389	<i>WRAP53</i>	5.58E-05	x	x	x	1.67E-01
	6461609	<i>TNFSF13</i>	5.99E-05	x	x	x	9.99E-01
	6482785	<i>CD68</i>	x	3.04E-04	x	x	6.38E-01
5	132209139	<i>LEAP2</i>	x	7.75E-05	x	5.15E-05	2.12E-01
16	85731409	<i>GINS2</i>	x	1.51E-04	x	4.50E-08	6.05E-01
1	46799469	<i>CMPK1</i>	6.60E-05	1.14E-04	x	x	6.66E-02
8	26630151	<i>CCDC25</i>	1.05E-04	x	x	x	9.60E-02
5	175853679	<i>GRK6</i>	x	x	x	1.15E-04	1.34E-01
1	204819245	<i>PM20D1</i>	3.32E-05	x	x	2.69E-04	3.96E-01
15	51030318	<i>LYSMD2</i>	1.44E-04	x	x	1.95E-04	6.40E-02
	51311398	<i>MAPK6</i>	2.63E-04	x	x	x	7.82E-02
17	16723265	<i>SREBF1</i>	x	1.16E-04	x	2.33E-04	8.86E-01
4	38460508	<i>RPL9</i>	5.43E-05	x	2.95E-04	x	4.75E-01
17	56232535	<i>SKA2</i>	2.47E-04	x	x	x	3.80E-01
4	56844989	<i>NOA1</i>	6.57E-04	x	x	x	6.79E-01
17	1206998	<i>SRR</i>	3.61E-04	6.13E-05	2.80E-03	x	6.28E-01

17	28624429	<i>OMG</i>	2.45E-03	x	x	1.13E-04	4.50E-02
17	65031635	<i>KPNA2</i>	2.90E-03	1.04E-04	x	x	3.40E-01
11	27494710	<i>LGR4</i>	x	9.25E-01	x	5.02E-05	7.25E-01
11	33185376	<i>CSTF3</i>	x	7.94E-02	x	4.96E-06	9.84E-01
2	178160898	<i>NFE2L2</i>	x	2.72E-05	x	1.73E-01	2.48E-01
2	44312696	<i>PPM1B</i>	x	5.25E-05	x	2.04E-01	5.01E-01
22	42090992	<i>A4GALT</i>	3.14E-01	8.26E-05	x	7.43E-05	7.10E-01
1	224997797	<i>EPHX1</i>	3.34E-01	x	x	1.20E-05	7.93E-01

Column names: Chr (Chromosome), BP (Start position of the gene), Gene (Candidate causal gene), p value (SMR P values for integration of DPW summary statistics with respective eQTL/mQTL annotation from adult or fetal brain). X denotes the missing values either due to non-significant results or sub-threshold expression or methylation values in the respective tissue.

Acknowledgements:

COGA: The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, T. Foroud; Scientific Director, A. Agrawal; Translational Director, D. Dick, includes 11 different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, T. Foroud, J. Nurnberger Jr., Y. Liu); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz, J. Meyers, C. Kamarajan, A. Pandey); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, A. Brooks, R. Hart); The Children's Hospital of Philadelphia, University of Pennsylvania (L. Almasy); Virginia Commonwealth University (D. Dick, J. Salvatore); Icahn School of Medicine at Mount Sinai (A. Goate, M. Kapoor, P. Slesinger); and Howard University (D. Scott). Other COGA collaborators include: L. Bauer (University of Connecticut); L. Wetherill, X. Xuei, D. Lai, S. O'Connor, M. Plawecki, S. Lourens (Indiana University); L. Acion (University of Iowa); G. Chan (University of Iowa; University of Connecticut); D.B. Chorlian, J. Zhang, S. Kinreich, G. Pandey (SUNY Downstate); M. Chao (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, S. Saccone (Washington University); F. Aliev, P. Barr (Virginia Commonwealth University); H. Chin and A. Parsian are the NIAAA Staff Collaborators. We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, Raymond Crowe, Wendy Reich, and Robert E. Taylor, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). The authors acknowledge Office of Research Infrastructure of the National Institutes of Health under award numbers S10OD018522 and S10OD026880. Dr. Kapoor's also receives part of his salary from NIAAA funded R21 grant R21AA026388.

Competing Financial Interests Statement

Alison Goate is on the scientific advisory board for Denali Therapeutics and has served as a consultant for AbbVie and Cognition Therapeutics. Alison Goate is also listed as an inventor on Issued U.S. Patent 8,080,371, "Markers for Addiction" covering the use of certain variants in determining the diagnosis, prognosis, and treatment of addiction.

Author Contributions:

M.K. and A.G. planned the study and acquired the data for the analysis. M.K. performed the data analysis, created the visualization tools and wrote the first draft of manuscript. E.M, M.C., E.C.J., G.N., D.L. and J.S. assisted and helped in the data analysis. COGA collaborators

including J.I.N., J.M., B.P., Y.L., T.F., H.J.E., and A.A. provide critical comments and suggestions on the manuscript.

References:

1. Heath, a.C. *et al.* Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychological medicine* **27**, 1381-1396 (1997).
2. Kendler, K.S., Myers, J., Dick, D. & Prescott, C.A. The relationship between genetic influences on alcohol dependence and on patterns of alcohol consumption. *Alcoholism: Clinical and Experimental Research* **34**, 1058-1065 (2010).
3. Kendler, K.S., Neale, M.C., Heath, A.C., Kessler, R.C. & Eaves, L.J. A twin-family study of alcoholism in women. *American Journal of Psychiatry* **151**, 707-715 (1994).
4. Prescott, C.A. & Kendler, K.S. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *Am J Psychiatry* **156**, 34-40 (1999).
5. Spillane, S. *et al.* Trends in Alcohol-Induced Deaths in the United States, 2000-2016. *JAMA Netw Open* **3**, e1921451 (2020).
6. Takahashi, T. *et al.* Comparison of DSM-IV and DSM-5 criteria for alcohol use disorders in VA primary care patients with frequent heavy drinking enrolled in a trial. *Addict Sci Clin Pract* **12**, 17 (2017).
7. Grant, B.F. *et al.* Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry* **72**, 757-66 (2015).
8. Grant, J.D. *et al.* Alcohol consumption indices of genetic risk for alcohol dependence. *Biol Psychiatry* **66**, 795-800 (2009).
9. Agrawal, A. *et al.* Developing a quantitative measure of alcohol consumption for genomic studies on prospective cohorts. *J Stud Alcohol Drugs* **70**, 157-68 (2009).
10. Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat Neurosci* **23**, 809-818 (2020).
11. Liu, M. *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* **51**, 237-244 (2019).
12. Walters, R.K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* **21**, 1656-1669 (2018).
13. Gupta, I., Dandavate, R., Gupta, P., Agarwal, V. & Kapoor, M. Recent Advances in Genetic Studies of Alcohol Use Disorders. *Current Genetic Medicine Reports* **8**, 27-34 (2020).
14. Sanchez-Roige, S. *et al.* Genome-wide association study of alcohol use disorder identification test (AUDIT) scores in 20 328 research participants of European ancestry. *Addict Biol* **24**, 121-131 (2019).
15. Huang, K.-l. *et al.* A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nature Neuroscience* **20**, 1052-1061 (2017).
16. Huckins, L.M. *et al.* Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. *Nat Genet* **51**, 659-674 (2019).

17. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371-375 (2014).
18. Andrews, S.J., Fulton-Howard, B. & Goate, A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol* **19**, 326-335 (2020).
19. Kranzler, H.R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun* **10**, 1499 (2019).
20. Lai, D. *et al.* Genome-wide association studies of alcohol dependence, DSM-IV criterion count and individual criteria. *Genes Brain Behav* **18**, e12579 (2019).
21. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
22. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
23. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat Genet* **37**, 129-37 (2005).
24. Natrajan, M.S. *et al.* Retinoid X receptor activation reverses age-related deficiencies in myelin debris phagocytosis and remyelination. *Brain* **138**, 3581-97 (2015).
25. Crews, F.T. Immune function genes, genetics, and the neurobiology of addiction. *Alcohol Res* **34**, 355-61 (2012).
26. Farris, S.P., Arasappan, D., Hunicke-Smith, S., Harris, R.A. & Mayfield, R.D. Transcriptome organization for chronic alcohol abuse in human brain. *Mol Psychiatry* **20**, 1438-47 (2015).
27. Farris, S.P. & Mayfield, R.D. RNA-Seq reveals novel transcriptional reorganization in human alcoholic brain. *Int Rev Neurobiol* **116**, 275-300 (2014).
28. Kapoor, M. *et al.* Analysis of whole genome-transcriptomic organization in brain to identify genes associated with alcoholism. *Transl Psychiatry* **9**, 89 (2019).
29. McClintick, J.N. *et al.* Stress-response pathways are altered in the hippocampus of chronic alcoholics. *Alcohol* **47**, 505-15 (2013).
30. Oldham, M.C. *et al.* Functional organization of the transcriptome in human brain. *Nat Neurosci* **11**, 1271-82 (2008).
31. Ponomarev, I., Rau, V., Eger, E.I., Harris, R.A. & Fanselow, M.S. Amygdala transcriptome and cellular mechanisms underlying stress-enhanced fear learning in a rat model of posttraumatic stress disorder. *Neuropsychopharmacology* **35**, 1402-11 (2010).
32. Baranger, D.A.A. *et al.* Convergent Evidence for Predispositional Effects of Brain Gray Matter Volume on Alcohol Consumption. *Biol Psychiatry* **87**, 645-655 (2020).
33. Zhang, J., Chang, L., Pu, Y. & Hashimoto, K. Abnormal expression of colony stimulating factor 1 receptor (CSF1R) and transcription factor PU.1 (SPI1) in the spleen from patients with major psychiatric disorders: A role of brain-spleen axis. *J Affect Disord* **272**, 110-115 (2020).
34. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
35. Socodato, R. *et al.* Daily alcohol intake triggers aberrant synaptic pruning leading to synapse loss and anxiety-like behavior. *Sci Signal* **13**(2020).
36. Joshi, P.C. *et al.* Chronic ethanol ingestion in rats decreases granulocyte-macrophage colony-stimulating factor receptor expression and downstream signaling in the alveolar macrophage. *J Immunol* **175**, 6837-45 (2005).

37. Mehta, A.J. *et al.* Zinc supplementation restores PU.1 and Nrf2 nuclear binding in alveolar macrophages and improves redox balance and bacterial clearance in the lungs of alcohol-fed rats. *Alcohol Clin Exp Res* **35**, 1519-28 (2011).
38. Desikan, R.S. *et al.* Genetic overlap between Alzheimer's disease and Parkinson's disease at the MAPT locus. *Mol Psychiatry* **20**, 1588-95 (2015).
39. Rademakers, R., Cruts, M. & van Broeckhoven, C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum Mutat* **24**, 277-95 (2004).
40. VandeVrede, L., Boxer, A.L. & Polydoro, M. Targeting tau: Clinical trials and novel therapeutic approaches. *Neurosci Lett* **731**, 134919 (2020).
41. Hoffman, J.L. *et al.* Alcohol drinking exacerbates neural and behavioral pathology in the 3xTg-AD mouse model of Alzheimer's disease. *Int Rev Neurobiol* **148**, 169-230 (2019).
42. Morikawa, Y. *et al.* Cerebrospinal fluid tau protein levels in demented and nondemented alcoholics. *Alcohol Clin Exp Res* **23**, 575-7 (1999).
43. Ray, L.A. *et al.* The CRHR1 gene, trauma exposure, and alcoholism risk: a test of G x E effects. *Genes Brain Behav* **12**, 361-9 (2013).
44. Jia, T. *et al.* Neural basis of reward anticipation and its genetic determinants. *Proc Natl Acad Sci U S A* **113**, 3879-84 (2016).
45. Nowell, S. & Falany, C.N. Pharmacogenetics of human cytosolic sulfotransferases. *Oncogene* **25**, 1673-8 (2006).
46. Boelen, A., Kwakkel, J. & Fliers, E. Thyroid hormone receptors in health and disease. *Minerva Endocrinol* **37**, 291-304 (2012).
47. Mello, T., Polvani, S. & Galli, A. Peroxisome proliferator-activated receptor and retinoic x receptor in alcoholic liver disease. *PPAR Res* **2009**, 748174 (2009).
48. Ferguson, D. *et al.* Essential role of SIRT1 signaling in the nucleus accumbens in cocaine and morphine action. *J Neurosci* **33**, 16088-98 (2013).
49. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**(2018).
50. De Jager, P.L. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data* **5**, 180142 (2018).
51. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* **9**, 2282 (2018).
52. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci* **20**, 1418-1426 (2017).
53. Hannon, E. *et al.* Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci* **19**, 48-54 (2016).
54. Jaffe, A.E. *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci* **19**, 40-7 (2016).
55. O'Brien, H.E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol* **19**, 194 (2018).
56. Fairfax, B.P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* **343**, 1246949-1246949 (2014).
57. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134-1139 (2019).
58. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).

59. Meng, X.H. *et al.* Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions. *Bone* **113**, 41-48 (2018).
60. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
61. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).

Fig 1: Overview of the study

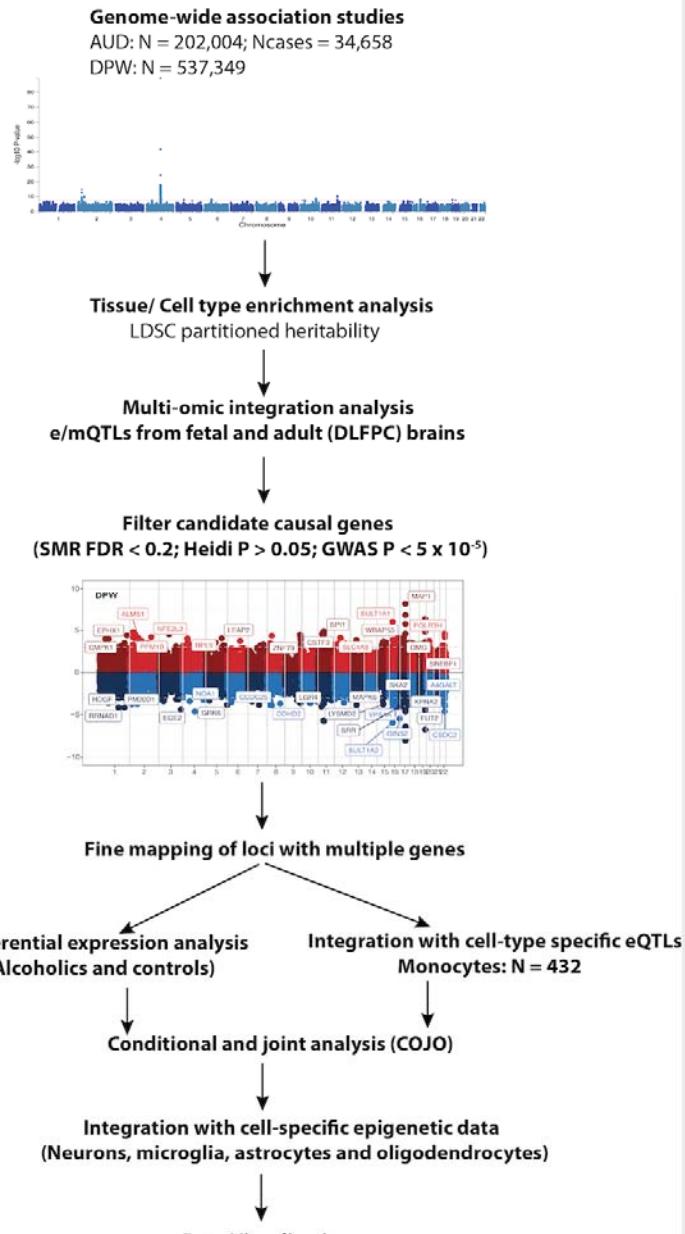
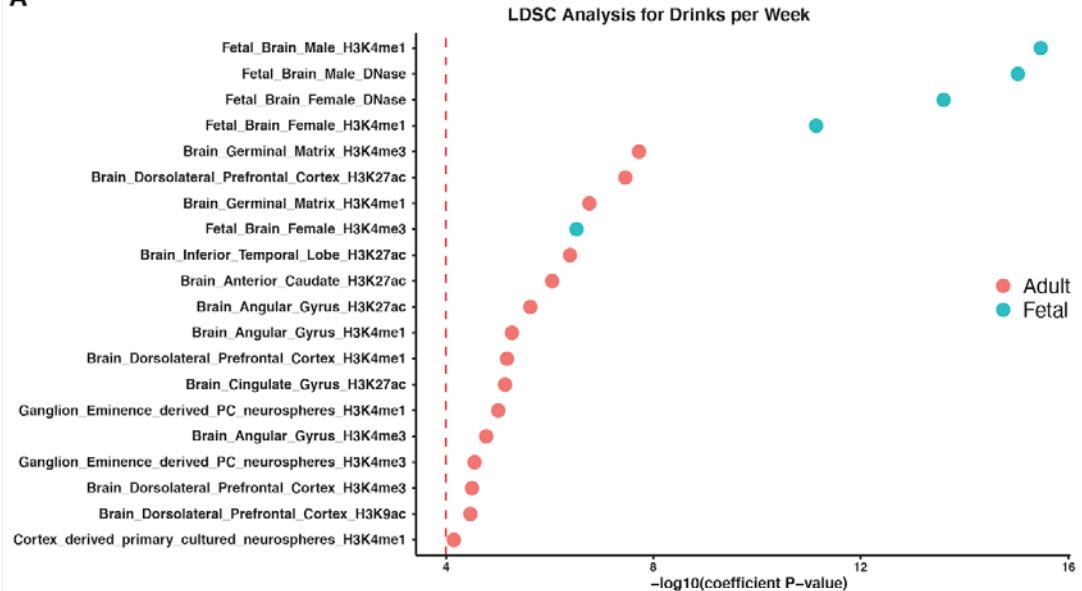


Figure 1: Overview of the study. Series of analyses were undertaken to identify the candidate causal genes associated with risk of AUD and DPW. We used the stratified linkage disequilibrium score (LDSC) regression to test whether the heritability of AUD and DPW is enriched in regions surrounding genes with chromatin markers in a specific tissue. This analysis helped us to identify the large eQTL/mQTLs datasets in the relevant tissues to perform the multi-omic integration analysis using SMR. The candidate causal SNPs and genes prioritized using SMR were further filtered according to threshold of association in GWAS and linkage disequilibrium (Heidi P and COJO). The complex loci with multiple genes were further validated and prioritized by exploring differential gene expression data from brains of alcoholics and controls. Integration of eQTL data from monocytes also helped to prioritize candidate genes specifically expressed in the myeloid cells. The cell type specific epigenetic data from the human brain was also used to identify the causal SNP/s associated with DPW and AUD.

Fig 2: LDSC partitioned heritability analysis of DPW and AUD GWAS by functional category

A



B

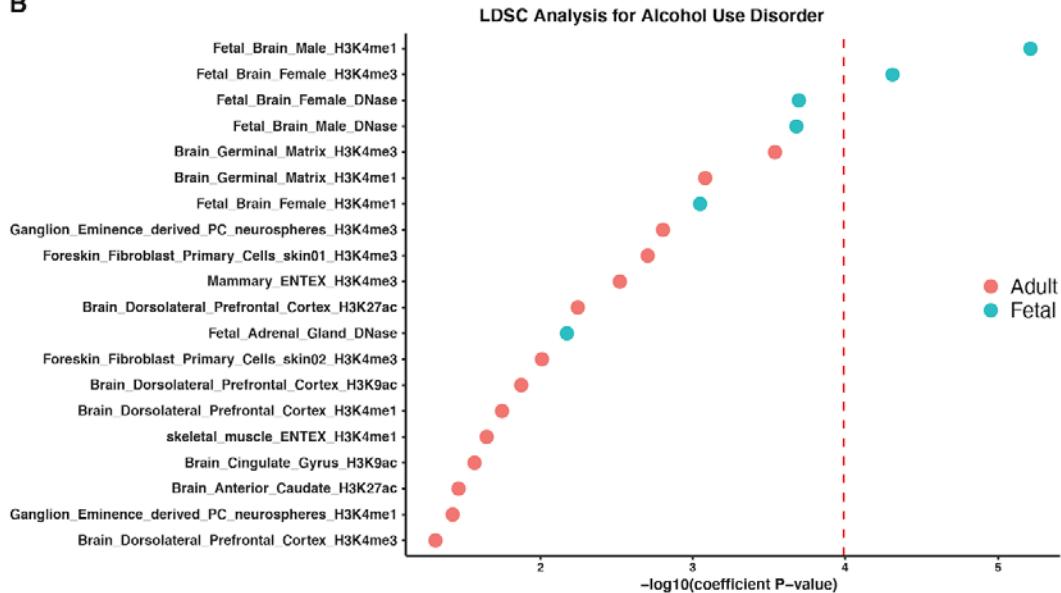


Figure 2: LDSC analysis using tissue specific chromatin data. LDSC analysis showed significant enrichment of promoter specific markers (H3K4me1/me3) in the fetal and adult brain for the SNPs identified in A) DPW and B) AUD GWAS analysis. Y-axis represents the annotations and X-axis represents the $-\log_{10}$ P value for enrichment. The dotted red line represents the threshold of multiple test correction according to Bonferroni.

Fig. 3 DPW and AUD associations with eQTLs/ mQTLs from fetal and adult brain

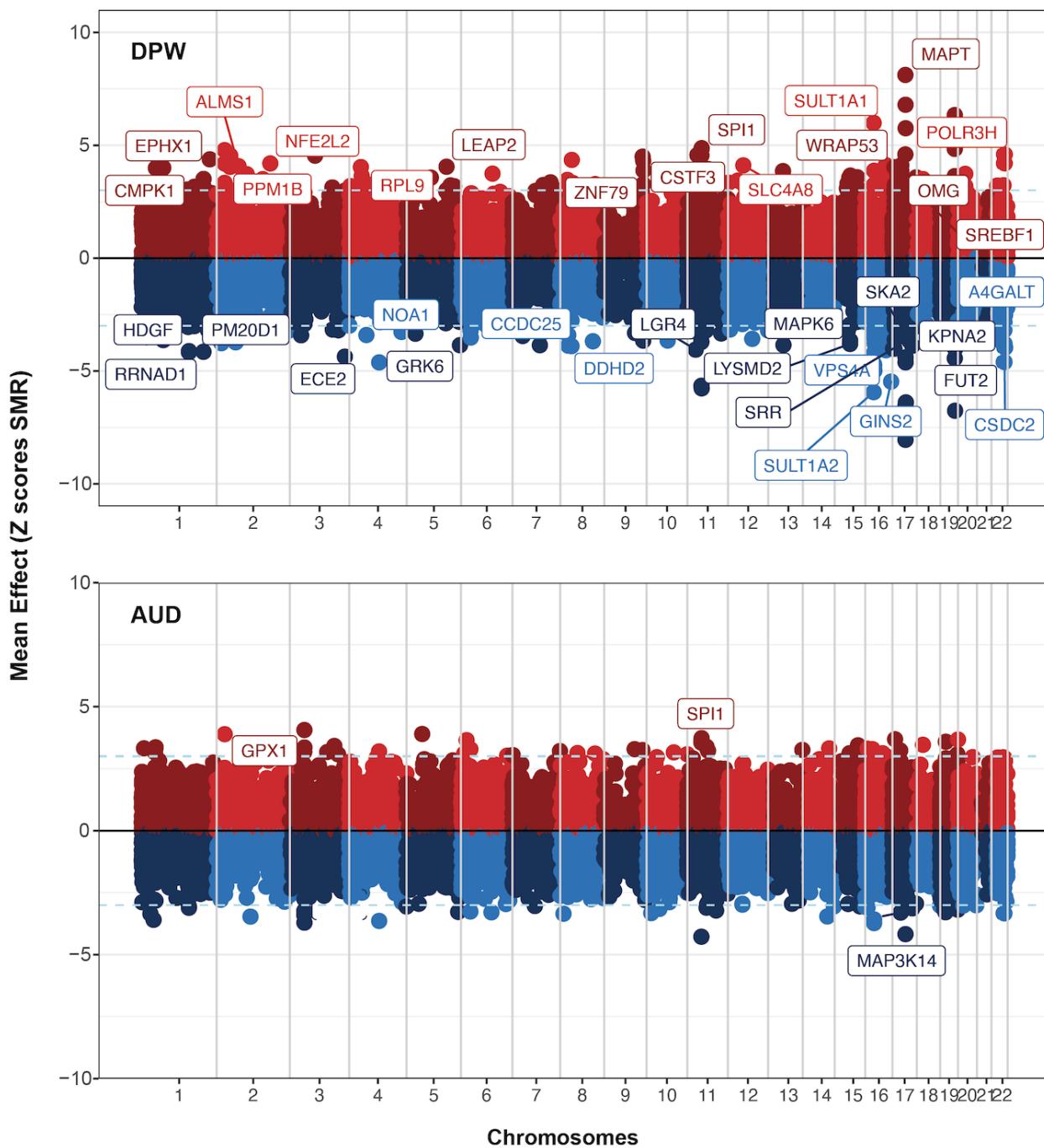
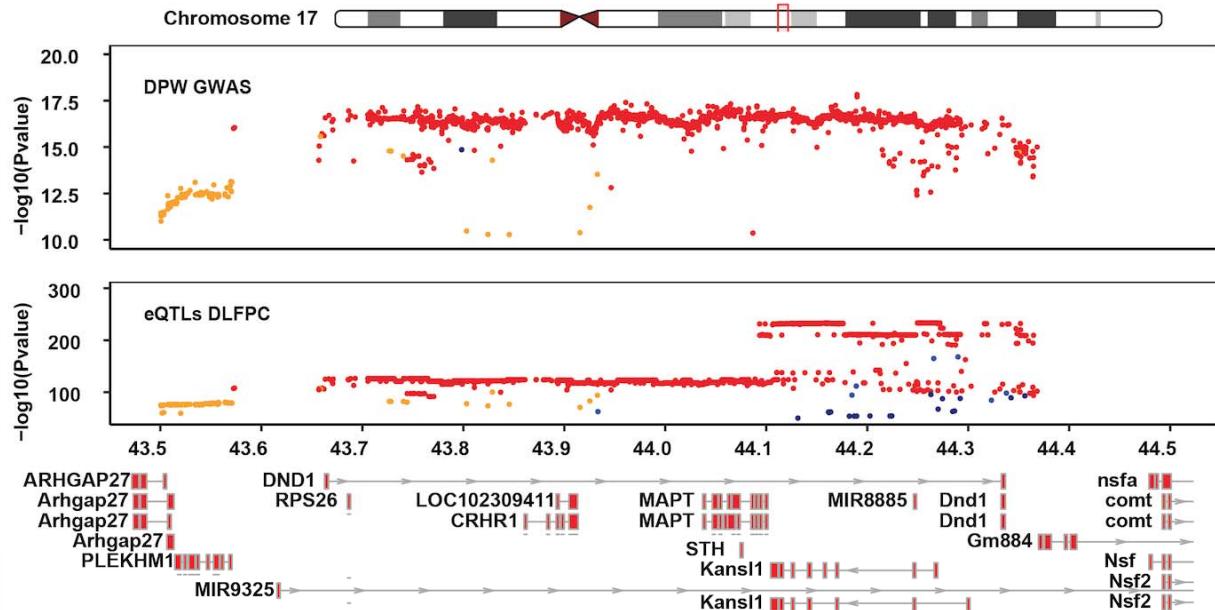


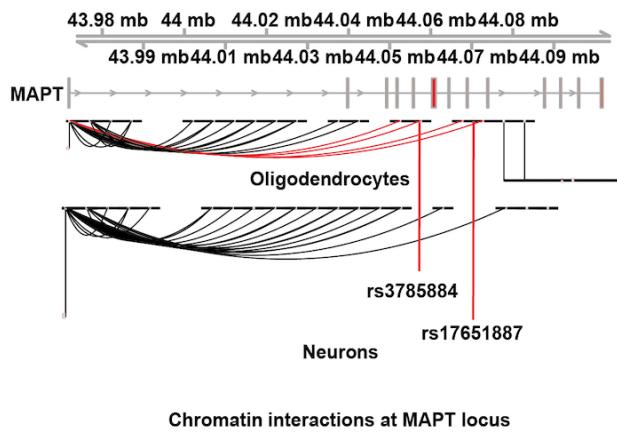
Figure 3: Results of SMR based integration analysis of DPW and AUD GWAS with eQTL/ mQTL from fetal and adult brain. X-axis represents the chromosomes and Y axis shows the direction of effect (Z scores) on gene expression/ methylation. Genes marked on the plots represent the genes nominated through strict threshold of co-localization (FDR < 20%; SMR_{Heidi} P > 0.05; GWAS P < 5 x 10⁻⁵) and/ or multiple levels of transcriptomic and epigenetic evidence.

Fig. 4

A Locus zoom plot showing DPW and eQTL associations at 17q.21.31



B Co-localized SNPs at 17q.21.31 overlap with chromatin interaction region in oligodendrocytes



C MAPT expression in brain was associated with amount of alcohol intake (gms/ day)

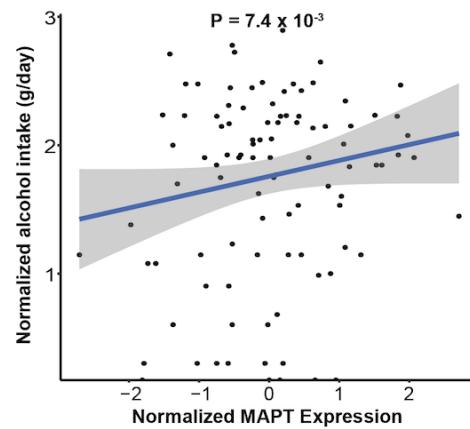


Figure 4: *MAPT* was identified as a candidate gene associated with increased DPW. A. Locus zoom plot showing DPW and eQTL (DLFPC) associations at 17q.21.3. X-axis represents the positions along chromosome 17 and y-axis represents the P values of each SNP at this locus. Color of each dot presents the R^2 for LD at the locus (Red = 0.8 - 1.0; Orange 0.6-0.79; Green 0.4-0.59; Blue 0.2-0.39 and dark blue < 0.2). B. The co-localized SNPs were found to be overlapping with the chromatin interaction region that loops back to the promoter of the *MAPT* gene. C. In independent transcriptomic data from the human brain (N = 92), mRNA expression of *MAPT* was found to be associated with the alcohol consumption.

Fig. 5

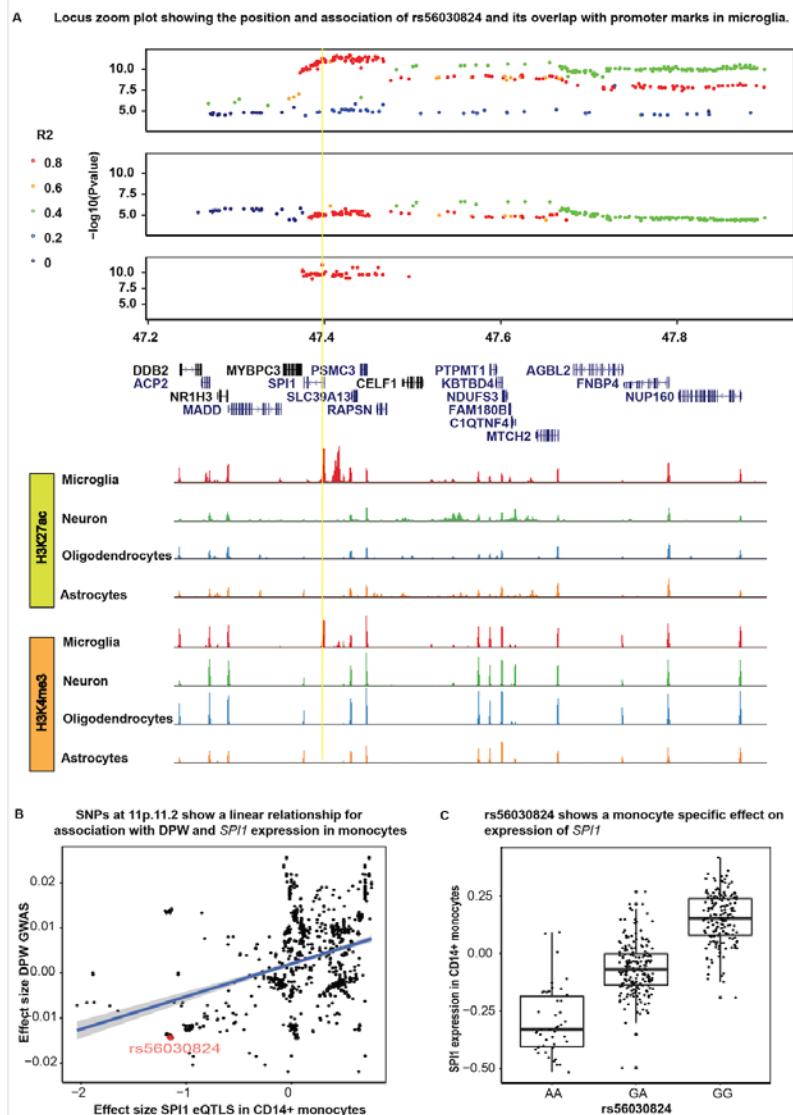


Figure 5: *SPI1* was nominated as candidate gene associated with increased DPW and AUD. Locus zoom plot showing DPW, AUD and mQTL (Fetal brain) associations at 11p.11.2. X-axis represents the positions along chromosome 11 and y-axis represents the P values of each SNP at this locus. Color of each dot presents the R^2 for LD at the locus (Red = 0.8 - 1.0; Orange 0.6-0.79; Green 0.4-0.59; Blue 0.2-0.39 and dark blue < 2.0). Yellow line represents the position of rs56030824 identified as a functional variant co-localized with AUD, DPW and mQTLs in the fetal brain. The tracks show the peaks for promoter marks in 4 major cell types of the brain. Rs56030824 was found to overlap with promoter specific marks (H3K4me3 and H3K27ac), specifically in microglia. B) Effect sizes for DPW GWAS and *SPI1* expression in CD14+ monocytes were found to be correlated i.e. decreased alcohol intake was associated with decreased *SPI1* expression. Rs56030824 showed the strongest association with DPW and mQTL in the common variant category. C) rs56030824 is a strong eQTL and associated with *SPI1* expression in CD14+ monocytes.

Figure 6: Genes nominated by multi-omic integration analysis were enriched in several pathways.

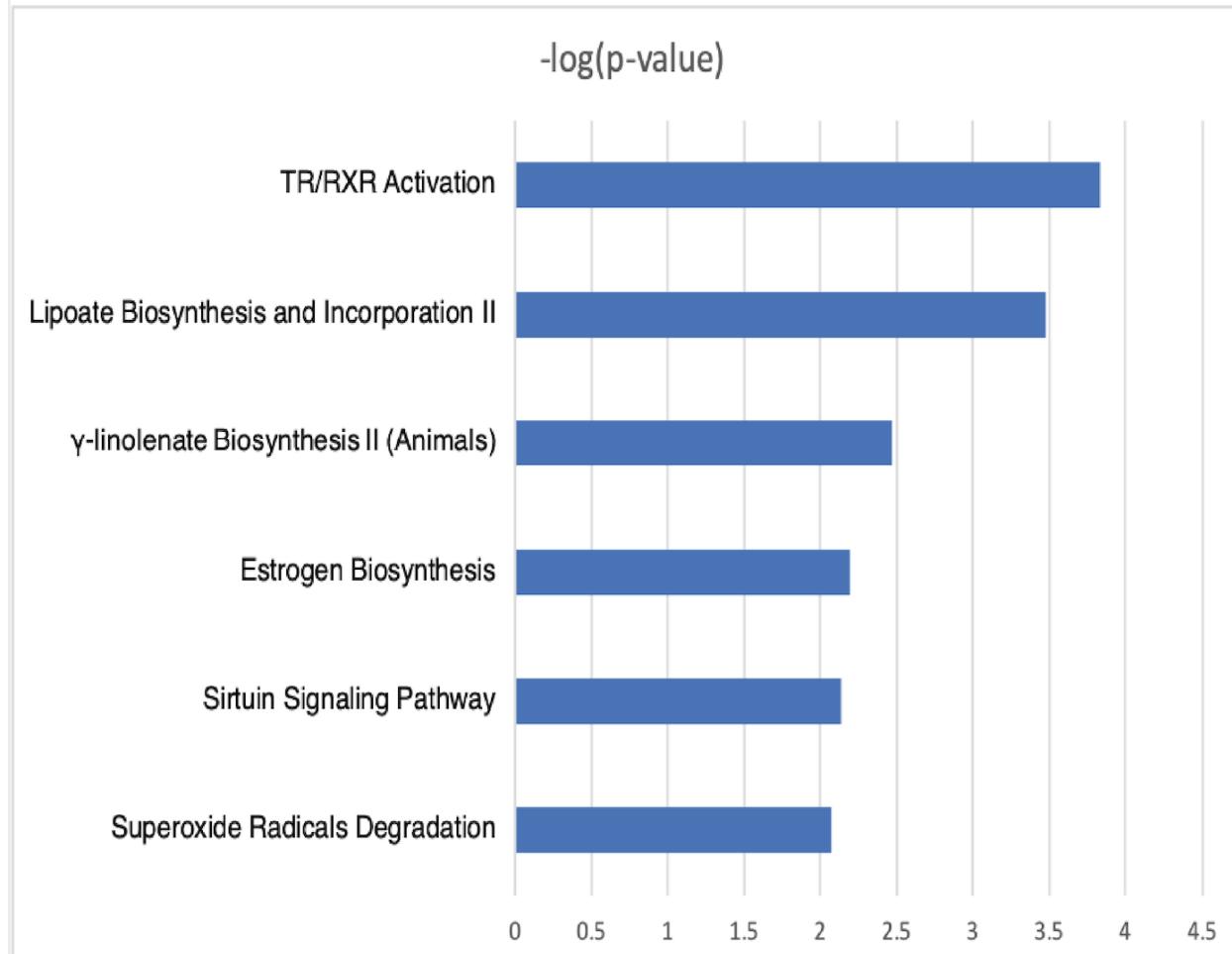


Figure 6: Genes nominated by multi-omic integration analysis were enriched in several pathways. The x-axis represents the -log 10 P value (Fisher's exact test) and y-axis shows the significant pathways.