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

Manav Kapoor1✉, Michael J. Chao1, Emma C. Johnson2, Gloriia Novikova1, Dongbing Lai3, Jacquelyn L. Meyers4, Jessica Schulman1, John I. Nurnberger Jr5, Bernice Porjesz4, Yunlong Liu3, The Collaborative Study on the Genetics of Alcoholism (COGA)*, Tatiana Foroud3,5, Howard J. Edenberg3,6, Edoardo Marcora1, Arpana Agrawal2 & Alison Goate1✉

Identification of causal variants and genes underlying genome-wide association study (GWAS) loci is essential to understand the biology of alcohol use disorder (AUD) and drinks per week (DPW). Multi-omics integration approaches have shown potential for fine mapping complex loci to obtain biological insights to disease mechanisms. In this study, we use multi-omics approaches, to fine-map AUD and DPW associations at single SNP resolution to demonstrate that rs56030824 on chromosome 11 significantly reduces SPI1 mRNA expression in myeloid cells and lowers risk for AUD and DPW. Our analysis also identifies MAPT as a candidate causal gene specifically associated with DPW. Genes prioritized in this study show overlap with causal genes associated with neurodegenerative disorders. Multi-omics integration analyses highlight, genetic similarities and differences between alcohol intake and disordered drinking, suggesting molecular heterogeneity that might inform future targeted functional and cross-species studies.
Alcohol use disorders (AUDs) are complex, moderately heritable (50–60%)\(^1\)–\(^4\), psychiatric disorders associated with heightened morbidity, and mortality\(^5\). 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\(^6\)–\(^9\). In contrast, typical alcohol intake, as assessed using measures such as drinks per week (DPW), represents the distribution of alcohol use from casual or social drinking to excessive drinking demarcating risk for AUD\(^9\)–\(^9\). While heritable, measures such as 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\(^10\). In comparison to other large GWAS, most genome-wide significant variants associated with AUD and DPW are intergenic and thus not directly mappable to a specific gene\(^15\),\(^16\),\(^18\). Furthermore, positionally mapping a non-coding variant to the nearest gene often does not influence AUD 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 positively 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\(^6\),\(^15\),\(^16\),\(^17\). 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\(^3\),\(^13\),\(^15\),\(^16\),\(^18\). Several recent studies have integrated GWAS data with expression QTLs (eQTLs) using co-localization or integration methods to identify causal variants and genes associated with schizophrenia, Alzheimer’s disease, and many other complex disorders\(^3\),\(^13\),\(^15\),\(^16\),\(^18\). 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 GTEx\(^10\)–\(^14\).

Here we present a multi-omics systems approach to identify causative variants associated with AUD and DPW. Using Mendelian Randomization-based methods on the largest available transcriptomic and epigenomic data for brain tissues (Supplementary Data 1) and myeloid cells, we prioritized regulatory variants that influence AUD and DPW (Fig. 1). The current manuscript explored the multi-omic integration results in AUD and DPW separately and subsequently focused on the overlapping genes between these two traits. Overlapping genes prioritized in current analysis are primarily driven by individual GWASs. Therefore, these signals are minimally influenced by the sample size bias that might arise due to the integration of two correlated traits with extreme differences in power. Results of predicted mRNA expression (integration analyses) were
signals, we focused on the loci where the strongest SNP was at least to test whether the heritability of AUD and DPW is significant. Among these 18 loci, SMR analysis nominated a single candidate gene. A total of 360 SNPs associated with AUD and DPW were in common (i.e., 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 (Fig. 4c). The association between AUD expression and alcohol consumption did not pass multiple test correction, most likely due to the small sample size of the brain dataset from people with AUD ($P = 7.4 \times 10^{-3}$; $P_{\text{Bonferroni}} = 0.46$). We did not observe any association between the expression of $\text{LRR37A}$ and the level of alcohol intake in this brain dataset. The co-localized SNPs within the 17q21.31 locus were also compared with the promoter (H3K27ac, H3K4me3), enhancer (ATAC-Seq), and promoter–enhancer interactome (PLAC-Seq) data from four specific brain cell types (microglia, neuron, astrocytes, and oligodendrocytes) to elucidate the functional significance of these variants. The co-localized mQTLs (rs3785884 and rs17651887) overlapped with the chromatin interaction region specifically in oligodendrocytes and these interactions looped at the MAPT promoter (Fig. 4b). 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 SPI1 and $NUP160$ respectively at 11p11.2 for association with DPW and AUD (Fig. 3). Both SPI1 and $NUP160$ are primarily expressed in myeloid cells. The predicted causal $\text{SPI1}$ mQTL (rs56030824) and $\text{NUP160}$ 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 ($4.90 \times 10^{-12}$) than rs10838753 (eQTL) ($\text{DPW} P = 1.28 \times 10^{-10}$; $\text{AUD} p = 8.45 \times 10^{-5}$). Adding rs56030824 as a covariate in conditional analyses had a larger effect on the association between rs10838753 and AUD ($P_{\text{org}} = 4.85 \times 10^{-5}$; $P_{\text{cond}} = 0.09$) than with DPW ($P_{\text{org}} = 1.28 \times 10^{-10}$; $P_{\text{cond}} = 8.1 \times 10^{-3}$). rs56030824 remained significantly associated with both AUD ($P_{\text{org}} = 8.91 \times 10^{-6}$; $P_{\text{cond}} = 2.0 \times 10^{-2}$) and DPW ($P_{\text{org}} = 4.90 \times 10^{-12}$; $P_{\text{cond}} = 7.0 \times 10^{-4}$) even after adding rs10838753 as a covariate. Rs56030824 overlapped the promoter marks (H3K27ac, H3K4me3) for SPI1 specifically in microglia (Fig. 5A). This SNP also alters the binding site regulatory motif for RXRA, a transcription factor which is involved in the promotion of myelin debris phagocytosis and remyelination by macrophages. Since SPI1 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 the expression of SPI1 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 (Fig. 5B, C). In fact, rs56030824 had the strongest effect size for SPI1 expression and DPW in the common variant category (Fig. 5C). These observations together established rs56030824 as a stronger candidate to be considered as a causal variant and SPI1 as a potential candidate gene associated with AUD and DPW.
Table 1 SMR analysis results with summary statistics of AUD GWAS meta-analysis.

| Chr | BP       | Gene       | Adult brain |                  | Fetal brain |                  |
|-----|----------|------------|-------------|-----------------|-------------|-----------------|
|     |          |            | eQTL p value | mQTL p value   |             | mQTL p value   |
| 11  | 46399942 | SPI1N      | x           | x               | x           | 1.91E−04       |
|     | 46664086 | MTCH2      | 1.89E-05    | x               | x           | 1.05E−04       |
|     | 46843734 | NUP160     | 1.89E-04    | x               | x           | 1.05E−04       |
|     | 48395716 | GPX1       | 4.93E-05    | x               | x           | 1.05E−04       |
|     | 48459884 | AMT        | 2.07E-04    | x               | x           | 4.47E−01       |
|     | 43361331 | MAP3K14N   | x           | x               | x           | 2.99E−05       |

The reported genes from the integration analyses survived four different \( P \) value thresholds to be nominated as potential causal candidate genes (GWAS \( P = 0.05 \); FDRSMR-\( P < 0.2 \)). SMR \( P \)-values for the co-localized SNPs are obtained using the Wald test. The superscript \( N \) in front of gene indicates the potential candidate causal gene prioritized using current multi-omic analysis.

**Fig. 3** Results of SMR based integration analysis of DPW and AUD GWAS meta-analysis with eQTL/mQTL from fetal and adult brain. X-axis represents the chromosomes and Y-axis shows the standardized direction of effect (Z scores) of co-localized SNP on gene expression/ methylation and GWAS phenotype. Z scores were derived from the effect size (betas) and standard errors (SE) from the SMR analyses. Positive Z score shows that increase in mRNA expression or methylation is associated with excessive drinking or increased risk of AUD, while negative Z score depicts the vice-versa. Genes marked on the plots represent the genes that passed the strict threshold of co-localization (FDRSMR-\( P < 20\% \); SMRHeidi-\( P > 0.05 \); GWAS \( P < 5 \times 10^{-5} \); eQTL \( P < 5 \times 10^{-5} \)) and multiple levels of transcriptomic and/or epigenetic evidence. SMR \( P \)-values for the co-localized SNPs are obtained using the Wald test.
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 (Supplementary Data 5; Supplementary Data 6). On chromosome 19, FUT2 was the strongest candidate; mRNA expression of FUT2 was also nominally 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 (Supplementary Data 5).

**Pathway and network analysis.** Ingenuity pathway analysis of the prioritized genes associated with DPW showed suggestively significant enrichment for pathways related to TR (Thyroid hormone receptor)/RXR (Retinoic X receptor) activation ($P = 1.45 \times 10^{-3}$), Lipote biosynthesis ($3.29 \times 10^{-4}$), Estrogen biosynthesis ($P = 6.39 \times 10^{-3}$) and Sirtuin signaling ($7.27 \times 10^{-3}$) (Supplementary Data 8). In all, 410 pathways were tested using IPA and none of the pathway survived the threshold for multiple test correction ($P_{\text{Bonferroni}} = 1.2 \times 10^{-4}$). 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 the 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 AUD. The
AUD GWAS meta-analysis used here specifically focused on the diagnosis rather than the disordered drinking. Importantly, our work highlights that GWAS variants for AUD and DPW 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 behaviors25–31. But these observations were never consistent with results from GWAS of AUD and DPW, most likely due to lack of power in these
Fig. 5 SPI1 was nominated as candidate gene associated with increased DPW and AUD. A: 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 −log10 (P) values of each SNP at this locus. P-values for the DPW GWAS (Liu et al, 2019) and AUD GWASs are obtained from aggregated weighted Z statistics. P-values for mQTL meta-analysis were obtained using conventional inverse-variance-weighted meta-analysis. Color of each dot presents the values for the DPW GWAS (Liu et al, 2019) and AUD GWASs obtained from aggregated weighted Z statistics.

-0.59; Blue 0.2; 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. The shaded area around trend-line depicts the 95% confidence level intervals plotted using “geom_smooth(method = “lm”)” from ggplot2. C: Rs56030824 is a strong eQTL and associated with SPI1 expression in CD14+ monocytes. The box in the Box Plot is extending from the 25th percentile to the 75th percentile, with median horizontal line within the box. The whiskers are extending to one and a half times the interquartile range.

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 (SPI1, 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 disorders.25 Identification of SPI1 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 SPI1 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 SPI1 in the cerebellum and spleen from subjects with Major depressive disorder and schizophrenia were significantly higher than in controls. In the past, we have demonstrated that functional variants related to SPI1 expression are associated with the risk of Alzheimer’s disease.15 Similar to this study, higher levels of expression of SPI1 is associated with increased risk for Alzheimer’s disease.

SPI1 (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.1 binding motif (PU).15 Given SPI1’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 the expression of this transcription factor in transcriptomic datasets from whole brains. Some studies using animal models have reported that chronic alcohol consumption can influence the 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 AUD.

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 disease,8 frontotemporal dementia,39 Parkinson’s disease,38 and the primary tauopathies (progressive supranuclear palsy) and corticobasal degeneration.40 Recently, Hoffman and colleagues31 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 people with alcohol use vs Alzheimer disease patients.26 CRHR1 (corticotropic releasing hormone type 1 receptor) is another gene on 17q21.31, that has been reported to be associated with alcoholism.43 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 MRI study.44 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 compounds.45 In IPA disease enrichment analysis, we observed a nominally significant overlap between genes implicated in DPW with other neurological, behavioral and immune-related disorders (Supplementary Fig 16). The genes associated with DPW also showed significant enrichment for pathways related to TR/ RXR activation, Lipoate biosynthesis, Estrogen biosynthesis, and Sir- tui signaling (Supplementary Data 8). 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 development.46 RXRs (Retinoic X Receptor) are known to potentially regulate the ethanol metabolizing enzymes after chronic alcohol consumption.47 It has been reported that the human aldehyde dehydrogenase-2 (ALDH2) promoter contains a retinoic response element, which might be contributing to the regulation of the gene.47 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.

We have identified a number of candidate causal genes for DPW and AUD, resulting from a multi-omic analysis of human genetic and expression data. Our resource in conjunction with data generated in animal studies will guide researchers to plan well-informed experiments. The current study also has some limitations. We want to emphasize that due to the limited availability of raw GWAS and e/mQTL data we were not able to perform sex-stratified analyses. There was also no data on alcohol...
consumption/ alcohol dependence in individuals contributing to the myeloid datasets. This limited us in comparisons of predicted gene expression changes (SMR analyses) in the myeloid cells to actual gene expression changes in monocytes or other myeloid cells. The differential expression results from the brains of people with alcohol use were also generated in a small dataset (Total N = 138; N_ALC_Cons = 92) (although this represents the largest dataset to date). Given the smaller effect sizes of GWAS signals it will require a very large brain dataset to detect associations of SNP mediated mRNA expression with phenotype. Still our data validated key genes at nominal association levels, which are encouraging for further targeted studies.

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.

Methods

Samples

**Alcohol use disorder.** We meta-analyzed three published GWAS: the Million Veteran Program (MVP)GWAS of AUD (European [EUR] N = 202,004; N_EUR = 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 dependence cases based on DSM-IV diagnoses; EUR unrelated genotyped N = 28,757; N_EUR = 8845 and the Collaborative Studies on Genetics of Alcoholism (COGA) GWAS of alcohol dependence cases based on DSM-IV diagnoses; EUR unrelated genotyped N = 4849; N_EUR = 2411)29. Genetically calculated principal component 1 (PC1) was added as additional covariate in analyses of individual GWAS data from individuals with European ancestry. The final meta-analysis in PGC was performed by combining summary statistics weighted on sample size from individual datasets. Appropriate PCs calculated using SNP data were also included as a covariate in MVP19 and COGA datasets.

**Drinks per week.** We used genome-wide summary statistics for 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 (DLPFC: PsychEncode (N = 1387)GWAS, ROSMAP (N = 461)GWAS, and COGA INIA (N = 136). ROSMAP eQTL analysis was limited to participants of European descent only and obtained from AMP-AD Knowledge Portal (https://www.synapse.org/#!Synapse:syn3219045). A full set of PsychEncode eQTL summary statistics without P-value threshold restriction was downloaded from the PsychEncode Integrative Analysis website (http://resource.psychencode.org/)39. The PsychENCODE Integrative Analysis contained adult brain prefrontal cortex data of 1387 individuals from the PsychENCODE and the Genotype-Tissue Expression (GTEx, https://www.gtexportal.org) data. Briefly, Psychencode consortium included data from gene expression matrix normalized using quantile normalization, followed by inverse quantile normalization to map to a standard normal distribution (and to remove outliers). In all QTL analyses, known confounding factors such as age, sex, population structure, and technical covariates were, and unidentified confounding factors were minimized through principal component analysis or similar methods. Detailed information about the data collection and analysis process can be found in the original study40. Human post-mortem brain samples (COGA-INIA) samples were obtained from the New South Wales Tissue Resource Centre at the University of Sydney (http://sydney.edu.au/medicine/pathology/btrc/). Fresh frozen samples of the superior frontal gyri were collected from each post-mortem sample after obtaining ethical approval for the COGA research project from Icahn School of Medicine and the Scientific Review Committee at New South Wales Tissue Resource Centre at the University of Sydney. The samples were genotyped using the UK Biobank Axiom array as part of the COGA-INIA collaboration.

eQTLs from adult brain. Brain-mMeta mQTL summary data in SME binary (BESS) format were obtained from the SME data resource (Supplementary Data 1). This is a set of mQTL data from a meta-analysis (N = 994) of ROSMAP, Hannon et al.35, and Jaffe et al.33.

eQTLs from fetal brain. Summary data for eQTLs from developing human brains (Supplementary Data 1) were obtained from an online repository shared by Heath O’Brien and Nicholas J. Bray85. The analyses were performed on 120 human fetal brains from the second trimester of gestation (12–19 post-conception weeks).

mQTLs from fetal brain. Summary data for mQTLs from developing human brains (Supplementary Data 1) were obtained from an online repository shared by Ellis Hannon and Jonathan Mill53. The mQTLs were characterized in a large collection (n = 160) 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 genes56.

Whole-genome transcriptomic data in the brain of people with AUD and with daily alcohol intake. mRNA expression data in the Dorso-lateral pre-frontal cortex (DLPFC) 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 inter-actome (PLAC-Seq) data for four specific cell types of brain (microglia, neuron, astrocytes, and oligodendrocytes) to elucidate the functional significance of co-localized SNPs57. The location of each epigenetic mark was intersected with the location of variants prioritized by summary MR. UCSC browser (https://genome.ucsc.edu/) was used to visualize the overlap of epigenetic markers with SNPs at 17q12.31 (MAPT) and 11p11.2 (SPP1) loci.

Analysis

eQTL meta-analysis in adult brain. RNA Sequencing data on the DLPFC region of the brain for 138 samples was generated as part of COGA-INIA collaboration. We also genotyped the brain samples using the UK Biobank Axiom array. More than 97% of subjects (N = 133) included in this study belonged to European ethnicity. All NSWBB samples were imputed to 1000 Genomes using the cosmo-politan reference panel (Phase 3, version 5, NCBI GRCh37) using SHAPEIT then Impute253 within each array. Only variants with non-A/T or C/G alleles, missing rates <5%, MAF ≥ 5%, and HWE P-values >0.001 were included for imputation. Imputed variants with R2 < 0.30 were excluded, and genotype probabilities were converted to genotypes if probabilities ≥ 0.90. All genotyped and imputed variants (4,615,871 SNPs) with missing rates <10%, MAF ≥ 5% and HWE P-values >0.1 were included in the downstream analyses using MatrixQTL. The gene expression was corrected for the batch, age, sex, RNA integrity number (RIN), Post-mortem Interval (PMI), and alcohol status using the "removeBatchEffect" option from the limma package. The genetically derived PC1 was also added as additional covariate in the eQTL analysis. The eQTL summary statistics from PsychEncode, ROSMAP, and COGA-INIA 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 (https://cnsgenomics.com/software/smrs#DataManagement). The meta-analysis for eQTLs was performed using the conventional inverse-variance-weighted meta-analysis assuming all cohorts are independent. 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 the LDSC program. We obtained the weights for the multi-cell and tissue chromatin covariates and performed the LDSC partition heritability analyses on munged summary statistics of AUD and DPW GWAS31. We specifically followed the default parameters to perform the LDSC analyses as listed in the tutorial for partition heritability analysis (https://github.com/bulik/ladsc/wiki/Partitioned-Heritability).

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) analysis39 on a set of mQTLs and eQTLs (Top SNP P < 5 x 10−7) from fetal and adult brains. SMR is a Mendelian randomization-based analysis which integrates GWAS summary statistics and 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 (BESP) to perform the SMR analysis. The scripts used to create the binary (BESP) files for each dataset have been deposited in GitHub. The European subset of ADGC GWAS (ph:000372.v1.p1) was used as the LD reference panel to perform the SMR analyses. The analysis was performed using default parameters as listed in the SMR tutorial. The genes below FDR threshold (FDR < 20%; GWAS P < 5 × 10\(^{-8}\); eQTL/mQTL P < 5 × 10\(^{-8}\)) and with heterogeneity P value > 0.05 were considered to be causal within each combination of analysis.

Conditional 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)\(^6\) approach is implemented in Genome-wide Complex Trait Analysis (GCTA) software\(^6\) 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, post-mortem interval [PMI], RNA integrity number [RIN]). Gene-level analyses started with the featureCounts-derived sample-by-gene 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, Body mass index (BMI), RIN, batch and severity of alcoholism (AUDIT scores).

Comparison of overlapping results between AUD and DPW integration analyses. The analyses presented in this manuscript performed multi-omic integration analyses separately for AUD and DPW and only later looked at the overlap. We specifically employed this approach as it minimizes the bias in results due to large sample size differences between the two GWASs. Hence, the summary statistics from the correlated meta-analysis of DPW and AUD GWASs will be primarily driven by DPW signals due to very large sample size for DPW GWAS and smaller standard errors. As a result, the multi-omics analysis using these summary statistics will predominantly prioritize genes associated with the DPW phenotype. In fact, we used the Multi-Trait Analysis of GWAS (MTAG) method to meta-analyze the summary statistics from DPW and AUD GWAs. Although, this method is robust to correlated multi-trait meta-analysis, the SMR analysis using the MTAG results was primarily driven by the DPW variants. To reduce this bias, we also focused on SMR results from MTAG using stricter threshold (GWAS AUD 5 × 10\(^{-8}\); eQTL/mQTL P < 10\(^{-8}\); SMR P < 20%; Heli P > 0.05). The results from this threshold were similar to our original analyses that focused on overlapping results after integration analysis (Supplementary Data 7 and Supplementary Data 9). In fact, many SNPs in the combined analysis were filtered out due to inflated summary statistics of the heterogeneity (Heidi P values) test (Supplementary Data 9).

Pathway analysis. The results of integration analysis for the DPW and AUD GWAS meta-analysis (SMR P \(\leq\) 0.05, Heidi P > 0.05, P\(\text{eQTL}\) < 5 × 10\(^{-8}\), P\(\text{mQTL}\) = 5 × 10\(^{-8}\)) 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 GWASs. Users will also be able to see whether the genes of interest are differentially expressed in the brains of people with AUD and controls.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The individual Manhattan plots with SNP meta-analysis for all conditions can be found in supplementary information Figures 7–15. Supplementary Figures 1–4 reports the AUD-meta association plots generated through FUMA. Additionally all the results can be visualized at our Shiny web app (https://iacd.shinyapps.io/acl_multomics/). No raw data was generated in this manuscript. AUD meta-analysis summary statistics along with summary statistics of all SMR analyses are available to download at figshare.com: AUD meta-analysis summary statistics: https://doi.org/10.6084/m9.figshare.10554198.v1. SMR Input BESP files for brain meta-analysis: https://doi.org/10.6084/m9.figshare.15054183.v1. mRNA expression analysis of alcohol consumption (corrected for BMI, AUDIT scores, age, sex, and PMI): 10.6084/m9.figshare.15054171.v1. SMR results (Complete summary statistics files): https://doi.org/10.6084/m9.figshare.15054120.

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. Psychol. Med. 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. Alcohol.: Clin. Exp. Res. 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. Am. J. 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, 34–49 (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. Pr. 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, 757–766 (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, 157–168 (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, 837–844 (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. Curr. Genet. Med. Rep. 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.-I. et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer’s disease. Nat. Neurosci. 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 PTO form long-range functional connections with IRX3. Nature 507, 371–375 (2014).
18. Andrews, S. J., Fulton-Howard, B. & Grote, 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, 291–295 (2015).
22. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).
23. Stefansson, H. et al. A common inversion under selection in Europeans. Nat. Genet. 37, 129–137 (2005).
24. Natrajian, M. S. et al. Retinoid X receptor activation reverses age-related deficiencies in myelin debris phagocytosis and remyelination. Brain 138, 3581–3597 (2015).
25. Crews, F. T. Immune function genes, genetics, and the neurobiology of addiction. Alcohol Res. 34, 355–361 (2012).
26. Farris, S. P., Arasappan, D., Hunnicke-Smith, S., Harris, R. A. & Mayfield, R. D. Transcriptome organization for chronic alcohol abuse in human brain. Mol. Psychiatry 20, 1438–1447 (2015).
27. Fahey, 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, 505–515 (2013).
30. Oldham, M. C. et al. Functional organization of the transcriptome in human brain. Nat. Neurosci. 11, 1271–1282 (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, 1402–1411 (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, 576–589 (2010).
35. Socodato, R. et al. Daily alcohol intake triggers aberrant synaptic pruning leading to synapse loss and anxiety-like behavior. Sci. Signal. 13, eaab5754 (2020).
36. Joshi, P. C. et al. Chronic ethanol ingestion in rats decreases granulocytic-macrophage colony-stimulating factor receptor expression and downstream signaling in the alveolar macrophage. J. Immunol. 175, 6837–6845 (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, 1519–1528 (2011).
38. Desikan, R. S. et al. Genetic overlap between Alzheimer’s disease and Parkinson’s disease at the MAPT locus. Mol. Psychiatry 20, 1588–1595 (2015).
39. Rademakers, R., Cruts, M. & van Broeckhoven, C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. Hum. Mutat. 24, 277–295 (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 neuronal 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, 575–577 (1999).
43. Ray, L. A. et al. The CRHR1 gene, trauma exposure, and alcoholism risk: a test of G × E effects. Genes Brain Behav. 12, 361–369 (2013).
44. Jia, T. et al. Neural basis of reward anticipation and its genetic determinants. J. Neurosci. 33, 16088–16098 (2013).
45. Nowell, S. & Falany, C. N. Pharmacogenetics of human cytosolic sulfotransferases. Oncogene 25, 1675–1678 (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 retinoid 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, 16088–16098 (2013).
49. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362, eaat4864 (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).
The Collaborative Study on the Genetics of Alcoholism (COGA)
Victor Hesselbrock, Samuel Kuperman, John Kramer, Chella Kamarajan, Ashwini Pandey, Laura Bierut, John P. Rice, Kathleen K. Bucholz, Marc Schuckit, Jay Tischfield, Andrew Brooks, Ronald P. Hart, Laura Almasy, Danielle Dick, Jessica Salvatore & Paul Slesinger

7Department of Psychiatry, University of Connecticut, Farmington, CT, USA. 8Department of Psychiatry, University of Iowa, Iowa City, IA, USA. 9Department of Psychiatry, Washington University in St. Louis, MO, USA. 10Department of Psychiatry, University of California at San Diego, San Diego, CA, USA. 11Department of Genetics, Rutgers University, Piscataway, NJ, USA. 12Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ, USA. 13Department of Biomedical and health Informatics, University of Pennsylvania, Philadelphia, PA, USA. 14Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA. 15Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA.