













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

Manav Kapoor <sup>1✉</sup>, Michael J. Chao<sup>1</sup>, Emma C. Johnson <sup>2</sup>, Glorria Novikova <sup>1</sup>, Dongbing Lai <sup>3</sup>, Jacquelyn L. Meyers <sup>4</sup>, Jessica Schulman<sup>1</sup>, John I. Nurnberger Jr <sup>5</sup>, Bernice Porjesz<sup>4</sup>, Yunlong Liu <sup>3</sup>, The Collaborative Study on the Genetics of Alcoholism (COGA)\*, Tatiana Foroud <sup>3,5</sup>, Howard J. Edenberg <sup>3,6</sup>, Edoardo Marcora <sup>1</sup>, Arpana Agrawal <sup>2</sup> & Alison Goate <sup>1✉</sup>

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.

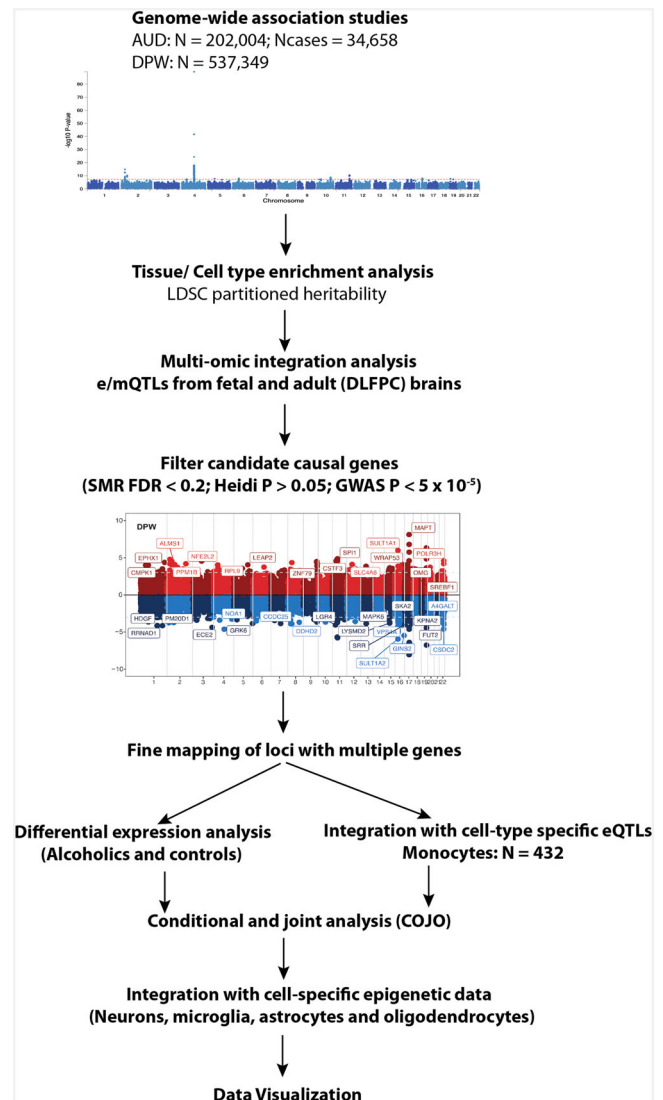
<sup>1</sup>Departments of Genetics and Genomic Sciences and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>2</sup>Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. <sup>3</sup>Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>4</sup>Department of Psychiatry, State University of New York, Downstate Medical Center, Brooklyn, NY, USA. <sup>5</sup>Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA. <sup>6</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA. \*A list of authors and their affiliations appears at the end of the paper. ✉email: [manav.kapoor@regeneron.com](mailto:manav.kapoor@regeneron.com); [alison.goate@mssm.edu](mailto:alison.goate@mssm.edu)

Alcohol use disorders (AUDs) are complex, moderately heritable (50–60%)<sup>1–4</sup>, psychiatric disorders associated with heightened morbidity, and mortality<sup>5</sup>. 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<sup>6,7</sup>. 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<sup>8,9</sup>. 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<sup>8,9</sup>.

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<sup>10</sup>. In comparison to Zhou and colleagues PAU GWAS, other comparatively large GWAS of alcoholism concentrated on the consumption aspect of addictive disorders. For example, 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<sup>11</sup>. 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<sup>10–14</sup>. 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)<sup>10–14</sup>. 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<sup>10,11,13,14</sup>.

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<sup>15,16</sup>. Furthermore, positionally mapping a non-coding variant to the nearest gene often does not identify the causal gene(s)<sup>15–17</sup>. 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<sup>13,15,16,18</sup>. 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<sup>13,15,16,18</sup>. 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<sup>10–14</sup>.

Here we present 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 (Supplementary Data 1) and myeloid cells, we prioritized regulatory



**Fig. 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 people with alcohol use disorder 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.

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

compared with the mRNA expression 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 the best of our knowledge, this is 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)<sup>19</sup>, the Psychiatric Genetics consortium (PGC-SUD)<sup>12</sup> and the Collaborative Studies on Genetics of Alcoholism (COGA)<sup>20</sup> identified 1157 SNPs (31 independent lead SNPs) within or near 79 genes at 10 independent loci associated with AUD (Supplementary Figs. 1–4). We did not include UKB-AUDIT-P in this meta-analysis to specifically focus on AUD. Many of these loci were shared between the AUD GWAS meta-analysis and the DPW GWAS by Liu et al.<sup>11</sup> 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) (Supplementary Fig. 5). A large and nominally significant proportion (45%) of AUD and DPW-associated SNPs were within intronic, UTR and non-coding regions of the genome (Supplementary Fig. 6).

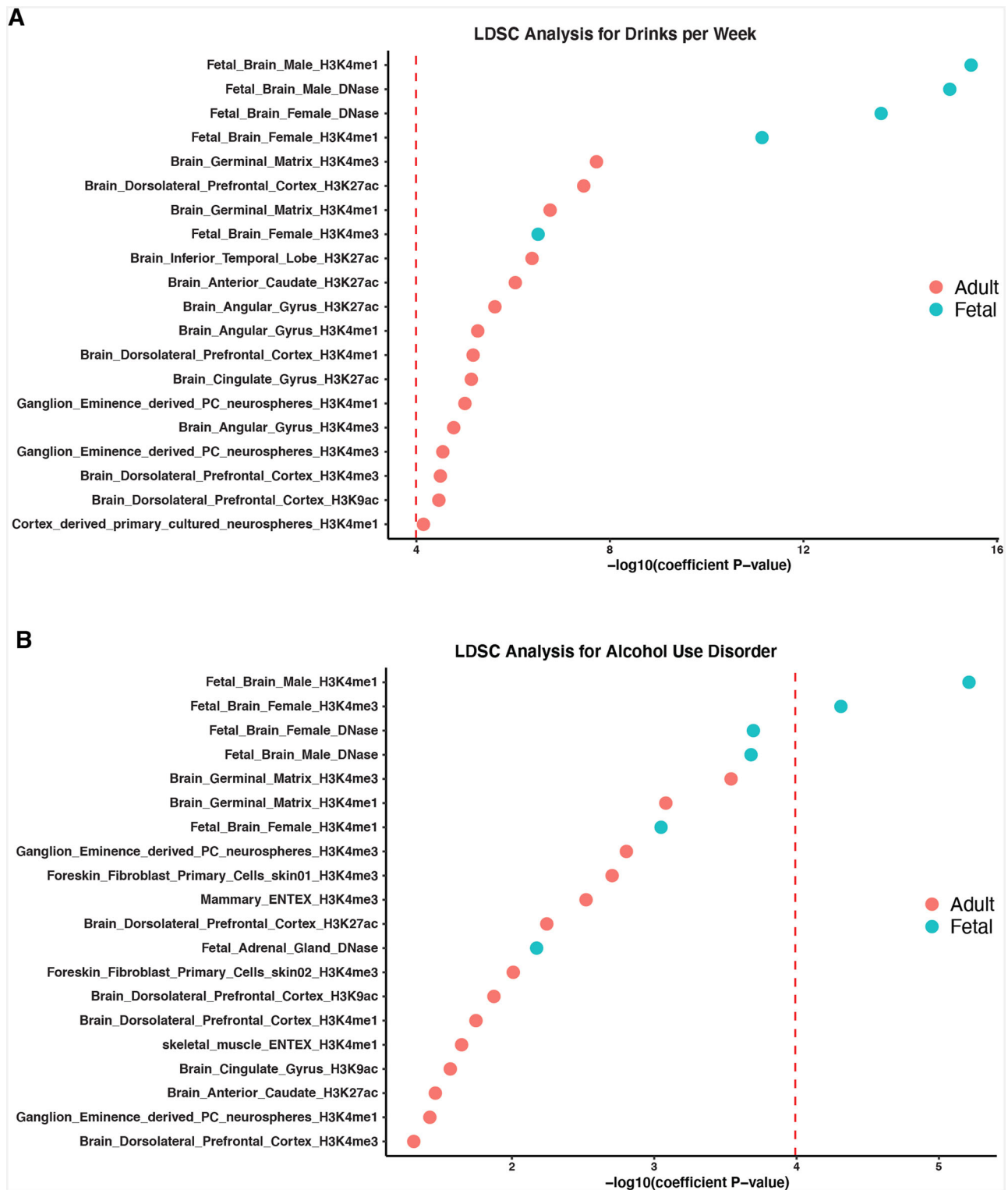
**LDSC analysis using tissue specific epigenetic annotations.** We used the stratified linkage disequilibrium score (LDSC) regression<sup>21</sup> to test whether the heritability of AUD and DPW is enriched in regulatory regions surrounding genes in a specific tissue. Using multi-tissue chromatin (ROADMAP and ENTEX) data<sup>22</sup>, we observed a significant enrichment of promoter-specific epigenetic markers (H3K4me1/me3) in the fetal and the adult (germinal matrix, frontal-cortex) brain ( $P < 5 \times 10^{-8}$ ) (Fig. 2; Supplementary Data 2 and 3) 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 ( $P_{\text{eQTL}} = 5 \times 10^{-8}$ ;  $P\text{-SMR FDR} < 20\%$ ;  $\text{Heidi} > 0.05$ ) [Supplementary Data 4; Supplementary Figs. 7–10]. 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 (*SPI1*, *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 ( $P\text{-GWAS} < 5 \times 10^{-5}$ ) and genome-wide significant in respective eQTL/mQTL datasets ( $P\text{-eQTL/mQTL} < 5 \times 10^{-8}$ ) [Table 1; Fig. 3]. Because of the much larger sample size of the DPW GWAS, 61 genes at nearly 31 loci passed the threshold for significance ( $P\text{-SMR FDR} < 20\%$ ;  $\text{Heidi} > 0.05$ ;  $\text{GWAS } P < 5 \times 10^{-5}$ ; eQTL/mQTL  $P < 5 \times 10^{-8}$ ) [Supplementary Data 5; Fig. 3; Supplementary Figs. 11–15]. On chromosome 11p11.2, our SMR based integration analysis co-localized a fetal brain specific mQTL (*SPI1*) and an adult brain specific eQTL (*NUP160*) with both traits (AUD and DPW) [Supplementary Data 7]. On chromosome 17q.21.31, the integration analysis prioritized different candidate genes for AUD (*MAP3K14*) and DPW (*MAPT*, *CRHR1*, and *LRRC37A*) [Supplementary Data 4–7]. 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, 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  $P < 5 \times 10^{-15}$ ) 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 ago<sup>23</sup>. 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 DPW, while decreased expression of *LRRC37A* (SMR beta = -0.02) was associated with a decrease in DPW. 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 (Fig. 4c). The association between *MAPT* 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 *LRRC37A* and the level of alcohol intake in this brain dataset. 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 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 *SPI1* 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 ( $4.90 \times 10^{-12}$ ) 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 *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<sup>24</sup>. Since *SPI1* is expressed in myeloid lineage cells, its mRNA expression in the bulk brain was too low to perform differential expression or



**Fig. 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 meta-analysis. Y-axis represents the annotations and X-axis represents the  $-\log_{10} P$  value for enrichment calculated using partition heritability method as implemented in LDSC. The dotted red line represents the threshold of multiple test correction according to Bonferroni.

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,

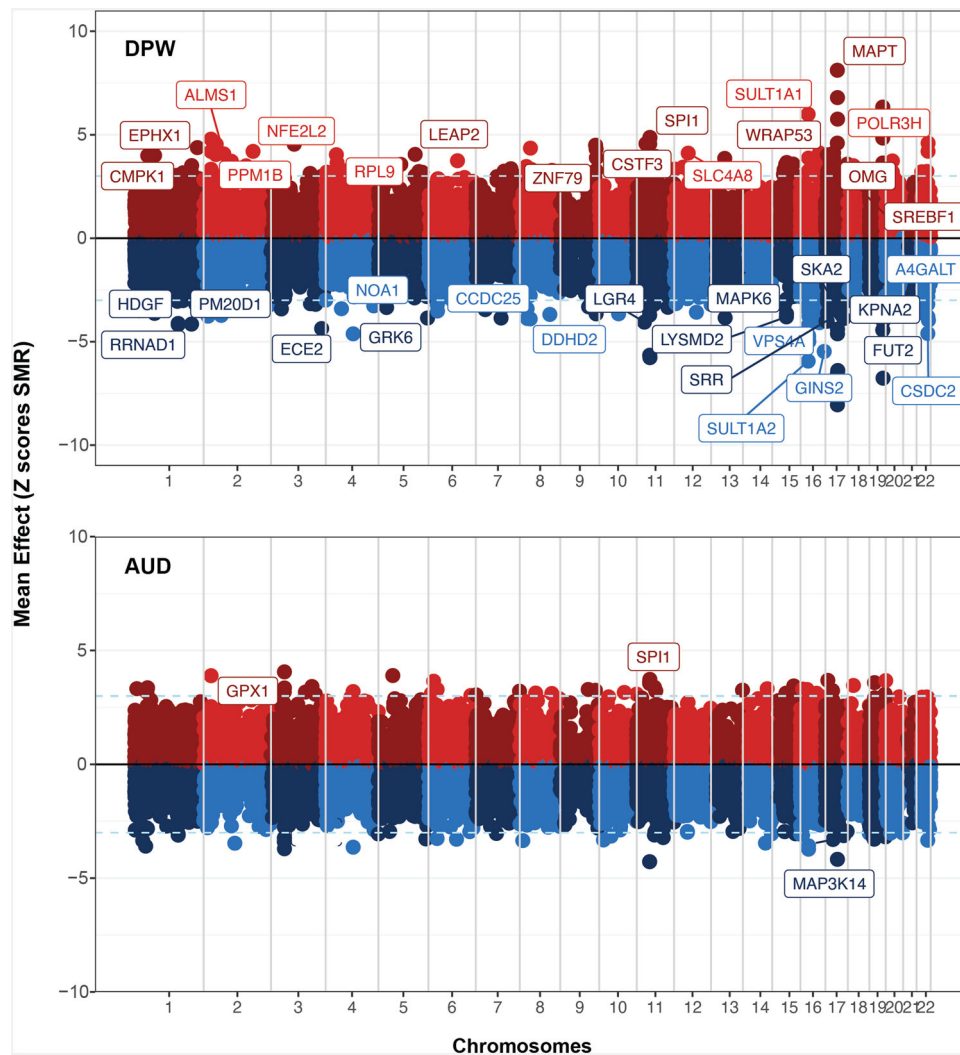
rs56030824 had the strongest effect size for *SPI1* expression and DPW in the common variant category (Fig. 5C). These observations together established rs56030824 as a stronger candidate to be considered as a causal variant and *SPI1* as a potential candidate gene associated with AUD and DPW.



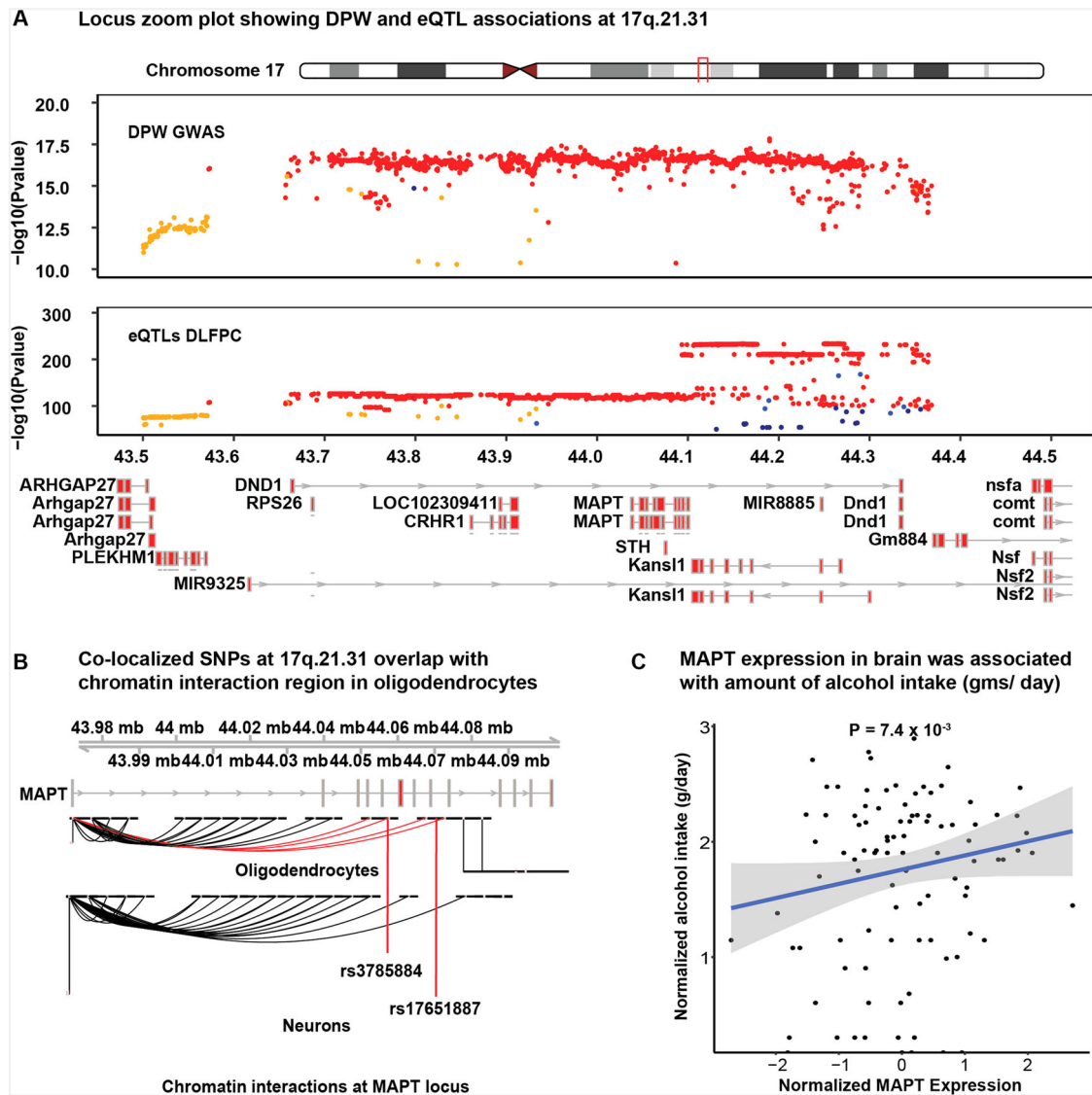
**Table 1 SMR analysis results with summary statistics of AUD GWAS meta-analysis.**

Chr	BP	Gene	Adult brain		Fetal brain	
			eQTL <i>p</i> value	mQTL <i>p</i> value	eQTL <i>p</i> value	mQTL <i>p</i> value
11	46399942	<i>SPI1</i> <sup>N</sup>	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> <sup>N</sup>	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. 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.



**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 (FDR<sub>SMR-P</sub> < 20%; SMR<sub>Heidi</sub> *P* > 0.05; GWAS *P* < 5 × 10<sup>-5</sup>; eQTL *P* < 5 × 10<sup>-8</sup>) and multiple levels of transcriptomic and/ or epigenetic evidence. SMR *P*-values for the co-localized SNPs are obtained using the Wald test.



**Fig. 4** *MAPT* was identified as a candidate gene associated with increased DPW. **A** Locus zoom plot showing DPW and eQTL (DLFCP) associations at 17q.21.3. X-axis represents the positions along chromosome 17 and the y-axis represents the  $P$  values of each SNP at this locus.  $P$ -values for the DPW GWAS are obtained from aggregated weighted Z statistics (Liu et al, 2019).  $P$ -values for DLFCP-eQTL meta-analysis were obtained using conventional inverse-variance-weighted meta-analysis as implemented in the SMR software package. 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. The analysis was performed using DeSeq2 program and  $P$ -value for association resulted from the Wald test. The shaded area around trend-line depicts the 95% confidence level intervals plotted using “geom\_smooth(method = “lm”)” from ggplot2.

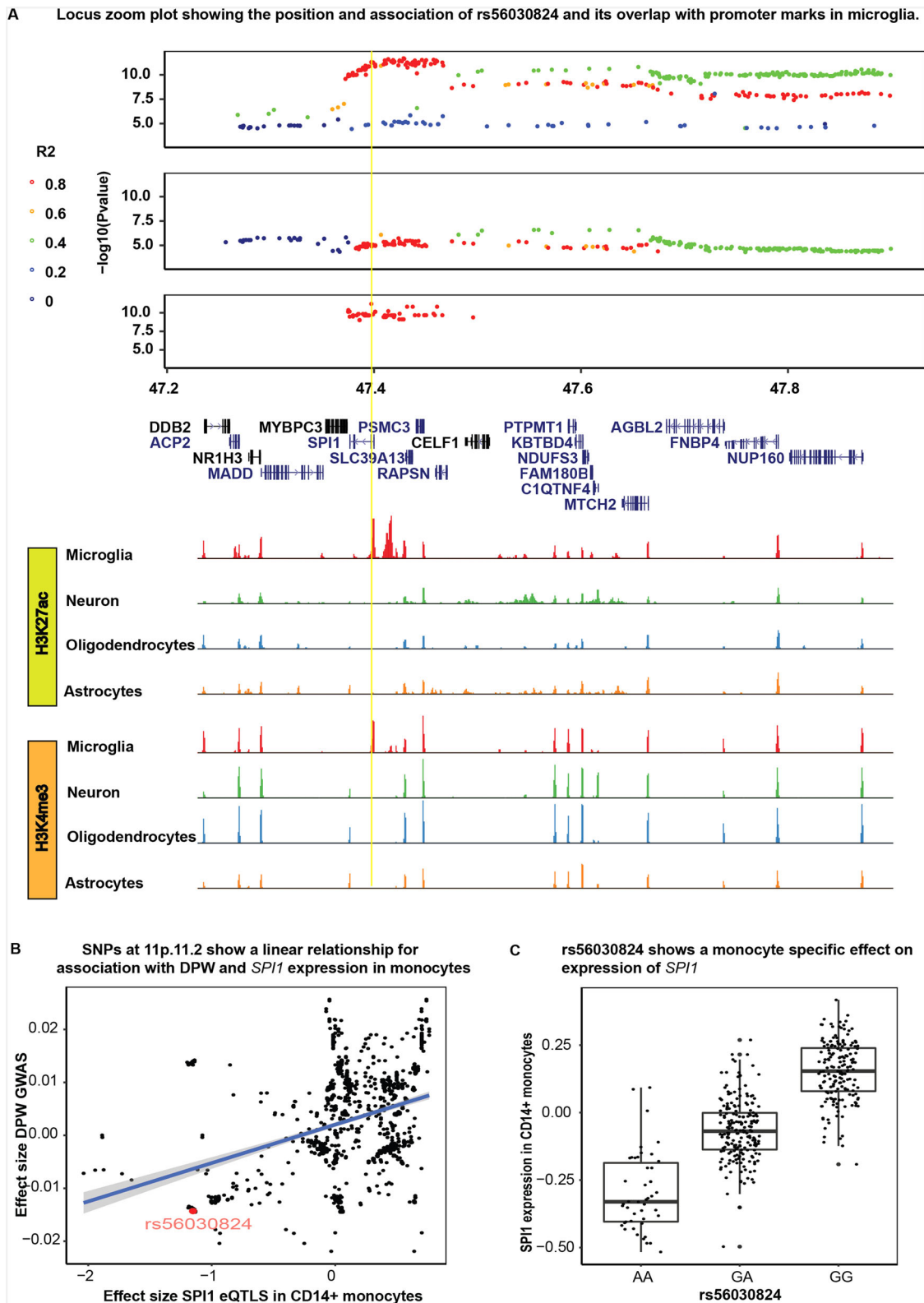
This study identified several other genes for DPW with multiple lines of evidence (eQTL, mQTL, differential expression;  $FDR < 20\%$ ; HEIDI  $P > 0.05$ ; GWAS  $P < 5 \times 10^{-5}$ ). For example, at locus 16p11.2, *SULT1A1* and *SULT1A2* were the strongest candidates with co-localization evidence 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 (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 ( $P = 1.45 \times 10^{-4}$ ), Lipoate 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 (*17q.21.31*,

*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<sup>25-31</sup>. 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  $-\log_{10}(P)$  values of each SNP at this locus.  $P$ -values for the DPW GWAS (Liu et al, 2019) and AUD GWAS 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  $R^2$  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). Yellow line represents the position of rs56030824 identified as a functional variant colocalized 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 (Sirtuin 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<sup>32</sup>. 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<sup>15,33</sup> 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 colleagues<sup>33</sup> 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<sup>15</sup>. 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-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 receptors<sup>34</sup>. In the brain, *SPI1* is specifically expressed in microglia<sup>15</sup>. 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 microglia<sup>35</sup> and peripheral lung macrophages<sup>36,37</sup>. 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<sup>38</sup>, frontotemporal dementia<sup>39</sup>, Parkinson’s disease<sup>38</sup>, and the primary tauopathies progressive supranuclear palsy and corticobasal degeneration<sup>40</sup>. Recently, Hoffman and colleagues<sup>41</sup> 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<sup>42</sup>. *CRHR1* (corticotropin-releasing hormone type I receptor) is another gene on 17q.21.31, that has been reported to be associated with alcoholism<sup>43</sup>. 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 study<sup>44</sup>. 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<sup>45</sup>. 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 Sirtuin 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<sup>46</sup>. RXRs (Retinoic X Receptor) are known to potentially regulate the ethanol metabolizing enzymes after chronic alcohol consumption<sup>47</sup>. 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 gene<sup>47</sup>. Sirtuins signaling has been shown to play an important role in cocaine and morphine Action in the Nucleus Accumbens. Ferguson and colleagues<sup>48</sup> 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  $N = 138$ ;  $N_{\text{Alc}} = 92$ ) (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)<sup>19</sup> GWAS of AUD (European [EUR]  $N = 202,004$ ;  $N_{\text{cases}} = 34,658$ ), 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<sup>12</sup> (cases based on DSM-IV diagnoses; EUR unrelated genotyped  $N = 28,757$ ;  $N_{\text{cases}} = 8485$ ) and the Collaborative Studies on Genetics of Alcoholism (COGA) GWAS of alcohol dependence (cases based on DSM-IV diagnoses; EUR unrelated genotyped  $N = 4849$ ;  $N_{\text{cases}} = 2411$ )<sup>20</sup>. 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 MVP<sup>19</sup> and COGA datasets.

**Drinks per week.** We used genome-wide summary statistics for DPW (EUR  $N = 537,349$  without the 23andMe samples) from GSCAN<sup>11</sup> to contrast with AUD.

**eQTLs from adult brain.** We meta-analyzed three eQTL datasets with data from the dorsolateral prefrontal cortex (DLFPC): PsychEncode ( $N = 1387$ )<sup>49</sup>, ROSMAP ( $N = 461$ )<sup>50</sup>, and COGA-INIA ( $N = 138$ ). ROSMAP eQTL analysis<sup>50</sup> 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 cis-eQTL summary statistics without  $P$ -value threshold restriction was downloaded from the PsychENCODE Integrative Analysis website (<http://resource.psychencode.org/>)<sup>49</sup>. 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 substructure, and technical covariates were corrected for, 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 study<sup>41</sup>. 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 gyrus 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.

**mQTLs from adult brain.** Brain-mMeta mQTL summary data<sup>51</sup> in SMR binary (BESD) format were obtained from the SMR data resource (Supplementary

Data 1). This is a set of mQTL data from a meta-analysis ( $N = 994$ ) of ROSMAP<sup>52</sup>, Hannon et al.<sup>53</sup>, and Jaffe et al.<sup>54</sup>.

**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. Bray<sup>55</sup>. 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 Mill<sup>53</sup>. 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 genes<sup>56</sup>.

**Whole-genome transcriptomic data in the brain of people with AUD and with daily alcohol intake.** mRNA expression data in the Dorsolateral Pre-frontal cortex (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 interaction (PLAC-Seq) data for four specific cell types of brain (microglia, neuron, astrocytes, and oligodendrocytes) to elucidate the functional significance of co-localized SNPs<sup>57</sup>. The location of each epigenetic mark was intersected with the location of variants prioritized by SMR analysis. UCSC browser (<https://genome.ucsc.edu/>) was used to visualize the overlap of epigenetic markers with SNPs at 17q.21.31 [MAPT] and 11p11.2 [SPI1] loci.

### Analysis

**eQTL meta-analysis in adult brain.** RNA Sequencing data on the DLFPC region of the brain for 138 samples was generated as part of COGA-INIA collaboration<sup>28</sup>. 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 cosmopolitan reference panel (Phase 3, version 5, NCBI GRCh37) using SHAPEIT then Impute2<sup>58</sup> 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  $R^2 < .30$  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. 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<sup>49</sup>, ROSMAP<sup>50</sup>, 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://cns.genomics.com/software/smr/#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 marks and performed the LDSC partition heritability analyses on munged summary statistics of AUD and DPW GWAS<sup>21</sup>. We specifically followed the default parameters to perform the LDSC analyses as listed in the tutorial for partition heritability analysis (<https://github.com/bulik/ldsc/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) analysis<sup>59</sup> on a set of mQTLs and eQTLs (Top SNP  $P < 5 \times 10^{-8}$ ) 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 scripts used to create the binary (BESD) files for each dataset have been deposited in GitHub. The European subset of ADGC GWAS (phs000372.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 ( $FDR < 20\%$ ;  $GWAS P < 5 \times 10^{-5}$ ;  $eQTL/mQTL P < 5 \times 10^{-8}$ ) 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)<sup>60</sup> approach is implemented in Genome-wide Complex Trait Analysis (GCTA) software<sup>61</sup> 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 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 log<sub>10</sub> 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 GWAS 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 GWASs. 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 ( $GWAS AUD P < 5 \times 10^{-5}$ ;  $e/mQTL P < 5 \times 10^{-8}$ ;  $SMR P < 20\%$ ;  $Heidi P > 0.05$ ). 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 ( $SMR P_{FDR} < 20\%$ ;  $Heidi P > 0.05$ ;  $P_{eQTL} < 5 \times 10^{-8}$ ;  $P_{mQTL} = 5 \times 10^{-5}$ ) 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 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 SMR analysis for all conditions can be found in supplementary information Figs. 7–15. Supplementary Figs. 1–4 reports the AUD-meta association plots generated through FUMA. Additionally all the results can be visualized at our Shiny web app ([https://lcaad.shinyapps.io/alc\\_multiomics/](https://lcaad.shinyapps.io/alc_multiomics/)). 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](https://figshare.com): AUD meta-analysis summary statistics: <https://doi.org/10.6084/m9.figshare.15054198.v1>. SMR Input

BESD 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](https://doi.org/10.6084/m9.figshare.15054171.v1). SMR results (Complete summary statistics files): <https://doi.org/10.6084/m9.figshare.15054120>.

## 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](https://github.com/kapoormanav/alc_multiomics)) [<https://doi.org/10.5281/zenodo.5076223>].

Received: 3 November 2020; Accepted: 3 August 2021;

Published online: 20 August 2021

## 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**, 34–40 (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. 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**, 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. *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 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**, 291–295 (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. Natrajan, 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. 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**, 505–515 (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**, eaba5754 (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**, 6837–6845 (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 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**, 575–577 (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**, 361–369 (2013).
44. Jia, T. et al. Neural basis of reward anticipation and its genetic determinants. *Proc. Natl Acad. Sci. USA* **113**, 3879–3884 (2016).
45. Nowell, S. & Falany, C. N. Pharmacogenetics of human cytosolic sulfotransferases. *Oncogene* **25**, 1673–1678 (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 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**, 16088–16098 (2013).
49. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (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**, 40–47 (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**, 369–375 (2012). S1–3.
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).

## 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 the 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. We also want to thank all investigators, volunteers, and participants of NSW Brain Tissue Resource Centre to provide the valuable brain samples of people with AUD and controls.

## 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.

## Competing interests

Alison Goate is on the scientific advisory board for Genentech 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. From February, 2021, Manav Kapoor is employee of Regeneron Pharmaceuticals. All other authors declare no competing interest.



**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-25392-y>.

**Correspondence** and requests for materials should be addressed to M.K. or A.G.

**Peer review information** *Nature Communications* thanks Ditte Demontis and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

---

**The Collaborative Study on the Genetics of Alcoholism (COGA)**

Victor Hesselbrock<sup>7</sup>, Samuel Kuperman<sup>8</sup>, John Kramer<sup>8</sup>, Chella Kamarajan<sup>4</sup>, Ashwini Pandey<sup>4</sup>, Laura Bierut<sup>9</sup>, John P. Rice<sup>9</sup>, Kathleen K. Bucholz<sup>9</sup>, Marc Schuckit<sup>10</sup>, Jay Tischfield<sup>11</sup>, Andrew Brooks<sup>11</sup>, Ronald P. Hart<sup>12</sup>, Laura Almasy<sup>13</sup>, Danielle Dick<sup>14</sup>, Jessica Salvatore<sup>14</sup> & Paul Slesinger<sup>15</sup>

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