Received: 23 February 2024 | Accepted: 7 June 2024

DOI: 10.1111/acer.15399

RESEARCH ARTICLE

Rare and common variants associated with alcohol consumption identify a conserved molecular network

Brittany S. Leger[1,2] | John J. Meredith[2] | Trey Ideker[3,4] | Sandra Sanchez-Roige[2,4,5] | Abraham A. Palmer[2,4]

[1]Program in Biomedical Sciences, University of California San Diego, La Jolla, California, USA

[2]Department of Psychiatry, University of California San Diego, La Jolla, California, USA

[3]Department of Medicine, University of California San Diego, La Jolla, California, USA

[4]Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA

[5]Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, Tennessee, USA

Correspondence
Sandra Sanchez-Roige and Abraham A. Palmer, Department of Psychiatry, University of California San Diego, Palmer Lab, 9500 Gilman Drive, 0667, La Jolla, CA 92093-0667, USA.
Email: sanchezroige@ucsd.edu and aap@ucsd.edu

Funding information
National Institute on Drug Abuse, Grant/Award Number: DP1 DA054394; National Institute on Alcohol Abuse and Alcoholism, Grant/Award Number: R01 AA029688; National Institute of General Medical Sciences, Grant/Award Number: GM008666

Abstract

Background: Genome-wide association studies (GWAS) have identified hundreds of common variants associated with alcohol consumption. In contrast, genetic studies of alcohol consumption that use rare variants are still in their early stages. No prior studies of alcohol consumption have examined whether common and rare variants implicate the same genes and molecular networks, leaving open the possibility that the two approaches might identify distinct biology.

Methods: To address this knowledge gap, we used publicly available alcohol consumption GWAS summary statistics (GSCAN, $N=666,978$) and whole exome sequencing data (Genebass, $N=393,099$) to identify a set of common and rare variants for alcohol consumption. We used gene-based analysis to implicate genes from common and rare variant analyses, which we then propagated onto a shared molecular network using a network colocalization procedure.

Results: Gene-based analysis of each dataset implicated 294 (common variants) and 35 (rare variants) genes, including ethanol metabolizing genes *ADH1B* and *ADH1C*, which were identified by both analyses, and *ANKRD12*, *GIGYF1*, *KIF21B*, and *STK31*, which were identified in only the rare variant analysis, but have been associated with other neuropsychiatric traits. Network colocalization revealed significant network overlap between the genes identified via common and rare variants. The shared network identified gene families that function in alcohol metabolism, including *ADH*, *ALDH*, *CYP*, and *UGT*. Seventy-one of the genes in the shared network were previously implicated in neuropsychiatric or substance use disorders but not alcohol-related behaviors (e.g. *EXOC2*, *EPM2A*, and *CACNG4*). Differential gene expression analysis showed enrichment in the liver and several brain regions.

Conclusions: Genes implicated by network colocalization identify shared biology relevant to alcohol consumption, which also underlie neuropsychiatric traits and substance use disorders that are comorbid with alcohol use, providing a more holistic understanding of two disparate sources of genetic information.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). Alcohol, Clinical and Experimental Research published by Wiley Periodicals LLC on behalf of Research Society on Alcohol.

KEYWORDS
alcohol consumption, ethanol metabolism, GWAS, network, rare variant

INTRODUCTION

Alcohol use disorder (AUD) is a highly heritable disease (Verhulst et al., 2015) with a significant public health burden (MacKillop et al., 2022). AUD can be viewed as the endpoint of a series of transitions, which begins with the initiation of use, continues with regular alcohol consumption, escalation to hazardous drinking, and culminates in compulsive harmful use that persists despite negative consequences (Sanchez-Roige et al., 2020). As such, alcohol consumption is frequently studied as a proxy for AUD, as it is a component of AUD, and is a quantitative trait that is widely measured, providing large sample sizes for genetic studies. In particular, genome-wide association studies (GWAS) have identified numerous common variants that contribute to AUD and related traits (Clarke et al., 2017; Kranzler et al., 2019; Liu et al., 2019; Mallard et al., 2022; Mallard & Sanchez-Roige, 2021; Sanchez-Roige et al., 2019; Saunders et al., 2022; Walters et al., 2018; Zhou et al., 2023), but the best powered GWAS of alcohol use are conducted in alcohol consumption (Clarke et al., 2017; Liu et al., 2019; Saunders et al., 2022). In this study, we focus on alcohol consumption both for its relationship with AUD, and its well-powered studies.

Recently, genetic studies of other psychiatric disorders have extended their reach to rare variants. One advantage of studies of rare exonic variants is that they unambiguously identify the causal gene (Sazonovs & Barrett, 2018). Such studies have already shown that rare exonic variants influence risk for multiple psychiatric disorders, including intellectual disability, autism spectrum disorder, and schizophrenia (Antaki et al., 2022; Charney et al., 2019; Fu et al., 2023; Ganna et al., 2018; Singh et al., 2022; Weiner et al., 2023).

Because they are uncommon, rare variants are not represented on genotyping microarrays and can be difficult or impossible to impute. However, they can be directly observed by sequencing (Backman et al., 2021; Karczewski et al., 2022; Manolio et al., 2009; Wang et al., 2021). Because few exome sequencing studies and rare variant studies for alcohol phenotypes have been undertaken (Ahangari et al., 2023; Curtis, 2022; Marees et al., 2018; Vrieze et al., 2014) the contribution of rare variation on alcohol behaviors remains poorly characterized, as does the relationship between common and rare variants. Even if the same genes are not identified, it is possible that shared biology might be implicated by studies of common and rare variants.

One way to identify this shared biology is by using biological knowledge networks. These networks contain information about the molecular interactions among genes and their products, both broadly and in disease contexts (Farris et al., 2015; Rosenthal et al., 2023). By defining a gene network from common and rare variants, we also produce a more holistic understanding of the biological mechanisms that are more actionable than lists of individual genes. While the interplay between rare and common variant-implicated genes has been studied in network space for other psychiatric traits (Ben-David & Shifman, 2012; Chang et al., 2018; Gilman et al., 2011), it has not been studied for alcohol-related traits or other substance use disorders (SUDs). Based on evidence from comparisons of common and rare variants for other psychiatric traits, we hypothesized that the same genes and molecular pathways would be identified by both approaches.

To test this hypothesis, we assembled data from UK Biobank (UKB) and other sources pertaining to both common and rare variants that are associated with alcohol consumption. We then used a network approach to investigate the biological overlap between common and rare variants for alcohol consumption. This approach allowed us to compare their relative contributions at the variant, gene, and molecular pathway levels.

MATERIALS AND METHODS

Please see Figure S1 for a diagram describing the generation of the alcohol consumption network from common and rare variant GWAS summary statistics.

Data and code availability

All code used for analysis and data visualization is freely available in public repositories. All original code is publicly available at https://github.com/BSLeger/rare_common_alcohol_comparison, DOI: https://zenodo.org/doi/10.5281/zenodo.11493942. The alcohol consumption network can be accessed at the network data exchange (NDEx, ndexbio.org) UUID:29b2d215-07fd-11ef-9621-005056ae23aa, DOI: https://doi.org/10.18119/N90P5K. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Data acquisition

Common variant experimental and control data acquisition

The common variant summary statistics for alcohol consumption were obtained from the *GWAS & Sequencing Consortium of Alcohol and Nicotine use*. Summary statistics were computed via a meta-analysis of GWAS results representing 666,978 individuals of

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH ISBRA

European ancestry (Saunders et al., 2022). Summary statistics for common variant negative control traits were obtained from the Neale Lab Round 2 GWAS (Abbott et al., 2018) (http://www.nealelab.is/uk-biobank). Phenotype codes of traits used are Forced Expiratory Volume per Second (FEV1, 20153_irnt) and Heel Quantitative Ultrasound Index (QUI, 4104_irnt). These negative control traits were selected as they have similar numbers of implicated genes to alcohol consumption ($250 < n < 350$), similar SNP heritability to alcohol consumption ($h^2 > 0.30$), and minimal genetic correlation with a comparable alcohol consumption trait (Amount of Alcohol Drunk on a Typical Drinking Day (phenotype code: 20403)) ($|r_g| < 0.38$). These estimates were obtained from the UKB Heritability browser (https://nealelab.github.io/UKBB_ldsc/h2_browser.html) and UKB Genetic Correlation browser (https://ukbb-rg.hail.is/), both generated by the Neale Lab (Abbott et al., 2018).

Rare variant and gene experimental and control data acquisition and filtering

Rare variant summary statistics data was downloaded from Genebass's Hail library (gs://ukbb-exome-public/500 k/results/variant_results.mt) and queried for alcohol consumption by phenotype code (alcohol_intake_custom) using Hail (https://hail.is). To control against statistical artifacts, we selected alcohol consumption rare variants that had both MAF < 0.05 and minor allele count > 2.

Rare variant gene-level summary statistics were downloaded from Genebass browser (https://app.genebass.org/). Due to limited power for computing genetic correlation and heritability from rare variants for some traits (Karczewski et al., 2022; Weiner et al., 2023), rare variant controls were chosen based on heritability and genetic correlations calculated from common variants, as described above. As such, rare variant control traits were chosen if they had non-zero heritability ($h^2 > 0.01$), minimal genetic correlation with a comparable alcohol consumption trait (The Amount of Alcohol Drunk on a Typical Drinking Day (phenotype code: 20403)) ($0.0 < |r_g| < 0.2$) and had the minimum number of rare seed genes recommended for network propagation using NetColoc ($n > 5$).

Phenotype codes for traits from Genebass that were used in this study are as follows: Alcohol Consumption (alcohol_intake_custom), FEV1: Forced Expiratory Volume per Second (20153), Pulse Rate (4194), Heel Bone Mineral Density (BMD) *T*-score, automated (78), Other Malignant Neoplasms of Skin (C44), and Malignant Neoplasm of Breast (C50).

For all traits, rare seed genes were defined as any gene that had False Discovery Rate-corrected $q_{\text{burden}} < 0.25$, or $q_{\text{SKAT-O}} < 0.25$, or $q_{\text{SKAT}} < 0.25$, calculated for each individual dataset. For alcohol consumption, these cutoffs were $p_{\text{SKAT-O}} < 1.5 \times 10^{-4}$, $p_{\text{burden}} < 1.1 \times 10^{-4}$, or $p_{\text{SKAT}} < 2.7 \times 10^{-5}$. The stringent alcohol consumption rare seed genes were defined as any genes that were significant based on the recommended significance threshold from Genebass ($p_{\text{SKAT-O}} < 2.5 \times 10^{-7}$; $p_{\text{burden}} < 6.7 \times 10^{-7}$; $p_{\text{SKAT}} < 2.5 \times 10^{-7}$) (Karczewski et al., 2022).

Common variant gene mapping

We generated gene-level significance values from the SNP-level summary statistics using MAGMA v1.10 (de Leeuw et al., 2015) using default parameters (https://ctg.cncr.nl/software/magma); annotation windows were 10 kb, the 1000 Genomes European sample was used for the reference panel (1000 Genomes Project Consortium, 2015), and the Hg38 Gene locations were used for mapping (Schneider et al., 2017). MAGMA projects the SNP matrix for a gene onto its principal components and uses the principal components as predictors for the phenotype using linear regression. Association of the gene to the phenotype using the principal component SNP matrix is used to calculate an *F* statistic, which is used to calculate the *p*-value for the individual genes. Bonferroni correction was used to define significant gene-level associations ($p < 2.6 \times 10^{-6}$).

Molecular interaction networks

The Parsimonious Composite Network (PCNet) (Huang et al., 2018; Wright et al., 2024) v1.4 was obtained from NDEx, UUID: c3554b4e-8c81-11ed-a157-005056ae23aa. PCNet is a molecular interaction resource formed from integrating 21 interaction databases that contain various evidence types, including physical protein–protein, genetic, co-expression, and co-citation evidence. Each interaction in PCNet is supported by at least two of the component databases, a threshold chosen to maximize the ability of PCNet to perform gene set recovery tasks via network propagation. All seed genes were mapped to the nodes of PCNet via gene symbols.

Generation of the alcohol consumption network

Network propagation and co-localization

We used the Python package NetColoc (Rosenthal et al., 2023) (https://pypi.org/project/netcoloc/) for network propagation and co-localization. The sets of significant trait-associated genes from GWAS were used as “seed” genes for network propagation using a Random Walk with Restart (Vanunu et al., 2010) algorithm. Following network propagation with $\alpha = 0.5$, we calculated a network proximity score (NPS) for each gene in the network by comparing the observed results to a null distribution, formed by propagating 1000 randomly selected seed gene sets. Each set was sampled to preserve the size and degree distribution of the original input set. As previously implemented (Rosenthal et al., 2021, 2023; Wright et al., 2023), we binned all genes in the network by degree with a minimum of 10 nodes per bin. For each gene, the NPS was calculated as a *z*-score comparing the observed heat at that gene after network propagation of the gene set, to the mean of the null distribution heats at that gene. All heat values are log-transformed to ensure the distributions are approximately normal.

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

From input seed genes calculated from common and rare variants, we independently calculated NPS_{common} and NPS_{rare} for each trait. We then defined a gene as colocalized between both if it had high proximity to both input sets. Therefore, we defined the combined network proximity $NPS_{common\text{-}rare}$ as the product of the independent species vectors: $NPS_{common-rare} = NPS_{common} * NPS_{rare}$.

NetColoc recommends fewer than 500 input seed genes given the sample space of PCNet (~18,000 genes). Therefore, as described previously (Rosenthal et al., 2023), we employed a weighted sampling procedure for any trait having more than 500 significantly associated genes. We sampled 500 genes from the set of all significant genes (weighted by $-\log_{10}(p)$ from GWAS) and ran the propagation analysis from this subset. After 100 repetitions, the 75% percentile NPS score was selected to approximate a consensus score for each gene.

Definition of the alcohol consumption network

From these NPS scores, we selected genes with high proximity scores from both common and rare sources to define the network using the following thresholds: $NPS_{common\text{-}rare} > 3$, $NPS_{common} > 1.5$, and $NPS_{rare} > 1.5$. NPS cutoffs were selected after a sensitivity analysis using seed genes from the alcohol consumption network (Figure S2). These cutoffs were used as they had the most significant network overlap compared with a permuted control ($p<0.05$, Bonferroni corrected), while having a moderate to high observed/expected network overlap, and moderate network size (N nodes > 100). As previously shown, these parameters performed well for other datasets as well (Rosenthal et al., 2023). To calculate the significance of the network co-localization, we compared the conserved network size and the mean $NPS_{common\text{-}rare}$ to a permuted null distribution. We permuted the labels of NPS_{rare} and NPS_{common} 10,000 times, each time calculating the mean $NPS_{common\text{-}rare}$ across all genes and the number of genes passing the above thresholds. For genes present in both input sets, labels were permuted separately to maintain the higher expected distribution for these genes. The significance of the conserved network size and mean $NPS_{common\text{-}rare}$ was calculated using the Z-test.

Network figures were formatted using Cytoscape v3.10.0 (Shannon et al., 2003).

Validation and functional annotation

GWAS catalog

To identify previously annotated genes, we used GWAS findings aggregated by the GWAS catalog (https://www.ebi.ac.uk/gwas/). We used the GWAS catalog's gene level associations version v1.0.2-associations_e111_r2024-04-22. We identified genes that had previously been associated with categories of interest by querying the Mapped Traits and the Disease/Traits, listed as "Mapped Trait: Disease/Trait (Pubmed ID)" in Tables S1, S3–S5. Traits were grouped into categories alcohol use, smoking and nicotine use, non-alcohol or smoking SUDs (for example, opioid use disorder), and non-SUD neurological and neuropsychiatric (including cognitive, mental health and psychiatric, and neuro-degenerative). All traits from the GWAS catalog identified in each group are listed in Table S1. All groups are mutually exclusive. In Tables S3–S5, traits are listed only once per gene for readability. Enrichment for each group was calculated using a hypergeometric test.

Rare variant PheWAS

To assess the association of network genes with other phenotypes through rare variants, gene level PheWAS results were downloaded from Genebass's Hail database (gs://ukbb-exome-public/500k/results/results.mt). PheWAS associations were mapped to network genes. Network genes were determined to be significantly associated with a phenotype using the same p-value cutoffs as used for lenient seed genes from alcohol consumption ($p_{SKAT\text{-}O}<1.5\times10^{-4}$; $p_{burden}<1.1\times10^{-4}$, $p_{SKAT}<2.7\times10^{-5}$).

Tissue enrichment

To assess the tissue-specific expression of network genes, we used the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) suite's in-browser gene to function tool (Watanabe et al., 2017). We used FUMA to calculate the enrichment of gene sets for 54 tissue types from human GTEx v8 using default parameters (GTEx Consortium, 2020). As described previously, this method takes normalized gene expressions (reads per kilobase per million, RPKM) from each GTEx tissue, and maps these genes to entrez ID (Watanabe et al., 2017). Pre-calculated differentially expressed genes (DEGs) sets were defined using a two-sided t-test per label versus all remaining tissue types. Genes with a Bonferroni corrected p-value < 0.05 and absolute log fold change ≥ 0.58 were selected as DEG. For the signed DEG, the direction of expression was taken into account. The $-\log_{10}$(p-values) in the graph were calculated by hypergeometric test (Watanabe et al., 2017).

RESULTS

Common and rare variant associations

We obtained GWAS summary statistics from GSCAN, which recently performed a meta-analysis of alcohol consumption, as measured in drinks per week, in Europeans (Saunders et al., 2022) ($n=666{,}978$, Figure 1A). In total, 501 independent common (MAF >0.05) variants were significantly associated with alcohol consumption ($p<5\times10^{-8}$) (Saunders et al., 2022). Genome-wide significant rare variants were obtained from Genebass's recent analysis of alcohol consumption,

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

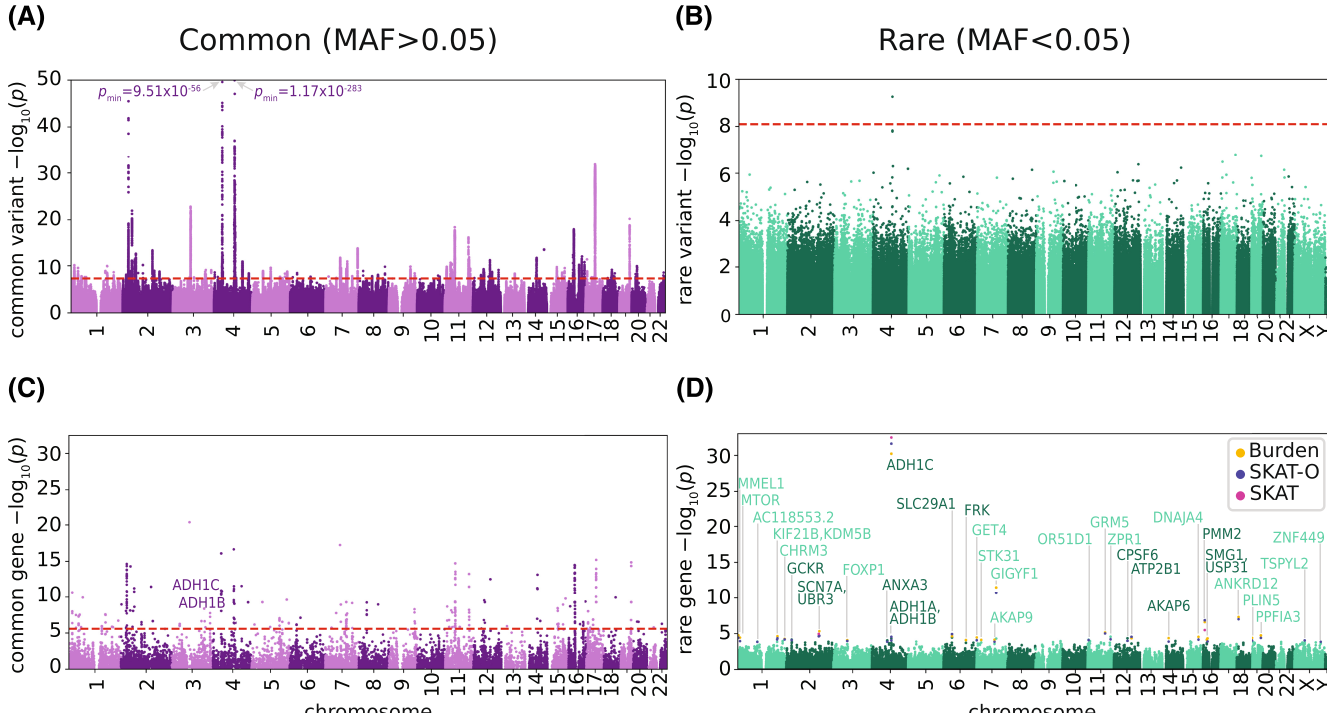

FIGURE 1 Common and rare variants mediate alcohol consumption. Manhattan plot of (A) common variants and (B) rare variants associated with alcohol consumption. Significance cutoff indicated in red (common: $p<5\times10^{-8}$; rare: $p<8\times10^{-9}$). *p*-values for peaks outside of range labeled. Only rare variants with minor allele count > 2 are shown. (C) Manhattan plot of alcohol consumption common variant-implicated genes. Significance cutoff ($p<2.6\times10^{-6}$) indicated in red. Significant genes that overlap with rare-variant-implicated genes are labeled. (D) Porcupine plot of genes calculated by burden test, SKAT-O, and SKAT algorithms from rare variants. Significantly associated genes ($q<0.25$) for each test are labeled and colored in yellow, blue, and pink, for burden, SKAT-O, and SKAT, respectively. See Figure S3A for individual Manhattan plots for each test.

as measured in grams/day, in 393,099 individuals from the UKB (Karczewski et al., 2022) (Figure 1B). Three rare variants were significantly associated with alcohol consumption ($p<8\times10^{-9}$): one potential loss of function (pLoF) variant in *ADH1C*, a missense variant in *ADH1B*, and a synonymous mutation in *C4orf54* (Table S2). The mutations in *ADH1C* and *C4orf54* were both protective. *ADH1C* and *ADH1B* both have known roles in ethanol metabolism (Le Daré et al., 2019; Tolstrup et al., 2008), but despite *C4orf54* being associated with addiction risk in prior GWAS (Hatoum et al., 2023), its function is poorly understood.

Common and rare gene-level associations

Common loci were assigned to genes based on proximity using MAGMA (de Leeuw et al., 2015), identifying 294 genes (Figure 1C; Table S3, $p<2.6\times10^{-6}$). Rare variants were previously aggregated (Karczewski et al., 2022) into gene level associations using SKAT, SKAT-O, and a gene burden test (Karczewski et al., 2022). These tests identified four genes that were significantly correlated with alcohol consumption via both SKAT-O and gene burden tests ($p_{\text{SKAT-O}}<2.5\times10^{-7}$; $p_{\text{burden}}<6.7\times10^{-7}$): *ADH1C*, *PMM2*, *GIGYF1*, and *ANKRD12*. Only *ADH1C* was significantly associated by SKAT ($p_{\text{SKAT}}<2.5\times10^{-7}$), and was the only one of these four genes that had been previously associated with alcohol-related traits by common gene analysis.

We also considered a more lenient cutoff for genes from rare variants ($q<0.25$, Figure 1D; Figure S3A; Table S4), which identified 35 genes across all tests. A total of 20 genes were identified by both SKAT-O and the burden test (Figure S3B), however, only *ADH1C* and *PMM2* were significant in all tests. 51% of genes were functionally annotated as loss of function, followed by missense and low confidence loss of function (40%), and the remaining 9% as synonymous (Figure S3C). Fourteen of these genes had previously been identified by common variants as mediating alcohol consumption and alcohol use traits in the GWAS catalog (Sollis et al., 2023) (Table S4; $p=8.24\times10^{-33}$, hypergeometric test). This includes alcohol dehydrogenase genes *ADH1A*, *ADH1B*, and *ADH1C*, and signaling genes *FOXP1*, *AKAP6*, *AKAP9*, and *GRM5*, highlighting the overlapping regulation of SUDs and psychiatric traits.

ADH1B and *ADH1C* were identified by both the rare and common gene-based analyses (Figure 2A, $p=0.01$, hypergeometric test).

Generation of the alcohol consumption network

Next, we examined the molecular pathways wherein these alcohol consumption genes function (Figure S1). We used PCNet, a resource

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

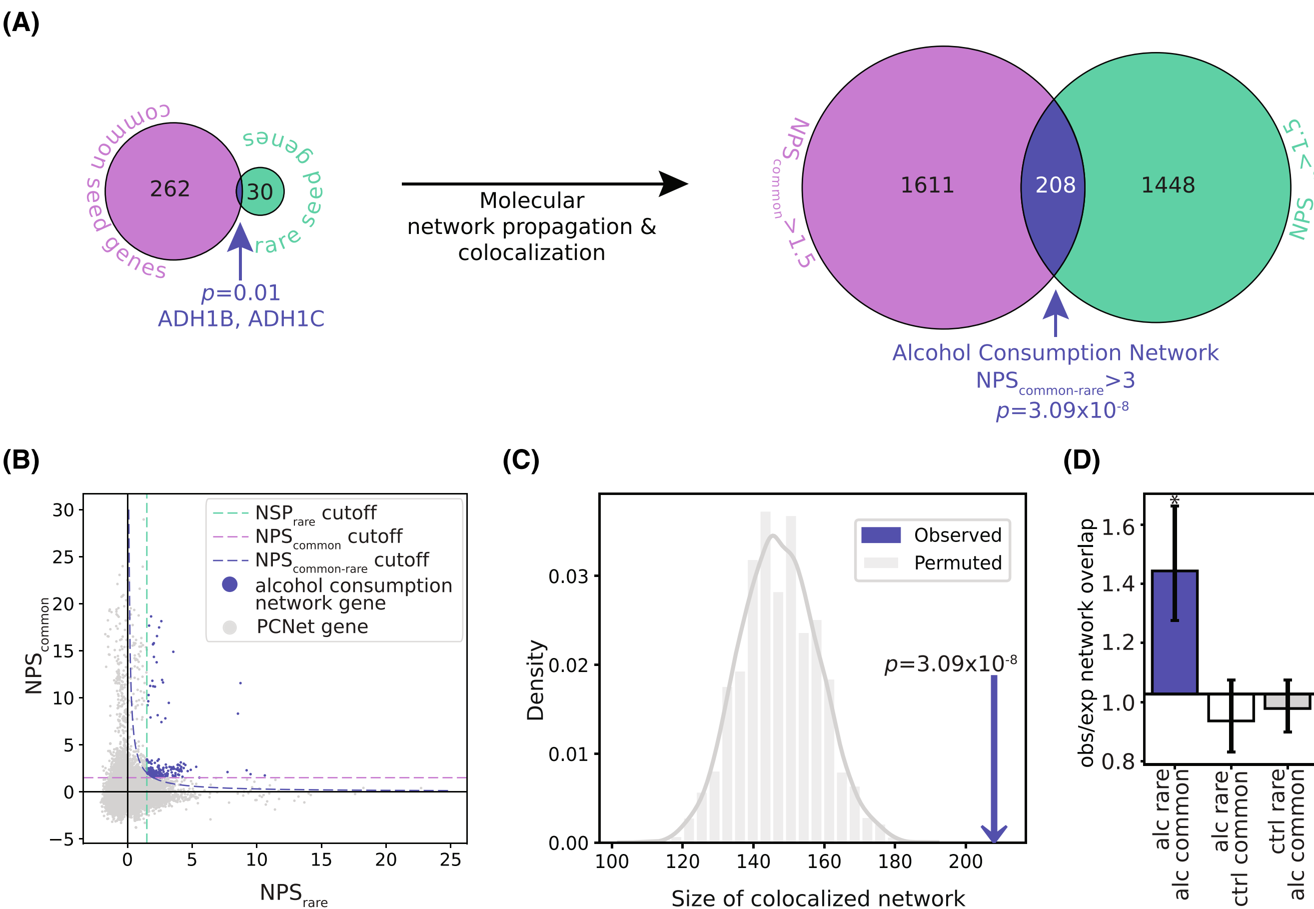

FIGURE 2 Convergence of rare and common variants on the network level. (A) Left, venn diagram showing overlap of common seed genes (purple) and rare seed genes (green). Overlapping genes are indicated in dark blue and labeled. Significance of overlap calculated via hypergeometric test. Right, venn diagram of genes passing NPS thresholds after network colocalization. Purple indicates genes with $NPS_{common} > 1.5$, green indicates $NPS_{rare} > 1.5$, dark blue indicates genes with $NPS_{common\text{-}rare} > 3$, $NPS_{common} > 1.5$, and $NPS_{rare} > 1.5$. Significance of intersection calculated in C. (B) NPS_{common} and NPS_{rare} for all genes in PCNet, with genes passing all thresholds for the alcohol consumption network ($NPS_{common\text{-}rare} > 3$, $NPS_{common} > 1.5$, and $NPS_{rare} > 1.5$) shown in dark blue. Dotted lines indicate NPS thresholds. (C) Observed (dark blue arrow) versus expected (gray distribution) size of the alcohol consumption network following 10,000 permutations of NPS labels. p-value calculated via Z-test. (D) The observed-to-expected ratio of colocalized network size for networks calculated from common and rare seed genes from alcohol consumption and from control trait FEV1 (Forced Expiratory Volume per Second). Vertical bars indicate 95% confidence intervals. Significance calculated by Z-test, Bonferroni corrected. * indicates $p = 3.09 \times 10^{-8}$. See also Figure S4C and Table S7 for additional controls.

of 2.7 million pairwise associations among genes (Huang et al., 2018; Wright et al., 2024). We found that 264 common seed genes and 32 rare seed genes were present in PCNet. NPS were generated for each gene in PCNet based on their proximity to the common (NPS_{common}) and rare (NPS_{rare}) seed genes (Figure S4A, Table S5). We used these NPS scores, and their product, $NPS_{common\text{-}rare}$, to define the alcohol consumption network (Figure 2B), thus identifying genes that were close in the molecular network to both common and rare seed genes, even if they were not identified by the individual studies (Table S6).

We found that the alcohol consumption network contained significantly more genes (Figure 2C, $p = 3.09 \times 10^{-8}$) and that the mean of $NPS_{common\text{-}rare}$ was significantly higher (Figure S4B, $p = 5.51 \times 10^{-6}$) than expected under the null. As a negative control, we produced networks using both the alcohol rare and common seed genes in conjunction with arbitrary traits; these negative controls did not produce networks that were larger than the permuted control (Figure 2D; Table S7). Additionally, when we considered a more stringent threshold for rare seed genes ($p_{SKAT\text{-}O} < 2.5 \times 10^{-7}$, $n = 4$) we obtained similar results (Figure S4C). However, exclusion of shared seed gene ADH1C removed the significance of the network overlap shown in Figure 2C. This reinforces the importance of genes that encode alcohol metabolizing enzymes in this network.

As shown in Figure 3, the alcohol consumption network contained 208 nodes, connected by 1226 edges. Only 27 of 264 seed genes from common and five of the 34 seed genes from rare variants were included in the alcohol consumption network. *ADH1B* and *ADH1C*, which were seed genes for both common and rare, were both present in the network (Table S6).

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH ISBRA

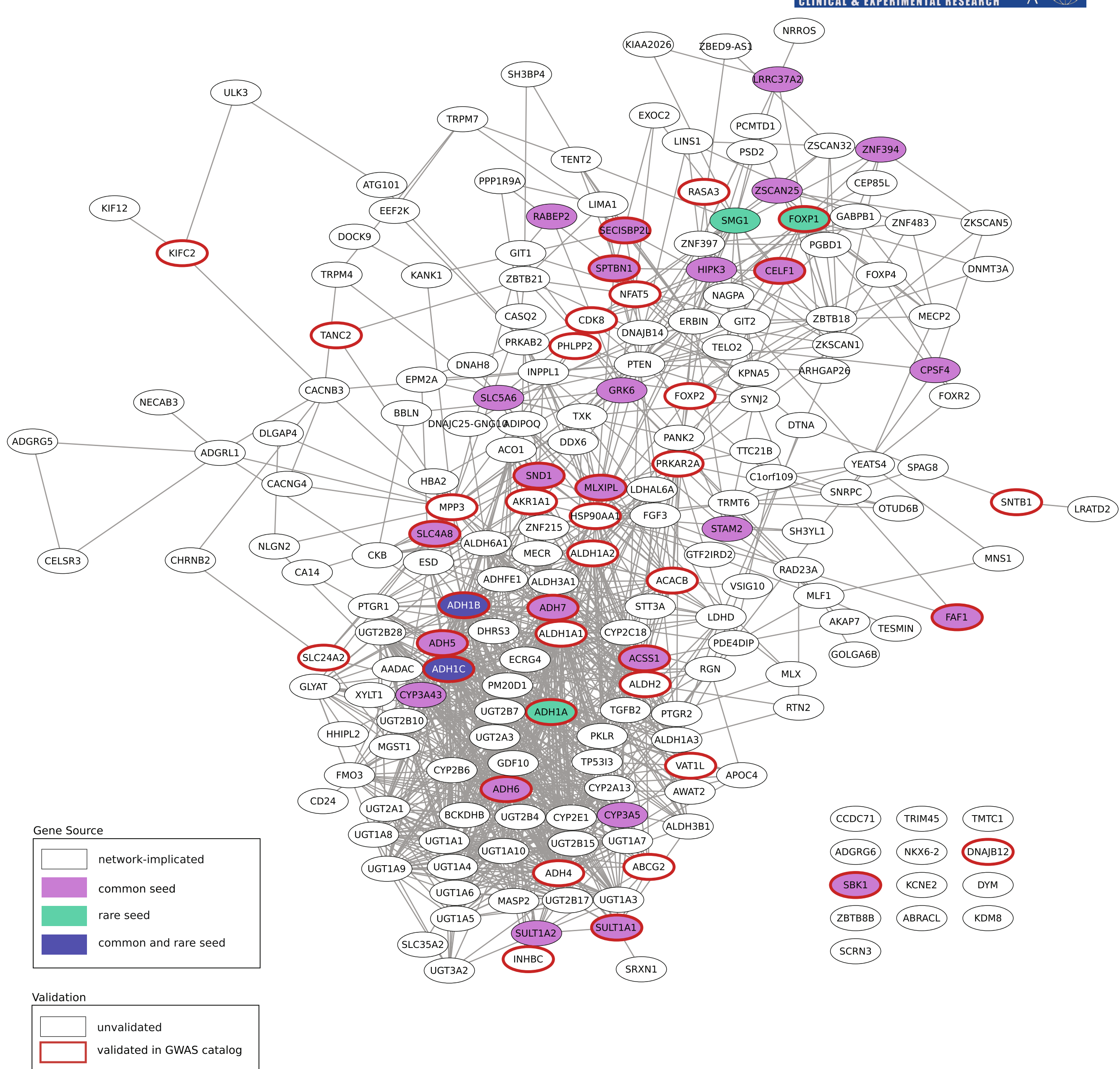

FIGURE 3 The alcohol consumption network. Subnetwork of PCNet including all genes proximal to both rare and common alcohol consumption seed genes. Purple nodes indicate common seed genes, green nodes indicate rare seed genes, dark blue nodes indicate seeds from both sources, and white nodes are network-implicated genes. Edges maintained from PCNet. Red outlined nodes have previously been annotated in the GWAS catalog for alcohol use traits.

The structure of the alcohol consumption network

One of the goals of generating the network shown in Figure 3 is to identify the underlying biology identified by common and rare seed genes. Several gene families previously known to play a role in ethanol metabolism were present in the network (Figure S5). For example, eight genes from the alcohol dehydrogenase (*ADH*) family (Le Daré et al., 2019) and seven aldehyde dehydrogenase (*ALDH*) family genes (Edenberg, 2007) were in the network. Six cytochrome P450 (*CYP*) genes, which mediate about 10% of alcohol metabolism via the microsomal pathway (Hamitouche et al., 2006), were also in the network. In addition, genes from the non-oxidative ethanol metabolism pathways, which primarily function in phase II drug metabolism (Le Daré et al., 2019), were also present. This includes two sulfotransferase (*SULT*) family genes, which metabolize ethanol into ethyl sulfate, and 18 genes in the UDP-Glycosyltransferase (*UGT*) superfamily, whose encoded proteins glucuronidate ethanol into ethyl glucuronide, a minor non-oxidative metabolite of ethanol (Walsham & Sherwood, 2014). Thus, the network recapitulates previously known biologies relevant to ethanol metabolism.

Another benefit of the network is the ability to identify relevant tissues. We found 25 tissues that were significantly enriched for

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

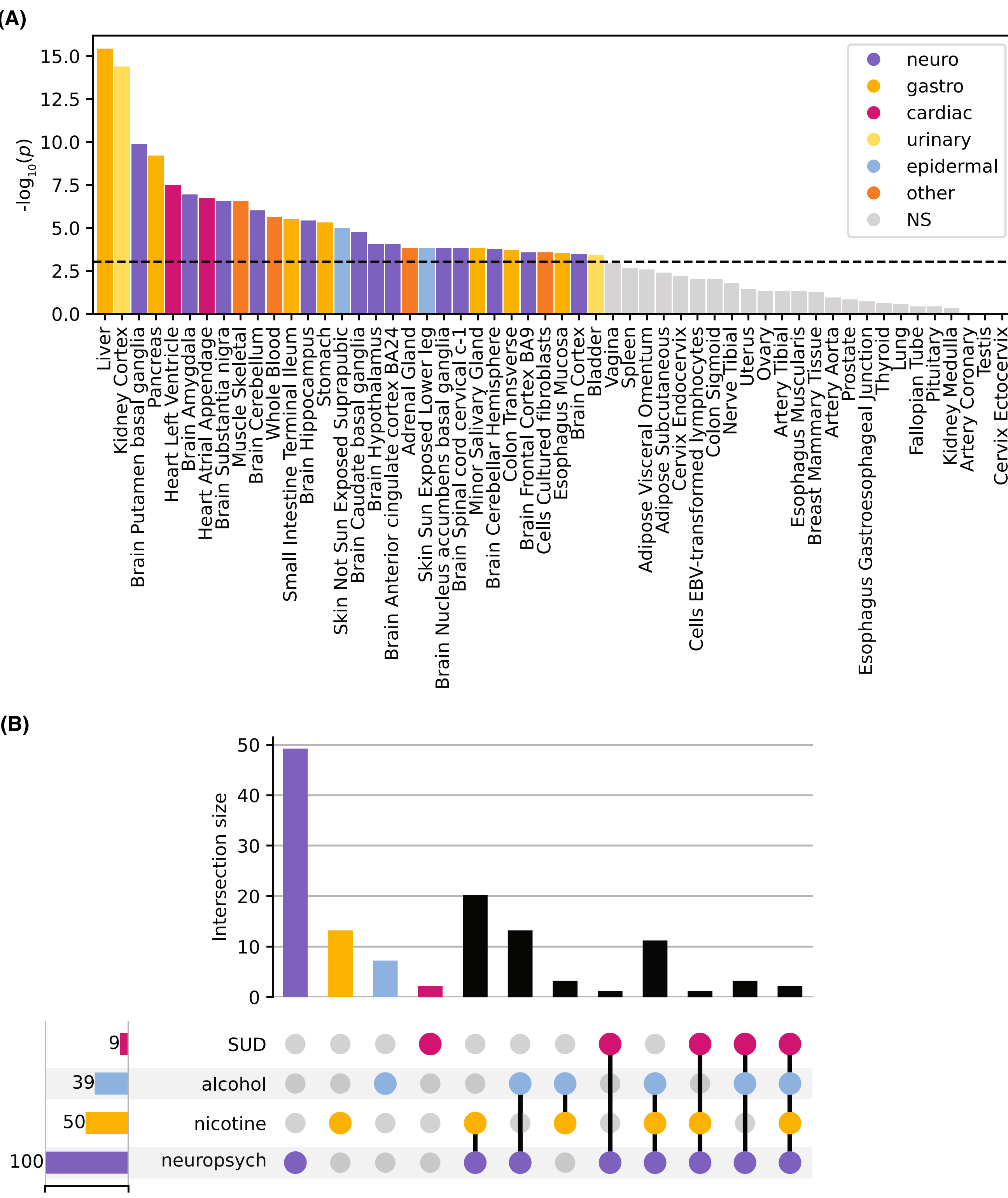

FIGURE 4 Validation of alcohol consumption network. (A) Enrichment of gene sets from the alcohol consumption network with bi-directional differential expression for 54 tissues from GTEx v8. DEG sets were defined by a two-sided *t*-tests per label, versus all remaining tissue types. Genes with $p < 9.26 \times 10^{-4}$ (Bonferroni corrected) and absolute log fold change ≥ 0.58 are selected as differentially expressed. Significance was calculated via hypergeometric test. Tissues are colored by type, non-significant (NS) associations are indicated in gray. (B) Upset plot showing the overlap of genes in the alcohol consumption network that have previously been annotated in the GWAS catalog for alcohol use traits, nicotine use and smoking traits, other SUDs, and neuropsychiatric traits.

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ALCOHOL CLINICAL & EXPERIMENTAL RESEARCH ISBRA

differential gene expression (Figure 4A; Table S6). Consistent with the presence of genes involved in ethanol metabolism in the network, the highest enrichment was in the liver and consisted of 115 genes, including 28 genes from the *ADH*, *ALDH*, *UGT*, *CYP*, and *SULT* families. In addition to the liver, numerous gastrointestinal tissues were also enriched: the gastrointestinal tract mediates absorption and gastric metabolism of alcohol, and chronic alcohol consumption may lead to inflammation and increased risk of gastrointestinal and esophageal cancers (Bode & Bode, 1997; Edenberg, 2007). As expected, all brain tissues were significantly enriched.

To determine whether the genes had been previously implicated by common variants in alcohol use, other SUDs, and related psychiatric disorders, we examined annotations from the GWAS catalog (Sollis et al., 2023). Specifically, we considered annotations for alcohol use, smoking and nicotine use, other SUDs, including opioid, cannabis, and polysubstance use, and neuropsychiatric disorders (Figure 4B). In total, 203 of the 208 genes in the network are annotated in the GWAS catalog. Of these, 39 have been previously associated in alcohol use ($p=9.2\times10^{-5}$, OR = 2.01, hypergeometric test), 50 for smoking traits ($p=3.4\times10^{-3}$, OR = 1.56, hypergeometric test) (Karlsson Linnér et al., 2021) 9 for SUDs ($p=0.048$, OR = 1.70, hypergeometric test), and 129 for other neuropsychiatric traits ($p=2.1\times10^{-3}$, OR = 1.45, hypergeometric test). Of the genes associated with these traits, many had annotations in multiple categories, such as *EPM2A*, *EXOC2*, *NFAT5*, and *SNTB1*. These findings highlight the neuropsychiatric function of the network and point to a shared underlying mechanism across alcohol and polysubstance use.

Finally, to determine whether these genes had been previously implicated by rare variants in alcohol use, we examined gene-level annotations from Genebass of genes in the network (Karczewski et al., 2022). In total, six of the 208 network genes (*ADH1C*, *AKAP7*, *ATG101*, *DTNA*, *NKX6-2*, and *SYNJ2*) were associated with secondary alcohol use traits by rare variants, including use status and frequency of use, negative societal impacts from use, and alcoholic liver disease (Table S6). Only *ADH1C* was also associated with alcohol use traits by common variants. Notably, all of these genes, excluding *ATG101*, were associated with other SUDs and neuropsychiatric traits through common variants.

DISCUSSION

The contribution of common variants in mediating alcohol consumption has been well documented, while rare variants represent a newer frontier that has recently become feasible due to the availability of large scale sequencing data. Prior rare variant analysis used in our study identified 4 genes at a stringent ($p_{\text{SKAT-O}}<2.5\times10^{-7}$; $p_{\text{burden}}<6.7\times10^{-7}$) and 35 genes at a lenient threshold ($q<0.25$), demonstrating the importance of rare variants for alcohol-related behaviors (Figure 1). We combined the findings from common and rare variants to determine whether they identify convergent biological networks (Figure 2). We identified a highly significant network (Figure 3). The network emphasized the role of ethanol metabolism, which was further supported by the tissue specific enrichment in both brain and liver (Figure 4), consistent with decades of research on the genetics of alcohol consumption.

The role of common variants in ethanol metabolizing enzymes is well established for alcohol consumption and related traits (Sanchez-Roige et al., 2020). Similarly, rare variant analysis of alcohol consumption identified *ADH1A*, *ADH1B*, and *ADH1C*, which have well documented roles in ethanol metabolism (Edenberg, 2007). By the joint analysis of common and rare variants, the network further identified genes for both oxidative and non-oxidative ethanol metabolism, including *ADH*, *ALDH*, *UGT*, *CYP*, and *SULT* family genes. Ethanol is primarily metabolized in the liver, but is also metabolized by the stomach and the brain (Zakhari, 2006), which was reflected in the high enrichment of network genes in the liver, gastrointestinal tissues, and the brain. Disulfiram is an effective treatment for AUD due to its inhibition of ALDH enzymes (Lanz et al., 2023), a gene family that is prevalent in the alcohol consumption network, suggesting the possibility that other genes identified by our network might also be viable pharmacological targets.

In addition to ethanol metabolism, genes found by our analyses have also been associated with neuropsychiatric conditions that are correlated (Walters et al., 2018) and highly comorbid with AUD, such as depression, schizophrenia, bipolar disorder, neuroticism, and cognitive dysfunction (Tables S3, S4 and S6). For example, the rare variant analysis identified *KIF21B*, which has been associated with smoking initiation (Saunders et al., 2022), ADHD (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013), and schizophrenia (Trubetskoy et al., 2022). *GIGYF1* has been associated with Alzheimer's disease and schizophrenia (Ding et al., 2023; Sollis et al., 2023). Finally, *SCN7A* has been associated with response to cognitive behavioral therapy in unipolar depression and educational attainment (Okbay et al., 2022; Rayner et al., 2019). Similarly, the alcohol consumption network identified genes that have also been associated with neuropsychiatric conditions, such as genes from the *FOXP* family (i.e., *FOXP1*, *FOXP2*, *FOXP4*; Davies et al., 2018; Sherva et al., 2023). Another example is *CACNB3* and *CACNG4*, calcium channel genes that have been associated with bipolar disorder and major depression (Marshe et al., 2021; Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2012). Finally, the gene *ADGRG6*, which was identified by the alcohol consumption network, has been associated with depression and smoking initiation (Saunders et al., 2022; Yao et al., 2021). Integrative analyses may help clarify the shared mechanisms of these conditions, but together this emphasizes shared genetic susceptibility across these traits.

While this study found that common and rare variants that were associated with alcohol consumption identified a shared network, there are several limitations to consider. We found that *ADH1C* is needed for network colocalization, showing that it is a hub gene for this network; this may reflect the need for increased power to detect additional rare variants. We only studied alcohol consumption, however as sample sizes grow, future studies should consider other AUD-relevant phenotypes such as AUD and problematic alcohol use. Similarly, methods for mapping common SNPs to genes

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

are imperfect; we used MAGMA, but other more or less stringent methods might produce different results. Additionally, we used a lenient significance threshold to select rare variants ($q > 0.25$), which likely introduced some false positives into the network analysis. However, we repeated this analysis with a more stringent cutoff for rare variants ($p_{SKAT\text{-}O} < 2.5 \times 10^{-7}$) and found little change in significance of network overlap. Additionally, NetColoc is robust to false positives, but functions best with a moderate number of input genes (Rosenthal et al., 2023). Larger sample sizes will increase power for both common and rare variant discovery, improving the ratio of true to false positive findings in the future.

While future improvements to our methodology and the underlying data will improve our ability to understand rare and common variant interaction, this work identified the first gene network from common and rare variants of alcohol consumption.

AUTHOR CONTRIBUTIONS

BSL, SSR, and AAP conceptualized the study. BSL and JJM performed the analysis. All authors wrote and edited the manuscript.

ACKNOWLEDGMENTS

Montana Kay Lara contributed scientific input to this manuscript. BSL was supported in part by NIGMS T32 GM008666. SSR was supported by NIH/NIDA DP1 DA054394. SSR and AAP were supported by NIAAA R01 AA029688.

FUNDING INFORMATION

BSL was supported in part by NIGMS T32 GM008666. SSR was supported by NIH/NIDA DP1DA054394. SSR and AAP were supported by NIAAA R01 AA029688.

CONFLICT OF INTEREST STATEMENT

TI is a co-founder, member of the advisory board, and has an equity interest in Data4Cure and Serinus Biosciences. TI is a consultant for and has an equity interest in Ideaya Biosciences. The terms of these arrangements have been reviewed and approved by the University of California San Diego in accordance with its conflict-of-interest policies.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Brittany S. Leger https://orcid.org/0000-0002-8467-5472
John J. Meredith https://orcid.org/0000-0001-6622-6414
Trey Ideker https://orcid.org/0000-0002-1708-8454
Sandra Sanchez-Roige https://orcid.org/0000-0001-6137-5699
Abraham A. Palmer https://orcid.org/0000-0003-3634-0747

REFERENCES

1000 Genomes Project Consortium. (2015) A global reference for human genetic variation. *Nature*, 526(7571), 68–74.

Abbott, L., Bryant, S., Churchhouse, C., Ganna, A., Howrigan, D., Palmer, D., et al. (2018) Nealelab GWAS Round 2. Available at: http://www.nealelab.is/uk-biobank [Accessed 17th May 2023].

Ahangari, M., Gentry, A.E., Hassan, M.F., Nguyen, T.H., Kendler, K.S., Bacanu, S.A. et al. (2023) Improving the discovery of rare variants associated with alcohol problems by leveraging machine learning phenotype prediction and functional information. *bioRxiv*. Available from: https://doi.org/10.1101/2023.09.11.557163

Antaki, D., Guevara, J., Maihofer, A.X., Klein, M., Gujral, M., Grove, J. et al. (2022) A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nature Genetics*, 54(9), 1284–1292.

Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D. et al. (2021) Exome sequencing and analysis of 454,787 UK biobank participants. *Nature*, 599(7886), 628–634.

Ben-David, E. & Shifman, S. (2012) Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. *PLoS Genetics*, 8(3), e1002556.

Bode, C. & Bode, J.C. (1997) Alcohol's role in gastrointestinal tract disorders. *Alcohol Health and Research World*, 21(1), 76–83.

Chang, X., Lima, L.A., Liu, Y., Li, J., Li, Q., Sleiman, P.M.A. et al. (2018) Common and rare genetic risk factors converge in protein interaction networks underlying schizophrenia. *Frontiers in Genetics*, 9, 434.

Charney, A.W., Stahl, E.A., Green, E.K., Chen, C.Y., Moran, J.L., Chambert, K. et al. (2019) Contribution of rare copy number variants to bipolar disorder risk is limited to schizoaffective cases. *Biological Psychiatry*, 86(2), 110–119.

Clarke, T.-K., Adams, M.J., Davies, G., Howard, D.M., Hall, L.S., Padmanabhan, S. et al. (2017) Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK biobank (N=112 117). *Molecular Psychiatry*, 22(10), 1376–1384.

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, 381(9875), 1371–1379.

Curtis, D. (2022) Investigation of Association of Rare, functional genetic variants with heavy drinking and problem drinking in exome sequenced UK biobank participants. *Alcohol and Alcoholism*, 57(4), 421–428.

Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D. et al. (2018) Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*, 9(1), 2098.

de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, 11(4), e1004219.

Ding, Z., Huang, G., Wang, T., Duan, W., Li, H., Wang, Y. et al. (2023) Genetic ablation of GIGYF1, associated with autism, causes behavioral and neurodevelopmental defects in zebrafish and mice. *Biological Psychiatry*, 94(10), 769–779.

Edenberg, H.J. (2007) The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Research & Health: The Journal of the National Institute on Alcohol Abuse and Alcoholism*, 30(1), 5–13.

Farris, S.P., Harris, R.A. & Ponomarev, I. (2015) Epigenetic modulation of brain gene networks for cocaine and alcohol abuse. *Frontiers in Neuroscience*, 9, 176.

Fu, S., Bury, L.A.D., Eum, J. & Wynshaw-Boris, A. (2023) Autism-specific PTEN p.Ile135Leu variant and an autism genetic background combine to dysregulate cortical neurogenesis. *American Journal of Human Genetics*, 110(5), 826–845.

Ganna, A., Satterstrom, F.K., Zekavat, S.M., das, I., Kurki, M.I., Churchhouse, C. et al. (2018) Quantifying the impact of rare and ultra-rare coding variation across the phenotypic Spectrum. *American Journal of Human Genetics*, 102(6), 1204–1211.

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M. & Vitkup, D. (2011) Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*, 70(5), 898–907.

GTEx Consortium. (2020) The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509), 1318–1330.

Hamitouche, S., Poupon, J., Dreano, Y., Amet, Y. & Lucas, D. (2006) Ethanol oxidation into acetaldehyde by 16 recombinant human cytochrome P450 isoforms: role of CYP2C isoforms in human liver microsomes. *Toxicology Letters*, 167(3), 221–230.

Hatoum, A.S., Colbert, S.M.C., Johnson, E.C., Huggett, S.B., Deak, J.D., Pathak, G.A. et al. (2023) Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nature Mental Health*, 1(3), 210–223.

Huang, J.K., Carlin, D.E., Yu, M.K., Zhang, W., Kreisberg, J.F., Tamayo, P. et al. (2018) Systematic evaluation of molecular networks for discovery of disease genes. *Cell Systems*, 6(4), 484–495.e5.

Karczewski, K.J., Solomonson, M., Chao, K.R., Goodrich, J.K., Tiao, G., Lu, W. et al. (2022) Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK biobank exomes. *Cell Genomics*, 2(9), 100168.

Karlsson Linnér, R., Mallard, T.T., Barr, P.B., Sanchez-Roige, S., Madole, J.W., Driver, M.N. et al. (2021) Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nature Neuroscience*, 24, 1367.

Kranzler, H.R., Zhou, H., Kember, R.L., Vickers Smith, R., Justice, A.C., Damrauer, S. et al. (2019) Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nature Communications*, 10(1), 1499.

Lanz, J., Biniaz-Harris, N., Kuvaldina, M., Jain, S., Lewis, K. & Fallon, B.A. (2023) Disulfiram: mechanisms, applications, and challenges. *Antibiotics*, 12(3), 524. Available from: https://doi.org/10.3390/antibiotics12030524

Le Daré, B., Lagente, V. & Gicquel, T. (2019) Ethanol and its metabolites: update on toxicity, benefits, and focus on immunomodulatory effects. *Drug Metabolism Reviews*, 51(4), 545–561.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F. et al. (2019) Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, 51(2), 237–244.

MacKillop, J., Agabio, R., Feldstein Ewing, S.W., Heilig, M., Kelly, J.F., Leggio, L. et al. (2022) Hazardous drinking and alcohol use disorders. *Nature Reviews. Disease Primers*, 8(1), 80.

Mallard, T.T. & Sanchez-Roige, S. (2021) Dimensional phenotypes in psychiatric genetics: lessons from genome-wide association studies of alcohol use phenotypes. *Complex Psychiatry*, 7(3–4), 45–48.

Mallard, T.T., Savage, J.E., Johnson, E.C., Huang, Y., Edwards, A.C., Hottenga, J.J. et al. (2022) Item-level genome-wide association study of the alcohol use disorders identification test in three population-based cohorts. *The American Journal of Psychiatry*, 179(1), 58–70.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J. et al. (2009) Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753.

Marees, A.T., Hammerschlag, A.R., Bastarache, L., de Kluiver, H., Vorspan, F., van den Brink, W. et al. (2018) Exploring the role of low-frequency and rare exonic variants in alcohol and tobacco use. *Drug and Alcohol Dependence*, 188, 94–101.

Marshe, V.S., Maciukiewicz, M., Hauschild, A.C., Islam, F., Qin, L., Tiwari, A.K. et al. (2021) Genome-wide analysis suggests the importance of vascular processes and neuroinflammation in late-life antidepressant response. *Translational Psychiatry*, 11(1), 127.

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S.M. et al. (2022) Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54(4), 437–449.

Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2012) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, 44(9), 1072.

Rayner, C., Coleman, J.R.I., Purves, K.L., Hodsoll, J., Goldsmith, K., Alpers, G.W. et al. (2019) A genome-wide association meta-analysis of prognostic outcomes following cognitive behavioural therapy in individuals with anxiety and depressive disorders. *Translational Psychiatry*, 9(1), 150.

Rosenthal, S.B., Willsey, H.R., Xu, Y., Mei, Y., Dea, J., Wang, S. et al. (2021) A convergent molecular network underlying autism and congenital heart disease. *Cell Systems*, 12(11), 1094–1107.e6.

Rosenthal, S.B., Wright, S.N., Liu, S., Churas, C., Chilin-Fuentes, D., Chen, C.H. et al. (2023) Mapping the common gene networks that underlie related diseases. *Nature Protocols*, 18(6), 1745–1759.

Sanchez-Roige, S., Palmer, A.A. & Clarke, T.K. (2020) Recent efforts to dissect the genetic basis of alcohol use and abuse. *Biological Psychiatry*, 87(7), 609–618.

Sanchez-Roige, S., Palmer, A.A., Fontanillas, P., Elson, S.L., the 23 and Me Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, Adams, M.J. et al. (2019) Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. *The American Journal of Psychiatry*, 176(2), 107–118.

Saunders, G.R.B., Wang, X., Chen, F., Jang, S.K., Liu, M., Wang, C. et al. (2022) Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, 612(7941), 720–724.

Sazonovs, A. & Barrett, J.C. (2018) Rare-variant studies to complement genome-wide association studies. *Annual Review of Genomics and Human Genetics*, 19, 97–112.

Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.C., Kitts, P.A. et al. (2017) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Research*, 27(5), 849–864.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504.

Sherva, R., Zhang, R., Sahelijo, N., Jun, G., Anglin, T., Chanfreau, C. et al. (2023) African ancestry GWAS of dementia in a large military cohort identifies significant risk loci. *Molecular Psychiatry*, 28(3), 1293–1302.

Singh, T., Poterba, T., Curtis, D., Akil, H., al Eissa, M., Barchas, J.D. et al. (2022) Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, 604(7906), 509–516.

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L. et al. (2023) The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, 51(D1), D977–D985.

Tolstrup, J.S., Nordestgaard, B.G., Rasmussen, S., Tybjærg-Hansen, A. & Grønbæk, M. (2008) Alcoholism and alcohol drinking habits predicted from alcohol dehydrogenase genes. *The Pharmacogenomics Journal*, 8(3), 220–227.

Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B. et al. (2022) Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, 604(7906), 502–508.

Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. (2010) Associating genes and protein complexes with disease via network propagation. *PLoS Computational Biology*, 6(1), e1000641.

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Verhulst, B., Neale, M.C. & Kendler, K.S. (2015) The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychological Medicine*, 45(5), 1061–1072.

Vrieze, S.I., Feng, S., Miller, M.B., Hicks, B.M., Pankratz, N., Abecasis, G.R. et al. (2014) Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biological Psychiatry*, 75(10), 783–789.

Walsham, N.E. & Sherwood, R.A. (2014) Chapter two—ethyl glucuronide and ethyl sulfate. In: Makowski, G.S. (Ed.) *Advances in clinical chemistry*. San Diego, CA: Elsevier, pp. 47–71.

Walters, R.K., Polimanti, R., Johnson, E.C., McClintick, J.N., Adams, M.J., Adkins, A.E. et al. (2018) Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, 21(12), 1656–1669.

Wang, Q., Dhindsa, R.S., Carss, K., Harper, A.R., Nag, A., Tachmazidou, I. et al. (2021) Rare variant contribution to human disease in 281,104 UK biobank exomes. *Nature*, 597(7877), 527–532.

Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. (2017) Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, 8(1), 1826.

Weiner, D.J., Nadig, A., Jagadeesh, K.A., Dey, K.K., Neale, B.M., Robinson, E.B. et al. (2023) Polygenic architecture of rare coding variation across 394,783 exomes. *Nature*, 614(7948), 492–499.

Wright, S.N., Colton, S., Schaffer, L.V., Pillich, R.T., Churas, C., Pratt, D. et al. (2024) State of the interactomes: an evaluation of molecular networks for generating biological insights. *bioRxiv*. Available from: https://doi.org/10.1101/2024.04.26.587073

Wright, S.N., Leger, B.S., Rosenthal, S.B., Liu, S.N., Jia, T., Chitre, A.S. et al. (2023) Genome-wide association studies of human and rat BMI converge on synapse, epigenome, and hormone signaling networks. *Cell Reports*, 42(8), 112873.

Yao, X., Glessner, J.T., Li, J., Qi, X., Hou, X., Zhu, C. et al. (2021) Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. *Translational Psychiatry*, 11(1), 69.

Zakhari, S. (2006) Overview: how is alcohol metabolized by the body? *Alcohol Research & Health: The Journal of the National Institute on Alcohol Abuse and Alcoholism*, 29(4), 245–254.

Zhou, H., Kember, R.L., Deak, J.D., Xu, H., Toikumo, S., Yuan, K. et al. (2023) Multi-ancestry study of the genetics of problematic alcohol use in over 1 million individuals. *Nature Medicine*, 29(12), 3184–3192.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Leger, B.S., Meredith, J.J., Ideker, T., Sanchez-Roige, S. & Palmer, A.A. (2024) Rare and common variants associated with alcohol consumption identify a conserved molecular network. *Alcohol: Clinical and Experimental Research*, 00, 1–12. Available from: https://doi.org/10.1111/acer.15399

29937175, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/acer.15399, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License