bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title: Genome-wide association study meta-analysis of the Alcohol Use Disorder Identification Test (AUDIT) in two population-based cohorts (N=141,958)

Authors:

Sandra Sanchez-Roige[1], Abraham A. Palmer[1,2], Pierre Fontanillas[3], Sarah L. Elson[3], The 23andMe Research Team[3], Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, Mark J. Adams[4], David M. Howard[4], Howard J. Edenberg[5], Gail Davies[6,7,] Richard C. Crist[8], Ian J. Deary[6-7], Andrew M. McIntosh[4,6] & Toni-Kim Clarke[4]

Affiliations:
[1] Department of Psychiatry, University of California San Diego, La Jolla, CA, 92093, USA
[2] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA
[3] Collaborator List for the 23andMe Research Team: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson
[4] Division of Psychiatry, University of Edinburgh, Edinburgh, UK
[5] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA.
[6] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
[7] Department of Psychology, University of Edinburgh, Edinburgh, UK
[8] Translational Research Laboratories, Center for Neurobiology and Behavior, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.

Corresponding author: Toni-Kim Clarke
email: toni.clarke@ed.ac.uk

Supplementary Material: Supplementary Tables (20) and Figures (21)

Word Count: 5349
Figures: 3
Tables: 1

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract [261 words]

Alcohol use disorders (**AUD**) are common conditions that have enormous social and economic consequences. We obtained quantitative measures using the Alcohol Use Disorder Identification Test (**AUDIT**) from two population-based cohorts of European ancestry: UK Biobank (**UKB**; N=121,630) and 23andMe (N=20,328) and performed a genome-wide association study (GWAS) meta-analysis. We also performed GWAS for AUDIT items 1-3, which focus on consumption (**AUDIT-C**), and for items 4-10, which focus on the problematic consequences of drinking (**AUDIT-P**). The GWAS meta-analysis of AUDIT total score identified 11 associated risk loci. Novel associations localized to genes including *JCAD* and *SLC39A13*; we also replicated previously identified signals in the genes *ADH1B, ADH1C*, *KLB*, and *GCKR*. The dimensions of AUDIT showed positive genetic correlations with alcohol consumption (r_g=0.78-0.96) and Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) alcohol dependence (r_g=0.33-0.64). AUDIT-P and AUDIT-C showed significantly different patterns of association across a number of traits, including psychiatric disorders. AUDIT-P was positively genetically correlated with schizophrenia (r_g=0.22, p=1.2x10^{-10}), major depressive disorder (r_g=0.26, p=4.5x10^{-5}), and ADHD (r_g=0.23, p=1.2x10^{-5}), whereas AUDIT-C was negatively genetically correlated with major depressive disorder (r_g=-0.23, p=3.2x10^{-3}) and ADHD (r_g=-0.10, p=1.8x10^{-2}). We also used the AUDIT data in the UKB to identify thresholds for dichotomizing AUDIT total score that optimize genetic correlations with DSM-IV alcohol dependence. Coding individuals with AUDIT total score of ≤4 as controls and ≥12 as cases produced a high genetic correlation with DSM-IV alcohol dependence (r_g=0.82, p=3.9x10^{-6}) while retaining most subjects. We conclude that AUDIT scores ascertained in population-based cohorts can be used to explore the genetic basis of both alcohol consumption and AUD.

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Alcohol use disorders (**AUD**) are modestly heritable, with twin-studies demonstrating that approximately 50% of the variance is attributed to genetic factors[1,2]. To date, genetic studies of AUD have identified genes that influence pharmacokinetic (e.g. *ADH1B, ADH1C, ALDH2*)[3–8], but not pharmacodynamic factors. The difficulty of assembling large, carefully diagnosed cohorts of AUD has stimulated additional studies of non-clinical phenotypes, such as alcohol consumption, in populations not ascertained for alcohol dependence. This approach has allowed for the relatively rapid collection of much larger sample sizes (e.g. >100,000s individuals) and has identified numerous loci associated with both pharmacokinetic and pharmacodynamic factors that influence alcohol consumption, including *ADH1B/ADH1C/ADH5*[9–11], *KLB* (encoding β-klotho)[10–12] and *GCKR*, encoding the glucokinase regulatory protein[10,11]. However, the genetic overlap between alcohol consumption (units per week) and DSM-IV diagnosed alcohol dependence is moderate ($r_g = 0.38$)[13], reinforcing the notion that alcohol consumption cannot be used as a surrogate for alcohol dependence or AUD.

The Alcohol Use Disorders Identification Test (**AUDIT**) is a screening tool designed to identify hazardous alcohol use in the past year[14]. The test consists of 10 items across 3 dimensions pertaining to alcohol consumption (items 1-3, often termed **AUDIT-C**), dependence symptoms (items 4-6), and harmful alcohol use (items 7-10) (collectively **AUDIT-P**). When the AUDIT was developed, a total score of 8 or higher was proposed to be indicative of harmful alcohol use[14] and a score of 20 or higher consistent with a diagnosis of alcohol dependence[15]; however, there is no clear consensus and subsequent studies have suggested that additional factors including sex and cultural and social contexts should be considered when deriving thresholds for alcohol dependence (reviewed in **Supplementary Table 1**).

A recent population-based GWAS of AUDIT in 20,328 research participants from the genetics company 23andMe, Inc., identified a locus near the gene *ADH1C* (rs141973904; $p = 4.4 \times 10^{-7}$)[9] nominally associated with AUDIT total score. AUDIT scores among 23andMe research participants were low and predominantly driven by alcohol consumption (**AUDIT-C**). The genetic correlation between AUDIT total score from 23andMe and alcohol consumption was much stronger ($r_g = 0.89$, $p = 9.01 \times 10^{-10}$) than the genetic correlation between AUDIT total score and alcohol dependence ($r_g = 0.08$; $p = 0.65$)[13].

In this study, we performed a GWAS meta-analysis using the UK Biobank (**UKB**; N = 121,630) and the previously published 23andMe cohort (N = 20,328)[9], yielding the largest GWAS meta-

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analysis of AUDIT total score to date (N = 141,958). Using only the UKB cohort, we also sought to determine whether the alcohol consumption component of the AUDIT had a genetic architecture distinct from the dependence and harmful use components by performing GWASs of consumption "AUDIT-C" (items 1-3) and problems "AUDIT-P" (items 4-10). Linkage Disequilibrium Score Regression Coefficient (**LDSC**)[16] was used to calculate genetic correlations between AUDIT measures and other substance use, psychiatric, and behavioral traits. We also calculated genetic correlations with obesity and blood lipid traits, as these have previously been shown to associate with alcohol consumption[9,10]. We hypothesized that AUDIT-P would correlate more strongly with measures of hazardous substance use, including alcohol dependence, and other psychiatric conditions. Finally, in order to determine the thresholds for dichotomizing AUDIT total score that would most closely approximate alcohol dependence, we converted continuous AUDIT total score into cases and controls using different thresholds, performed GWAS on each, and calculated the genetic correlation with DSM-IV alcohol dependence[13].

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and Methods

UK Biobank sample

The UK Biobank (**UKB**) is a population-based sample of 502,629 individuals who were recruited from 22 assessment centers across the United Kingdom from 2006-2010[17]. 157,366 individuals filled out a mental health questionnaire as part of an online follow-up over a one-year period in 2017. The Alcohol Use Disorder Identification Test (**AUDIT**)[14] was administered to assess past year alcohol use, using gating logic (see **Supplementary Figure 1**). After performing quality control to remove participants with missing data, and keeping only white British unrelated individuals, 121,630 individuals with AUDIT total scores were available. AUDIT total score was created by taking the sum of items 1-10 for all participants, including those who endorsed currently never drinking alcohol (as they could still endorse past alcohol harm on items 9 and 10). We also created AUDIT subdomain scores by aggregating the scores from items 1-3, which include the information pertaining to alcohol consumption (**AUDIT-C**, N = 121,630), and from items 4-10 which indexes the information pertaining to alcohol problems (**AUDIT-P**, N = 121,630). These traits were log transformed to approximate a normal distribution (**Supplementary Figure 2**).

Genotyping, quality control and imputation

Genotype imputation was performed on 487,409 individuals by the UKB team using IMPUTE4[18] and the Haplotype Reference Consortium reference panel. After quality control, 11,785,427 SNPs remained for GWAS analyses. Additional details on genotyping and quality control are shown in the **Supplementary Material**.

Discovery GWASs using UKB

GWAS analyses were performed using BGENIE v1.1[18] with AUDIT scores (AUDIT total score, AUDIT-C, and AUDIT-P, tested independently) as the outcome variable and age, sex, genotyping array and 20 principal components of genotype as covariates. See the Life Sciences Reporting Summary for extended details. In order to identify independently-associated variants ("index variants"), clump-based pruning was applied in PLINK v1.9[19] using an r^2 of 0.1 and a 1 Mb sliding window. LD was calculated using UKB as the reference panel. A 1 Mb window was

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

used due to the regions of extended linkage disequilibrium on chromosomes 4q23 and 17q21.31, which were associated with AUDIT score in this study.

In addition, we performed a series of 18 case-control GWAS analyses of AUDIT total score using different thresholds (cases: ≥8, 10, 12, 15, 18, 20 vs controls: ≤2, 3, 4). The sample size at each threshold is shown in **Supplementary Table 2**. The results of these analyses were used to determine which thresholds would produce the greatest genetic correlation with DSM-IV defined alcohol dependence[13].

SNP-Heritability analyses

The SNP-heritability of UKB AUDIT scores (total, AUDIT-C, AUDIT-P) was calculated using a genomic restricted maximum likelihood (**GREML**) method implemented in Genetic Complex Trait Analysis (**GCTA**)[20] on a subset of 117,072 unrelated individuals using a relatedness cut-off of 0.05 and controlling for age and sex. GREML analyses were run using genotyped SNPs with a MAF greater than 0.01 to construct the GRM.

GWAS meta-analysis of AUDIT total score using the UKB and 23andMe cohorts

Because the genetic correlation of AUDIT total score between the UKB and 23andMe cohorts was high (r_g = 0.76, SE = 0.11, p = 2.26 x 10^{-11}), we performed a sample-size based meta-analysis of AUDIT total score from the UKB and 23andMe cohorts using METAL (version 2011-03-25) based on effect sizes[21]. This meta-analysis comprises a total of 141,958 research participants of European ancestry and 9,136,657 genetic variants that passed quality control. We used clump-based pruning (see 'Discovery GWAS') to identify independently-associated variants. For each GWAS signal we defined a set of credible variants using a Bayesian refinement method developed by Maller et al[22]. These credible sets are considered to have a 99% probability of containing the 'causal' variant at each locus. Credible set analyses were performed in R (https://github.com/hailianghuang/FM-summary) for each of the index variants associated with AUDIT score in the GWAS meta-analysis using SNPs within 1Mb with an r^2 >0.4 to the index variant.

All downstream genetic analyses of AUDIT total score were performed using the GWAS meta-analysis summary statistics.

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Functional mapping and annotation of GWAS meta-analysis

We used FUMA v1.2.8[23] to study the functional consequences of the index SNPs, and of the SNPs contained in each credible set, which included ANNOVAR categories, Combined Annotation Dependent Depletion (**CADD**) scores, RegulomeDB scores, eQTLs, and chromatin states. We also studied the regulatory consequences of the index SNPs using the Variant Effect Predictor (**VEP**; Ensembl GRCh37).

Gene-set and pathway analyses

We performed MAGMA[23] competitive gene-set and pathway analyses using the summary statistics from the GWAS meta-analysis of AUDIT total score and the AUDIT-C and AUDIT-P subsets. SNPs were mapped to 18,133 protein-coding genes from Ensembl build 85. Gene-sets were obtained from Msigdb v5.2 ("Curated gene sets", "GO terms").

Gene-based association using transcriptomic data with S-PrediXcan

We used S-PrediXcan[24] to predict gene expression levels in 10 brain tissues, and to test whether the predicted gene expression correlates with AUDIT scores. We used pre-computed tissue weights from the Genotype-Tissue Expression (**GTEx** v7) project database (https://www.gtexportal.org/) as the reference transcriptome dataset. Further details are provided in the **Supplementary Material**.

Genetic correlation analysis

We used LD Score regression (**LDSC**) to identify genetic correlations between traits[25]. This method was used to calculate genetic correlations (r_g) between AUDIT total score, AUDIT-C, and AUDIT-P and 39 other traits and diseases (see **Supplementary Tables 3, 4** and **5**). We did not constrain the intercepts in our analysis, as we could not quantify the exact amount of sample overlap between cohorts. We used False Discovery Rate (**FDR**) to correct for multiple testing[26]. We also used LDSC to examine genetic correlations between various dichotomized versions of AUDIT and DSM-IV defined alcohol dependence[13]. To test for significant differences between the genetic correlations, z-score statistics were calculated (see **Supplementary Table 6)**.

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

UKB sample demographics and characteristics

In the UKB cohort, there were 121,630 individuals with AUDIT scores available for GWAS analysis (**Supplementary Table 7**). The UKB sample was 56.2% female (N = 68,403) and the mean age was 56.1 years (S.D. = 7.7). The mean AUDIT total score was 5.0 (S.D. = 4.18, range = 0-40); a histogram showing the distribution of the scores is shown in **Supplementary Figure 2**. Over the prior year, 91.9% of the participants reported drinking 1 or 2 drinks on a single day. Over the prior year, 6.3% of the participants reported they were not able to stop drinking once they started, and 10.7% felt guilt or remorse after drinking alcohol (**Supplementary Table 7**). Males had significantly higher AUDIT total mean scores than females (6.09 ± 4.45 vs. 4.15 ± 3.72, respectively; $\beta = 0.47$, $p < 2 \times 10^{-6}$; **Supplementary Figure 3**). In addition, age was negatively correlated with AUDIT scores ($\beta = -0.02$, $p < 2 \times 10^{-6}$; **Supplementary Table 8**). Therefore, both sex and age were used as covariates in the GWAS analyses. The mean AUDIT-C score was 4.24 (S.D. = 2.83) and the mean AUDIT-P score was 0.75 (S.D. = 2.0). As expected, there was a moderate positive phenotypic correlation between AUDIT-C and AUDIT-P ($r = 0.478$, 95% C.I. = 0.473-0.481, $p < 2 \times 10^{-16}$; **Supplementary Table 8**).

SNP-heritability in UKB

We estimated the SNP-heritability of AUDIT total score to be 12% (**GCTA**: ± 0.48%, $p = 4.6 \times 10^{-273}$; **LDSC:** 8.6% ± 0.50%), which is similar to the estimate from Sanchez-Roige et al.[9]. The SNP-heritability for AUDIT-C was 11% (**GCTA**: ±0.47%, $p = 1.5 \times 10^{-211}$; **LDSC**: 8.5% ±0.50%), and 9% for AUDIT-P (**GCTA**: ±0.46%, $p = 2.0 \times 10^{-178}$; **LDSC**: 5.9% ±0.48%).

GWAS of AUDIT scores in UKB

The significant results ($p < 5 \times 10^{-8}$) of the GWAS of AUDIT total score in the UKB cohort are shown in **Supplementary Table 9**; this analysis revealed 13 independent GWAS signals located in 8 loci. The UKB GWAS of AUDIT-C and AUDIT-P subsets are summarized in **Supplementary Tables 10** and **11** and **Supplementary Figure 17**. Nine of these 13 independent GWAS signals also significantly associated with AUDIT-C; interestingly, the same index variants were identified in the two analyses. An additional GWAS signal was also

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

identified close to *FNBP4*. For AUDIT-P, 5 independent GWAS signals were significantly associated, including 4 signals already present in the total AUDIT and AUDIT-C, and a new independent signal close to ADH1B.

GWAS meta-analysis of AUDIT total score

The GWAS meta-analysis of the UKB and 23andMe samples found 13 independent GWAS signals (**Supplementary Table 12**) associated with AUDIT total score spanning 11 genomic loci (**Table 1**). **Figure 1** shows the Manhattan and QQ plots of the GWAS meta-analysis of AUDIT total score and **Supplementary Figures 5-17** show the regional association plots for the independent signals. The inflation factor of the meta-analysis GWAS was $\lambda_{GC} = 1.22$ with an LDSC intercept of 1.008 (SE = 0.007), suggesting that the majority of the inflation is due to polygenicity. The 13 independent SNPs show 100% sign concordance for association with AUDIT total score across UKB and 23andMe (**Table 1**); 8 of these SNPs were nominally associated with AUDIT total score in 23andMe ($p \leq 0.05$), and all index SNPs were associated with AUDIT total score in UKB ($p < 1.6 \times 10^{-6}$).

The top hit for the GWAS meta-analysis of AUDIT total score was an intronic variant (rs138495951) of the gene *ADH1B* ($p = 10.7 \times 10^{-36}$). rs138495951 is in moderate LD ($r^2 = 0.54$) with the well-known functional SNP in *ADH1B*, rs1229984, which is known to alter alcohol metabolism[27]. rs1229984 was not available in the 23andMe cohort and therefore was not included in the meta-analysis; however, rs1229984 was strongly associated with AUDIT total score in the UKB cohort alone ($\beta = 0.15$, S.E. = 0.008, $p = 5.8 \times 10^{-72}$). In addition, two other regions in 4q23 were associated with AUDIT total score in the meta-analysis: the index SNPs were located in the *ADH1B* and *ADH1C* genes, suggesting that there are a number of independent loci in this region associated with AUDIT scores. This region has been previously associated with alcohol consumption, AUD, and AUDIT scores[6,7,28,29]

We also replicated the association between *KLB* (**Supplementary Table 12**), on chromosome 4q14, and alcohol use[10,11,30]; the index SNP rs11940694, which is located in the intron of *KLB*, was associated with AUDIT total score in the present study. Credible set analyses revealed 7 SNPs at this locus, all located within *KLB* (**Supplementary Table 13**). AUDIT total score was also associated with SNPs that localized to *GCKR* on chromosome 2p23.3, which has been previously associated with alcohol consumption[10,11]. Seven SNPs comprised the credible set at

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the *GCKR* locus, including the missense variant (rs1260326) in *GCKR* that was identified as the index SNP.

We identified GWAS signals in several regions that have not been previously implicated in the genetics of AUD, including 2p21, 17q21, 3q25, 8q22, 10p11, 11p11 and 19q13. The index SNP in the 4q24 region is located in an intron of *SLC39A8,* and the remainder of the credible set for this locus, including the missense variant rs13107325, are located in this gene. *SLC39A8* is highly pleiotropic[31] but it is a novel association in relation to alcohol. A region of association on 2p21 contains 17 SNPs that are localized to the non-coding RNA, *LINC01833*. A novel region of association was also detected on chromosome 10p11.23; this region contains 9 credible SNPs that localize to the *JCAD* (junctional cadherin 5 associated) gene. *JCAD* encodes an endothelial cell junction protein, and has previously been associated with coronary heart disease[32]. Lastly, the index SNP on 19q13.3 is a synonymous variant in *FUT2. FUT2* encodes galactoside 2-alpha-L-fucosyltransferase 2, which controls the expression of ABO blood group antigens. A missense SNP in *FUT2*, rs601338, was strongly associated with susceptibility to diarrhoeal episodes[33] and this SNP is part of the credible set at this locus.

The remaining novel associations on 3q25, 8q, 11p11 and 17q21 were more complex. The index variants on chromosomes 3q25.33 and 8q22.1 were not localized to any genes, and it is unclear from the credible set analyses what the causal variants may be at these loci. The credible SNP sets for the 11p11 and 17q21 regions contained over 60 SNPs each, which spanned several genes. For example, the index SNP on chromosome 17q21.31 was an intronic SNP in *MAPT*, which encodes the tau protein and has been robustly associated with Parkinson's disease[34,35] (**Supplementary Table 14**) and other neurodegenerative tauopathies[36], and more recently with neuroticism[37]. However, we note that the region of association on chromosome 17q21.31 spans the corticotrophin receptor gene (*CRHR1*), which has been associated with alcohol use in animals and humans[38]. Thus, due to the extended complex LD in this region, we are unable to determine the likely causal variant. Similarly, the index SNP (rs2293576) at chromosome 11p11.2 is a synonymous SNP of the zinc transporter gene *SLC39A13*; however, this region includes 90 associated SNPs, which map to four additional genes.

We used FUMA to functionally annotate all 2,298 SNPs in the credible sets (see **Supplementary Table 13**). The majority of the SNPs were intronic (83.6%; N = 1,922) and intergenic (10.5%; N = 240), and only 77 SNPs (4.0%) were exonic. Furthermore, 79 SNPs showed CADD scores >12.37, which is the suggested threshold to be considered deleterious[39].

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The exonic SNPs (rs492602, rs62062288, rs13135092) of *FUT2*, *MAPT* and *SLC39A8*, respectively, had the highest CADD scores (>34), suggesting potential deleterious protein effects. 300 SNPs had RegulomeDB scores of 1a-1f, showing evidence of potential regulatory effects. 91.0% of the SNPs were in open chromatin regions (minimum chromatin state 1-7).

Gene-based and pathway analyses

We used MAGMA[23] to perform a gene-based association analysis; which identified 59 genes that were significantly associated with AUDIT total score ($p < 2.76 \times 10^{-6}$; **Supplementary Table 15, Supplementary Figure 18**). As expected, the majority of these genes were in the 11 GWAS loci (i.e. *KLB, WNT3, RFC1, GCKR, CRHR1*); *DRD2* was also among the top hits. In addition, the analysis revealed a strong burden signal in *CADM2* ($p = 1.05 \times 10^{-9}$), where the index variant in GWAS meta-analysis was only $p = 3.09 \times 10^{-7}$. We did not identify any canonical pathways that were significantly associated with AUDIT (**Supplementary Table 16**).

Gene-based (**MAGMA**) analyses for the AUDIT-C and AUDIT-P subsets (**Supplementary Figures 19** and **20**, respectively) revealed evidence of overlap (**Supplementary Figure 21, Supplementary Table 17**). Seven genes were associated with all 3 AUDIT traits (AUDIT total score, AUDIT-C, and AUDIT-P). There was considerable overlap between AUDIT total score and AUDIT-C, with 34 overlapping genes associated at the gene-based level. Only 5 genes were associated with both AUDIT total score and AUDIT-P.

S-PrediXcan

S-PrediXcan identified a positive correlation ($p < 1.71 \times 10^{-6}$) between AUDIT total score and the predicted expression of 9 genes across multiple brain tissues (full results are presented in **Supplementary Table 18**), including *MAPT* (cerebellum) and *FUT2* (caudate and nucleus accumbens). SNPs in the region of *MAPT* and *FUT2* were associated with AUDIT total score in the GWAS. *MAPT* (cerebellum) and *FUT2* (nucleus accumbens) were also associated with AUDIT-C. S-PrediXcan for AUDIT-C and AUDIT-P (**Supplementary Tables 19** and **20**, respectively) revealed lower predicted *RFC1* expression in the cerebellar hemisphere was associated with both higher AUDIT-C ($p = 1.76 \times 10^{-10}$) and AUDIT-P ($p = 1.54 \times 10^{-6}$) scores.

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genetic correlations

We used LDSC to evaluate evidence for genetic correlations between our three primary traits (AUDIT total score, AUDIT-C, and AUDIT-P) and numerous other traits for which GWAS summary statistics were available; these included alcohol and substance use traits, personality and behavioral traits, psychiatric disorders, blood lipids, and brain structure volumes (**Supplementary Tables 3-5 and Figure 2**).

As expected, AUDIT-C and AUDIT-P were highly genetically correlated (r_g = 0.70, p = 3.7 x 10^{-68}). AUDIT scores (AUDIT total score, AUDIT-C, and AUDIT-P) showed strong genetic correlations with alcohol consumption from two other studies (r_g = 0.79-0.85, p < 3.5 x 10^{-9}). The AUDIT-C had a significantly stronger (p = 9.28 x 10^{-3}) correlation with alcohol consumption (r_g = 0.92, p = 1.4 x 10^{-159}) than did AUDIT-P (r_g = 0.77, p = 3.4 x 10^{-53}). In contrast, AUDIT total and AUDIT-C scores were only modestly correlated with alcohol dependence (r_g = 0.39 & 0.33 respectively, p < 7.7 x 10^{-5}), whereas AUDIT-P showed a nominally stronger genetic correlation with alcohol dependence (r_g = 0.64, p = 2.5 x 10^{-8}; AUDIT-P vs AUDIT-C, p = 0.038; see **Supplementary Table 6**).

We detected positive genetic correlations between AUDIT scores (AUDIT total, AUDIT-C, AUDIT-P) and other substance use phenotypes, including lifetime smoking (r_g = 0.24-0.41, p < 1.8 x 10^{-5}) and cannabis use (r_g = 0.26-0.46, p < 9.7 x 10^{-5}). We also observed a positive genetic correlation between AUDIT-P and cigarettes per day (r_g = 0.28, p = 2.9 x 10^{-3}).

Several psychiatric disorders and related traits were positively genetically correlated with AUDIT-P scores, including schizophrenia (r_g = 0.22, p = 1.2 x 10^{-10}), bipolar disorder (r_g = 0.27, p = 6.8 x 10^{-5}), ADHD (r_g = 0.23, p = 1.2 x 10^{-5}), and major depressive disorder (MDD, r_g = 0.26, p = 4.6 x 10^{-3}). Intriguingly, AUDIT-C was negatively correlated with MDD (r_g = -0.23, p = 3.3 x 10^{-3}) and ADHD (r_g = -0.10, p = 1.8 x 10^{-2}), whereas AUDIT-P showed positive genetic correlations with these same disease traits.

We observed a positive genetic correlation between AUDIT-P scores and depressive symptoms (r_g = 0.30, p = 2.1 x 10^{-8}) and neuroticism (r_g = 0.23, p = 1.1 x 10^{-5}), and a negative genetic correlation with subjective well-being (r_g = -0.24, p = 4.1 x 10^{-5}). We also observed a positive genetic correlation between AUDIT-P and caudate brain volume (r_g = 0.21, p = 1.3 x 10^{-2}). The genetic correlations with AUDIT-P were significantly different from the correlations with AUDIT-C (**Supplementary Table 6**).

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We observed positive genetic correlations between AUDIT total score, AUDIT-C and education, college attainment, and cognitive ability (r_g = 0.19-0.24, $p < 1.9 \times 10^{-5}$). The AUDIT-P genetic correlations with the same traits were near to zero, and were significantly lower than AUDIT C and AUDIT total or education traits (**Supplementary Table 6**).

There were negative genetic correlations with obesity (r_g = -0.16-0.17, $p < 7.6 \times 10^{-6}$), similar to previous reports regarding AUDIT total score[9] and alcohol consumption[10]. In contrast, obesity was not significantly genetically correlated with AUDIT-P scores (r_g = 0.03, 0.005, respectively). Similarly, HDL cholesterol and triglycerides were genetically correlated with AUDIT total score and AUDIT-C (r_g = 0.18-21, $p < 9.8 \times 10^{-6}$, r_g = -0.16, $p < 1.0 \times 10^{-4}$), but this association was not found for AUDIT-P (r_g = 0.11, $p = 2.1 \times 10^{-2}$, rg = -0.02, $p = 6.5 \times 10^{-1}$). Obesity showed significantly different correlations with both AUDIT-P and AUDIT-C (**Supplementary Table 6**)

Dichotomizing AUDIT total score to more closely approximate alcohol dependence

As AUDIT can be rapidly ascertained in large populations, we explored methods for dichotomizing AUDIT total score that optimized the genetic correlation with DSM-IV alcohol dependence[13]. Higher genetic correlations with alcohol dependence were observed as the control threshold was increased from 2 to 4, and with increasingly stringent case cut-offs (**Figure 3** and **Supplementary Table 2**). The highest genetic correlation was observed for cases with AUDIT total score ≥20 and controls ≤4 (r_g = 0.90, SE = 0.26, $p = 4.5 \times 10^{-4}$), however, this highly stringent threshold produced very few cases (N=1,290). Defining cases as ≥ 12 yielded an r_g of 0.82 (SE = 0.18, $p = 3.9 \times 10^{-6}$) while retaining more than 7 times as many cases (N=9,130), these genetic correlations were not significantly different from those obtained using cases ≥20 and controls ≤4 (p = 0.82).

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

We have presented the largest GWAS meta-analysis of AUDIT total score to date, using large population-based cohorts from UKB and 23andMe. We identified novel associations with AUDIT total score; the genes located in these regions include *JCAD* and *SLC39A8*. We found evidence for association in several loci previously associated with alcohol use via single-variant and gene-based analyses (i.e. *KLB, GCKR, ADH1B, CADM2*). Furthermore, we showed that there is shared genetic architecture between AUDIT scores and other alcohol and substance use phenotypes. AUDIT-P showed a positive genetic correlation with several psychiatric diseases, distinguishing AUDIT-P from AUDIT-C. Finally, using LDSC, we identified thresholds for dichotomizing AUDIT total score (AUDIT score ≥12 to define cases, and ≤4 to define controls) that maximize the genetic correlation with alcohol dependence while retaining a large number of participants.

Our top GWAS hits replicated previous association signals for alcohol use traits. The strongest associations with AUDIT score in this study spanned the alcohol metabolism genes on chromosome 4q23[40]. Variants in this region were associated with AUDIT total score, AUDIT-C and AUDIT-P, demonstrating that alcohol metabolism is a risk factor for both alcohol consumption and problematic use. The second strongest signal, also associated with the three AUDIT phenotypes, is located in the *KLB*, confirming the robust association of this gene with both alcohol consumption[10–12] in humans, and in mice[12]. However, the biology of this locus could be more complex than previously described. Although the credible set analysis suggested that the 7 more probable causal variants are all located on the first intron of *KLB*, one of these variants, rs11940694, is an eQTL for *RFC1* expression in the brain, and S-PrediXcan analysis predicted that lower expression of *RFC1* in the cerebellar hemisphere is associated with higher predicted AUDIT (AUDIT-C and AUDIT-P) scores. Interestingly, a gene in the complex GWAS signal on chromosome 19, *FGF21* (Fibroblast growth factor 21), was associated with AUDIT (AUDIT total score, AUDIT-C, AUDIT-P) at the gene-based level (**Supplementary Table 17**). Fibroblast growth factor 21 regulates sweet and alcohol preference in mice as part of a receptor complex with β-Klotho (*KLB*) in the central nervous system[37]. Additionally, we replicated the association between a genetic variant (rs1260326) in the gene *GCKR* and alcohol use[10,11], here associated with AUDIT total score and AUDIT-C. *GCKR* encodes the glucokinase regulatory protein, is involved in the phosphorylation of glucose in the liver, and has previously been associated with multiple metabolic phenotypes[41,42] (**Supplementary Table 14**). Other loci previously associated with alcohol use include *CADM2*[10], which was associated at the gene-

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

based level for all three AUDIT traits. Here, the burden analysis suggests that multiple (rare and common) variants are necessary to explain the association signal. Intriguingly, several of the novel associations with AUDIT scores were mapped to highly pleiotropic genes (*MAPT*, *FUT2*, *SLC39A8)*[31].

Genetic analysis of the AUDIT subsets revealed evidence of distinct genetic architecture between AUDIT-C and AUDIT-P (alcohol consumption vs. problem use), with support from the gene-based (**Supplementary Figures 18** and **19**), S-PrediXcan, (**Supplementary Tables 19** and **20**) and genetic correlation analyses (**Figure 2**). Furthermore, AUDIT-P showed a strong genetic correlation with alcohol dependence[13]. In contrast, AUDIT-C had a stronger genetic correlation with alcohol consumption. Thus, partitioning AUDIT scores into different subsets (alcohol consumption vs problem use) may disentangle genetic factors that contribute to different aspects of AUD vulnerability.

Polygenic overlap was observed for all measures of AUDIT and other substance use traits, including lifetime tobacco and cannabis use, as we previously reported[28,29,43,44], demonstrating that genetic risk factors for high AUDIT scores overlap with increased consumption of multiple drug types.

We found several significant differences between the genetic correlations with AUDIT-P and AUDIT-C. These differences were particularly pronounced for psychiatric and behavioral traits. AUDIT-P was positively genetically correlated with psychopathology (schizophrenia, bipolar disorder, MDD, ADHD), personality traits including neuroticism and regional brain volumes. These associations have previously been observed at the phenotypic level; AUDs commonly co-occur in individuals with schizophrenia[45], bipolar disorder[46], MDD[47] and adults with ADHD[48]. Intriguingly, genetic risk for high AUDIT-C score was *negatively* correlated with MDD and ADHD demonstrating that a distinct genetic component of AUDIT-P is shared with genetic risk for psychiatric disease. Regional volume abnormalities in subcortical brain regions of AUD individuals have been reported[49–51], however, it is unclear whether these alterations are a result of high alcohol drinking or a pre-existing vulnerability. We identified a positive genetic correlation between AUDIT-P and increased caudate volume; however, the majority of studies report reductions in regional brain volumes associated with AUD.

For AUDIT total score and AUDIT-C we showed positive genetic correlations with educational attainment and cognitive ability and negative genetic correlations with obesity, consistent with earlier reports[9,10]. These associations were not observed for AUDIT-P. Similarly, HDL

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cholesterol showed a significant positive correlation, and triglycerides a negative correlation, with AUDIT total score and AUDIT-C, but not AUDIT-P. These patterns were previously observed for alcohol consumption[10]. We could speculate that these differences may be linked to socioeconomic status (SES). Alcohol consumption is often higher in individuals with higher SES[52], whereas alcohol-related problems, such as binge drinking[53] and alcohol related mortality[54], are more prevalent in individuals with lower SES. Furthermore, individuals with low SES are more likely to have AUDs with psychiatric co-morbidities[55]. Consistent with this idea, we find positive genetic correlations between AUDIT-C and education, a trait correlated with SES[56], and negative genetic correlations between AUDIT-P and psychopathology. Our findings provide further evidence that different dimensions of alcohol use associate differently with behavior and that these differences may have a biological underpinning.

A clinical diagnosis of AUD is often required to define cases for genetic studies. An alternative strategy would be to use AUDIT to infer AUD case status; however, it has not been clear whether and how to perform meta-analyses between AUDIT scores and alcohol dependence. A GWAS meta-analysis for AUDIT and alcohol dependence would be simplified if a threshold could be used to define cases and controls based on AUDIT scores, an approach that was used by Mbarek et al.[57]. We have provided empirical evidence about genetic correlations between AUDIT and alcohol dependence using dichotomized AUDIT scores, and found thresholds for AUDIT that produced high genetic correlations with AUD (**Figure 3**). Genetic correlations increased as the upper threshold was made more stringent, although the standard errors for all of these estimates were overlapping. The genetic correlation with alcohol dependence appeared to asymptote when case status was defined as ≥12; therefore, this threshold could we used to define case status. We also considered various thresholds for defining controls and found that ≤4 produced a high genetic correlation with alcohol dependence while also retaining the largest number of subjects.

Our study is not without limitations. AUDIT specifically asks about the past year, and thus may not capture information on lifetime alcohol use and misuse. This is suboptimal for genetic studies because it effectively measures a recent state rather than a stable trait. Measures capturing drinking and AUD across the lifespan may be preferable. Also, although mean scores for the AUDIT-C dimension were 4.24, the mean of the AUDIT-P dimension was considerably lower (0.75). Thus, we were not able to perform a more refined categorization (e.g. 3 subsets: consumption [items 1-3], dependence [items 4-6], hazardous use [items 7-10]) as fewer individuals endorsed the items comprising AUDIT-P (see **Supplementary Table 7**, **items 9** and

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10). Furthermore, our study uses data from UKB and 23andMe research participants, who were volunteers not ascertained for AUD, and hence our findings may not generalize to other populations showing higher rates of alcohol use and dependence. Additional alcohol-related phenotypes (e.g. age at first use; patterns of alcohol drinking, including binge drinking) could be used in subsequent genetic studies to identify additional sources of genetic vulnerability for AUD. Lastly, we offered guidelines to identify cases to use in genetic studies of AUD (i.e. AUDIT score ≥12); however, these recommendations were based on genetic correlations and may not determine thresholds for diagnosing dependence in a clinical setting. Future studies will be able to test whether using AUDIT as a surrogate for AUD will be beneficial for gene discovery. In addition, several studies have argued that lower thresholds should be used for females, which has not been addressed in the present study.

We have reported the largest GWAS of AUDIT ever undertaken. We replicated previously identified signals (i.e. *ADH1B/ADH1C; KLB; GCKR*), and identified novel GWAS signal (i.e. *JACD, SLC39A8*) associated with AUDIT. We show that different portions of the AUDIT (AUDIT-C, AUDIT-P) correlate with distinct traits, which will aid in dissecting genetic vulnerability towards alcohol use and abuse. The genetic factors that predispose to high alcohol consumption inevitably overlap with those for problem drinking, as heavy drinking is generally a prerequisite for the development of hazardous use. However, not everyone who consumes alcohol experiences the same level of harmful consequences. By studying the different subsets of AUDIT, we identify genetic factors that may be specific to problem drinking. Larger studies of cohorts with a wider range of AUDIT-P scores are required to both replicate and expand these findings. Finally, we describe an alternative strategy to rigorous ascertainment for genetic studies of AUD, i.e. AUDIT score ≥12 to define cases and ≤4 to define controls, which could be used to achieve large sample sizes in a cost-efficient manner.

URLs:

UK Biobank: http://www.ukbiobank.ac.uk/

FUMA: http://fuma.ctglab.nl/

LD score software: https://github.com/bulik/ldsc/

LDHub: http://ldsc.broadinstitute.org/

METAL: https://genome.sph.umich.edu/wiki/METAL

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

S-PrediXcan: https://github.com/hakyimlab/S-PrediXcan-Working

The NHGRI GWAS Catalog: http://www.genome.gov/gwastudies/

Regulome DB database: http://www.regulomedb.org/

PredictDB Data Repository: http://predictdb.hakyimlab.org/

Credible set estimation method, R script: https://github.com/hailianghuang/FM-summary

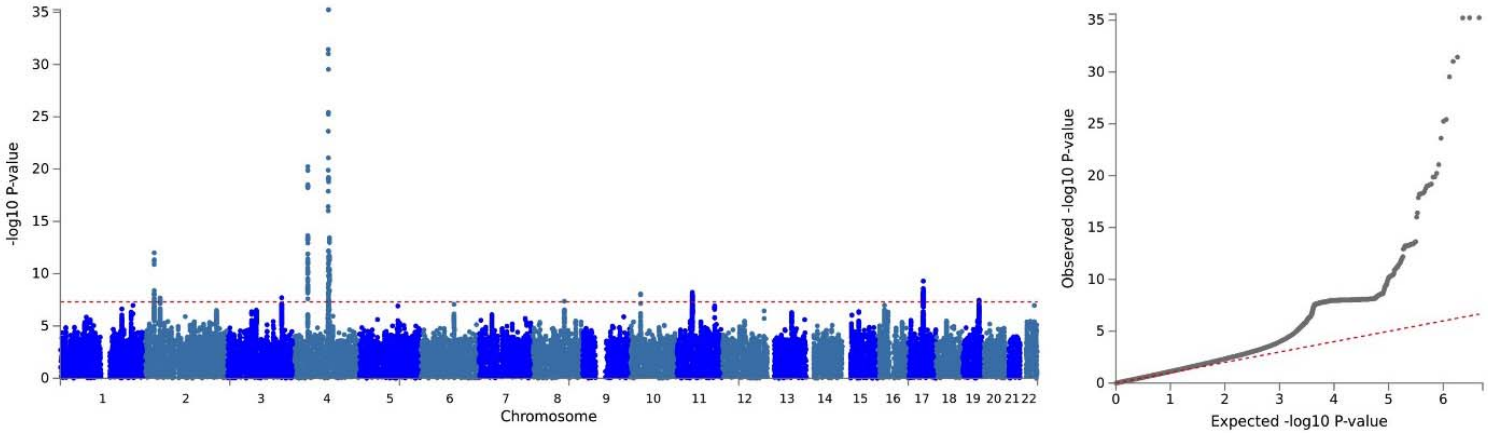
Data availability:

We have provided summary statistics for the top 10,000 SNPs (Supplementary Data Set). Full GWAS summary statistics for the 23andMe dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Interested investigators should email dataset-request@23andme.com for more information. GWAS summary statistics for the UK Biobank GWAS of AUDIT scores will be available on request.

Acknowledgements: We would like to thank the research participants and employees of 23andMe for making this work possible. S.S-R was supported by the Frontiers of Innovation Scholars Program (FISP; #3-P3029), the Interdisciplinary Research Fellowship in NeuroAIDS (IRFN; MH081482) and a pilot award from DA037844. This research has been conducted using the UK Biobank Resource: application number 4844 and was supported by a Wellcome Trust Strategic Award 'Stratifying Resilience and Depression Longitudinally' (STRADL) (Reference 104036/Z/14/Z), and by the Medical Research Council- and Biotechnology and Biological Sciences Research Council-funded Centre for Cognitive Ageing and Cognitive Epidemiology (Reference MR/K026992/1).

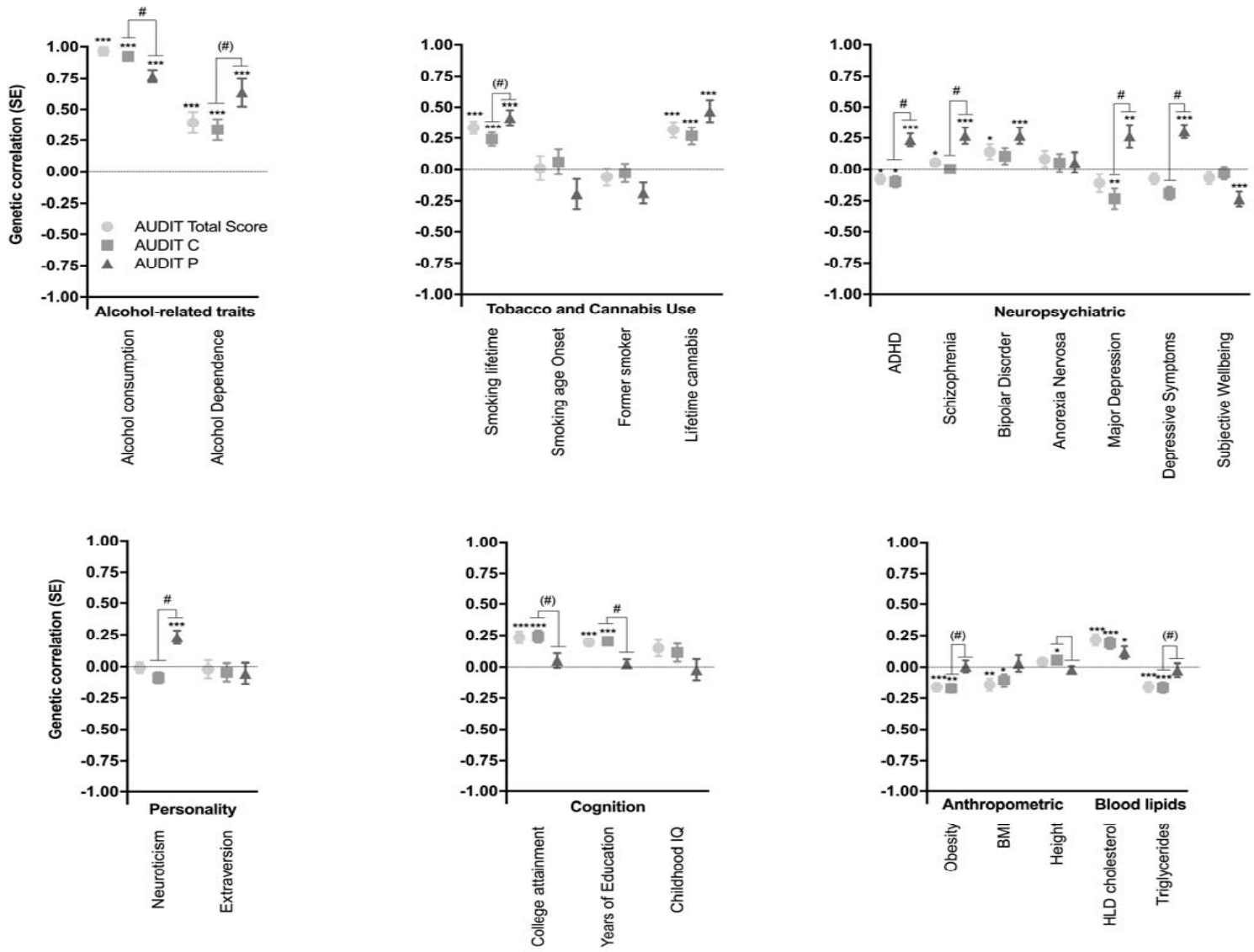
Figures

Figure 1. Manhattan and QQ plots for the SNP-based GWAS meta-analysis of AUDIT total score (N = 141,958)



bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

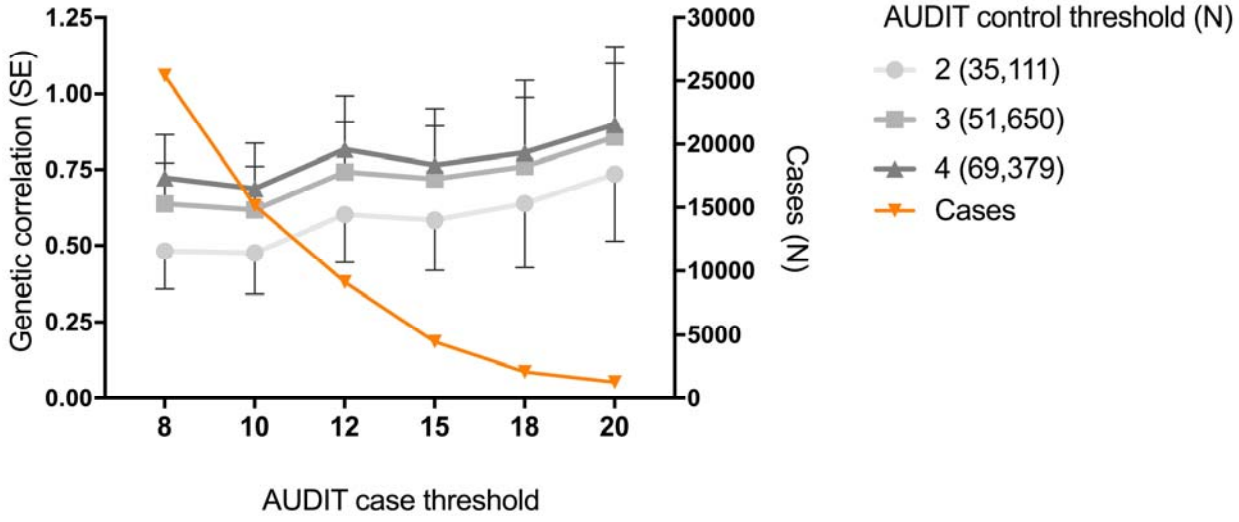
Figure 2. Genetic correlations between the three AUDIT phenotypes (total score, AUDIT-C, AUDIT-P) and several traits measured in independent cohorts as described in the Supplementary Tables 3-5: alcohol-related traits, tobacco and cannabis use, neuropsychiatric, personality, cognition, anthropomorphic and blood lipids. ADHD, attention-deficit/hyper-activity disorder; BMI, body mass index; SE, standard error; IQ, intelligence quotient; HDL, high-density lipoprotein. * p < 0.05, ** p < 0.01, *** p < 0.0001; # AUDIT-P vs AUDIT-C, p < 0.01 FDR 5%, (#) AUDIT-P vs AUDIT-C, p < 0.05



bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. Genetic correlations between AUDIT cases (8 [N = 25,423], 10 [N = 15,151], 12 [N = 9,130], 15 [N = 4,471], 18 [N = 2,099], 20 [N = 1,290]) vs controls (2, 3, 4) in the UK Biobank and DSM-IV derived alcohol dependence from the Psychiatric Genetics Consortium



bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Heath, A. C. *et al.* Genetic differences in alcohol sensitivity and the inheritance of alcoholism risk. *Psychol. Med.* **29,** 1069–1081 (1999).
2. Verhulst, B., Neale, M. C. & Kendler, K. S. The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol. Med.* **45,** 1061–1072 (2015).
3. Frank, J. *et al.* Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster: ADH gene cluster and alcohol dependence. *Addict. Biol.* **17,** 171–180 (2012).
4. Kilcoyne, B. *et al.* Alcohol Consumption Mediates the Relationship Between *ADH1B* and DSM-IV Alcohol Use Disorder and Criteria. *J. Stud. Alcohol Drugs* **75,** 635–642 (2014).
5. Quillen, E. E. *et al. ALDH2* is associated to alcohol dependence and is the major genetic determinant of “daily maximum drinks” in a GWAS study of an isolated rural chinese sample. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **165,** 103–110 (2014).
6. Gelernter, J. *et al.* Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Mol. Psychiatry* **19,** 41–49 (2014).
7. Bierut, L. J. *et al.* ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol. Psychiatry* **17,** 445–450 (2012).
8. Macgregor, S. *et al.* Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum. Mol. Genet.* **18,** 580–593 (2009).
9. 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.* (2017). doi:10.1111/adb.12574
10. Clarke, T.-K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112□117). *Mol. Psychiatry* (2017). doi:10.1038/mp.2017.153
11. Jorgenson, E. *et al.* Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Mol. Psychiatry* **22,** 1359–1367 (2017).
12. Schumann, G. *et al. KLB* is associated with alcohol drinking, and its gene product β-Klotho is necessary for FGF21 regulation of alcohol preference. *Proc. Natl. Acad. Sci.* **113,** 14372–14377 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

13. Walters, R. Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Unpublished* (2018).
14. Saunders, J. B. & Babor, T. F. AUDIT questionnaire□: screen for. *Addiction* **88,** 1–2 (1993).
15. Babor, T. F. The Alcohol Use Disorders Identification Test - Guidelines for Use in Primary Care. *World Health Organ. - Dep. Ment. Health Subst. Depend.* (2001).
16. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
17. Allen, N. *et al.* UK Biobank: Current status and what it means for epidemiology. *Health Policy Technol.* **1,** 123–126 (2012).
18. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *doi.org* 166298 (2017). doi:10.1101/166298
19. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
20. 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).
21. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
22. Maller, J. B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44,** 1294–1301 (2012).
23. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8,** 1826 (2017).
24. Barbeira, A. *et al.* MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results. *bioRxiv* 045260 (2016). doi:10.1101/045260
25. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).
26. Benjamini, Y. & Hochberg, M. T. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *J. R. Stat. Soc.* **57,** 289–300 (1995).
27. Hurley, T. D. & Edenberg, H. J. Genes encoding enzymes involved in ethanol metabolism. *Alcohol Res. Curr. Rev.* **34,** 339–44 (2012).
28. 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: GWAS of AUDIT. *Addict. Biol.* (2017). doi:10.1111/adb.12574

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

29. Clarke, T.-K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol. Psychiatry* **22,** 1376–1384 (2017).
30. Schumann, G. *et al.* Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc. Natl. Acad. Sci.* **108,** 7119–7124 (2011).
31. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48,** 709–717 (2016).
32. Erdmann, J. *et al.* Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur. Heart J.* **32,** 158–168 (2011).
33. Bustamante, M. *et al.* A genome-wide association meta-analysis of diarrhoeal disease in young children identifies FUT2 locus and provides plausible biological pathways. *Hum. Mol. Genet.* **25,** 4127–4142 (2016).
34. Satake, W. *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.* **41,** 1303–1307 (2009).
35. Simón-Sánchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* **41,** 1308–1312 (2009).
36. Pittman, A. M., Fung, H. C. & de Silva, R. Untangling the tau gene association with neurodegenerative disorders. *Human Molecular Genetics* **15,** (2006).
37. Talukdar, S. *et al.* FGF21 regulates sweet and alcohol preference. *Cell Metab.* **23,** 344–349 (2016).
38. Zorrilla, E. P., Logrip, M. L. & Koob, G. F. Corticotropin releasing factor: A key role in the neurobiology of addiction. *Frontiers in Neuroendocrinology* **35,** 234–244 (2014).
39. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–315 (2014).
40. Hurley, T. D. & Edenberg, H. J. Genes encoding enzymes involved in ethanol metabolism. *Alcohol Res. Curr. Rev.* **34,** 339–344 (2012).
41. Raimondo, A., Rees, M. G. & Gloyn, A. L. Glucokinase regulatory protein: Complexity at the crossroads of triglyceride and glucose metabolism. *Current Opinion in Lipidology* **26,** 88–95 (2015).
42. Rasheed, H., Stamp, L. K., Dalbeth, N. & Merriman, T. R. Interaction of the GCKR and A1CF loci with alcohol consumption to influence the risk of gout. *Arthritis Res. Ther.* **19,** (2017).

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

43. Nivard, M. G. *et al.* Connecting the dots, genome-wide association studies in substance use. *Molecular Psychiatry* (2016). doi:10.1038/mp.2016.14
44. Vink, J. M. *et al.* Polygenic risk scores for smoking: predictors for alcohol and cannabis use?: Polygenic risk scores for smoking. *Addiction* **109,** 1141–1151 (2014).
45. Drake, R. E. & Mueser, K. T. Co-Occurring Alcohol Use Disorder and Schizophrenia. *Alcohol Res. Health* **26,** (2002).
46. Sonne, S. C. & Brady, K. T. Substance abuse and bipolar comorbidity. *Psychiatric Clinics of North America* **22,** 609–627 (1999).
47. Petrakis, I. L., Gonzalez, G., Rosenheck, R. & Krystal, J. H. Comorbidity of alcoholism and psychiatric disorders - an overview. *Alcohol Res. Health* **26,** 81–89 (2002).
48. Ohlmeier, M. D. *et al.* Comorbidity of alcohol and substance dependence with attention-deficit/hyperactivity disorder (ADHD). *Alcohol Alcohol. Oxf. Oxfs.* **43,** 300–304 (2008).
49. Pfefferbaum, A. *et al.* Longitudinal Changes in Magnetic Resonance Imaging Brain Volumes in Abstinent and Relapsed Alcoholics. *Alcohol. Clin. Exp. Res.* **19,** 1177–1191 (1995).
50. Pfefferbaum, A. *et al.* Brain gray and white matter volume loss accelerates with aging in chronic alcoholics: A quantitative MRI study. *Alcohol. Clin. Exp. Res.* **16,** 1078–1089 (1992).
51. Cardenas, V. A., Studholme, C., Meyerhoff, D. J., Song, E. & Weiner, M. W. Chronic active heavy drinking and family history of problem drinking modulate regional brain tissue volumes. *Psychiatry Res. - Neuroimaging* **138,** 115–130 (2005).
52. Collins, S. E. Associations Between Socioeconomic Factors and Alcohol Outcomes. *Alcohol Res. Curr. Rev.* **38,** 83–94 (2016).
53. Fone, D. L., Farewell, D. M., White, J., Lyons, R. A. & Dunstan, F. D. Socioeconomic patterning of excess alcohol consumption and binge drinking: A cross-sectional study of multilevel associations with neighbourhood deprivation. *BMJ Open* **3,** (2013).
54. Probst, C., Roerecke, M., Behrendt, S. & Rehm, J. Socioeconomic differences in alcohol-attributable mortality compared with all-cause mortality: A systematic review and meta-analysis. *Int. J. Epidemiol.* **43,** 1314–1327 (2014).
55. Lee, S. J. *et al.* Functional limitations, socioeconomic status, and all-cause mortality in moderate alcohol drinkers. *J. Am. Geriatr. Soc.* **57,** 955–962 (2009).
56. Marioni, R. E. *et al.* Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* **44,** 26–32 (2014).
57. Mbarek, H. *et al.* The genetics of alcohol dependence: Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **168,** 739–748 (2015).

			UK Biobank (n=121,630)			23andMe (n=20,328)			Meta (n=141,958)		
Chrom	**Index variant**	**Genes (Credible SNPs)**	**A1**	**Beta (S.E.)**	**P**	**A1**	**Beta (S.E.)**	**P**	**Z Score**	**P**	**Direction**
2p23.3	rs1260326	*SNX17, GCKR* (4)	C	0.02 (0.003)	2.5×10^{-10}	T	-0.10 (0.034)	2.7×10^{-3}	-7.12	1.0×10^{-12}	-/-
2p21	rs4953148	LINC01833 (17)	T	0.02 (0.003)	2.6×10^{-8}	T	0.03 (0.036)	3.5×10^{-1}	-5.60	2.1×10^{-8}	-/-
4p14	rs11940694	*KLB* (7)	G	0.03 (0.003)	8.9×10^{-20}	G	0.08 (0.034)	2.5×10^{-2}	-9.39	5.9×10^{-21}	-/-
17q21.31	rs62062288	*CRHR1, KANSL1, MAPT, NSF, PLEKHM1, SPPL2C, STH, WNT3* (1874)	A	-0.02 (0.003)	1.6×10^{-9}	G	0.09 (0.042)	2.1×10^{-2}	-6.22	5.1×10^{-10}	-/-
3q25.33	rs1920650	*SMC4, TRIM59, B3GAT3P1, KPNA4, SCARNA7, KRT8P12, RPL6P8, ARL14* (132)	C	0.01 (0.003)	1.2×10^{-6}	T	-0.07 (0.037)	4.2×10^{-2}	-5.61	2.0×10^{-8}	-/-
4q23	rs138495951	*ADH1B, ADH1C (3)*	A	-0.14 (0.01)	2.5×10^{-30}	G	0.05 (0.009)	3.6×10^{-7}	-12.51	5.7×10^{-36}	-/-
4q23	rs11733695	*RP11-696N14.1 (1)*	A	-0.16 (0.02)	7.9×10^{-26}	G	0.05 (0.012)	6.7×10^{-6}	-11.43	2.9×10^{-30}	-/-
4q24	rs13135092	*SLC39A8* (6)	G	-0.04 (0.005)	4.9×10^{-13}	G	-0.09 (0.060)	1.2×10^{-1}	7.57	3.7×10^{-14}	+/+
4q23	rs3114045	*ADH1B, ADH1C* (12)	C	0.03 (0.004)	3.1×10^{-10}	T	-0.15 (0.048)	1.6×10^{-3}	-7.19	6.5×10^{-13}	-/-
8q22.1	rs35040843	- (78)	T	0.02 (0.003)	2.5×10^{-7}	T	0.09 (0.038)	8.9×10^{-3}	5.48	4.3×10^{-8}	+/+
10p11.23	rs7078436	*JCAD* (9)	G	-0.01 (0.003)	1.6×10^{-6}	G	-0.10 (0.035)	3.1×10^{-3}	5.75	8.9×10^{-9}	+/+
11p11.2	rs2293576	*SPI1, SLC39A13, PSMC3, RAPSN, CELF1* (90)	A	0.02 (0.003)	5.8×10^{-9}	G	-0.06 (0.035)	1.2×10^{-1}	5.81	6.2×10^{-9}	+/+
19q13.33	rs492602	*FUT2, MAMSTR, RASIP1, IZUMO1, FUT1* (67)	G	0.01 (0.003)	9.5×10^{-8}	G	0.05 (0.033)	1.7×10^{-1}	-5.52	3.4×10^{-8}	-/-

Table 1. Index SNPs associated with AUDIT scores in GWAS meta-analysis of AUDIT scores and corresponding p-values in UK Biobank and 23andMe. Protein coding genes listed are those spanned by SNPs in each credible set.

bioRxiv preprint doi: https://doi.org/10.1101/275917; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.