

TITLE PAGE

Genome-wide association study of Alcohol Use Disorder Identification Test (AUDIT) scores in 20,328 research participants of European ancestry

Running Title: GWAS of AUDIT

Authors: Sandra Sanchez-Roige¹, Pierre Fontanillas², Sarah L. Elson², the 23andMe Research Team², Joshua C. Gray^{3,4}, Harriet de Wit⁵, Lea K. Davis⁶, James MacKillop⁷, Abraham A. Palmer^{1,8*}

Affiliations:

¹ Department of Psychiatry, University of California San Diego, La Jolla, CA, 92093, USA

² 23andMe, Inc., Mountain View, CA, USA

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, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Carrie A.M. Northover, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Catherine H. Wilson, Steven J. Pitts

³ Department of Psychology, University of Georgia, USA

⁴ Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, RI 02903, USA

⁵ Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL 60637, USA

⁶ Vanderbilt Genetics Institute; Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA

⁷ Peter Boris Centre for Addictions Research, McMaster University/St. Joseph's Healthcare Hamilton, Hamilton, ON L8N 3K7, Canada; Homewood Research Institute, Guelph, ON N1E 6K9, Canada

⁸ Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

* **Corresponding author:** Professor Abraham A. Palmer, Department of Psychiatry, University of California San Diego, 9500 Gilman Drive, Mail Code 0667, BRF2 3A24, La Jolla, CA 92093, USA. E-mail: aap@ucsd.edu

Number of words in abstract: 245

Number of words in article body: 3986

Number of figures: 2

Supplementary information: supplemental tables (9) and figures (4)

ABSTRACT

Genetic factors contribute to the risk for developing alcohol use disorder (**AUD**). In collaboration with the genetics company 23andMe, Inc., we performed a genome-wide association (**GWAS**) study of the Alcohol Use Disorder Identification Test (**AUDIT**), an instrument designed to screen for alcohol misuse over the past year. Our final sample consisted of 20,328 research participants of European ancestry (55.3% females; mean age = 53.8, SD = 16.1) who reported ever using alcohol. Our results showed that the ‘chip-heritability’ of AUDIT score, when treated as a continuous phenotype, was 12%. No loci reached genome-wide significance. The gene *ADH1C*, which has been previously implicated in AUD, was among our most significant associations (4.4×10^{-7} ; rs141973904). We also detected a suggestive association on chromosome 1 (2.1×10^{-7} ; rs182344113) near the gene *KCNJ9*, which has been implicated in mouse models of high ethanol drinking. Using LD score regression, we identified positive genetic correlations between AUDIT score and AUD, high alcohol consumption, and cigarette smoking. We also observed an unexpected positive genetic correlation between AUDIT and educational attainment, and additional unexpected negative correlations with BMI/obesity and attention-deficit/hyperactivity disorder (**ADHD**). We conclude that conducting a genetic study using data from a population unselected for AUD and responding to an online questionnaire may represent a cost-effective strategy for elucidating the etiology of AUD.

INTRODUCTION

The heritability of AUD and associated symptomatology such as high alcohol consumption has been estimated at ~50% by family and twin studies (Mbarek et al., 2015; Verhulst, Neale, & Kendler, 2015), with a smaller proportion being attributable to additive effects of common genetic variation (e.g. 33% for AUD (Mbarek et al., 2015); 18% (Vrieze, McGue, Miller, Hicks, & Iacono, 2013) and 13% (Clarke et al., 2017) for alcohol consumption).

The search for specific genes that convey risk for AUD has been an active area of research for several decades. There have been numerous family-based linkage studies of AUD (Agrawal & Bierut, 2012; Edenberg & Foroud, 2014; Enoch, 2013; Rietschel & Treutlein, 2013), as well as candidate gene association studies. Robust linkage signals have been found near the cluster of aldehyde dehydrogenase (*ALDH*) genes on chromosome 12 (Wall, Luczak, & Hiller-Sturmhöfel, 2016), and alcohol dehydrogenase (*ADH*) genes on chromosome 4 (Long et al., 1998; Williams et al., 1999), a result that has also been replicated by candidate gene association studies (Li, Zhao, & Gelernter, 2011, 2012; Luo et al., 2006; Macgregor et al., 2009; Thomasson et al., 1991; van Beek, Willemsen, de Moor, Hottenga, & Boomsma, 2010).

More recently, genome-wide association studies (**GWAS**) have been used to explore the genetic basis of AUD (Hart & Kranzler, 2015; Tawa, Hall, & Lohoff, 2016). The most robust and replicated risk alleles in European, African American, and Asian ancestry populations map to alcohol-metabolizing enzyme genes on chromosome 4q22-23 and 12q24: *ADH1B* (Gelernter et al., 2014; Clarke et al., 2017; Xu et al., 2015), *ADH1C* (Clarke et al., 2017; Edenberg et al., 2010; Frank et al., 2012; Gelernter et al., 2014; Treutlein et al., 2009), *ADH5* (Clarke et al., 2017), *ADH7* (Park et al., 2013) and *ALDH2* (Jorgenson et al., 2017; Park et al., 2013; Quillen et al., 2014; Takeuchi et al., 2011; Yang et al., 2013). More recent GWAS that target alcohol consumption rather than AUD have identified novel genes including *KLB*, which influence both high alcohol consumption in humans (Clarke et al., 2017; Jorgenson et al., 2017; Schumann et al., 2016) and ethanol preference in mice (Schumann et al., 2016).

In collaboration with the genetics company 23andMe, Inc., we performed a GWAS for alcohol misuse using the Alcohol Use Disorders Identification Test (**AUDIT**), a questionnaire developed to screen for alcohol misuse in the past year (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993a). The estimated heritability of AUDIT score is 60%, similar to the heritability of AUD (Mbarek et al 2015). Self-reported AUDIT scores are predictive of future problematic drinking and higher AUD risk (Allen, Litten, Fertig, & Babor, 1997; Boschloo et al., 2010), perhaps because the AUDIT includes questions that are related to the criteria for AUD (e.g. DSM-V, criterion 2: “More than once wanted to cut down or stop drinking, or tried to, but couldn’t?” versus AUDIT, item 4: “How often during the last year have you found that you were not able to stop drinking once you had started?”). A previous GWAS of dichotomized AUDIT scores in 7,842

individuals in an unselected Dutch population did not reveal any significant associations (Mbarek et al 2015). Here, to maximize power and to capture the dimensionality of alcohol misuse, we treated AUDIT scores from 20,328 research participants as a continuous trait rather than dichotomizing it by using a threshold score. We hypothesized that a GWAS for AUDIT scores might identify some of the same alleles that influence AUD, even though our cohort had relatively modest AUDIT scores.

MATERIALS AND METHODS

Sample

All participants included in the analyses were drawn from the customer base of 23andMe, Inc., a consumer genetics company. Participants provided informed consent and answered surveys online under a protocol approved by Ethical and Independent Review Services, an independent AAHRPP-accredited institutional review board (<http://www.eandireview.com>). We restricted the original sample (~25,000 individuals) to a set of unrelated participants of European ancestry (>97% as determined through an analysis of local ancestry [Durand, Do, Mountain, Macpherson, 2014]; see **Supplementary** for additional details) for whom AUDIT data were available. Participants were excluded if they reported that they never drank alcohol ($N = 1,376$). The final number of participants included in the analysis was 20,328. Recruitment occurred over an approximately four-month period in 2015. Sociodemographic details are described in the **Supplementary Table 1**.

AUDIT scores

To evaluate alcohol misuse in the past year, participants completed the AUDIT (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993). We only included subjects who answered yes to the question “Have you ever in your life used alcohol” (i.e., “ever drinkers” vs “never drinkers”). The ten-item AUDIT questionnaire yields scores from 0 to 40. Since the scores were not normally distributed (by visual inspection), we used a log-10 transformation, which is frequently employed to approximate a normal distribution for AUDIT (**Supplementary Table 2**).

Genotyping, quality control and imputation

DNA extraction and genotyping were performed on saliva samples by the National Genetics Institute, a CLIA-certified laboratory. Samples were genotyped on 23andMe custom genotyping array platforms (Illumina HumanHap550+ Bead chip V1 V2, OmniExpress+ Bead chip V3, Custom array V4). Quality control of genetic variants and imputation were performed by 23andMe (see **Supplementary Table 3**). A full description of the methods have been reported elsewhere (Hyde et al., 2016; Lo et al., 2016).

Estimation of variance in AUDIT scores explained by the genotyped SNPs

To estimate the proportion of phenotypic variance explained (‘chip heritability’; h_g^2), we used a genomic

restricted maximum likelihood (GREML) method implemented in Genetic Complex Trait Analysis (GCTA; Yang, Lee, Goddard, & Visscher, 2013). In brief, the GREML method estimates the proportion of variation in a phenotype that is due to all SNPs, and exploits the fact that genotypic similarity (i.e., “relatedness”, measured using genotyped SNPs) will be correlated with phenotypic similarity for heritable traits. Individual-level quality control was implemented, and distantly related individuals with pair-wise relationships were filtered at two thresholds ($K_{IBS} < 0.05$ and $K_{IBS} < 0.03$). We included age (inverse-normalized), self-reported sex (male/female), genotyping platform and top four principal components as covariates. GREML analyses were run using only directly genotyped SNPs to construct the GRM.

Chip-heritability using LD Score Regression

We used a second method to measure chip heritability of AUDIT that is implemented by Linkage Disequilibrium Score Regression Coefficient (LDSC; Bulik-Sullivan et al., 2015a). To standardize the input file (GWAS summary statistics), we followed quality controls as implemented by the LDSC python software package. We used pre-calculated LD scores (“eur_w_ld_chr/” files (Finucane et al., 2015); MHC region excluded) for each SNP using individuals of European ancestry from the 1000 Genomes project, suitable for LD score analysis in European populations. We restricted the analysis to well-imputed SNPs: the SNPs were filtered to HapMap3 SNPs (International HapMap 3 Consortium et al., 2010), and were required to have a minor allele frequency (**MAF**) above 1%. InDels, structural variants, strand-ambiguous SNPs, and SNPs with extremely large effect sizes ($\chi^2 > 80$) were removed. In addition, this approach allowed us to distinguish between genomic inflation attributed to polygenic signal, from confounding biases such as population stratification or polygenicity (LD Score regression intercept > 1 ; Bulik-Sullivan et al., 2015a; Bulik-Sullivan et al., 2015b). As expected under polygenicity, we observed inflation of the median test statistic (Mean $\chi^2 = 1.05$), and adjusted for a genomic control inflation factor λ (the ratio of the observed median χ^2 to that expected by chance) = 1.02. LD score intercept of 1.01 (SE = 0.01) suggested that deviation from the null was due to a polygenic structure rather than inflation due to population structure biases.

Genome-wide association analysis

For quality control of genotyped GWAS results, we removed SNPs with MAF of $< 0.1\%$, a Hardy-Weinberg $P < 10^{-20}$ in Europeans, or a call rate of $< 90\%$. We also removed SNPs that were only genotyped on the 23andMe V1 platform, due to limited sample size, and SNPs on chrM or chrY. Using trio data, we removed SNPs that failed a test for parent-offspring transmission; specifically, we regressed the child’s allele count against the mean parental allele count and removed SNPs with fitted $\beta < 0.6$ and $P < 10^{-20}$ for a test of $\beta < 1$. We also tested genotyped SNPs for genotype date effects, and removed SNPs with $P < 10^{-50}$ by ANOVA of SNP genotypes against a factor dividing genotyping date into 20 roughly equal-sized buckets. For imputed GWAS results, we removed SNPs with average $r^2 < 0.50$ or minimum $r^2 < 0.30$ in any imputation

batch, as well as SNPs that had strong evidence of an imputation batch effect. The batch effect test is an F test from an ANOVA of the SNP dosages against a factor representing imputation batch; we removed results with $P < 1 \times 10^{-50}$. We also removed linear regression results for SNPs with MAF $< 0.1\%$ because tests of low frequency variants can be sensitive to violations of the regression assumption of normally distributed residuals. We performed association tests by linear regression assuming an additive model. We included age (inverse-normal transformed), sex, the top four principal components of genotype, and indicator variables for genotype platforms as covariates (**Supplementary Table 4**).

Phenotypic and genetic correlation analyses

We examined two distinct types of correlations: phenotypic correlations, where both variables were measured in the same individuals, and genetic correlations, where we used AUDIT data from this cohort in conjunction with summary statistics for GWAS conducted in other cohorts (**Supplementary Table 9**). The interpretation of these is different, since phenotypic correlations can be due to a combination of genetic and non-genetic factors, whereas genetic correlations measure only genetically driven correlations.

We used bivariate correlations to examine the direct phenotypic correlations between AUDIT and several variables of interest (age, gender, race, education, annual household), and to identify significant covariates for inclusion in GWAS analysis (**Supplementary Table 5**).

We calculated genetic correlations (r_g) between AUDIT and 30 other complex traits or diseases using LDSC. References for the datasets used are identified in **Supplementary Table 9**. Files were standardized using the steps described in the section above (“Chip-heritability using LD Score Regression”). We did not constrain the intercepts in our analysis because the degree of sample overlap was unknown. We used False Discovery Rate (FDR) to correct for multiple testing (Benjamini & Hochberg, 1995).

Query for expression quantitative trait loci (eQTL)

We queried eQTL evidence for our top ($P < 10^{-7}$) GWAS SNPs using public online resources. We used the Genotype-Tissue Expression Portal (GTEx) to identify eQTLs associated with the SNPs; and the RegulomeDB (Boyle et al., 2012) to identify regulatory DNA elements in non-coding and intergenic regions of the genome in normal cell lines and tissues.

RESULTS

Demographics

Demographic data are shown in **Supplementary Table 1**. Mean age was 53.8 years (SD = 16.1), and 55.3% were women. The annual household income ranged from less than \$14,999 (13.5%) to greater than \$75,000 (21.5%), and the mean years of education completed was 16.8 (SD = 2.6). About half of the participants (49.3%) were married/partnered. Participants showed low to moderate alcohol use, average frequency of alcohol use was ~9 days per month (mean = 8.78, SD = 9.82); during the period of heaviest

lifetime use, subjects reported reaching an average of 13.78 ± 10.96 days over a 30-day period. Over the prior year, 78% of the participants reported drinking 1 or 2 drinks on a single day, and only 28% reported drinking more than 6 drinks on one occasion. Also over the prior year, 92% of the participants were able to stop drinking once they started, and 85% drank alcohol without feeling guilt or remorse.

AUDIT scores

The distribution of the AUDIT scores is shown in **Supplementary Table 2**. The phenotypic correlations between AUDIT and demographic variables measured in the same cohort are shown in the **Supplementary Table 4**. Age and sex were negatively correlated with AUDIT scores; younger individuals and males showed higher AUDIT scores ($r = -0.15$, $P < 0.0001$; $r = -0.17$, $P < 0.0001$, respectively). BMI was *negatively* associated with AUDIT ($r = -0.07$, $P < 0.0001$), whereas household income was positively correlated with AUDIT ($r = 0.07$, $P < 0.0001$). AUDIT scores were slightly higher in unmarried individuals ($r = -0.02$, $P = 0.013$) but we did not observe significant correlations with years of education ($r = 0.01$, $P = 0.085$). Measures of alcohol use were positively correlated with AUDIT scores ($r = 0.50-0.52$, $P < 0.0001$).

Chip-heritability estimates

We estimated the chip-heritability of AUDIT at 12.05% ($\pm 1.91\%$, $P = 2.70 \times 10^{-11}$), which is lower than previous chip-heritability estimates based on dichotomized AUDIT data (30% $\pm 12\%$; Mbarek et al., 2015), and considerably lower than twin based heritabilities of alcohol abuse, dependence and alcoholism (~50%, (Enoch, 2013; Goldman, Oroszi, & Ducci, 2005).

GWAS of AUDIT

The Manhattan and quantile-quantile (QQ) plots for AUDIT are shown in **Figure 1** and **Supplementary Figure 4**, respectively. The most significant association was at rs182344113, located on chromosome 1 ($P = 2.10 \times 10^{-7}$; $\beta = 0.168$, SE = 0.03; MAF = 0.002; **Supplementary Fig. 1**). The association was in an intergenic region of the gene *PIGM*, and near *KCNJ9* (*GIRK3*), which has been implicated in preclinical models of ethanol sensitivity. G-protein-gated inwardly rectifying potassium (GIRK) channels, which are coupled to GABA-B receptors, can be activated by ethanol (Aryal, Dvir, Choe, & Slesinger, 2009; Bodhinathan & Slesinger, 2013). Interestingly, *Kcnj9* knock-out mice exhibit excessive alcohol drinking (Dere et al., 2015).

Several other SNPs also showed suggestive associations (**Supplementary Table 6**), including rs141973904 ($P = 4.40 \times 10^{-7}$, $\beta = -0.05$, SE = 0.01; MAF = 0.02; **Supplementary Fig. 2**) in an intron of *ADH1C*, replicating previous findings for that same SNP in a GWAS of alcohol consumption in males (Clarke et al., 2017), and broadly consistent with numerous previous genetic studies of AUD (Biernacka et al., 2013; Edenberg, 2007; Hn et al., 2013; Thomasson et al., 1991).

Another suggestive association was at rs8059260 ($P = 1.6 \times 10^{-6}$, $\beta = 0.017$, SE = 0.004; MAF = 0.160;

Supplementary Fig. 3), which is near the first exon of *CLEC16A*. Using the Genotype-Tissue Expression Portal (**GTEx**) database, we identified a *cis*-expression quantitative trait loci (**eQTLs**) for *CLEC16A* that co-localized with rs8059260 ($r^2 > 0.79$; see **Supplementary Table 7**). We also found evidence of regulatory elements associated with rs8059260 using the RegulomeDB (Boyle et al., 2012; **Supplementary Table 7**).

Previously studied candidate genes

Our results did not strongly support any of the previously published candidate gene studies of AUD (reviewed in Bühler et al., 2015; **Supplementary Table 8**); various differences including the distinction between AUD and AUDIT, demographics characteristics and especially the low prevalence of AUD in our cohort could partially account for the lack of replication.

Genetic correlations

LD score regression (Bulik-Sullivan et al., 2015b) showed a genetic overlap between AUDIT and numerous traits (**Fig. 2** and **Supplementary Table 9**). Both alcohol consumption and AUD showed the steepest correlations with AUDIT score ($r_g = 0.68$ for both), while the better powered alcohol consumption trait ($N = 70,460$) yielded a significant result ($P = 3.40 \times 10^{-3}$) the P value for the genetic correlation with AUD ($N = 7,280$) fell just short of nominal significance ($P = 6.41 \times 10^{-2}$). We detected a significant negative genetic correlation between AUDIT and ADHD ($r_g = -0.29$; $P = 1.43 \times 10^{-3}$). We observed a positive genetic correlation between AUDIT and lifetime tobacco use ($r_g = 0.42$; $P = 1.52 \times 10^{-3}$). Unexpectedly, we identified a *positive* correlation between AUDIT and years of education ($r_g = 0.27$; $P = 3.14 \times 10^{-5}$), college attainment ($r_g = 0.26$; $P = 8.11 \times 10^{-3}$) and childhood IQ ($r_g = 0.42$; $P = 6.26 \times 10^{-3}$). Also surprisingly, AUDIT was *negatively* genetically correlated with BMI ($r_g = -0.25$; $P = 1.48 \times 10^{-4}$) and adulthood obesity ($r_g = -0.23$; $P = 2.06 \times 10^{-3}$). Height, which is not strongly influenced by individual behavior and thus can be viewed as a negative control, was not genetically correlated with AUDIT ($r_g = 0.02$; $P = 7.12 \times 10^{-1}$).

DISCUSSION

With over 20,000 research participants, ours is by far the largest genetic study of AUDIT. By using a self-report measure of alcohol misuse, as opposed to recruiting a clinically-diagnosed population, we were able to rapidly and inexpensively ascertain a large number of participants. We identified rs141973904 (**Supplementary Fig. 2**) in the *ADH* cluster on chromosome 4q23, which has been previously associated with AUD (Edenberg et al., 2010; Frank et al., 2012a; Gelernter et al., 2014; Treutlein et al., 2009a). The same SNP has recently been associated ($P = 1.22 \times 10^{-15}$) with alcohol consumption using 53,089 males of European ancestry (Clarke et al., 2017). Furthermore, the most associated signal, rs182344113, which resides near the *KCNJ9* (*GIRK3*) gene, was unknown, and is consistent with mouse studies of the homologous gene. The signal at rs182344113 was not significant and it will have to be replicated. We also identified a number of genetic correlations that have behavioral precedents, such as lifetime tobacco use, and several others that were unexpected, including lower BMI and obesity rates, and higher education. We

found that AUDIT was genetically correlated with AUD and alcohol consumption, suggesting that a non-clinical population can be used as an alternative approach to study the genetics of AUD.

We identified modestly associated variants in alcohol metabolizing genes. Genes influencing pharmacokinetics have previously been identified through linkage, candidate gene and genome-wide association studies for AUD and related traits (reviewed in Tawa et al., 2016). The most robust signal was located in the *ADH1C* gene, which contributes to ethanol oxidation. This signal was also identified in earlier GWAS studies for alcohol consumption (Clarke et al., 2017) and AUD status in both European (Clarke et al., 2017; Frank et al., 2012) and African American (Gelernter et al., 2014) populations, suggesting that pharmacokinetic factors are an important contributor to differences in both AUDIT score and AUD.

In addition to SNPs in the alcohol metabolizing genes, linkage and candidate gene studies have identified the *GABRA2*, *OPRM1*, *DRD2* and *ANKK1* genes, as candidate genes associated with AUD phenotypes (Bühler et al., 2015). However, we did not find robust signals for any of them (**Supplementary Table 8**), suggesting that previous studies may have overestimated the effects of these genes, or that these genes are associated with AUD but not AUDIT.

The strongest association we observed resides near *KCNJ9*; the frequency of the implicated allele was very low (MAF = 0.002). *KCNJ9* encodes one of the G protein-activated inwardly rectifying K⁺ channels (GIRK3), which are expressed in the brain (Aguado et al., 2008; Koyrakh et al., 2005), and can be directly activated by ethanol (Herman et al., 2015), even at low concentrations. In humans, two linkage studies have mapped this region for AUD (Hill et al., 2004), age of onset of drinking, harm avoidance, and novelty seeking (Dick et al., 2002). Additionally, DNA methylation levels of CpG in the promoter region of the *GRIK3* gene showed altered expression in postmortem prefrontal cortex tissue of male alcoholics (Wang, Xu, Zhao, Gelernter, & Zhang, 2016). In mice, *Kcnj9* also harbors a QTL for a variety of alcohol-related behaviors, including: ethanol preference (Tarantino, McClearn, Rodriguez, & Plomin, 1998), ethanol aversion (Risinger & Cunningham, 1998), acute sensitivity to ethanol (Tipps, Raybuck, Kozell, Lattal, & Buck, 2016), and hypersensitivity to ethanol withdrawal (Kozell, Walter, Milner, Wickman, & Buck, 2009). Mice lacking GIRK3 in the brain have elevated alcohol drinking, without affecting the sensitivity to ethanol intoxication (Tipps, Raybuck, Kozell, Lattal, & Buck, 2016). Collectively, these results could provide an example of convergent results from humans and mice; however, until this non-significant observation is replicated it should be viewed with caution.

We hypothesized that the genetic risk for AUD is likely to overlap with numerous traits relevant to addiction and psychiatric phenotypes, based on previous epidemiological data (Compton, Thomas, Stinson, & Grant, 2007), twin studies (Kendler, Heath, Neale, Kessler, & Eaves, 1993; Knopik, Heath, Bucholz, Madden, & Waldron, 2009; Pickens, Svikis, McGue, & LaBuda, 1995) and recent genetic correlations between alcohol consumption and neuropsychiatric traits (Clarke et al., 2017). We showed positive genetic correlations

between AUDIT and lifetime cigarette smoking, as previously observed between alcohol consumption and daily cigarettes, and tobacco initiation (Nivard et al., 2016; Vink et al., 2014). We also observed shared genetic architecture across AUDIT and two other AUD traits: alcohol consumption (Schumann et al., 2016), and AUD diagnosis ($P = 0.062$; Gelernter et al., 2014). With regards to other psychiatric traits, we found a negative correlation with ADHD; but there were no other significant genetic correlations between AUDIT and psychiatric traits. This result is generally consistent with a recent GWAS by Clarke and colleagues (2017), where they did not observe strong correlations between alcohol consumption and psychiatric conditions, with the exception of schizophrenia. This difference between our results and those of Clarke et al (2017) may be due to differences in cohorts used.

Unexpectedly, we found *positive* genetic correlations between AUDIT and years of education, college attainment and childhood IQ. This association was suggestive for the within-sample phenotypic correlation (**Supplementary Table 5**) and was significant for the genetic correlation (**Supplementary Table 9**).

Consistent with this finding, Clarke and colleagues (2017) reported that college attainment and years of education were positively genetically correlated with alcohol consumption in females but not males.

Also unexpectedly, we observed *negative genetic* correlations between AUDIT and BMI and obesity. We also observed a *phenotypic* correlation between high AUDIT scores and low BMI. Previous studies have shown both positive and negative phenotypic correlations between alcohol use and BMI and obesity (Breslow & Smothers, 2005; Green et al., 2016; Hn et al., 2013; Sobczyk-Kopciol et al., 2011; Tolstrup et al., 2005), which may reflect differences in the populations used. The generalizability and biological meaning of these observations will require further research.

Our study is not without limitations. Cumulative AUDIT scores reflect two distinct constructs: one measuring alcohol consumption and another measuring alcohol-related problems; thus, AUDIT scores may conflate multiple genetic signals (Bergman & Källmén, 2002; Shevlin & Smith, 2007). In addition, our study focused on a cohort with relatively low levels of alcohol use; the unexpected positive genetic correlation between AUDIT and educational attainment, and the negative genetic correlation between AUDIT and both BMI/obesity and ADHD, may not generalize to cohorts with higher levels of alcohol use (Goldman et al., 2005). Another limitation of this study is the reliance on self-reported alcohol consumption, which may have induced biases and result in a more heterogeneous sample (Agrawal et al., 2012). Finally, AUDIT explicitly asks about alcohol use in the past year (i.e. state rather than trait), this temporal specificity is suboptimal for a genetic study.

Although we used a unique and potentially powerful technique to examine the genetic basis of alcohol misuse, we recognize that alcohol consumption, misuse and dependence are influenced by numerous factors, both genetic and environmental (e.g. availability of alcohol, social norms, laws, psychosocial and personality factors, expectancies, health factors). Further, individuals vary in susceptibility at every stage of

alcohol use from initiation to severe dependence, including the continued use after a first drink, the direct subjective and behavioral effects of the drug, withdrawal severity, tolerance and susceptibility to relapse, among others. Genetic factors are likely to influence variability at each of these stages, but different stages may be influenced by different sources of genetic variability. Thus, whereas we have investigated genetic variance related to an intermediate outcome measure of alcohol misuse (as measured by the AUDIT), it remains to be determined exactly how genetic sources of variation influence alcohol consumption. This ambiguity limits our ability to elucidate the underlying molecular mechanisms identified by GWAS of alcohol use and abuse.

Nonetheless, unlike studies of disease traits, which require careful diagnosis and ascertainment, we rapidly obtained a large cohort for which genotype data were available. We replicated a previously identified signal (*ADH1C*), and identified a novel GWAS signal (near *KCNJ9*) that has preclinical correlates. Our approach shows that genetic studies of AUDIT in community-based samples are an economical and effective alternative to rigorously diagnosed AUD cohorts that can nevertheless be used to gain insight into the biology of AUD, and comorbid psychopathology.

FUNDING AND DISCLOSURE

J.M.'s contributions were partially supported by the Peter Boris Chair in Addictions Research. Pierre Fontanillas, Sarah Elson, and members of the 23andMe Research Team are employees of 23andMe, Inc.

ACKNOWLEDGEMENTS

We would like to thank the 23andMe research participants and employees for making this work possible.

Supplementary Information with nine additional tables are available on the *Neuropsychopharmacology* website

REFERENCES

- Agrawal, A., & Bierut, L. J. (2012). Identifying genetic variation for alcohol dependence. *Alcohol Research: Current Reviews*, 34(3), 274–281.
- Agrawal, A., Freedman, N. D., Cheng, Y.-C., Lin, P., Shaffer, J. R., Sun, Q., ... GENEVA Consortium. (2012). Measuring alcohol consumption for genomic meta-analyses of alcohol intake: opportunities and challenges. *The American Journal of Clinical Nutrition*, 95(3), 539–547. <https://doi.org/10.3945/ajcn.111.015545>
- Aguado, C., Colón, J., Ciruela, F., Schlaudraff, F., Cabañero, M. J., Perry, C., ... Luján, R. (2008). Cell type-specific subunit composition of G protein-gated potassium channels in the cerebellum. *Journal of Neurochemistry*, 105(2), 497–511. <https://doi.org/10.1111/j.1471-4159.2007.05153.x>
- Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A Review of Research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research*, 21(4), 613–619. <https://doi.org/10.1111/j.1530-0277.1997.tb03811.x>
- Aryal, P., Dvir, H., Choe, S., & Slesinger, P. A. (2009). A discrete alcohol pocket involved in GIRK channel activation. *Nature Neuroscience*, 12(8), 988–995. <https://doi.org/10.1038/nn.2358>
- Benjamini, Y., & Hochberg, M. T. (1995). Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *Journal of the Royal Statistical Society*, 57(1), 289–300.
- Bergman, H., & Källmén, H. (2002). Alcohol use among Swedes and a psychometric evaluation of the alcohol use disorders identification test. *Alcohol and Alcoholism (Oxford, Oxfordshire)*, 37(3), 245–251.
- Biernacka, J. M., Geske, J. R., Schneekloth, T. D., Frye, M. A., Cunningham, J. M., Choi, D.-S., ... Karpyak, V. M. (2013). Replication of genome wide association studies of alcohol dependence: support for association with variation in ADH1C. *PloS One*, 8(3), e58798. <https://doi.org/10.1371/journal.pone.0058798>
- Bodhinathan, K., & Slesinger, P. A. (2013). Molecular mechanism underlying ethanol activation of G-protein-gated inwardly rectifying potassium channels. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45), 18309–18314. <https://doi.org/10.1073/pnas.1311406110>
- Boschloo, L., Vogelzangs, N., Smit, J. H., van den Brink, W., Veltman, D. J., Beekman, A. T. F., & Penninx, B. W. J. H.

- (2010). The performance of the Alcohol Use Disorder Identification Test (AUDIT) in detecting alcohol abuse and dependence in a population of depressed or anxious persons. *Journal of Affective Disorders*, *126*(3), 441–446. <https://doi.org/10.1016/j.jad.2010.04.019>
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., ... Snyder, M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*, *22*(9), 1790–1797. <https://doi.org/10.1101/gr.137323.112>
- Breslow, R. A., & Smothers, B. A. (2005). Drinking patterns and body mass index in never smokers: National Health Interview Survey, 1997–2001. *American Journal of Epidemiology*, *161*(4), 368–376. <https://doi.org/10.1093/aje/kwi061>
- Bühler, K.-M., Giné, E., Echeverry-Alzate, V., Calleja-Conde, J., de Fonseca, F. R., & López-Moreno, J. A. (2015). Common single nucleotide variants underlying drug addiction: more than a decade of research. *Addiction Biology*, *20*(5), 845–871. <https://doi.org/10.1111/adb.12204>
- Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ... Neale, B. M. (2015a). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. <https://doi.org/10.1038/ng.3406>
- Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., ... Neale, B. M. (2015b). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. <https://doi.org/10.1038/ng.3211>
- Clarke, T.-K., Adams, M. J., Davies, G., Howard, D. M., Hall, L. S., Padmanabhan, S., ... McIntosh, A. M. (2017). Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112,117). *bioRxiv*, 116707. <https://doi.org/10.1101/116707>
- Compton, W. M., Thomas, Y. F., Stinson, F. S., & Grant, B. F. (2007). Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Archives of General Psychiatry*, *64*(5), 566–576. <https://doi.org/10.1001/archpsyc.64.5.566>

- Dere, E., Winkler, D., Ritter, C., Ronnenberg, A., Poggi, G., Patzig, J., ... Werner, H. B. (2015). Gpm6b deficiency impairs sensorimotor gating and modulates the behavioral response to a 5-HT_{2A/C} receptor agonist. *Behavioural Brain Research*, 277, 254–263. <https://doi.org/10.1016/j.bbr.2014.04.021>
- Dick, D. M., Nurnberger, J., Edenberg, H. J., Goate, A., Crowe, R., Rice, J., ... Foroud, T. (2002). Suggestive linkage on chromosome 1 for a quantitative alcohol-related phenotype. *Alcoholism, Clinical and Experimental Research*, 26(10), 1453–1460. <https://doi.org/10.1097/01.ALC.0000034037.10333.FD>
- 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.
- Edenberg, H. J., & Foroud, T. (2014). Chapter 32 - Genetics of alcoholism. In E. V. S. and A. Pfefferbaum (Ed.), *Handbook of Clinical Neurology* (Vol. 125, pp. 561–571). Elsevier. <https://doi.org/10.1016/B978-0-444-62619-6.00032-X>
- Edenberg, H. J., Koller, D. L., Xuei, X., Wetherill, L., McClintick, J. N., Almasy, L., ... Foroud, T. (2010). Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcoholism, Clinical and Experimental Research*, 34(5), 840–852. <https://doi.org/10.1111/j.1530-0277.2010.01156.x>
- Enoch, M.-A. (2013). Genetic influences on the development of alcoholism. *Current Psychiatry Reports*, 15(11), 412. <https://doi.org/10.1007/s11920-013-0412-1>
- Durand, E. Y., Do, X. B., Mountain, J. L., Macpherson, M. (2014). Ancestry Composition: A Novel, Efficient Pipeline for Ancestry Deconvolution. <https://doi.org/https://doi.org/10.1101/010512>
- Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., ... Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11), 1228–1235. <https://doi.org/10.1038/ng.3404>
- Frank, J., Cichon, S., Treutlein, J., Ridinger, M., Mattheisen, M., Hoffmann, P., ... Rietschel, M. (2012). Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addiction Biology*, 17(1), 171–180. <https://doi.org/10.1111/j.1369-1600.2011.00395.x>

- Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Koesterer, R., Smith, A. H., ... Farrer, L. A. (2014). Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Molecular Psychiatry*, *19*(1), 41–49. <https://doi.org/10.1038/mp.2013.145>
- Goldman, D., Oroszi, G., & Ducci, F. (2005). The genetics of addictions: uncovering the genes. *Nature Reviews. Genetics*, *6*(7), 521–532. <https://doi.org/10.1038/nrg1635>
- Green, M. A., Strong, M., Razak, F., Subramanian, S. V., Relton, C., & Bissell, P. (2016). Who are the obese? A cluster analysis exploring subgroups of the obese. *Journal of Public Health*, *38*(2), 258–264. <https://doi.org/10.1093/pubmed/fdv040>
- Hart, A. B., & Kranzler, H. R. (2015). Alcohol dependence genetics: Lessons learned from genome-wide association studies (GWAS) and post-GWAS analyses. *Alcoholism, Clinical and Experimental Research*, *39*(8), 1312–1327. <https://doi.org/10.1111/acer.12792>
- Herman, M. A., Sidhu, H., Stouffer, D. G., Kreifeldt, M., Le, D., Cates-Gatto, C., ... Contet, C. (2015). GIRK3 gates activation of the mesolimbic dopaminergic pathway by ethanol. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), 7091–7096. <https://doi.org/10.1073/pnas.1416146112>
- Hill, S. Y., Shen, S., Zezza, N., Hoffman, E. K., Perlin, M., & Allan, W. (2004). A genome wide search for alcoholism susceptibility genes. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *128B*(1), 102–113. <https://doi.org/10.1002/ajmg.b.30013>
- Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., ... Winslow, A. R. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*, *48*(9), 1031–1036. <https://doi.org/10.1038/ng.3623>
- International HapMap 3 Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., ... McEwen, J. E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52–58. <https://doi.org/10.1038/nature09298>
- Jorgenson, E., Thai, K. K., Hoffmann, T. J., Sakoda, L. C., Kvale, M. N., Banda, Y., ... Choquet, H. (2017). Genetic

- contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. *Molecular Psychiatry*. <https://doi.org/10.1038/mp.2017.101>
- Kendler, K. S., Heath, A. C., Neale, M. C., Kessler, R. C., & Eaves, L. J. (1993). Alcoholism and major depression in women. A twin study of the causes of comorbidity. *Archives of General Psychiatry*, *50*(9), 690–698.
- Knopik, V. S., Heath, A. C., Bucholz, K. K., Madden, P. A. F., & Waldron, M. (2009). Genetic and environmental influences on externalizing behavior and alcohol problems in adolescence: a female twin study. *Pharmacology, Biochemistry, and Behavior*, *93*(3), 313–321. <https://doi.org/10.1016/j.pbb.2009.03.011>
- Koyrakh, L., Luján, R., Colón, J., Karschin, C., Kurachi, Y., Karschin, A., & Wickman, K. (2005). Molecular and cellular diversity of neuronal G-protein-gated potassium channels. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(49), 11468–11478. <https://doi.org/10.1523/JNEUROSCI.3484-05.2005>
- Kozell, L. B., Walter, N. A. R., Milner, L. C., Wickman, K., & Buck, K. J. (2009). Mapping a barbiturate withdrawal locus to a 0.44 Mb interval and analysis of a novel null mutant identify a role for Kcnj9 (GIRK3) in withdrawal from pentobarbital, zolpidem, and ethanol. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(37), 11662–11673. <https://doi.org/10.1523/JNEUROSCI.1413-09.2009>
- Li, D., Zhao, H., & Gelernter, J. (2011). Strong Association of the Alcohol Dehydrogenase 1B Gene (ADH1B) with Alcohol Dependence and Alcohol-Induced Medical Diseases. *Biological Psychiatry*, *70*(6), 504–512. <https://doi.org/10.1016/j.biopsych.2011.02.024>
- Li, D., Zhao, H., & Gelernter, J. (2012). Strong Protective Effect of The Aldehyde Dehydrogenase Gene (ALDH2) 504Iys (*2) Allele Against Alcoholism And Alcohol-Induced Medical Diseases in Asians. *Human Genetics*, *131*(5), 725–737. <https://doi.org/10.1007/s00439-011-1116-4>
- Lo, M.-T., Hinds, D. A., Tung, J. Y., Franz, C., Fan, C.-C., Wang, Y., ... Chen, C.-H. (2016). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nature Genetics*, *49*(1), 152–156. <https://doi.org/10.1038/ng.3736>
- Long, J. C., Knowler, W. C., Hanson, R. L., Robin, R. W., Urbanek, M., Moore, E., ... Goldman, D. (1998). Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an

American Indian population. *American Journal of Medical Genetics*, 81(3), 216–221.

Luo, X., Kranzler, H. R., Zuo, L., Lappalainen, J., Yang, B., & Gelernter, J. (2006). ADH4 gene variation is associated with alcohol dependence and drug dependence in European Americans: results from HWD tests and case-control association studies. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 31(5), 1085–1095. <https://doi.org/10.1038/sj.npp.1300925>

Macgregor, S., Lind, P. A., Bucholz, K. K., Hansell, N. K., Madden, P. A. F., Richter, M. M., ... Whitfield, J. B. (2009). Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Human Molecular Genetics*, 18(3), 580–593. <https://doi.org/10.1093/hmg/ddn372>

Mbarek, H., Milaneschi, Y., Fedko, I. O., Hottenga, J.-J., de Moor, M. H. M., Jansen, R., ... Vink, J. M. (2015). The genetics of alcohol dependence: Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 168(8), 739–748. <https://doi.org/10.1002/ajmg.b.32379>

Nivard, M. G., Verweij, K. J. H., Minică, C. C., Treur, J. L., International Cannabis Consortium, Vink, J. M., & Boomsma, D. I. (2016). Connecting the dots, genome-wide association studies in substance use. *Molecular Psychiatry*, 21(6), 733–735. <https://doi.org/10.1038/mp.2016.14>

Park, B. L., Kim, J. W., Cheong, H. S., Kim, L. H., Lee, B. C., Seo, C. H., ... Choi, I.-G. (2013). Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. *Human Genetics*, 132(6), 657–668. <https://doi.org/10.1007/s00439-013-1281-8>

Pickens, R. W., Svikis, D. S., McGue, M., & LaBuda, M. C. (1995). Common genetic mechanisms in alcohol, drug, and mental disorder comorbidity. *Drug and Alcohol Dependence*, 39(2), 129–138.

Quillen, E. E., Chen, X.-D., Almasy, L., Yang, F., He, H., Li, X., ... Gelernter, J. (2014). 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. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official*

Publication of the International Society of Psychiatric Genetics, 165B(2), 103–110.

<https://doi.org/10.1002/ajmg.b.32213>

- Rietschel, M., & Treutlein, J. (2013). The genetics of alcohol dependence. *Annals of the New York Academy of Sciences, 1282*, 39–70. <https://doi.org/10.1111/j.1749-6632.2012.06794.x>
- Risinger, F. O., & Cunningham, C. L. (1998). Ethanol-induced conditioned taste aversion in BXD recombinant inbred mice. *Alcoholism, Clinical and Experimental Research, 22(6)*, 1234–1244.
- Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993a). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction (Abingdon, England), 88(6)*, 791–804.
- Schumann, G., Liu, C., O'Reilly, P., Gao, H., Song, P., Xu, B., ... Elliott, P. (2016). KLB is associated with alcohol drinking, and its gene product β -Klotho is necessary for FGF21 regulation of alcohol preference. *Proceedings of the National Academy of Sciences of the United States of America, 113(50)*, 14372–14377.
<https://doi.org/10.1073/pnas.1611243113>
- Shevlin, M., & Smith, G. W. (2007). The factor structure and concurrent validity of the alcohol use disorder identification test based on a nationally representative UK sample. *Alcohol and Alcoholism (Oxford, Oxfordshire), 42(6)*, 582–587. <https://doi.org/10.1093/alcalc/aggm045>
- Sobczyk-Kopciol, A., Broda, G., Wojnar, M., Kurjata, P., Jakubczyk, A., Klimkiewicz, A., & Ploski, R. (2011). Inverse association of the obesity predisposing FTO rs9939609 genotype with alcohol consumption and risk for alcohol dependence. *Addiction (Abingdon, England), 106(4)*, 739–748. <https://doi.org/10.1111/j.1360-0443.2010.03248.x>
- Takeuchi, F., Isono, M., Nabika, T., Katsuya, T., Sugiyama, T., Yamaguchi, S., ... Kato, N. (2011). Confirmation of ALDH2 as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circulation Journal: Official Journal of the Japanese Circulation Society, 75(4)*, 911–918.
- Tarantino, L. M., McClearn, G. E., Rodriguez, L. A., & Plomin, R. (1998). Confirmation of quantitative trait loci for alcohol preference in mice. *Alcoholism, Clinical and Experimental Research, 22(5)*, 1099–1105.

- Tawa, E. A., Hall, S. D., & Lohoff, F. W. (2016). Overview of the Genetics of Alcohol Use Disorder. *Alcohol and Alcoholism (Oxford, Oxfordshire)*, 51(5), 507–514. <https://doi.org/10.1093/alcalc/agw046>
- Thomasson, H. R., Edenberg, H. J., Crabb, D. W., Mai, X. L., Jerome, R. E., Li, T. K., ... Yin, S. J. (1991). Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *American Journal of Human Genetics*, 48(4), 677–681.
- Tipps, M. E., Raybuck, J. D., Kozell, L. B., Lattal, K. M., & Buck, K. J. (2016). G Protein-Gated Inwardly Rectifying Potassium Channel Subunit 3 Knock-Out Mice Show Enhanced Ethanol Reward. *Alcoholism, Clinical and Experimental Research*, 40(4), 857–864. <https://doi.org/10.1111/acer.13012>
- Tolstrup, J. S., Heitmann, B. L., Tjønneland, A. M., Overvad, O. K., Sørensen, T. I. A., & Grønbaek, M. N. (2005). The relation between drinking pattern and body mass index and waist and hip circumference. *International Journal of Obesity (2005)*, 29(5), 490–497. <https://doi.org/10.1038/sj.ijo.0802874>
- van Beek, J. H. D. A., Willemsen, G., de Moor, M. H. M., Hottenga, J. J., & Boomsma, D. I. (2010). Associations between ADH gene variants and alcohol phenotypes in Dutch adults. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, 13(1), 30–42. <https://doi.org/10.1375/twin.13.1.30>
- 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. <https://doi.org/10.1017/S0033291714002165>
- Vink, J. M., Hottenga, J. J., de Geus, E. J. C., Willemsen, G., Neale, M. C., Furberg, H., & Boomsma, D. I. (2014). Polygenic risk scores for smoking: predictors for alcohol and cannabis use? *Addiction (Abingdon, England)*, 109(7), 1141–1151. <https://doi.org/10.1111/add.12491>
- Vrieze, S. I., McGue, M., Miller, M. B., Hicks, B. M., & Iacono, W. G. (2013). Three mutually informative ways to understand the genetic relationships among behavioral disinhibition, alcohol use, drug use, nicotine use/dependence, and their co-occurrence: twin biometry, GCTA, and genome-wide scoring. *Behavior Genetics*, 43(2), 97–107. <https://doi.org/10.1007/s10519-013-9584-z>
- Wall, T. L., Luczak, S. E., & Hiller-Sturmhöfel, S. (2016). Biology, Genetics, and Environment. *Alcohol Research*:

Current Reviews, 38(1), 59–68.

Wang, F., Xu, H., Zhao, H., Gelernter, J., & Zhang, H. (2016). DNA co-methylation modules in postmortem prefrontal cortex tissues of European Australians with alcohol use disorders. *Scientific Reports*, 6, 19430.

<https://doi.org/10.1038/srep19430>

Williams, J. T., Begleiter, H., Porjesz, B., Edenberg, H. J., Foroud, T., Reich, T., ... Blangero, J. (1999). Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. II. Alcoholism and event-related potentials.

American Journal of Human Genetics, 65(4), 1148–1160. <https://doi.org/10.1086/302571>

Xu, K., Kranzler, H. R., Sherva, R., Sartor, C. E., Almasy, L., Koesterer, R., ... Gelernter, J. (2015). Genomewide Association Study for Maximum Number of Alcoholic Drinks in European Americans and African Americans.

Alcoholism, Clinical and Experimental Research, 39(7), 1137–1147. <https://doi.org/10.1111/acer.12751>

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2013). Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods in Molecular Biology* (Clifton, N.J.), 1019, 215–236.

https://doi.org/10.1007/978-1-62703-447-0_9

Yang, X., Lu, X., Wang, L., Chen, S., Li, J., Cao, J., ... Gu, D. (2013). Common variants at 12q24 are associated with drinking behavior in Han Chinese. *The American Journal of Clinical Nutrition*, 97(3), 545–551.

<https://doi.org/10.3945/ajcn.112.046482>

FIGURE LEGENDS

Figure 1 Results of GWAS on AUDIT. **(a)** Manhattan plot of GWAS results indicating the strongest associations between the 22 autosomes, X chromosome, and AUDIT. Line denotes genome-wide significance ($P < 5 \times 10^{-8}$).

Figure 2 Genetic correlations between AUDIT and several traits: **(a)** neuropsychiatric, **(b)** smoking, **(c)** personality, **(d)** cognition, **(e)** anthropomorphic. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$.

FIGURE 1

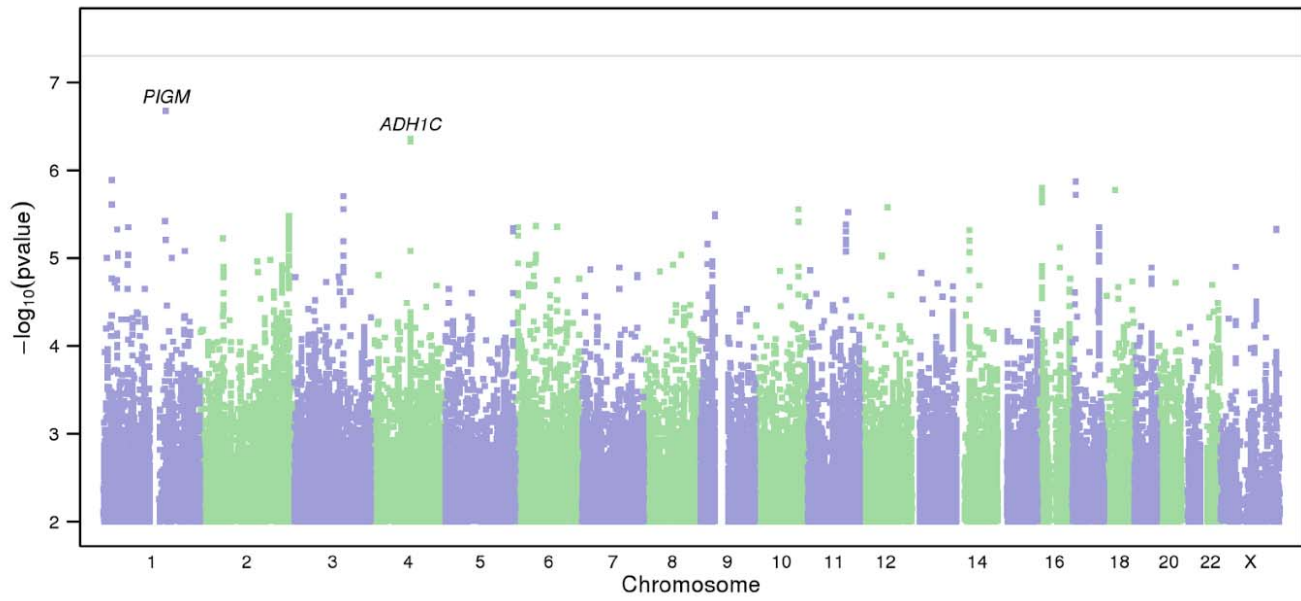
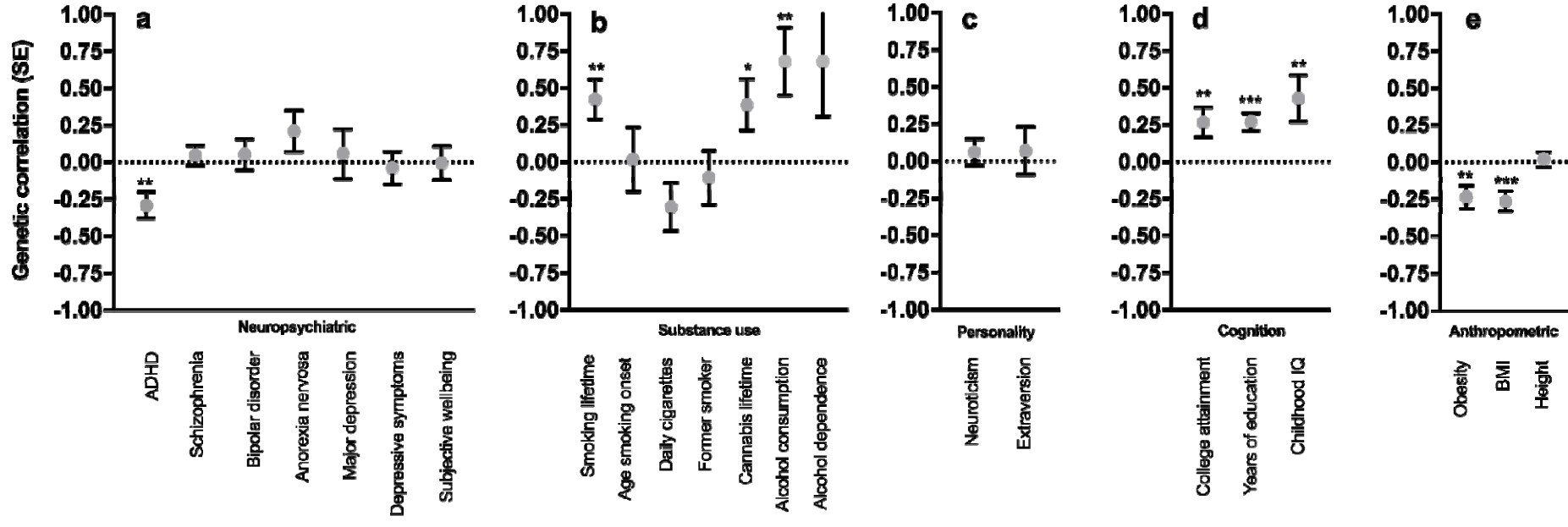
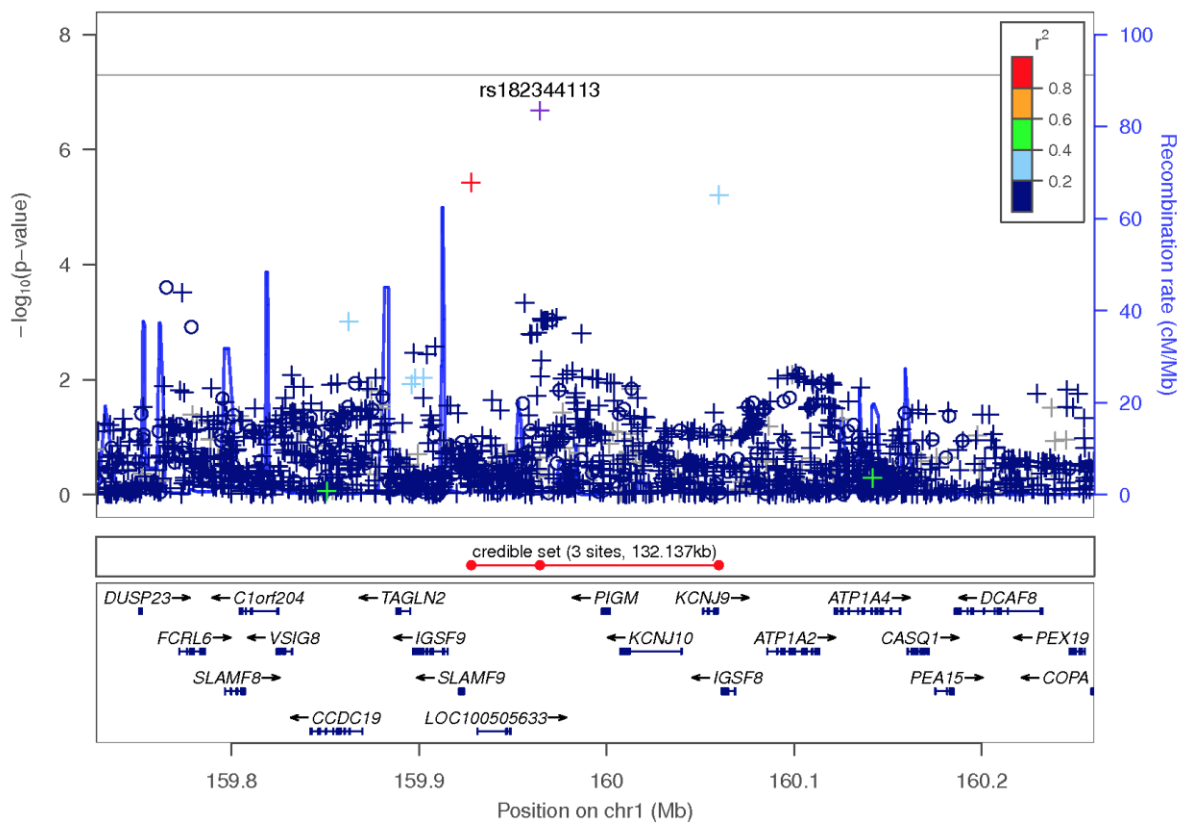
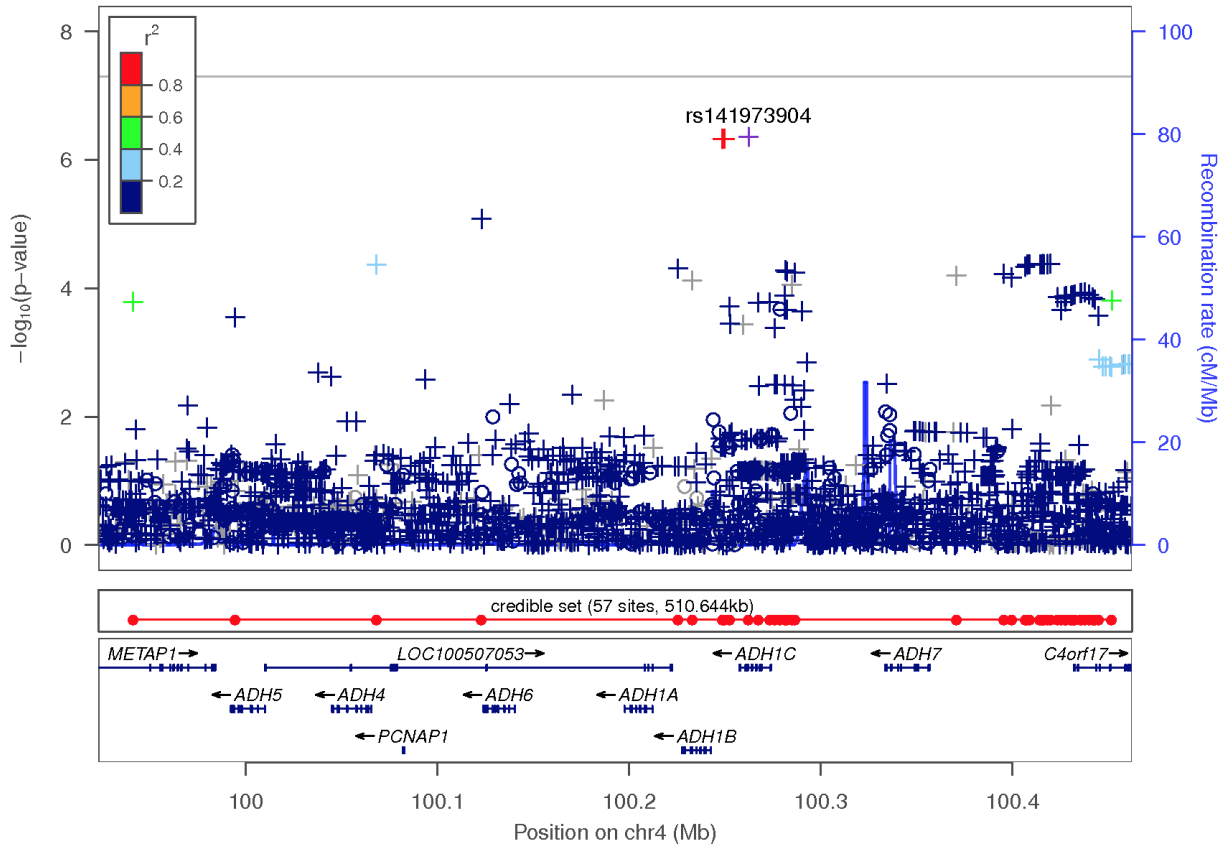


FIGURE 2

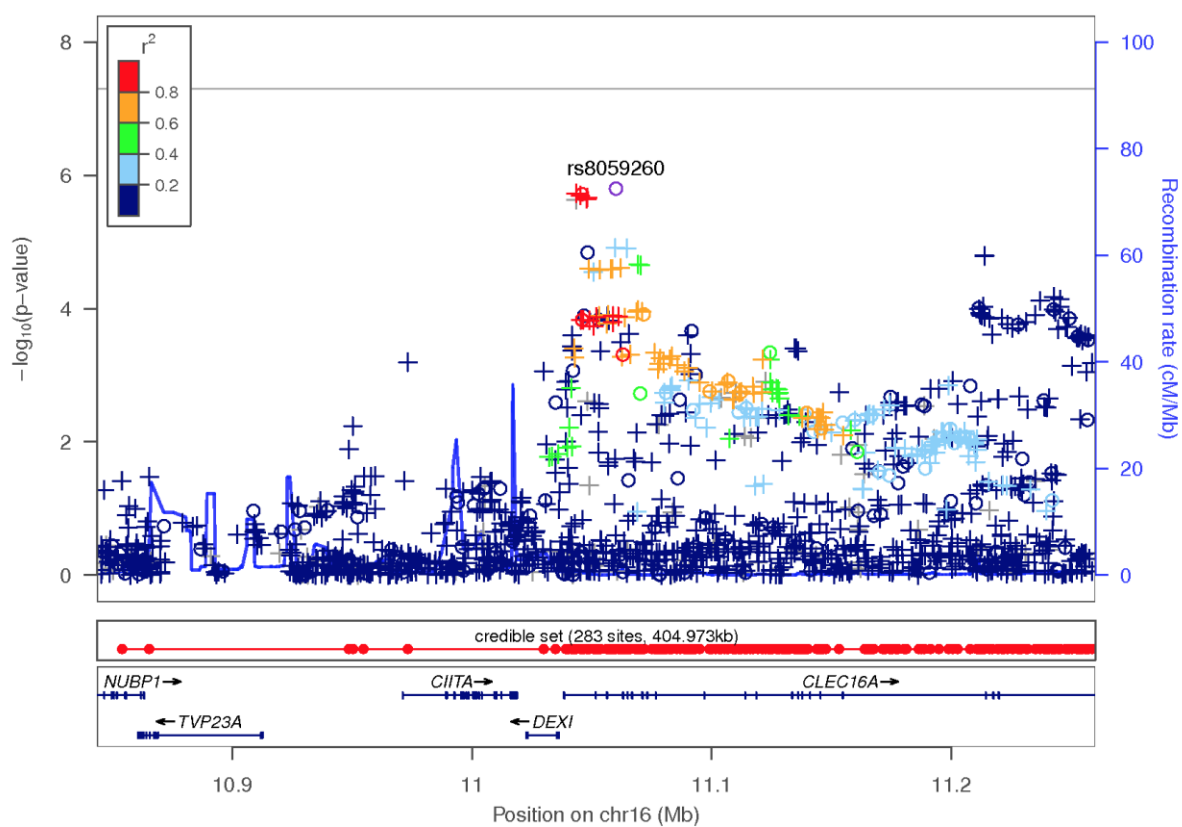




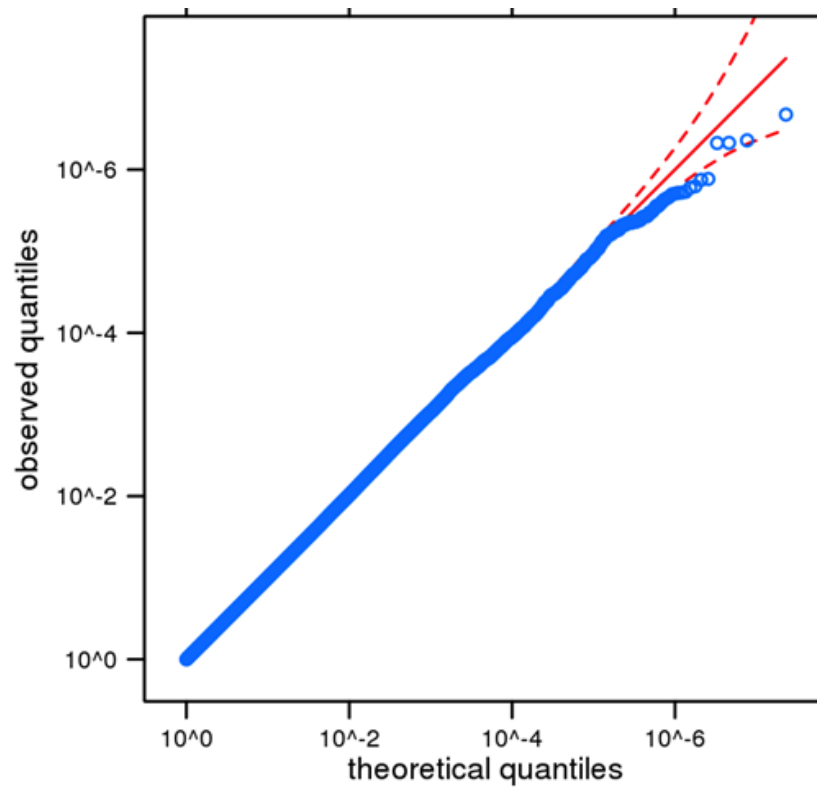
Supplementary Figure 1 Regional association plot focusing on top SNP rs182344113 on chromosome 1.



Supplementary Figure 2 Regional association plot showing the second index SNP, rs141973904, located in the gene *ADH1C* on chromosome 4.



Supplementary Figure 3 Regional association plot showing the index SNP rs8059260, located in the gene *CLEC16A* on chromosome 16.



Supplementary Figure 4 Quantile-Quantile (QQ) plot of AUDIT. The results have been adjusted for a genomic control inflation factor $\lambda=1.021$ (sample size = 20,328).

Supplementary Table 1

Demographic characteristics of the 23andMe cohort

| | 23andMe Cohort * | N ** |
|---|------------------|--------|
| Demographics | | |
| Age in years (mean, S.D) | 53.79 (16.08) | 23,677 |
| Age in years (range) | 18-101 | 23,677 |
| Female | 55.30% | 23,677 |
| BMI (mean, S.D) | 27.00 (5.79) | 22,889 |
| Years of education (mean, S.D) | 16.75 (2.62) | 20,460 |
| Household income (mean, S.D) | 6.07 (1.98) | 17,919 |
| Marital status (married vs. unmarried) | 49.30% | 21,862 |
| Alcohol use ¹ | | |
| Alcohol lifetime use (N, never/ever) | 1,376 / 23,108 | 24,484 |
| Days of alcohol use (past 30 days) | 8.78 (9.82) | 21,727 |
| Days of alcohol use (heaviest, lifetime, 30-day period) | 13.78 (10.96) | 21,229 |

Abbreviations * European Ancestry only; ** N prior to GWAS analysis (final sample size = 20,328); N sample size; *BMI* Body Mass Index; *Household income* 9 categories from 10K to 500K

References

¹ <https://www.phenxtoolkit.org>

Supplementary Table 2

Distribution (%) of AUDIT scores

| AUDIT scores | Total | Sex | | Age | | | |
|---------------|-------|--------|--------|--------|--------|--------|--------|
| | | M | F | 18-30 | 30-45 | 45-60 | >60 |
| 0 - 0.301 | 5085 | 32.60% | 67.40% | 6.60% | 19.20% | 28.80% | 45.30% |
| 0.301 - 0.602 | 6452 | 42.30% | 57.70% | 8.50% | 21.10% | 28.80% | 41.50% |
| 0.602 - 0.778 | 4742 | 51.70% | 48.30% | 9.50% | 22.90% | 24.00% | 43.70% |
| 0.778 - 1.61 | 4049 | 59.20% | 40.80% | 15.40% | 32.10% | 25.00% | 27.50% |

Higher values indicate higher AUDIT (log₁₀ transformed) scores. Sample range from 0-40 (Mean= 3.84, Standard Deviation= 3.47). *Abbreviations* AUDIT Alcohol Use Disorder Identification Test *Total* sample size *M* males *F* females *18-30 / 30-45 / 45-60 / >60* years of age

References

Saunders JB, Aasland OG, Babor TF, Fuente JR de la, Grant M (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addict Abingdon Engl* **88**: 791–804.

Supplementary Table 3

Quality Control filters at SNP-level for genotyped and imputed dosage data, and association test results data

| | Genotyped | Imputed Dosage | Association Results |
|--------------------------------------|-----------|----------------|---------------------|
| no filters | 1,050,074 | 15,574,220 | - |
| not V1/V2-only platforms, chrM, chrY | 1,015,755 | - | - |
| parent-offspring test | 1,012,649 | - | - |
| Minor Allele Frequency > 0% | 1,008,228 | 15,574,220 | - |
| Hardy-Weinberg Equilibrium > 1e-20 | 954,373 | - | - |
| genotype call rate > 90% | 934,855 | - | - |
| imputation quality | - | 13,230,301 | - |
| batch effects | 926,512 | 13,204,909 | - |
| imputed only | - | - | 12,354,282 |
| both passed | - | - | 13,204,463 |
| genotyped only | - | - | 13,280,794 |
| no test result | - | - | 12,010,653 |
| Minor Allele Frequency < 0.1% * | - | - | 11,508,740 |

*Notes * poor residual behavior*

Supplementary Table 4
Null model with covariates

| Covariates | Estimate | SE | t | Pr(> t) |
|--------------------------|----------|--------|-------|------------------------|
| Sex | -0.1093 | 0.0037 | -29.3 | 2.1×10^{-184} |
| Age (inverse normalized) | -0.0516 | 0.0023 | -23.0 | 2.3×10^{-115} |
| PC 0 | 0.0199 | 0.0019 | 10.8 | 6.3×10^{-27} |
| PC 1 | -0.0005 | 0.0019 | -0.3 | 0.76 |
| PC 2 | -0.0042 | 0.0019 | -2.3 | 0.023 |
| PC 3 | -0.0064 | 0.0019 | -3.5 | 5.4×10^{-5} |
| PC 4 | 0.0066 | 0.0019 | 3.5 | 4.4×10^{-4} |
| Platform v2 | 0.0221 | 0.0103 | 2.1 | 0.032 |
| Platform v3.1 | 0.0183 | 0.0063 | 2.9 | 3.7×10^{-3} |
| Platform v4 | 0.0174 | 0.0059 | 3.0 | 0.003 |

Notes Results of fitting a model for delay discounting based on covariates. Principal components (PC) computed from all 23andMe research participants with European ancestry. Effect sizes are in units of standard deviations.

Supplementary Table 5

Phenotypic correlations between AUDIT scores and demographic, drug (alcohol, tobacco, marijuana; current/heaviest use) and caffeine use, and impulsivity traits

| Phenotype | Pearson r | P value | N |
|---|-------------|-------------------------|--------|
| Demographics | | | |
| Age in years * | -0.15 | 4.71×10^{-109} | 20,746 |
| Sex * | -0.17 | 3.83×10^{-136} | 20,754 |
| BMI * | -0.07 | 1.46×10^{-21} | 20,098 |
| Household income * | 0.07 | < 0.0001 | 15,818 |
| Marital status | -0.02 | 0.013 | 19,177 |
| Years of education | 0.01 | 0.085 | 17,948 |
| Alcohol Use | | | |
| Days of use (Past 30 days) * | 0.52 | < 0.0001 | 20,185 |
| Days of heaviest use (Lifetime, 30-days period) * | 0.50 | < 0.0001 | 20,521 |

Notes Pairwise phenotypic correlations (r) Abbreviations * $P < 0.001$ ** Prior to GWAS analysis (Final sample size (N) = 20,328)

Supplementary Table 6Common genetic variants for strongest associations ($P < 10^{-7}$) with AUDIT

| SNP id | chr | cytoband | bp | alleles | EAF | <i>P</i> value | β | 95% CI | localization |
|-------------|-----|----------|-----------|---------|-------|----------------------|---------|-----------------|--------------|
| rs182344113 | 1 | 1q23.2 | 159964442 | C/G | 0.002 | 2.1×10^{-7} | 0.168 | [0.105,0.232] | intergenic |
| rs141973904 | 4 | 4q23 | 100262242 | C/T | 0.023 | 4.4×10^{-7} | -0.047 | [-0.065,-0.029] | intron |
| rs183472573 | 1 | 1p36.12 | 22469072 | A/C | 0.976 | 1.3×10^{-6} | -0.052 | [-0.074,-0.031] | intron |
| rs4791735 | 17 | 17p13.1 | 8599007 | C/T | 0.276 | 1.3×10^{-6} | -0.015 | [-0.021,-0.009] | intron |
| rs8059260 | 16 | 16p13.13 | 11060151 | A/G | 0.160 | 1.6×10^{-6} | -0.017 | [-0.024,-0.010] | intron |

Abbreviations SNP id marker name chr chromosome bp position alleles Other Allele/Effect Allele EAF effect allele frequency β beta effect size CI confidence interval

Supplementary Table 7 SNP-gene expression associations for the lead SNPs: **a** nearby expression QTL **b** nearby coding SNPs **c** nearby DNase Hotspots

| SNP id | region | bp | a Nearby expression QTL | | | | | b Coding SNPs | | | | c DNase Hotspots | | | | | |
|-------------|---------|-----------|--------------------------------|-----------|---------------|-------------------------------------|----------------------|----------------------|--------------------|-------|---------|-------------------------|--|-------|--|------------|-------------|
| | | | r^2 | SNP id | distance (bp) | gene | P value | r^2 | tissue | r^2 | gene | hgvs | consequence | r^2 | best tissue | best score | hotspot end |
| | | | 0.79 | rs7192695 | -53398 | CLEC16A | 1.3×10^{-7} | | Whole Blood, Brain | 0.81 | CLEC16A | c.344-7N>T | splice region variant & intron variant | 1 | gum tissue fibroblasts (AG09319) | 871 | 11060310 |
| rs8059260 | 16p13.3 | 11060151 | 0.81 | rs7195509 | -14722 | CLEC16A, LOC100130276, LOC100132071 | 6.5×10^{-7} | 0.19 | Temporal Cortex | NA | NA | NA | NA | | NA | NA | NA |
| | | | 0.81 | rs2302558 | 2860 | CLEC16A | 9.6×10^{-4} | | Whole Blood | NA | NA | NA | NA | | NA | NA | NA |
| rs182344113 | 1p23.2 | 159964442 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1 | fetal buttock/thigh fibroblast (AG04449) | 224 | 159964450 |
| rs183472573 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1 | cervical carcinoma (IFNa4h) | 1000 | 22469475 |

Notes Association results between **a** index SNP within 500 kb and $r^2 > 0.50$ with a SNP associated with expression of a nearby gene (expression QTL or eQTL) **c** index SNP within 500 kb from reported DNase hotspots. *Abbreviations* hgvs Human Genome Variation Society annotation bp position

Supplementary Table 8

See Online Material for a comparison between previous single nucleotide polymorphisms associated with alcohol-related traits, as reviewed by (Bühler *et al*, 2015), and AUDIT scores

Supplementary Table 9 Extension of Figure 2

Genetic correlations between AUDIT scores and other relevant traits using LD Regression Score (LDSC)

| Phenotype | Sample Size | Reference | Genetic correlation (r_g) | Standard error | <i>P</i> value | FDR < 0.05 |
|----------------------------|-------------|------------------------------------|-------------------------------|----------------|----------------|------------|
| Years of education | 293,723 | (Okbay <i>et al</i> , 2016b) | 0.27 | 0.06 | 3.14E-05 | Yes |
| BMI | 339,224 | (Locke <i>et al</i> , 2015) | -0.25 | 0.06 | 1.48E-04 | Yes |
| ADHD | 53,293 | (Demontis <i>et al</i> , 2017) | -0.29 | 0.01 | 1.43E-03 | Yes |
| Smoking lifetime | 74,035 | (Furberg <i>et al</i> , 2010) | 0.42 | 0.13 | 1.52E-03 | Yes |
| Obesity class 1 | 98,697 | (Berndt <i>et al</i> , 2013) | -0.23 | 0.08 | 2.06E-03 | Yes |
| Alcohol consumption | 70,460 | (Schumann <i>et al</i> , 2016) | 0.68 | 0.23 | 3.40E-03 | Yes |
| Years of education | 101,069 | (Rietveld <i>et al</i> , 2013) | 0.27 | 0.09 | 6.08E-03 | Yes |
| Childhood IQ | 12,441 | (Benyamin <i>et al</i> , 2014) | 0.42 | 0.16 | 6.26E-03 | Yes |
| College attainment | 95,427 | (Rietveld <i>et al</i> , 2013) | 0.26 | 0.10 | 8.11E-03 | Yes |
| Cannabis lifetime use | 32,330 | (Stringer <i>et al</i> , 2016) | 0.38 | 0.17 | 2.60E-02 | No |
| Daily cigarettes | 68,028 | (Furberg <i>et al</i> , 2010) | -0.31 | 0.16 | 6.19E-02 | No |
| Alcohol dependence | 7,280 | (Gelernter <i>et al</i> , 2014) | 0.68 | 0.37 | 6.41E-02 | No |
| Intracranial brain volume | 13,171 | (Hibar <i>et al</i> , 2015) | 0.33 | 0.20 | 8.91E-02 | No |
| Anorexia nervosa | 17,767 | (Boraska <i>et al</i> , 2014) | 0.21 | 0.14 | 1.29E-01 | No |
| Loneliness | 10,760 | (Gao <i>et al</i> , 2017) | -0.39 | 0.31 | 2.08E-01 | No |
| Hippocampus | 30,717 | (Hibar <i>et al</i> , 2015) | -0.21 | 0.18 | 2.38E-01 | No |
| Childhood obesity | 13,848 | (Bradfield <i>et al</i> , 2012) | -0.12 | 0.12 | 3.04E-01 | No |
| Schizophrenia | 77,096 | (Ripke <i>et al</i> , 2014) | 0.05 | 0.07 | 4.95E-01 | No |
| Neuroticism | 170,911 | (Okbay <i>et al</i> , 2016a) | 0.06 | 0.09 | 5.08E-01 | No |
| Former smoker | 70,675 | (Furberg <i>et al</i> , 2010) | -0.11 | 0.18 | 5.58E-01 | No |
| Bipolar disorder | 16,731 | (Sklar <i>et al</i> , 2011) | 0.05 | 0.10 | 6.11E-01 | No |
| Extraversion | 63,030 | (van den Berg <i>et al</i> , 2016) | 0.07 | 0.16 | 6.43E-01 | No |
| Neuroticism | 106,716 | (Smith <i>et al</i> , 2016) | -0.04 | 0.09 | 6.57E-01 | No |
| Height | 253,288 | (Wood <i>et al</i> , 2014) | 0.02 | 0.05 | 7.12E-01 | No |
| Major depressive disorder | 18,759 | (Ripke <i>et al</i> , 2013) | 0.06 | 0.17 | 7.22E-01 | No |
| Depressive symptoms | 161,460 | (Okbay <i>et al</i> , 2016a) | -0.04 | 0.11 | 7.38E-01 | No |
| Nucleus accumbens | 30,717 | (Hibar <i>et al</i> , 2015) | 0.08 | 0.24 | 7.43E-01 | No |
| Caudate | 30,717 | (Hibar <i>et al</i> , 2015) | -0.02 | 0.13 | 8.87E-01 | No |
| Neo-openness to experience | 17,375 | (de Moor <i>et al</i> , 2012) | 0.02 | 0.18 | 8.93E-01 | No |
| Smoking age onset | 47,961 | (Furberg <i>et al</i> , 2010) | 0.01 | 0.22 | 9.40E-01 | No |
| Subjective wellbeing | 298,420 | (Okbay <i>et al</i> , 2016a) | -0.01 | 0.11 | 9.64E-01 | No |

Supplementary Table 9 Continued

References

- Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK, Brion M-JA, *et al* (2014). Childhood intelligence is heritable, highly polygenic and associated with FBNP1L. *Mol Psychiatry* **19**: 253–258.
- Berg SM van den, Moor MHM de, Verweij KJH, Krueger RF, Luciano M, Arias Vasquez A, *et al* (2016). Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav Genet* **46**: 170–182.
- Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, *et al* (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* **45**: 501–512.
- Boraska V, Franklin CS, Floyd J a. B, Thornton LM, Huckins LM, Southam L, *et al* (2014). A genome-wide association study of anorexia nervosa. *Mol Psychiatry* **19**: 1085–1094.
- Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoecur C, Warrington NM, *et al* (2012). A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* **44**: 526–531.
- Bühler K-M, Giné E, Echeverry-Alzate V, Calleja-Conde J, Fonseca FR de, López-Moreno JA (2015). Common single nucleotide variants underlying drug addiction: more than a decade of research. *Addict Biol* **20**: 845–871.
- Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, *et al* (2017). Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. doi:10.1101/145581.
- Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, *et al* (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**: 441–447.
- Gao J, Davis LK, Hart AB, Sanchez-Roige S, Han L, Cacioppo JT, *et al* (2017). Genome-Wide Association Study of Loneliness Demonstrates a Role for Common Variation. *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol* **42**: 811–821.
- Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH, *et al* (2014). Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry* **19**: 41–49.
- Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivieres S, Jahanshad N, *et al* (2015). Common genetic variants influence human subcortical brain structures. *Nature* **520**: 224–229.
- Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, *et al* (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**: 197–206.
- Moor MHM de, Costa PT, Terracciano A, Krueger RF, Geus EJC de, Toshiko T, *et al* (2012). Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* **17**: 337–349.
- Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, *et al* (2016a). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**: 624–633.
- Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, *et al* (2016b). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**: 539–542.
- Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, *et al* (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science* **340**: 1467–1471.
- Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, *et al* (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**: 421–427.
- Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, *et al* (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**: 497–511.
- Schumann G, Liu C, O'Reilly P, Gao H, Song P, Xu B, *et al* (2016). KLB is associated with alcohol drinking, and its gene product β -Klotho is necessary for FGF21 regulation of alcohol preference. *Proc Natl Acad Sci U S A* **113**: 14372–14377.
- Sklar P (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**: 977–983.
- Smith DJ, Escott-Price V, Davies G, Bailey MES, Colodro-Conde L, Ward J, *et al* (2016). Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Mol Psychiatry* **21**: 749–757.
- Stringer S, Minicá CC, Verweij KJH, Mbarek H, Bernard M, Derringer J, *et al* (2016). Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. *Transl Psychiatry* **6**: e769.

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, *et al* (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**: 1173–1186.

Extended Methods

A maximal set of unrelated individuals was chosen for the analysis using a segmental identity-by-descent (IBD) estimation algorithm (Henn *et al*, 2012) to ensure that only unrelated individuals were included in the sample. Individuals were defined as related if they shared more than 700 cM IBD, including regions where the two individuals shared either one or both genomic segments identical-by-descent. This level of relatedness (~20% of the genome) and corresponds approximately to the minimal expected sharing between first cousins in an outbred population.

We imputed participant genotype data against the September 2013 release of 1000 Genomes phase 1 version 3 reference haplotypes. We phased and imputed data for each genotyping platform separately. We phased using an internally developed phasing tool, Finch, which implements the Beagle haplotype graph-based phasing algorithm (Browning and Browning, 2007), modified to separate the haplotype graph construction and phasing steps. In preparation for imputation, we split phased chromosomes into segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs. We excluded SNPs with Hardy-Weinberg equilibrium $P < 10^{-20}$, call rate $< 95\%$, or with large allele frequency discrepancies compared to European 1000 Genomes reference data. Frequency discrepancies were identified by computing a 2x2 table of allele counts for European 1000 Genomes samples and 2000 randomly sampled 23andMe research participants with European ancestry, and identifying SNPs with a $\chi^2 P < 10^{-15}$. We imputed each phased segment against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac2 (Fuchsberger *et al*, 2015), using 5 rounds and 200 states for parameter estimation. After quality control, we analyzed 11,508,740 SNPs.

For the X chromosome, we built separate haplotype graphs for the non-pseudoautosomal region and each pseudoautosomal region, and these regions were phased separately. We then imputed males and females together using Minimac2, as with the autosomes, treating males as homozygous pseudo-diploids for the non-pseudoautosomal region.

For tests using imputed data, we use the imputed dosages rather than best-guess genotypes. We imputed HLA allele dosages from SNP genotype data using HIBAG (Zheng *et al*, 2014). We imputed alleles for HLA-A, B, C, DPB1, DQA1, DQB1, and DRB1 loci at four-digit resolution. To test associations between HLA allele dosages and phenotypes, we performed linear regression using the same set of covariates used in the SNP based GWAS. We performed separate association tests for each imputed allele.

References

- Browning SR, Browning BL (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am J Hum Genet* **81**: 1084–1097.
- Fuchsberger C, Abecasis GR, Hinds DA (2015). minimac2: faster genotype imputation. *Bioinformatics* **31**: 782–784.

Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, *et al* (2012). Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. *PLoS ONE* **7**: e34267.

Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, *et al* (2014). HIBAG—HLA genotype imputation with attribute bagging. *Pharmacogenomics J* **14**: 192–200.