Title

Genome-wide association studies of the self-rating of effects of ethanol (SRE)

Journal

Addiction Biology, 25(2)

ISSN

1355-6215

Authors

Lai, Dongbing
Wetherill, Leah
Kapoor, Manav
et al.

Publication Date

2020-03-01

DOI

10.1111/adb.12800

Peer reviewed

Published in final edited form as:

Addict Biol. 2020 March ; 25(2): e12800. doi:10.1111/adb.12800.

Genome wide association studies of the Self-Rating of Effects of Ethanol (SRE)

Dongbing Lai, Ph.D.[1], **Leah Wetherill, Ph.D.**[1], **Manav Kapoor, Ph.D.**[2], **Emma C. Johnson, Ph.D**[3], **Melanie Schwandt, Ph.D.**[4], **Vijay A. Ramchandani, Ph.D.**[5], **David Goldman, M.D.**[4], **Geoff Joslyn, Ph.D.**[6], **Xi Rao, Ph.D.**[1], **Yunlong Liu, Ph.D.**[1], **Sean Farris, Ph.D.**[7], **R. Dayne Mayfield, Ph.D.**[7], **Danielle Dick, Ph.D.**[8], **Victor Hesselbrock, Ph.D.**[9], **John Kramer, Ph.D.**[10], **Vivia V. McCutcheon, Ph.D.**[3], **John Nurnberger, M.D., Ph.D**[1,11], **Jay Tischfield, Ph.D.**[12], **Alison Goate, D. Phil.**[2], **Howard J. Edenberg, Ph.D.**[1,13], **Bernice Porjesz, Ph.D.**[14], **Arpana Agrawal, Ph.D.**[3], **Tatiana Foroud, Ph.D.**[1], **Marc Schuckit, M.D.**[15]

[1.]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN

[2.]Department of Neuroscience, Icahn School of Medicine at Mt. Sinai, New York, NY

[3.]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO

[4.]Office of the Clinical Director, National Institute on Alcohol Abuse & Alcoholism, Bethesda, MD

[5.]Section on Human Psychopharmacology, Division of Intramural Clinical and Biological Research, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD

[6.]Ernest Gallo Clinic and Research Center, Emeryville, CA

[7.]Waggoner Center for Alcohol and Addiction Research, The University of Texas at Austin, Austin, TX

[8.]Department of Psychology, Virginia Commonwealth University, Richmond, VA

[9.]Department of Psychiatry, University of Connecticut, Farmington, CT

[10.]Department of Psychiatry, Roy Carver College of Medicine, University of Iowa, Iowa City, IA

[11.]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN

[12.]Department of Genetics and the Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ

[13.]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN

Corresponding Author: Dongbing Lai, Ph.D., 410 W. 10th Street, HS 4000, HITS, Indianapolis, IN 46202-3002, dlai@iu.edu, phone: 317-278-9544, fax: 317-278-1100.

Author contributions

DL, AA, TF, and MS were responsible for the study concept and design. MK, MS, VAR, DG, GJ, XR, YL, SF, and RDM contributed to the acquisition of data. DL, LW, MK, EJ performed the analysis. DL drafted the manuscript. DD, VH, JK, VVM, JN, JT, AG, HJE, BP, AA, TF, and MS provided critical revision of the manuscript for important intellectual content. All authors critically reviewed content and approved final

[14]Henri Begleiter Neurodynamics Lab, Department of Psychiatry, State University of New York, Downstate Medical Center, Brooklyn, NY

[15]Department of Psychiatry, University of California, San Diego Medical School, San Diego, CA

Abstract

The level of response (LR) to alcohol as measured with the Self-Report of the Effects of alcohol retrospective questionnaire (SRE) evaluates the number of standard drinks usually required for up to four effects. The need for a higher number of drinks for effects is genetically influenced and predicts higher risks for heavy drinking and alcohol problems. We conducted GWAS in the African-American (COGA-AA, N=1,527 from 309 families) and European-American (COGA-EA, N=4,723 from 956 families) subsamples of the Collaborative Studies on the Genetics of Alcoholism (COGA) for two SRE scores: SRE-T (average of first 5 times of drinking, the period of heaviest drinking, and the most recent three months of consumption) and SRE-5 (the first 5 time of drinking). We then meta-analyzed the two COGA subsamples (COGA-AA+EA). Both SRE-T and SRE-5 were modestly heritable (h^2: 21%-31%) and genetically correlated with alcohol dependence (AD) and DSM-IV AD criterion count (r_g: 0.35-0.76). Genomewide significant associations were observed (SRE-T: chromosomes 6, rs140154945, COGA-EA p=3.30E-08 and 11, rs10647170, COGA-AA+EA p=3.53E-09; SRE-5: chromosome13, rs4770359, COGA-AA p=2.92E-08). Chromosome 11 was replicated in an EA dataset from the National Institute on Alcohol Abuse and Alcoholism intramural program. *In silico* functional analyses and RNA expression analyses suggest that the chromosome 6 locus is an eQTL for *KIF25*. Polygenic risk scores derived using the COGA SRE-T and SRE-5 GWAS predicted 0.47-2.48% of variances in AD and DSM-IV AD criterion count in independent datasets. This study highlights the genetic contribution of alcohol response phenotypes to the etiology of alcohol use disorders.

Keywords

Genomewide association study (GWAS); genetic correlation; heritability; RNA expression; polygenic risk score; Self-Rating of the Effects of Ethanol (SRE)

INTRODUCTION

A low level of response (LR) to alcohol, a measure of the intensity of reaction to alcohol as evaluated by either alcohol challenges or a self-report questionnaire regarding the number of standard drinks needed for a range of effects, is an important predictor of the future development of alcohol dependence (AD) (Schuckit et al., 1997a). AD is associated with considerable morbidity and affects about 12.5% of the U.S. population during their lifetime (Hasin and Grant, 2015). Typically, individuals who experience fewer intoxicating or sedating effects of alcohol are more likely to escalate their drinking behaviors (Schuckit et al., 2012). Thus, low LR has been related to future problems with alcohol primarily via heavy consumption (Schuckit et al., 2012), even after accounting for histories of heavy drinking (Schuckit et al., 2008b) and family history (Schuckit and Smith, 2000). Multiple putative pathways including overly positive expectations of the effects of alcohol, heavier peer drinking, and using alcohol to cope with stress (Schuckit et al., 2012) have also been

found to contribute to the heightened risk for AD in those with low LR. Interestingly, recent research suggests college-based prevention programs that target LR are effective in individuals with low, but not high, LR (Schuckit et al., 2016, Savage et al., 2015). Thus, even though low LR is a marker for predisposition to AD, largely via heavy consumption, individual differences in LR are also associated with unique sources of variance.

A person's LR can be measured by administering an alcohol challenge, either through oral ingestion or intravenously, and recording the individual's level of reaction over several hours. Such alcohol-administration paradigms are time consuming and costly, making it difficult to evaluate large numbers of subjects (Schuckit et al., 1997a). LR can also be measured by the retrospective Self-Report of the Effects of alcohol (SRE) questionnaire, in which participants estimate the number of standard drinks usually required to experience up to four different effects of alcohol (feel an effect; feel dizzy or slurring speech; stumble or walk in an uncoordinated manner; pass out or fall sleep) at three time periods (first 5 times of drinking, the period of heaviest drinking, and the most recent three months of consumption). SRE scores and LR measured during an alcohol challenge are correlated (r = 0.3-0.6), suggesting that the easier to collect SRE can be used in large-scale studies to measure LR (Schuckit et al., 1997a, Schuckit et al., 2009b, Schuckit et al., 1997b). SRE scores also have an one year test-retest correlation between 0.72-0.82 and five-year retest of 0.66, indicating high levels of reliability (Schuckit et al., 2009, Schuckit et al., 1997b). Their predictive validities regarding future heavy drinking and alcohol problems across different populations and sexes in studies conducted in the U.S., U.K., and Switzerland are also consistent (Daeppen et al., 2000, Kalu et al., 2012, Pedersen and McCarthy, 2013, Schuckit et al., 2007, Schuckit et al., 2008a, Schuckit et al., 2008b, Wall et al., 1999). These studies cumulatively indicate that an individual's LR, as measured by the SRE, could serve as an additional risk marker for the development of problem drinking, and can be a useful tool in efforts to personalize prevention and treatment (Schuckit et al., 2016).

Both AD and LR (self-report or alcohol challenge results) are heritable (Viken et al., 2003, Kalu et al., 2012, Heath et al., 1999). While there have been numerous efforts to identify genetic variants for AD (Walters et al., 2018), there have been only a few studies examining the genetic contributions to SRE. Two linkage studies, one using 745 European ancestry individuals (included in this study) (Schuckit et al., 2001) and another in American Indians (Ehlers et al., 2010), detected modest evidence of linkage. Recently, a meta-analysis of genomewide association studies (GWAS) of SRE score of the first 5 times of drinking (SRE-5) did not identify genomewide significant loci (Edwards et al., 2018). Genetic analyses of the average SRE scores across the three time periods (SRE-T) have been limited to candidate gene studies (Schuckit, 2018). A recent analysis showed that polygenic risk scores derived from a GWAS of alcohol consumption in a large U.K. population of older adults explained minimal variance in SRE-T (Johnson and Agrawal, 2018) suggesting that even at a polygenic level, SRE scores might be associated with unique genetic variation. Given the importance of SRE as a predictor of problem drinking, and its role in intervention research, the identification of such variants is necessary.

We conducted genomewide association studies of SRE scores in the Collaborative Study on the Genetics of Alcoholism (COGA). First, using variance decomposition approaches for

family data, we estimated the heritability of the SRE scores in the African-American and European-American COGA subsamples. The covariance between SRE-T, SRE-5, AD (defined by DSM-IV) and DSM-IV AD criterion count (used as a measurement of severity of AD) was decomposed into its genetic and environmental sources. We then performed GWAS to identify genetic variants contributing to the variation in SRE-T and SRE-5 scores. Genomewide significant loci were examined in two independent datasets and in RNA expression analysis using brain tissue from individuals with AD and controls. Finally, we derived polygenic risk scores (PRS) from these discovery GWAS and evaluated whether the polygenic effects of SRE-related variants predicted AD and DSM-IV AD criterion count in four independent AA and EA datasets.

METHODS

Sample and Assessments

AD probands from alcohol use disorder inpatient and outpatient treatment facilities and their family members were recruited from the seven participating COGA sites. Comparison individuals and their families were ascertained from the community in the same seven areas (Bucholz et al., 1994). Institutional review boards at all sites approved the study and all participants provided informed consent.

The Self-Report of Effects of alcohol questionnaire (Schuckit et al., 1997b) is a self-report instrument used to measure the number of standard drinks required to produce up to four effects of alcohol at three different time periods: a) the first five times the subject used alcohol; b) the period of heaviest drinking; and c) the most recent three months of consumption. For each time period, the subject is asked: a) “how many (standard) drinks did it take for you to begin to feel an effect?”; b) “how many drinks did it take for you to feel a bit dizzy or begin to slur your speech?”; c) “how many drinks did it take you to begin to stumble or walk in an uncoordinated manner?”; d) “how many drinks did it take you to pass out or fall asleep when you did not want to?” Extreme observations were winsorized at the mean plus 2 standard deviations and only individuals who drank at least 2 drinks on one occasion were included in the analysis. Based on their distributions, the square root of SRE-T and the natural logarithm of SRE-5 were used in analyses. Distributions of original and transformed SRE scores are in supplemental material.

The Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) interview and an adolescent version of the SSAGA (Bucholz et al., 1994) were administered to adults and individuals age 17 or under, respectively. An AD case was defined as an individual age 15 or older who met the DSM-IV criteria for AD, to exclude early onset AD, which may have different etiology (Dick et al., 2014). Control subjects were at least 21 years old and had consumed at least one full drink of alcohol in their lifetime, but who did not endorse more than one DSM-IV AD criterion. We set the minimum age as 21 to exclude individuals who had not have passed through the period of major risk for the onset of AD (Grant et al., 2004, Hingson et al., 2006). All others were not classified. Although these unclassified individuals were not used in any analysis related to DSM-IV AD, they were included in all other analysis (e.g. GWAS of SRE scores).

Genotyping, Quality Review and Imputation

COGA samples were generated on four arrays: Illumina Human1M array (Illumina, San Diego, CA, USA) genotyped at the Center for Inherited Disease Research (CIDR), Johns Hopkins University; Illumina Human OmniExpress 12V1 array (Illumina, San Diego, CA, USA) genotyped at the Genome Technology Access Center, Washington University School of Medicine in St. Louis; Illumina 2.5M array (Illumina, San Diego, CA, USA) at CIDR; and Smokescreen genotyping array (Biorealm LLC, Walnut, CA, USA) genotyped at Rutgers University. Two to 127 samples were genotyped on at least two different arrays and pairwise concordance rates within subjects typed on more than one array exceeded 99.18% for those common SNPs, confirming limited batch or array effects. Sex and heterogeneity checks were performed and any discrepancies were removed.

To confirm pedigree structure, a set of 47,000 variants genotyped on all four arrays were used. These variants met the following criteria: 1. not in linkage disequilibrium (LD, defined as r^2 <0.5); 2) minor allele frequency (MAF) >10%; 3) Hardy-Weinberg Equilibrium (HWE) p-value >0.001; and 4) missing rate <2%. Pairwise identity by descent was computed using PLINK (Purcell et al., 2007) to confirm family relationships. When necessary, family structures were altered accordingly. The same set of 47,000 variants was also used to estimate principal components (PCs) of population stratification using Eigenstrat (Price et al., 2006). Based on the first two PCs, each individual was assigned an ancestry (i.e. African-American, European-American, or Other). Family ancestry was assigned based on the individual ancestry of the greatest proportion of family members.

Prior to imputation, Mendelian inconsistences were detected using Pedcheck (O'Connell and Weeks, 1998), and genotype inconsistences were set to missing. Variants with missing rates >5%, MAF <3% and HWE p values <0.0001 were excluded. In addition, variants with A/T or C/G alleles were also removed to avoid ambiguities in strand designation. SHAPEIT2 (Delaneau et al., 2013), which utilizes pedigree information to increase accuracy, was used to phase haplotypes of each sample. All samples were imputed to 1000 Genomes the cosmopolitan reference panel (Phase 3, version 5, http://www.internationalgenome.org/) using Minimac3 (Das et al., 2016). Due to the differences in the variants on each array, samples were imputed separately by array. After imputation, variants with R^2 <0.30 were excluded. To allow for further Mendelian error checking, genotype probabilities were converted to genotypes if they were ≥0.90 (similar to procedures for QC and analysis of family data in Walters et al., 2018 (Walters et al., 2018); see also Kranzler et al., 2019 that used converted genotypes (Kranzler et al., 2019)). All genotype and imputed variants with missing rates <20%, MAF ≥1% and HWE p values >0.000001 were used in analyses. 15,464,468 and 7,823,558 variants passed QC and were included in the AA and EA subsamples analyses, respectively. All variants were mapped to NCBI GRCh37.

Heritability, genetic correlations, and GWAS

All analyses were performed separately in COGA AA (COGA-AA) and EA (COGA-EA) subsamples, then meta-analyzed (COGA-AA+EA). Due to the complex pedigree structure in COGA, methods using genomewide genotype data such as Genomewide Complex Traits Analysis (Yang et al., 2011) could not be used to estimate heritability or genetic correlations.

Instead, using an approach similar to those applied in classical twin studies, the Sequential Oligogenic Linkage Analysis Routines (SOLAR8.3.1) (Almasy and Blangero, 1998) package, which was designed to partition variance in large pedigrees, was used to estimate heritability of SRE scores (i.e., proportion of total phenotypic variance attributable to latent genetic influences), as well as the genetic correlations between the SRE measures and AD and DSM-IV AD criterion count (Almasy and Blangero, 1998). Sex and birth cohorts (birth year: 1890-1929; 1930-1949; 1950-1969; >=1970) were important covariates as shown in previous studies (Grucza et al., 2008); therefore both of them and the first four PCs were used as covariates in SOLAR analyses.

GWAS was conducted using linear mixed models with random effects adjusting for relatedness within families using the R package GWAF (Chen and Yang, 2010). Sex, birth cohorts, array indicators, and the first four PCs were included as covariates. METAL (Willer et al., 2010) was used to perform meta-analysis with effect size of each variant weighted by the inverse of the estimated standard error from each subsample. Genomic control was applied to the results of each subsample as well as to the meta-analysis results. As SRE-T and SRE-5 were highly correlated (correlation coefficients were 0.75 and 0.78 in COGA-AA and COGA-EA, respectively), we used matSpD (Nyholt, 2004) to spectrally decompose the correlation matrix of SRE scores, resulting 1.44 effective test. Thus, the genomewide significance was set at $p < 3.5 \times 10^{-8}$. In regions of genomewide significance, conditional analyses were performed by including the most significant variant in the region as a covariate.

Gene prioritization

FUMA (Watanabe et al., 2017), a web-based tool that incorporates information from multiple public databases, was used to prioritize potentially causal genes. Three gene prioritization strategies are implemented in FUMA: positional mapping, expression Quantitative Trait Locus (eQTL) mapping, and chromatin interaction mapping. Positional mapping annotates variants based on their physical positions and functional consequences. For any region in which at least one variant met genomewide significance, all variants in that region with p values ≤1.0E-04 were annotated. If an annotated variant was in the promoter or untranslated regions (UTRs) of a gene, or in exonic or splicing regions of a gene and predicted to be deleterious, that gene was prioritized as a potential causal gene. Ensembl (build 85; http://www.ensembl.org/) was used to map variants to genes. For eQTL mapping, genes within 1 Mb of the most significant variant were tested, and those with significant associations (defined as FDR <0.05) were considered as potential causal genes. eQTL data from brain tissues in BRAINEAC (http://www.braineac.org/), CommonMind Consortium (https://www.synapse.org), and GTEx v7 (https://www.gtexportal.org/) were used. Genes prioritized by any of these three strategies were considered as potential causal genes and further examined by subsequent RNA expression analysis.

RNA expression analysis

To test the associations between the most significant variants and gene expressions, human autopsy brain samples were obtained from the New South Wales Brain Tissue Resource Centre (NSWBTRC) at the University of Sydney (http://sydney.edu.au/medicine/pathology/

btrc/). Fresh frozen samples of the superior frontal cortex (Brodmann area 8) were used. This region was selected because a prior enrichment analysis suggests aggregation of AD-associated GWAS signals in this brain region (Kapoor et al., 2018). In addition, there was an adequate sample size available for this region. The samples were sequenced on Illumina Hi-Seq (Illumina, CA, USA) at the New York Sequencing center (N=83) and at the Waggoner Center for Alcohol and Addiction Research (N=60). Total RNA was extracted from frozen tissue and samples with contamination or degraded quality RNA (RNA integrity numbers (RIN) <5.0) were excluded. 138 samples passed quality review and were used in analysis (65 AD, 73 age and sex matched controls). The KAPA Stranded RNA-Seq Kit with RiboErase was used for library preparation. The library QC included a measurement of the average size of library fragments using the Fragment Analyzer (Advanced Analytical Technologies, IA, USA), and estimation of the total concentration of DNA by PicoGreen (Thermo Fisher Scientific, MA, USA). Raw reads were aligned to human genome 19 (hg19) using STAR aligner (Dobin et al., 2013). Quality control was assessed using RSeQC (Wang et al., 2012) and Picard (% GC, % duplicates, gene body coverage, unsupervised clustering, and library complexity, http://broadinstitute.github.io/picard/). The Picard "MarkDuplicates" option was used to flag and remove the duplicate reads. Gene quantification was performed with featureCounts using Gencode annotations. We filtered out all genes with lower expression in a substantial fraction of the cohort, with 18,463 genes with at least 1 CPM (counts per million) in at least 50% of the individuals; note that only these genes were carried forward in all subsequent analyses.

DNA was extracted from the same brain tissue samples used in the RNA analysis. Genotyping was performed on the Axiom™ Biobank Plus Genotyping Array (Catalog number: 000854). Initial QC was performed using PLINK (Purcell et al., 2007) to remove variants with: 1) no alternate alleles; 2) genotyping call rate ≤0.98; 3) Hardy-Weinberg Equilibrium p value <5E-5; and 4) samples with genotyping call rate <0.90. Phasing was performed on each chromosome using SHAPEIT2 (Delaneau et al., 2013) then imputation was done using IMPUTE2 (Marchini et al., 2007) with the 1,000 Genomes Phase 1 integrated reference panel excluding singleton variants. Analysis was performed on the imputed variants with imputation score (INFO) ≥0.8 and estimated MAF ≥0.05.

eQTLs were computed using a linear model on the imputed genotype dosages using Matrix EQTL69 (Shabalin, 2012). If the most significant variant was filtered out in this sample due to QC, the next available variant with the lowest p value and in LD (r^2>0.9) with the most significant variant was tested instead. The gene expression data were adjusted for sex, age, post-mortem interval, RIN, batch, and decedent phenotype (AD or control)

Replication samples

The first replication dataset was from the San Diego Sibling Pair investigation and genotyped on the Illumina HumanCNV370-duo array (Illumina, San Diego, CA, USA). This study was approved by the Human Subjects Protection Committee of the University of California, San Diego. The same quality review and imputation used in COGA were performed. This dataset included both unrelated and related individuals. Empirical kinships were estimated from genomewide genotype data using the vcf2kinship tool as implemented

in RVTESTS and adjusted in mixed models using RVTESTS (Zhan et al., 2016). The majority of the dataset had European ancestry; therefore, only EA samples was used (SD-EA, N=206, 68 AD cases and 138 controls). Only sex was available and included as a covariate.

Two datasets from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) Intramural program were also used as replication. Participants were recruited under two NIH Institutional Review Board-approved screening and assessment protocols and were comprehensively assessed at the National Institutes of Health Clinical Center (Bethesda, Maryland, USA) between 2005 and 2015. All participants provided written informed consent. Genotyping was conducted at the NIAAA Laboratory of Neurogenetics (Rockville, MD, USA) using the Illumina Human OmniExpress and Human OmniExpressExome arrays. The same quality review and imputation as in COGA were performed. 169 (130 AD cases, 27 controls, 12 not classified) and 100 (36 AD cases, 27 controls, 12 not classified) unrelated AA (NIAAA-AA) and EA (NIAAA-EA) individuals were used in analyses, respectively. Sex and age were included as covariates. Only AA samples were adjusted for the first three PCs because unlike COGA-EA, EA samples in replication and PRS analysis didn't include non EA individuals.

Polygenic risk scores analyses

Summary statistics from both SRE scores in COGA-AA and COGA-EA were used to derive PRS using PRSice-2 (Euesden et al., 2015). Four test datasets from the database of Genotypes and Phenotypes (dbGaP) were included: 1) AA subsample (SAGE-AA) in the Study of Addiction: Genetics and Environment (SAGE, phs000092.v1.p1); 2) AA subsample (Yale-Penn-AA) in Alcohol Dependence GWAS in European and African Americans (phs000425.v1.p1); 3) EA subsample in SAGE (SAGE-EA); and 4) EA subsample (OZALC-EA) in The Australian Twin-family Study of Alcohol Use Disorder (OZALC, phs000181.v1.p1). Overlapping COGA individuals were excluded from these independent datasets. PRSice-2 requires the analysis of unrelated samples; therefore, in datasets with related individuals, one individual was randomly selected from each family. Summary of each test dataset is shown in supplemental material. Any variant that was within 500 kb from the index variant and had r^2 >=0.25 with the index variant was clumped using PRSice-2. PRS was calculated based on effect size (Euesden et al., 2015). To reduce the burden of multiple testing, only variants that had p values <0.05 were included. For SAGE, sex and the same birth cohort as defined in COGA were included as covariates, while for Yale-Penn-AA and OZALC-EA, sex and age were used. The first 3 PCs were also adjusted in AA samples.

RESULTS

Sample summary, heritability, and genetic correlations

Table 1 summarizes the COGA-AA and COGA-EA subsamples. Compared to those in COGA-AA, individuals in COGA-EA had higher mean DSM-IV AD criterion count in both AD cases and controls, SRE-T, and SRE-5 scores. Latent genetic influences (familial h^2) contributed to 21-31% of the variance in SRE scores, indicating moderate heritability (Table

2). Overall, genetic correlations (r_g) with AD and DSM-IV AD criterion count were higher for SRE-T (r_g=0.49 -0.76) than for SRE-5 (r_g=0.35-0.48) (Table 2).

GWAS results

Two loci for SRE-T (chromosome 6 for COGA-EA; chromosome 11 for COGA-AA+EA) and one locus for SRE-5 (chromosome 13 for COGA-AA) reached genomewide significance (Table 3, Figure S2). Tables S1 **and** S2 provide a broader list of association results (P<1E-04) for SRE-T and SRE-5, respectively.

SRE-T: The most significant association results were with variants on chromosome 11 in COGA-AA+EA (most significant: rs10647170, p=3.53E-09), with both subsamples providing supporting evidence for the association. Evaluation of rs10647170 in independent datasets provided nominal evidence of replication in the NIAAA-AA (p=0.05), but not in the SD-EA or NIAAA-EA (Table 3) samples. Meta-analysis of COGA-EA+AA and the 3 replication datasets resulted in a lower p-value of 6.45E-10.

In addition, a genomewide significant result was noted with a variant on chromosome 6 (rs140154945, p=3.30E-08) in COGA-EA (Table 3) that not supported in COGA-AA. There was also no evidence of replication for rs140154945 in the EA replication datasets.

Regional association plots demonstrated strong LD support in both the chromosomes 6 and 11 regions (Figures S3A, S3B, **and** S3C). Conditional analyses using the top variant in each region as a covariate confirmed that there were no additional association signals in these chromosomal regions associated with SRE-T.

SRE-5: Genomewide significant associations were identified in COGA-AA on chromosome 13 (rs4770359, p=2.92E-08), although it was not associated in COGA-EA (Table 3). The variant was not replicated in the NIAAA-AA. There was modest LD support in this region possibly due to unusual LD pattern as shown in 1000 Genomes data using LDlink (https://ldlink.nci.nih.gov/?tab=home) (Figures S3D, S4). Unlike SRE-T, conditional analyses found that some variants demonstrated modest association (p<0.001) after adjustment of the top variant, suggesting that additional independent associations might exist in each region, consistent with the LD patterns showing in Figure S3D. There were no genomewide significant association findings in COGA-AA+EA and COGA-EA GWAS.

The effect sizes of the most significant variants in each locus in discovery and replication samples are shown in the forest plots in Figure S5.

Gene prioritization and RNA expression analysis

For chromosome 11, one variant (rs2270678, COGA-AA+EA p=1.81E-08) was located in the 5' UTR of *PRMT3* (protein arginine methyltransferase 3). Another variant (rs3781679, COGA-AA+EA p = 3.58E-05, Table S1) was located in the 3' UTR of *HTATIP2* (HIV-1 Tat interactive protein 2). Therefore, both genes were prioritized in FUMA. Chromatin interaction mapping further prioritized *DBX1* (developing brain homebox 1) and *SLC6A5* (solute carrier family 6 member 5) in this region. On chromosome 6, three variants were located in the 3' UTR of *AFDN* (adherens junction formation factor) (Table S1) and this

gene was therefore prioritized by FUMA. eQTL analysis further prioritized the gene *KIF25* (kinesin family member 25) in this region (multiple variants had FDR<0.01). No genes were prioritized in the chromosome 13 region.

For the six genes prioritized by FUMA (Chromosome 11: *PRMT3, HTATIP2, DBX1, SLC6A5;* Chromosome 6: *AFDN, KIF25*), RNA expression in brain tissue obtained from NSWBRTC was examined. After adjusting for six tests, the Bonferroni correction threshold was set at p<0.0083. Rs140154945 did not pass QC in this smaller postmortem tissue sample, therefore the next most significant variant in high LD (rs3778662, COGA-EA p=7.24E-08, r^2=0.92 with rs140154945) was used. After adjustment for covariates, including AD status of decedent, rs3778662 was identified as an eQTL of *KIF25* (p=6.62E-05).

PRS analyses

PRS analysis results are summarized in Table 4. For both SRE-T and SRE-5, there are four discovery-test dataset pairs (i.e., Bonferroni p<0.0125). Overall, PRS from SRE-T outperformed those from SRE-5 in prediction of AD phenotypes; PRS of both SRE-T and SRE-5 predicted greater variance in DSM-IV AD criterion count than AD; prediction was also higher in EA than in AA samples. The strongest finding was for polygenic scores based on the COGA-EA SRE-T GWAS, which predicted 2.48% of the variance in AD in the SAGE-EA.

DISCUSSION

This is the first GWAS to identify common variants associated with SRE-T and the first GWAS of SRE scores in AA samples. This is also the first study to report genomewide significant findings for any SRE score. These findings are not under linkage regions previously published (Ehlers et al., 2010, Schuckit et al., 2001) using SRE-5. Of the genomewide significant loci, the locus on chromosome 11 was nominally replicated in one sample, although the replication sample was small. Of the genes prioritized for SRE-T in the EA sample, *KIF25* was supported by data from both *in silico* functional analyses and RNA expression analysis. In addition, PRS derived from these discovery GWAS of SRE predicted variance in AD-related phenotypes in independent datasets. Broadly, these GWAS analyses suggest that larger samples of SRE phenotypes, particularly SRE-T, might be highly informative in the identification of loci related to LR, which influences risk for development of AD and is emerging as a notable target for prevention.

SRE-T and SRE-5 represent self-reported assessments of LR. SRE-T and SRE-5 were moderately heritable, but our estimates were lower than some prior reports. For instance, a smaller (n=101) sibling study of non-alcohol-dependent individuals estimated the heritability of SRE-T at 67-71% (Kalu et al., 2012), although the corresponding estimate for an “early drinking period” was lower and statistically non-significant. Similar high heritability estimates have been noted for other assessments of subjective ratings (e.g., 40-60%) (Heath and Martin, 1991, Viken et al., 2003). It is possible that our estimates are less precise than those from larger twin studies, as we relied on pedigree data. We did not utilize genomic data to estimate the SNP-heritability of SRE scores as our large pedigrees,

when reduced to a smaller cohort of unrelated individuals, is not well-powered to make such estimates. Nonetheless, our estimates suggest that future larger efforts should be successful in estimating the extent to which common variants explicate the heritability of SRE (Ray et al., 2010).

Our discovery GWAS in both COGA-AA and COGA-EA identified genomewide significant loci. The association signal on chromosome 11 was supported by both COGA-AA and COGA-EA. Genes prioritized in this region have been supported by other GWAS. In a recent GWAS of alcohol dependence in a Thai population, the region between *DBX1* and *PRMT3* showed suggestive association with AD (Gelernter et al., 2018). Another study reported an association between *PRMT3* and smoking (Park et al., 2015). Variants upstream of *PRMT3* demonstrated suggestive association with marijuana dependence criterion count in an African-American cohort (Sherva et al., 2016). Although *HTAPIP2* in this region was also prioritized by FUMA, it was not reported to be related to any neuropsychiatric diseases. Two additional genes in this region, *DBX1* and *SLC6A5*, were prioritized by chromatin interaction mapping. *DBX1* has been linked to educational achievement (Rietveld et al., 2014), which is negatively correlated with genetic liability to alcohol dependence (Walters et al., 2018), while *SLC6A5* was reported to be related to schizophrenia (Deng et al., 2008), which is positively genetically correlated with AD (Walters et al., 2018). Thus, there is increasing evidence for the involvement of variants in this region to liability to problem drinking.

Analyses of RNA expression data in brain tissue identified *KIF25* on chromosome 6 to be a gene of interest for SRE-T in the EA sample. The minor alleles of the associated variants in this region increased the SRE-T score, risk of AD, and expression of *KIF25* in both curated brain tissue data (e.g., from GTEx) and RNA expression analysis that accounted for AD status. Due to the small sample size, there was insufficient statistical power to analyze the effects of variants stratified by cases and controls. This gene encodes a protein that belongs to the kinesin-like protein family. A region that is approximately 40 kb away from *KIF25* was shown to be associated with Attention Deficit Hyperactivity Disorder (Ebejer et al., 2013). However, there are few functional studies of this gene and its exact role in alcohol-related behaviors remains unknown. Nonetheless, these results highlight the importance of *KIF25* expression in LR.

Individuals with low LR, and consequently, higher SRE scores are at increased likelihood of drinking heavily and developing problems AD (Schuckit et al., 1997a). The present study found that a preponderance of this correlation is attributable to genetic factors, as indexed by strong correlations in latent heritable factors influencing SRE and AD phenotypes. However, these high correlations should be interpreted within the context of our ascertainment strategy, which oversampled for individuals with family history of AD. It should also be distinguished from more recent SNP-based genetic correlations that examine the extent of similarity between genomewide effect sizes for two traits. In contrast to the high r_g from the pedigree data, our study also added a novel exploration of the extent to which polygenic liability to SRE relates to AD. Overall, PRS mostly explained modest proportions of variance in AD phenotypes. Nearly 2.5% and 1.4% of the variance in AD diagnosis and criterion count, respectively, were explained by polygenic liability to SRE-T in the SAGE-

EA data. Predictions were less significant in other target cohorts, potentially due to the similar ascertainment for alcohol dependence of SAGE and COGA. For example, even though Yale-Penn-AA was ascertained for alcohol dependence, it has a high rate of other illicit drug dependence which may have impacted predictions. Polygenic liability to SRE-5 was poorly related to AD phenotypes, consistent with its lower genetic correlation with AD diagnosis and criterion count.

Our results suggest that some polygenic liability underlying SRE-T and AD is common but also highlight that each phenotype is likely to be influenced by unique genetic factors. For instance, the current strongest signal for AD, rs1229984 in *ADH1B* (Walters et al., 2018), was not associated at genomewide significant levels to SRE-T (EA p= 0.00027; AA p=0.019; rs2066702, strongest signal in AA, p=0.012). Similarly, the most significant variant for SRE-T in chromosome 11, rs10647170, was not associated with AD (p>0.18 in COGA-AA, COGA-EA, and COGA-AA-EA). Thus, the genetic variants associated with SRE and those relating to AD are likely to only modestly overlap. For example, the AD risk is associated with high impulsivity and sensation seeking, low conscientiousness, and the risk for some psychiatric disorders (e.g., schizophrenia), characteristics that have little, if any, links to LR.

This study has several limitations. Despite the size of the COGA discovery samples, power to detect association with low frequency variants is modest. Likewise, sample sizes of the replication datasets were small, raising the possibility that true findings were not replicated. Despite this, one association was confirmed by the NIAAA samples, and meta-analyses of all datasets resulted in increased significance of genomewide significant signals. Additionally, we used hard-called genotypes as necessitated by our family data requiring rigorous Mendelian error checking, which is not straightforward with dosage data (see also Walters et al., 2018 for a similar approach to family data (Walters et al., 2018)). Nonetheless, results for our top loci are similar when dosage data were used for a confirmatory analysis (e.g., rs10647170 P=1.53E-08). Another limitation inherent in both *in silico* functional analyses and RNA expression analysis is the source of the data, which are largely from subjects of European ancestry. As a result, conclusions from these functional data may be limited when analyzing associated variants identified in the COGA-AA sample. Nonetheless, this represents an important first step towards characterizing genetic risk in a key index of AD liability. This study is also further strengthened by investigation of both EA and AA individuals, who are known to differ in SRE scores.

We identified novel loci associated with SRE scores in both EA and AA samples. As additional samples with SRE data, specifically SRE-T, become available, individual loci and polygenic risk scores from such GWAS might explain even greater variance in LR. It will set the stage for future interventions that might aim to interrupt progression to AD (Roche and Ray, 2015, Schuckit et al., 2016). This boost in power necessitates further collection of SRE data. Since the assessment is brief and easily administered, and is included the PhenX tool kit (https://www.phenxtoolkit.org/), large biobanks and nation-wide data collection efforts are encouraged to gather data on SRE to capture greater variability to susceptibility to future problem drinking.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

COGA: The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. Edenberg, L. Bierut, includes eleven different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, J. Nurnberger Jr., T. Foroud; Y. Liu); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. Bierut, J. Rice, K. Bucholz, A. Agrawal); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield, A. Brooks); Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA (L. Almasy), Virginia Commonwealth University (D. Dick), Icahn School of Medicine at Mount Sinai (A. Goate), and Howard University (R. Taylor). Other COGA collaborators include: L. Bauer (University of Connecticut); J. McClintick, L. Wetherill, X. Xuei, D. Lai, S. O'Connor, M. Plawecki, S. Lourens (Indiana University); G. Chan (University of Iowa; University of Connecticut); J. Meyers, D. Chorlian, C. Kamarajan, A. Pandey, J. Zhang (SUNY Downstate); J.-C. Wang, M. Kapoor, S. Bertelsen (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, S. Saccone (Washington University); J. Salvatore, F. Aliev, B. Cho (Virginia Commonwealth University); and Mark Kos (University of Texas Rio Grande Valley). A. Parsian and H. Chen are the NIAAA Staff Collaborators.

We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

Funding and Disclosure

COGA is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

A. Agrawal receives additional funding support from NIDA (DA032573).

Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201200008I.

D. Goldman, ML Schwandt and VA Ramchandani are supported by the NIAAA Division of Intramural Clinical and Biological Research.

R. Dayne Mayfield is supported from NIH/NIAAA: AA020926 and AA012404

Alison Goate is listed as an inventor on Issued U.S. Patent 8080,371, "Markers for Addiction" covering the use of certain variants in determining the diagnosis, prognosis, and treatment of addiction.

John Nurnberger receives support as an investigator for Janssen.

All other authors have no potential conflicts of interest.

References:

ALMASY L & BLANGERO J 1998 Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet, 62, 1198–211. [PubMed: 9545414]

BUCHOLZ KK, CADORET R, CLONINGER CR, DINWIDDIE SH, HESSELBROCK VM, NURNBERGER JI JR., REICH T, SCHMIDT I & SCHUCKIT MA 1994 A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol, 55, 149–58. [PubMed: 8189735]

CHEN MH & YANG Q 2010 GWAF: an R package for genome-wide association analyses with family data. Bioinformatics, 26, 580–1. [PubMed: 20040588]

DAEPPEN JB, LANDRY U, PECOUD A, DECREY H & YERSIN B 2000 A measure of the intensity of response to alcohol to screen for alcohol use disorders in primary care. Alcohol Alcohol, 35, 625–7. [PubMed: 11093971]

DAS S, FORER L, SCHONHERR S, SIDORE C, LOCKE AE, KWONG A, VRIEZE SI, CHEW EY, LEVY S, MCGUE M, SCHLESSINGER D, STAMBOLIAN D, LOH PR, IACONO WG, SWAROOP A, SCOTT LJ, CUCCA F, KRONENBERG F, BOEHNKE M, ABECASIS GR & FUCHSBERGER C 2016 Next-generation genotype imputation service and methods. Nat Genet, 48, 1284–1287. [PubMed: 27571263]

DELANEAU O, ZAGURY JF & MARCHINI J 2013 Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods, 10, 5–6. [PubMed: 23269371]

DENG X, SAGATA N, TAKEUCHI N, TANAKA M, NINOMIYA H, IWATA N, OZAKI N, SHIBATA H & FUKUMAKI Y 2008 Association study of polymorphisms in the neutral amino acid transporter genes SLC1A4, SLC1A5 and the glycine transporter genes SLC6A5, SLC6A9 with schizophrenia. BMC Psychiatry, 8, 58. [PubMed: 18638388]

DICK DM, CHO SB, LATENDRESSE SJ, ALIEV F, NURNBERGER JI JR., EDENBERG HJ, SCHUCKIT M, HESSELBROCK VM, PORJESZ B, BUCHOLZ K, WANG JC, GOATE A, KRAMER JR & KUPERMAN S 2014 Genetic influences on alcohol use across stages of development: GABRA2 and longitudinal trajectories of drunkenness from adolescence to young adulthood. Addict Biol, 19, 1055–64. [PubMed: 23692184]

DOBIN A, DAVIS CA, SCHLESINGER F, DRENKOW J, ZALESKI C, JHA S, BATUT P, CHAISSON M & GINGERAS TR 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29, 15–21. [PubMed: 23104886]

EBEJER JL, DUFFY DL, VAN DER WERF J, WRIGHT MJ, MONTGOMERY G, GILLESPIE NA, HICKIE IB, MARTIN NG & MEDLAND SE 2013 Genome-wide association study of inattention and hyperactivity-impulsivity measured as quantitative traits. Twin Res Hum Genet, 16, 560–74. [PubMed: 23527680]

EDWARDS AC, DEAK JD, GIZER IR, LAI D, CHATZINAKOS C, WILHELMSEN KP, LINDSAY J, HERON J, HICKMAN M, WEBB BT, BACANU SA, FOROUD TM, KENDLER KS, DICK DM & SCHUCKIT MA 2018 Meta-Analysis of Genetic Influences on Initial Alcohol Sensitivity. Alcohol Clin Exp Res.

EHLERS CL, GIZER IR, SCHUCKIT MA & WILHELMSEN KC 2010 Genome-wide scan for self-rating of the effects of alcohol in American Indians. Psychiatr Genet, 20, 221–8. [PubMed: 20505555]

EUESDEN J, LEWIS CM & O'REILLY PF 2015 PRSice: Polygenic Risk Score software. Bioinformatics, 31, 1466–8. [PubMed: 25550326]

GELERNTER J, ZHOU H, NUNEZ YZ, MUTIRANGURA A, MALISON RT & KALAYASIRI R 2018 Genomewide Association Study of Alcohol Dependence and Related Traits in a Thai Population. Alcohol Clin Exp Res, 42, 861–868. [PubMed: 29460428]

GRANT BF, DAWSON DA, STINSON FS, CHOU SP, DUFOUR MC & PICKERING RP 2004 The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. Drug Alcohol Depend, 74, 223–34. [PubMed: 15194200]

GRUCZA RA, BUCHOLZ KK, RICE JP & BIERUT LJ 2008 Secular trends in the lifetime prevalence of alcohol dependence in the United States: a re-evaluation. Alcohol Clin Exp Res, 32, 763–70. [PubMed: 18336633]

HASIN DS & GRANT BF 2015 The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) Waves 1 and 2: review and summary of findings. Soc Psychiatry Psychiatr Epidemiol, 50, 1609–40. [PubMed: 26210739]

HEATH AC, MADDEN PA, BUCHOLZ KK, DINWIDDIE SH, SLUTSKE WS, BIERUT LJ, ROHRBAUGH JW, STATHAM DJ, DUNNE MP, WHITFIELD JB & MARTIN NG 1999 Genetic differences in alcohol sensitivity and the inheritance of alcoholism risk. Psychol Med, 29, 1069–81. [PubMed: 10576299]

HEATH AC & MARTIN NG 1991 Intoxication after an acute dose of alcohol: an assessment of its association with alcohol consumption patterns by using twin data. Alcohol Clin Exp Res, 15, 122–8. [PubMed: 2024724]

HINGSON RW, HEEREN T & WINTER MR 2006 Age at drinking onset and alcohol dependence: age at onset, duration, and severity. Arch Pediatr Adolesc Med, 160, 739–46. [PubMed: 16818840]

JOHNSON EC & AGRAWAL A 2018 The genetic relationship between alcohol consumption and aspects of social and problem drinking. Alcoholism: Clinical and Experimental Research, Submited.

KALU N, RAMCHANDANI VA, MARSHALL V, SCOTT D, FERGUSON C, CAIN G & TAYLOR R 2012 Heritability of level of response and association with recent drinking history in nonalcohol-dependent drinkers. Alcohol Clin Exp Res, 36, 1034–41. [PubMed: 22235947]

KAPOOR M, FARRIS SP, WANG JC, BERTELSEN S, EDENBERG HJ, LIU Y, MAYFIELD D, GOATE A, COLLABORATORS C & COLLABORATORS I 2018 Analysis of Whole Genome-Transcriptomic Organization in Brain to Identify Genes Associated with Alcoholism. Alcoholism-Clinical and Experimental Research, 42, 283a–283a.

KRANZLER HR, ZHOU H, KEMBER RL, VICKERS SMITH R, JUSTICE AC, DAMRAUER S, TSAO PS, KLARIN D, BARAS A, REID J, OVERTON J, RADER DJ, CHENG Z, TATE JP, BECKER WC, CONCATO J, XU K, POLIMANTI R, ZHAO H & GELERNTER J 2019 Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nature Communications, 10, 1499.

MARCHINI J, HOWIE B, MYERS S, MCVEAN G & DONNELLY P 2007 A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet, 39, 906–13. [PubMed: 17572673]

NYHOLT DR 2004 A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. American Journal of Human Genetics, 74, 765–769. [PubMed: 14997420]

O'CONNELL JR & WEEKS DE 1998 PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet, 63, 259–66. [PubMed: 9634505]

PARK SL, CARMELLA SG, CHEN M, PATEL Y, STRAM DO, HAIMAN CA, LE MARCHAND L & HECHT SS 2015 Mercapturic Acids Derived from the Toxicants Acrolein and Crotonaldehyde in the Urine of Cigarette Smokers from Five Ethnic Groups with Differing Risks for Lung Cancer. PLoS One, 10, e0124841. [PubMed: 26053186]

PEDERSEN SL & MCCARTHY DM 2013 Differences in acute response to alcohol between African Americans and European Americans. Alcohol Clin Exp Res, 37, 1056–63. [PubMed: 23398190]

PRICE AL, PATTERSON NJ, PLENGE RM, WEINBLATT ME, SHADICK NA & REICH D 2006 Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet, 38, 904–9. [PubMed: 16862161]

PURCELL S, NEALE B, TODD-BROWN K, THOMAS L, FERREIRA MA, BENDER D, MALLER J, SKLAR P, DE BAKKER PI, DALY MJ & SHAM PC 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet, 81, 559–75. [PubMed: 17701901]

RAY LA, MACKILLOP J & MONTI PM 2010 Subjective responses to alcohol consumption as endophenotypes: advancing behavioral genetics in etiological and treatment models of alcoholism. Subst Use Misuse, 45, 1742–65. [PubMed: 20590398]

RIETVELD CA, ESKO T, DAVIES G, PERS TH, TURLEY P, BENYAMIN B, CHABRIS CF, EMILSSON V, JOHNSON AD, LEE JJ, DE LEEUW C, MARIONI RE, MEDLAND SE, MILLER MB, ROSTAPSHOVA O, VAN DER LEE SJ, VINKHUYZEN AA, AMIN N, CONLEY D, DERRINGER J, VAN DUIJN CM, FEHRMANN R, FRANKE L, GLAESER EL, HANSELL NK, HAYWARD C, IACONO WG, IBRAHIM-VERBAAS C, JADDOE V, KARJALAINEN J, LAIBSON D, LICHTENSTEIN P, LIEWALD DC, MAGNUSSON PK, MARTIN NG, MCGUE M, MCMAHON G, PEDERSEN NL, PINKER S, PORTEOUS DJ, POSTHUMA D, RIVADENEIRA F, SMITH BH, STARR JM, TIEMEIER H, TIMPSON NJ, TRZASKOWSKI M, UITTERLINDEN AG, VERHULST FC, WARD ME, WRIGHT MJ, DAVEY SMITH G, DEARY IJ, JOHANNESSON M, PLOMIN R, VISSCHER PM, BENJAMIN DJ, CESARINI D & KOELLINGER PD 2014 Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. Proc Natl Acad Sci U S A, 111, 13790–4. [PubMed: 25201988]

ROCHE DJ & RAY LA 2015 Subjective response as a consideration in the pharmacogenetics of alcoholism treatment. Pharmacogenomics, 16, 721–36. [PubMed: 25950242]

SAVAGE JE, NEALE Z, CHO SB, HANCOCK L, KALMIJN JA, SMITH TL, SCHUCKIT MA, DONOVAN KK & DICK DM 2015 Level of Response to Alcohol as a Factor for Targeted Prevention in College Students. Alcoholism-Clinical and Experimental Research, 39, 2215–2223.

SCHUCKIT MA 2018 A Critical Review of Methods and Results in the Search for Genetic Contributors to Alcohol Sensitivity. Alcohol Clin Exp Res, 42, 822–835. [PubMed: 29623680]

SCHUCKIT MA, EDENBERG HJ, KALMIJN J, FLURY L, SMITH TL, REICH T, BIERUT L, GOATE A & FOROUD T 2001 A genome-wide search for genes that relate to a low level of response to alcohol. Alcohol Clin Exp Res, 25, 323–9. [PubMed: 11290841]

SCHUCKIT MA, KALMIJN JA, SMITH TL, SAUNDERS G & FROMME K 2012 Structuring a college alcohol prevention program on the low level of response to alcohol model: a pilot study. Alcohol Clin Exp Res, 36, 1244–52. [PubMed: 22309202]

SCHUCKIT MA & SMITH TL 2000 The relationships of a family history of alcohol dependence, a low level of response to alcohol and six domains of life functioning to the development of alcohol use disorders. J Stud Alcohol, 61, 827–35. [PubMed: 11188488]

SCHUCKIT MA, SMITH TL, CLAUSEN P, FROMME K, SKIDMORE J, SHAFIR A & KALMIJN J 2016 The Low Level of Response to Alcohol-Based Heavy Drinking Prevention Program: One-Year Follow-Up. J Stud Alcohol Drugs, 77, 25–37. [PubMed: 26751352]

SCHUCKIT MA, SMITH TL, DANKO GP, PIERSON J, HESSELBROCK V, BUCHOLZ KK, KRAMER J, KUPERMAN S, DIETIKER C, BRANDON R & CHAN G 2007 The ability of the Self-Rating of the Effects of Alcohol (SRE) Scale to predict alcohol-related outcomes five years later. J Stud Alcohol Drugs, 68, 371–8. [PubMed: 17446976]

SCHUCKIT MA, SMITH TL & TIPP JE 1997a The Self-Rating of the Effects of alcohol (SRE) form as a retrospective measure of the risk for alcoholism. Addiction, 92, 979–88. [PubMed: 9376780]

SCHUCKIT MA, SMITH TL, TRIM R, FUKUKURA T & ALLEN R 2009 The overlap in predicting alcohol outcome for two measures of the level of response to alcohol. Alcohol Clin Exp Res, 33, 563–9. [PubMed: 19120060]

SCHUCKIT MA, SMITH TL, TRIM R, HERON J, HORWOOD J, DAVIS JM, HIBBELN JR & TEAM AS 2008a The performance of elements of a ‘level of response to alcohol’-based model of drinking behaviors in 13-year-olds. Addiction, 103, 1786–92. [PubMed: 18778389]

SCHUCKIT MA, SMITH TL, TRIM RS, HERON J, HORWOOD J, DAVIS J, HIBBELN J & TEAM AS 2008b The self-rating of the effects of alcohol questionnaire as a predictor of alcohol-related outcomes in 12-year-old subjects. Alcohol Alcohol, 43, 641–6. [PubMed: 18845530]

SCHUCKIT MA, TIPP JE, SMITH TL, WIESBECK GA & KALMIJN J 1997b The relationship between Self-Rating of the Effects of alcohol and alcohol challenge results in ninety-eight young men. J Stud Alcohol, 58, 397–404. [PubMed: 9203121]

SHABALIN AA 2012 Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics, 28, 1353–8. [PubMed: 22492648]

SHERVA R, WANG Q, KRANZLER H, ZHAO H, KOESTERER R, HERMAN A, FARRER LA & GELERNTER J 2016 Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. JAMA Psychiatry, 73, 472–80. [PubMed: 27028160]

VIKEN RJ, ROSE RJ, MORZORATI SL, CHRISTIAN JC & LI TK 2003 Subjective intoxication in response to alcohol challenge: heritability and covariation with personality, breath alcohol level, and drinking history. Alcohol Clin Exp Res, 27, 795–803. [PubMed: 12766624]

WALL TL, JOHNSON ML, HORN SM, CARR LG, SMITH TL & SCHUCKIT MA 1999 Evaluation of the self-rating of the effects of alcohol form in Asian Americans with aldehyde dehydrogenase polymorphisms. J Stud Alcohol, 60, 784–9. [PubMed: 10606490]

WALTERS RK, POLIMANTI R, JOHNSON EC, MCCLINTICK JN, ADAMS MJ, ADKINS AE, ALIEV F, BACANU S-A, BATZLER A, BERTELSEN S, BIERNACKA JM, BIGDELI TB, CHEN L-S, CLARKE T-K, CHOU Y-L, DEGENHARDT F, DOCHERTY AR, EDWARDS AC, FONTANILLAS P, FOO JC, FOX L, FRANK J, GIEGLING I, GORDON S, HACK LM, HARTMANN AM, HARTZ SM, HEILMANN-HEIMBACH S, HERMS S, HODGKINSON C, HOFFMANN P, HOTTENGA JJ, KENNEDY MA, ALANNE-KINNUNEN M, KONTE B,

LAHTI J, LAHTI-PULKKINEN M, LAI D, LIGTHART L, LOUKOLA A, MAHER BS, MBAREK H, MCINTOSH AM, MCQUEEN MB, MEYERS JL, MILANESCHI Y, PALVIAINEN T, PEARSON JF, PETERSON RE, RIPATTI S, RYU E, SACCONE NL, SALVATORE JE, SANCHEZ-ROIGE S, SCHWANDT M, SHERVA R, STREIT F, STROHMAIER J, THOMAS N, WANG J-C, WEBB BT, WEDOW R, WETHERILL L, WILLS AG, ANDME RESEARCH T, BOARDMAN JD, CHEN D, CHOI D-S, COPELAND WE, CULVERHOUSE RC, DAHMEN N, DEGENHARDT L, DOMINGUE BW, ELSON SL, FRYE MA, GÄBEL W, HAYWARD C, ISING M, KEYES M, KIEFER F, KRAMER J, KUPERMAN S, LUCAE S, LYNSKEY MT, MAIER W, MANN K, MÄNNISTÖ S, MÜLLER-MYHSOK B, MURRAY AD, NURNBERGER JI, PALOTIE A, PREUSS U, RÄIKKÖNEN K, REYNOLDS MD, RIDINGER M, SCHERBAUM N, SCHUCKIT MA, SOYKA M, TREUTLEIN J, WITT S, et al. 2018 Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nature Neuroscience.

WANG L, WANG S & LI W 2012 RSeQC: quality control of RNA-seq experiments. Bioinformatics, 28, 2184–5. [PubMed: 22743226]

WATANABE K, TASKESEN E, VAN BOCHOVEN A & POSTHUMA D 2017 Functional mapping and annotation of genetic associations with FUMA. Nat Commun, 8, 1826. [PubMed: 29184056]

WILLER CJ, LI Y & ABECASIS GR 2010 METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26, 2190–1. [PubMed: 20616382]

YANG J, LEE SH, GODDARD ME & VISSCHER PM 2011 GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet, 88, 76–82. [PubMed: 21167468]

ZHAN X, HU Y, LI B, ABECASIS GR & LIU DJ 2016 RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics, 32, 1423–6. [PubMed: 27153000]

Table 1:

Summary of COGA-AA and COGA-EA subsamples.

Sample	# Family	# Individual	% AD	# AD case[1] (%male)	# AD control[2] (%male)	Mean AD DSM-IV criterion count (stdev) in cases	Mean AD DSM-IV criterion count (stdev) in controls	Mean SRE-T (stdev)	Mean SRE-5 (stdev)
COGA-AA	309	1,527	32.74	500 (58.60)	421 (28.50)	5.05(1.46)	0.19 (0.39)	4.55 (2.86)	3.36 (2.08)
COGA-EA	956	4,723	36.35	1,717 (60.51)	1,336(27.77)	5.14(1.54)	0.21 (0.41)	5.26 (2.78)	3.73 (1.98)
Total	1,265	6,250							

1: Individuals who met the DSM-IV criteria for AD but younger than 15 were not classified.

2: Individuals who had two DSM-IV criteria, or never drank, or younger than 21 were not classified.

Table 2:

Heritability, genetic correlations between AD, DSM-IV AD criterion count and SRE scores.

			AD		DSM-IV AD criterion count	
Sample	**Phenotype**	**Heritability (p value)**	**Genetic correlation**	**P value**	**Genetic correlation**	**P value**
COGA-AA	**SRE-T**	0.21 (4.10E-06)	0.76	1.25E-07	0.63	4.20E-06
COGA-AA	**SRE-5**	0.20 (6.35E-05)	0.48	1.88E-03	0.40	6.13E-03
COGA-EA	**SRE-T**	0.31 (1.14E-35)	0.61	4.55E-25	0.49	8.26E-19
COGA-EA	**SRE-5**	0.31 (7.01E-34)	0.41	4.52E-11	0.35	9.44E-10

Table 3:

Summary of genome wide significant variants. Genome wide significant p values are in bold.

								SRE-T P values						
								Discovery dataset			Replication dataset			
CHR	**BP**	**Variants**	**Gene**	**Function**	**Minor Allele**	**COGA-AA MAF**	**COGA-EA MAF**	**COGA-AA**	**COGA-EA**	**COGA-AA+EA**	**NIAAA-AA**	**NIAAA-EA**	**SD-EA**	**Meta***
6	168,329,922	rs140154945	*AFDN*	intronic	TA	0.23	0.31	0.22	**3.30E-08**	7.19E-06	NA	0.14	0.24	1.82E-07
11	20,401,835	rs7934001	*HTATIP2*	intronic	T	0.20	0.05	2.46E-06	1.66E-03	**4.49E-08**	0.73	0.02	0.40	**4.53E-08**
11	20,401,940	rs7936865	*HTATIP2*	intronic	T	0.20	0.05	3.47E-06	9.34E-04	**2.97E-08**	0.74	0.03	0.41	**2.64E-08**
11	20,402,702	rs2055244	*HTATIP2*	intronic	T	0.20	0.05	2.64E-06	1.78E-03	**4.35E-08**	0.92	0.06	0.47	**4.88E-08**
11	20,402,886	rs2055246	*HTATIP2*	intronic	C	0.13	0.05	7.51E-08	3.21E-03	**1.57E-08**	0.16	0.05	0.47	**8.31E-09**
11	20,409,169	rs2270678	*PRMT3*	UTR5	G	0.21	0.04	3.18E-06	6.73E-04	**1.81E-08**	0.38	0.11	0.37	**8.20E-09**
11	20,411,516	rs2403592	*PRMT3*	intronic	A	0.13	0.04	2.72E-07	1.42E-03	**9.17E-09**	0.03	0.19	0.39	**1.64E-09**
11	20,420,005	rs7938365	*PRMT3*	intronic	G	0.12	0.05	4.46E-07	2.04E-03	**2.65E-08**	0.05	0.16	0.57	**7.45E-09**
11	20,422,730	rs7480236	*PRMT3*	intronic	T	0.13	0.04	5.45E-07	1.04E-03	**1.12E-08**	0.03	0.16	0.41	**1.96E-09**
11	20,422,731	rs7480181	*PRMT3*	intronic	T	0.13	0.04	5.45E-07	1.04E-03	**1.12E-08**	0.03	0.16	0.41	**1.96E-09**
11	20,424,851	rs11025551	*PRMT3*	intronic	G	0.13	0.05	4.10E-07	1.18E-03	**9.66E-09**	0.03	0.18	0.40	**1.72E-09**
11	20,427,324	rs12221482	*PRMT3*	intronic	C	0.12	0.05	1.66E-07	1.09E-03	**5.49E-09**	0.05	0.16	0.42	**1.26E-09**
11	20,432,046	rs10647170	*PRMT3*	intronic	TAGTC	0.12	0.04	1.69E-07	7.25E-04	**3.53E-09**	0.05	0.26	0.42	**6.45E-10**
11	20,432,797	rs74761974	*PRMT3*	intronic	A	0.12	0.04	2.24E-07	6.93E-04	**3.93E-09**	0.05	0.15	0.42	**9.04E-10**
11	20,435,147	rs74183993	*PRMT3*	intronic	C	0.12	0.04	2.36E-07	7.04E-04	**4.69E-09**	0.05	0.14	0.42	**9.84E-10**
								SRE-5 P values						
13	23,552,108	rs4770359	*LINC00621,SGCG*	intergenic	G	0.19	0.46	**2.92E-08**	0.75	0.08	0.82	NA	NA	6.33E-08

*For meta-analysis, if genomewide significant findings were from COGA-AA, then COGA-AA, NIAAA-AA were meta-analyzed; if genomewide significant findings were from COGA-EA, then COGA-EA, NIAAA-EA, and SD-EA were meta-analyzed; otherwise, all datasets were meta-analyzed.

Table 4:

Results of PRS derived using COGA SRE GWAS summary statistics. P<0.0125 are in bold.

		AD		DSM-IV AD criterion count	
Discovery dataset	**Test dataset**	**R^2**	**p value**	**R^2**	**p value**
SRE-T COGA-EA	SAGE-EA	2.48%	**4.13E-07**	1.37%	**8.55E-07**
SRE-T COGA-EA	OZALC-EA	0.34%	0.05	0.60%	**1.14E-03**
SRE-T COGA-AA	SAGE-AA	1.49%	**6.10E-03**	1.04%	**1.72E-03**
SRE-T COGA-AA	Yale-Penn-AA	0.36%	0.03	0.05%	0.33
SRE-5 COGA-EA	SAGE-EA	0.88%	**2.37E-03**	0.61%	**1.08E-03**
SRE-5 COGA-EA	OZALC-EA	0.14%	0.21	0.47%	**3.85E-03**
SRE-5 COGA-AA	SAGE-AA	0.53%	0.10	0.34%	0.07
SRE-5 COGA-AA	Yale-Penn-AA	0.0007%	0.92	0.03%	0.43