Multivariate GWAS elucidates the genetic architecture of alcohol consumption and misuse, corrects biases, and reveals novel associations with disease

Travis T Mallard[1], Jeanne E Savage[2], Emma C Johnson[3], Yuye Huang[4], Alexis C Edwards[5], Jouke J Hottenga[6], Andrew D Grotzinger[1], Daniel E Gustavson7, Mariela V Jennings[4], Andrey Anokhin[3], Danielle M Dick[8], Howard J Edenberg[9], John R Kramer[10], Dongbing Lai[11], Jacquelyn L Meyers[12], Ashwini K Pandey[12], Kathryn Paige Harden[1], Michel G Nivard[6], Eco JC de Geus[6], Dorret I Boomsma[6], Arpana Agrawal[3], Lea K Davis[7,13,14], Toni-Kim Clarke[15], Abraham A Palmer[4,16], & **Sandra Sanchez-Roige (Presenter)**[4,7]

[1] Department of Psychology, University of Texas at Austin, Austin, TX, 78712, [2] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Netherlands, 1081HV, [3] Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO 63110, [4] Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, [5] Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA 23298, [7] Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, [8] Department of Psychology, Virginia Commonwealth University, Richmond, VA 23220, [9] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, [10] Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, [11] Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 4622, [12] Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY 11203, [13] Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA, [14] Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA, [15] Division of Psychiatry, University of Edinburgh, Scotland, UK, EH8 9YL, [16] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093

Motivation for this study

Genome-wide association studies (GWASs) of the Alcohol Use Disorder Identification Test (AUDIT), a ten-item screener for alcohol use disorder (AUD), have elucidated novel loci for alcohol consumption and misuse. However, these studies also revealed that GWASs can be influenced by numerous biases (e.g., measurement error, selection bias), which have led to inconsistent genetic correlations between alcohol involvement and AUD, as well as paradoxically negative genetic correlations between alcohol involvement and psychiatric disorders/medical conditions. To explore these unexpected differences in genetic correlations, we conducted the first item-level and largest GWAS of AUDIT items (N=160,824), and applied a multivariate framework to mitigate previous biases.

Sanchez-Roige et al, *Am J Psych,* 2019
Sanchez-Roige et al, *Biol Psych,* 2019

GWAS of AUDIT scores in UK Biobank, NTR, ALSPAC

We obtained genotype and phenotypic data from the AUDIT, which is a 10-item screening questionnaire that measures both aspects of alcohol consumption (items 1-3) and problematic use (items 4-10), from three population-based cohorts, UK Biobank (UKB, N=147,267), the Netherlands Twin Register (NTR, N=9,975), and the Avon Longitudinal Study of Parents and Children (ALSPAC, N=3,582), and performed ten GWAS for each of the AUDIT items. We used METAL to perform sample-sized weighted meta-analyses of the cohort-level GWAS for each AUDIT item.

Consumption	Problems
How often do you have a drink containing alcohol?	**How often** have you found that you were **not able to stop** drinking once you had started?
How many drinks containing alcohol do you have on a typical day when you are drinking?	How often have you **failed to do what was expected** from you because of drinking?
How often do you have **six or more drinks on one occasion?**	How often have you needed a **first drink in the morning** to get yourself going after a heavy drinking session?
	How often have you had a feeling of **guilt or remorse** after drinking?
	How often have you been **unable to remember** what happened the night before because you had been drinking?
	Have you or someone else been **injured** as a result of your drinking?
	Has a relative or friend or a doctor or another health worker been **concerned** about your drinking or suggested you cut down?

Genomic Structural Equation Modeling

We used the lavaan v0.6.5 and Genomic SEM v0.0.2 packages in R to conduct phenotypic and genetic confirmatory factor analyses, respectively, using weighted least squares estimation. We used Genomic SEM to perform multivariate GWAS for the latent genetic factors from the best-fitting model.

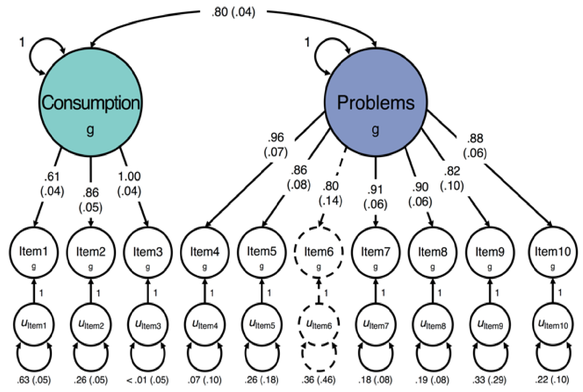
Polygenic Analyses

We created polygenic scores (PRS), using the PRS-CS "auto" version, for the latent genetic AUDIT factors (Consumption, Problems), from two independent samples: (i) an independent subset in the UKB who did not fill out the AUDIT, and (ii) a subset from the Collaborative Study on the Genetics of Alcoholism (COGA). Using the 'score' algorithm in PLINK v1.90, we computed individual-level PRS to predict additional alcohol phenotypes (drinking quantity, drinking frequency, and AUD diagnosis) in UKB and COGA. In UKB, we included sex, age at first assessment, Townsend Deprivation Index score, and the first ten ancestry PCs as covariates. In COGA, we included age, sex, array type, income, and the first 10 ancestry PCs as fixed effect covariates, with family ID included as a random effect (i.e., allowing the intercept to vary by family).

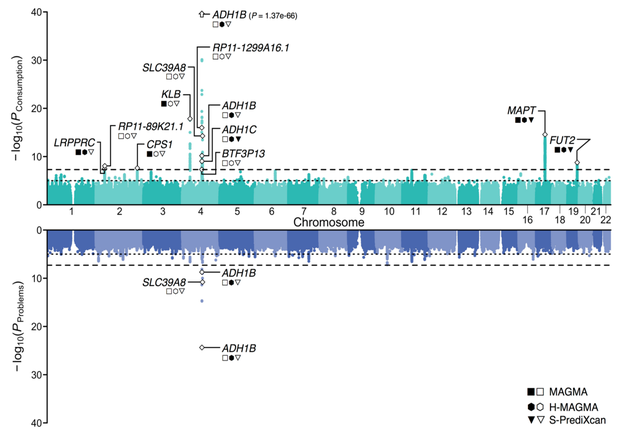
Phenome-Wide Association Analyses

PRS were computed using the PRS-CS method described above for each of the 66,915 unrelated genotyped individuals of European ancestry from the Vanderbilt University Medical Center (BioVU). We performed phenome-wide association analyses (PheWAS) for Consumption and Problems using the PheWAS v0.12 package in R. We fitted a logistic regression model to each of 1,335 case/control phenotypes to estimate the odds of each diagnosis given the PRS, after adjustment for sex, median age of the longitudinal electronic health record measurements, and the first 10 ancestry PCs. We repeated the PheWAS analyses using AUD diagnoses (phecodes 317, 317.1) as additional covariates. We used the standard Benjamini–Hochberg false discovery rate (FDR 5%) correction to account for multiple testing.

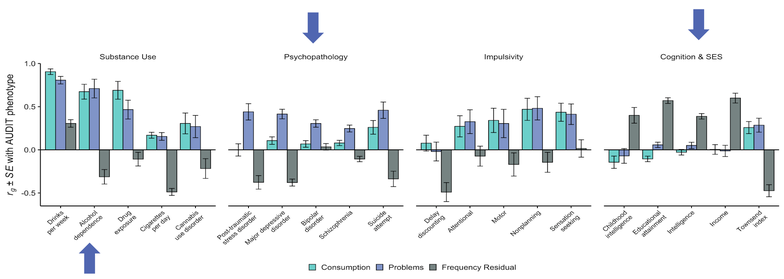
We found evidence of a correlated two-factor structure at the genetic level (*Consumption* and *Problems*, rg=.80)



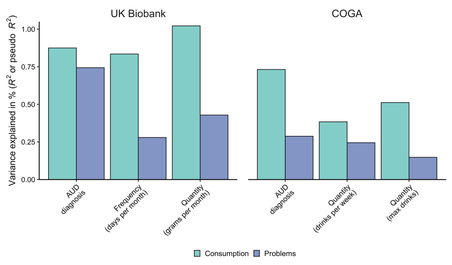
Consumption and *Problems* have a distinct genetic basis



Our method corrects previous spurious associations with *positive* health and socioeconomic outcomes (in particular, the bias present in item 1, Frequency)

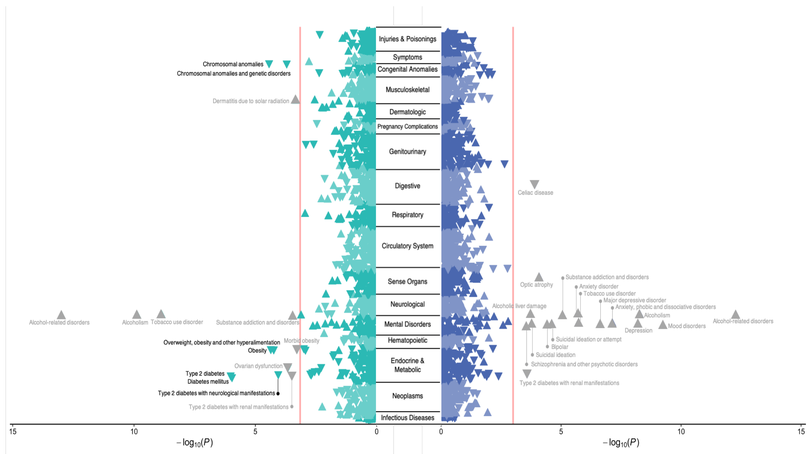


Consumption and *Problems* were highly genetically correlated with alcohol dependence



Consumption is now a valid predictor of alcohol use disorders in low-risk (UKB) and higher-risk samples (COGA)

Problems was most robustly genetically associated with psychopathology



We identified that the associations between ***Problems PRS*** and mental health in BioVU did not persist in the absence of the clinical manifestation of AUD (traits highlighted in gray). These findings suggest that the associations with mental health are not the result of horizontal pleiotropy. Instead, they may be either (1) a consequence of AUD, (2) correlated with other risk factors for AUD (along and/or aside from genetic risk), or (3) related to ascertainment of patients with diagnosed AUD in the medical record. These results also encouragingly suggest that treating AUD could have widespread improvements in overall health.

Conclusions

In the present study, we have performed the first item-level and largest GWAS of AUDIT to date, and used Genomic SEM to elucidate the genetic etiology of alcohol consumption and problematic alcohol use. By conducting genetic factor analysis of the individual AUDIT items, we provide evidence that two correlated latent factors (*Consumption* and *Problems*) parsimoniously explained the covariance in measures of alcohol consumption and problematic alcohol use. Moreover, by applying empirically-derived weights to the AUDIT items in a Genomic SEM framework, we demonstrated that our method ameliorates confounding biases that have complicated previous work with consumption phenotypes (in particular, the bias present in item 1). Notably, both *Consumption* and *Problems* share a strong, positive genetic correlation with alcohol dependence (both rg~0.7), and we show, for the first time, that the polygenic signal of *Consumption* is strongly associated with several AUD phenotypes in three independent cohorts. Finally, the results of our bioinformatic analyses further illustrate that *Consumption* and *Problems* have unique components of their genetic etiology. Collectively, our novel framework provides a means to study two genetic liabilities that are more closely related to AUD, and advances our understanding of the associated biology.

We need more data!

We aim to expand the current discovery sample to >500,000 research participants by the end of 2021

sanchezroige@health.ucsd.edu