

Item-Level Genome-Wide Association Study of the Alcohol Use Disorders Identification Test in Three Population-Based Cohorts

Travis T. Mallard, M.A., Jeanne E. Savage, Ph.D., Emma C. Johnson, Ph.D., Yuye Huang, Alexis C. Edwards, Ph.D., Jouke J. Hottenga, Ph.D., Andrew D. Grotzinger, M.A., Daniel E. Gustavson, Ph.D., Mariela V. Jennings, B.Sc., Andrey Anokhin, Ph.D., Danielle M. Dick, Ph.D., Howard J. Edenberg, Ph.D., John R. Kramer, Ph.D., Dongbing Lai, Ph.D., Jacquelyn L. Meyers, Ph.D., Ashwini K. Pandey, Ph.D., Kathryn Paige Harden, Ph.D., Michel G. Nivard, Ph.D., Eco J.C. de Geus, Ph.D., Dorret I. Boomsma, Ph.D., Arpana Agrawal, Ph.D., Lea K. Davis, Ph.D., Toni-Kim Clarke, Ph.D., Abraham A. Palmer, Ph.D., Sandra Sanchez-Roige, Ph.D.

Objective: Genome-wide association studies (GWASs) of the Alcohol Use Disorders Identification Test (AUDIT), a 10-item screen 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 may have led to inconsistent genetic correlations between alcohol involvement and AUD, as well as paradoxically negative genetic correlations between alcohol involvement and psychiatric disorders and/or medical conditions. The authors used genomic structural equation modeling to elucidate the genetics of alcohol consumption and problematic consequences of alcohol use as measured by AUDIT.

Methods: To explore these unexpected differences in genetic correlations, the authors conducted the first item-level and the largest GWAS of AUDIT items ($N=160,824$) and applied a multivariate framework to mitigate previous biases.

Results: The authors identified novel patterns of similarity (and dissimilarity) among the AUDIT items and found evidence of a correlated two-factor structure at the genetic level ("consumption" and "problems," $r_g=0.80$). Moreover, by applying empirically derived weights to each of the AUDIT items, the authors constructed an aggregate measure of alcohol consumption that was strongly associated with alcohol dependence ($r_g=0.67$), moderately associated with several other psychiatric disorders, and no longer positively associated with health and positive socioeconomic outcomes. Lastly, by conducting polygenic analyses in three independent cohorts that differed in their ascertainment and prevalence of AUD, the authors identified novel genetic associations between alcohol consumption, alcohol misuse, and health.

Conclusions: This work further emphasizes the value of AUDIT for both clinical and genetic studies of AUD and the importance of using multivariate methods to study genetic associations that are more closely related to AUD.

Am J Psychiatry 2022; 179:58–70; doi: 10.1176/appi.ajp.2020.20091390

Over the past decade, genome-wide association studies (GWASs) have advanced our understanding of alcohol use disorder (AUD) (1). Many of these studies have relied on a categorical approach to AUD phenotypes, comparing clinically ascertained case and control subjects (e.g., 2), but recent studies have increasingly employed a complementary approach leveraging dimensional measures of alcohol consumption and screen-based AUD symptoms in population-based cohorts (e.g., 3–6). Compared to clinical diagnostic phenotypes, these dimensional measures can

often be administered more easily at scale via self-report questionnaires, thus accelerating genetic discovery through drastic increases in sample size. The Alcohol Use Disorders Identification Test (AUDIT) (7), a 10-item questionnaire that screens for drinking habits and problems by measuring aspects of alcohol use and misuse in the past year, is one such measure. A recent GWAS meta-analysis of AUD and AUDIT phenotypes identified 29 novel loci (5), representing one of the biggest advances of AUD genetics to date (2–4, 6).

See related feature: **Editorial** by Dr. Kranzler and colleagues (p. 14)

Notably, several studies using self-report instruments have revealed that not all aspects of alcohol involvement are interchangeable. While AUDIT can be used as a unidimensional screen (i.e., AUDIT total score), previous research has shown that AUDIT can differentiate between two related but distinct facets of AUD: alcohol consumption (sum of items 1–3, AUDIT-C), which is necessary but not sufficient for a diagnosis of AUD, and problematic consequences of alcohol consumption (sum of items 4–10, AUDIT-P), which more closely resemble the diagnostic criteria of AUD. We previously found that AUDIT-C and AUDIT-P have distinct genetic relationships with clinically defined AUD (6) as well as other forms of psychopathology. Surprisingly, AUDIT-C was positively associated with socioeconomic variables, negatively associated with some forms of psychopathology, and only moderately positively associated with alcohol dependence, whereas AUDIT-P exhibited strong positive associations with alcohol dependence and numerous other psychiatric disorders. Although this divergence may reflect true differences between the biological mechanisms underlying alcohol consumption and problems, it may be confounded by other factors, such as sources of selection bias, genetic heterogeneity among the individual items, and measurement error (1, 8).

Because AUDIT-C and AUDIT-P are computed using an unweighted composite score approach, they inherently rely on the assumptions that the scale is unidimensional and that each item is equally informative of the construct being measured. This approach is not based on any empirical evidence but rather reflects a holdover from the original use of AUDIT as a screen for use in primary health care settings. Therefore, it is possible that the lack of item-specific weights introduces error in downstream analyses. While these issues have been thoroughly studied at the phenotypic level via factor analysis (see Table S1 in the online supplement), they have not yet been investigated at the genetic level. Using methods that can account for, or mitigate, such measurement problems will allow researchers to better capitalize on the potential of dimensional measures like AUDIT for genetic discovery.

In the present study, we sought to elucidate the genetics of alcohol consumption and problematic consequences of alcohol use measured via AUDIT using genomic structural equation modeling (9), a novel multivariate framework that allows structural equation modeling techniques to be applied to genetic covariance matrices based on GWAS results. Accordingly, we undertook the first item-level and the largest-to-date GWAS meta-analyses of AUDIT (N=160,824), using data from three population-based cohorts of European ancestry. We then used genomic structural equation modeling to analyze the item-level GWAS results with the aims of 1) investigating the latent genetic factor structure of AUDIT, based on prior knowledge (see Table S1 in the online supplement), and 2) conducting multivariate GWASs of the resulting latent genetic factor(s). We posited that applying this approach would lead to more nuanced, empirically derived weights to each of the AUDIT items when constructing our

aggregate measures (as opposed to giving each item equivalent weight), which is a novel approach for GWASs of AUD phenotypes. Finally, to characterize the biology and liability associated with each latent genetic factor, we used a variety of *in silico* tools and polygenic analyses spanning three independent cohorts that varied in method of ascertainment and prevalence of AUD.

We hypothesized that a higher resolution of each of the alcohol phenotypes measured in AUDIT would further our understanding of the differences among indices of alcohol consumption (items 1–3) and problematic alcohol use (items 4–10) and how they relate to health. We anticipated that the genetic contributions to alcohol consumption and problematic use would not be completely overlapping and that genomic modeling using item-level data would ameliorate the confounding issues between alcohol consumption, AUD, and indices of health that complicated previous GWAS efforts.

METHODS

Discovery Samples and Phenotype Construction

We collected AUDIT and genotype data from three population-based cohorts: the UK Biobank (maximum N=147,267), the Netherlands Twin Register (maximum N=9,975), and the Avon Longitudinal Study of Parents and Children (ALSPAC; maximum N=3,582). We used the same phenotyping strategies across the three cohorts, which are described in section 2 of the online supplement. AUDIT scores and demographic characteristics for each cohort are reported in Table S2 in the online supplement. Genotyping, imputation, and quality control procedures have been extensively described in previous publications (10–12). Because AUDIT was administered with skip logic in UK Biobank, we used multiple imputation by chained equations to minimize the impact of missing data on our item-level GWAS (see section 2.1 in the online supplement for details).

Univariate Genome-Wide Association and Meta-Analyses

In UK Biobank, we used BOLT-LMM, version 2.3.2 (13), to conduct GWASs for each of the 10 AUDIT items with the first 40 ancestry principal components, sex, age, sex-by-age interactions, and batch as covariates. In the Netherlands Twin Register, we used the *fastgwa* function of GCTA (14) and included the first five ancestry principal components, sex, birth year, and genotyping platform as covariates. In ALSPAC, we used PLINK, version 2.0 (15), to analyze unrelated participants, including the first 10 ancestry principal components, sex, and age as covariates. Note that both BOLT-LMM and *fastgwa* are capable of analyzing related individuals. Further details are included in section 3 in the online supplement and in previous work (16). We then used METAL (17) to conduct sample-size-weighted meta-analyses of the cohort-level GWAS summary statistics for each AUDIT item after quality control procedures (see section 4 in the online supplement). A total of 8,596,116 single-nucleotide polymorphisms (SNPs) were included in the meta-analyses.

Phenotypic and Genetic Correlations

We used the *lavaan* package, version 0.6.5 (18), in R to estimate polychoric phenotypic correlations (r_p) among AUDIT items. We used the GenomicSEM package, version 0.0.2, in R, which is based on linkage disequilibrium (LD) score regression (19), to estimate the heritability of each of the 10 AUDIT items and the genetic correlations between them. We applied standard quality control procedures prior to all analyses (e.g., use of precomputed LD scores, exclusion of the major histocompatibility region, restriction of SNPs to HapMap 3, application of minor allele frequency $\geq 1\%$, and information score > 0.90 filters). Lastly, we used GenomicSEM (9) to estimate genetic correlations between latent genetic factors and complex traits and disorders broadly related to human health (see section 5.1.2 in the online supplement). We applied a standard Benjamini-Hochberg false discovery rate correction (FDR 5%) to account for multiple testing.

Phenotypic and Genetic Factor Analysis

To empirically model the phenotypic and genetic relationships among AUDIT items, we used *lavaan* and GenomicSEM to conduct phenotypic and genetic confirmatory factor analyses, respectively, using weighted least squares estimation. This process has been described extensively elsewhere (9, 16, 20, 21), and further details are provided in section 5.1 in the online supplement. We tested three models: a parallel factor model (i.e., a sum-score model), a common factor model, and a correlated factors model. The common and correlated factors models were selected based on prior research (see Table S1 in the online supplement), and the parallel factor model served to test the restrictive assumptions of sum-score approaches. We assessed model fit using conventional indices that were available in both the *lavaan* and GenomicSEM software packages (see section 5 in the online supplement). Only data from UK Biobank (the largest sample) were included in the phenotypic factor analyses. For the genetic factor analyses, GWAS summary statistics from the meta-analyses for each AUDIT item were subjected to standard quality control practices, as described above. GenomicSEM's multivariable version of LD score regression was then used to estimate the genetic covariance and sampling covariance matrices for the AUDIT items, which were used to test the specified confirmatory factor models. The sampling covariance matrix was smoothed beforehand, as it was slightly non-positive-definite. Factor extension analysis was used to estimate the expected factor loading of item 6 (i.e., "eye opener"; see section 5.1.1 in the online supplement), as it was excluded from the final genetic confirmatory factor model because of nonsignificant SNP heritability.

Multivariate GWASs

Using GenomicSEM (9), we conducted multivariate GWAS analyses by estimating SNP associations with the AUDIT latent genetic factors from the best-fitting model. The details of these analyses are provided in section 5.1 in the online supplement. Individual SNP effects were estimated for the latent

genetic factors in each model if they were available in all univariate summary statistics, had a minor allele frequency $\geq 0.5\%$, and were present in the 1000 Genomes Phase 3 (version 5) reference panel. The effective sample size for each latent factor was estimated using the approach described by Mallard et al. (16).

Biological Annotation, Gene, and Transcriptome-Based Association Analyses

We performed multiple in silico analyses to compare the results from each of the AUDIT latent genetic factors. First, we used FUMA, version 1.2.8 (22), to identify independent SNPs and study their functional consequences, which included ANNOVAR categories, Combined Annotation Dependent Depletion scores, and RegulomeDB scores. Second, we used MAGMA, version 1.08 (22, 23), to conduct competitive gene-set and pathway analyses for each of the AUDIT genetic latent factors. SNPs were mapped to 18,546 protein coding genes from Ensembl, build 85. Gene sets were obtained from MSigDB, version 7.0 ("curated gene sets," "GO terms"). We also used an extension of this method, Hi-C-coupled MAGMA (H-MAGMA) (24), to assign noncoding (intergenic and intronic) SNPs to genes based on their chromatin interactions. Exonic and promoter SNPs are assigned to genes based on physical position. We used four Hi-C data sets, which were derived from fetal brain, adult brain, and induced pluripotent stem cell-derived neurons and astrocytes (<https://github.com/thewonlab/H-MAGMA>). Lastly, we used S-PrediXcan, version 0.6.2 (25), to predict gene expression levels in 13 brain tissues and to test whether the predicted gene expression showed divergent correlation patterns with each of the AUDIT latent genetic factors. Precomputed tissue weights from the Genotype-Tissue Expression (GTEx, version 8) project database (<https://www.gtexportal.org/>) were used as the reference transcriptome data set. Further details are provided in section 6 in the online supplement.

Polygenic Risk Score Analyses

Prediction of alcohol phenotypes in UK Biobank and COGA. We used the PRS-CS "auto" version (26) to compute polygenic risk scores (PRSs) for the latent genetic AUDIT factors ("consumption" and "problems") and their sum-score counterparts (AUDIT-C and AUDIT-P) in two independent samples: an independent subset of unrelated individuals of European ancestry in the UK Biobank who did not fill out the AUDIT and a subset of individuals of European ancestry from the Collaborative Study on the Genetics of Alcoholism (COGA) (27), which includes probands meeting criteria for alcohol dependence, their family members, and community control families. Using the *score* algorithm in PLINK, version 1.90, we computed individual-level PRSs to predict additional alcohol phenotypes (drinking quantity, drinking frequency, and lifetime AUD diagnosis) measured in UK Biobank and COGA (see section 7 in the online supplement). We tested for associations between AUDIT PRSs and alcohol phenotypes using linear (quantity and frequency phenotypes) or logistic (AUD) regression models in R, version 3.6.3. In UK

Biobank, we included sex, age at first assessment, Townsend deprivation index score (28), and the first 10 ancestry principal components as covariates. In COGA, we included age, sex, array type, income, and the first 10 ancestry principal components as fixed-effect covariates, with family identity included as a random effect (i.e., allowing the intercept to vary by family).

We sought to compare the performance of the latent factor-based PRSs (the consumption and problems PRSs) against the performance of their sum-score counterparts (the AUDIT-C and AUDIT-P PRSs) in predicting different alcohol phenotypes. To this end, we applied two approaches to our PRS analyses: cross-dimension PRS models (i.e., the consumption and problems PRSs included as simultaneous predictors) and cross-method PRS models (i.e., the consumption and AUDIT-C PRSs included as simultaneous predictors in a model, and the problems and AUDIT-P PRSs included as simultaneous predictors in a model). We corrected for the total number of outcome phenotypes across the validation samples using a conservative Bonferroni p value of 8.33×10^{-3} , since the same PRSs were used as predictors across models (and were correlated with each other).

Phenome-wide association study in BioVU. To examine exploratory associations between PRSs and hundreds of medical diagnoses, we used the PRS-CS method (26) described above to compute consumption and problems PRSs for each of the 66,915 unrelated genotyped individuals of European ancestry from the Vanderbilt University Medical Center biobank (BioVU) (29). Using electronic health record data in BioVU, we performed phenome-wide association studies (PheWASs) for consumption and problems PRSs using the PheWAS package, version 0.12 (30), in R. Specifically, we fitted a logistic regression model to each of the 1,335 case/control phenotypes in BioVU (“phcodes”; see section 7.3 in the online supplement) in order to estimate the effect of a given PRS on each diagnosis. Sex, median age of the longitudinal electronic health record measurements, and the first 10 principal components were included as covariates. We then repeated the PheWAS analyses using AUD diagnoses (phcodes 317, 317.1) as additional covariates. A standard Benjamini-Hochberg false discovery rate (FDR 5%) correction was applied to account for multiple testing.

RESULTS

Phenotypic and Genetic Analyses Reveal a Consistent Two-Factor Structure of Alcohol Consumption and Problematic Use

Phenotypic and genetic analyses showed that AUDIT items were positively correlated with each other, with correlation estimates ranging from moderate to large (see Tables S3 and S4 in the online supplement). The one exception to this pattern was item 1 (frequency of consumption), which was generally less correlated with the other AUDIT items. Moreover, we found that genetic correlations tended to be

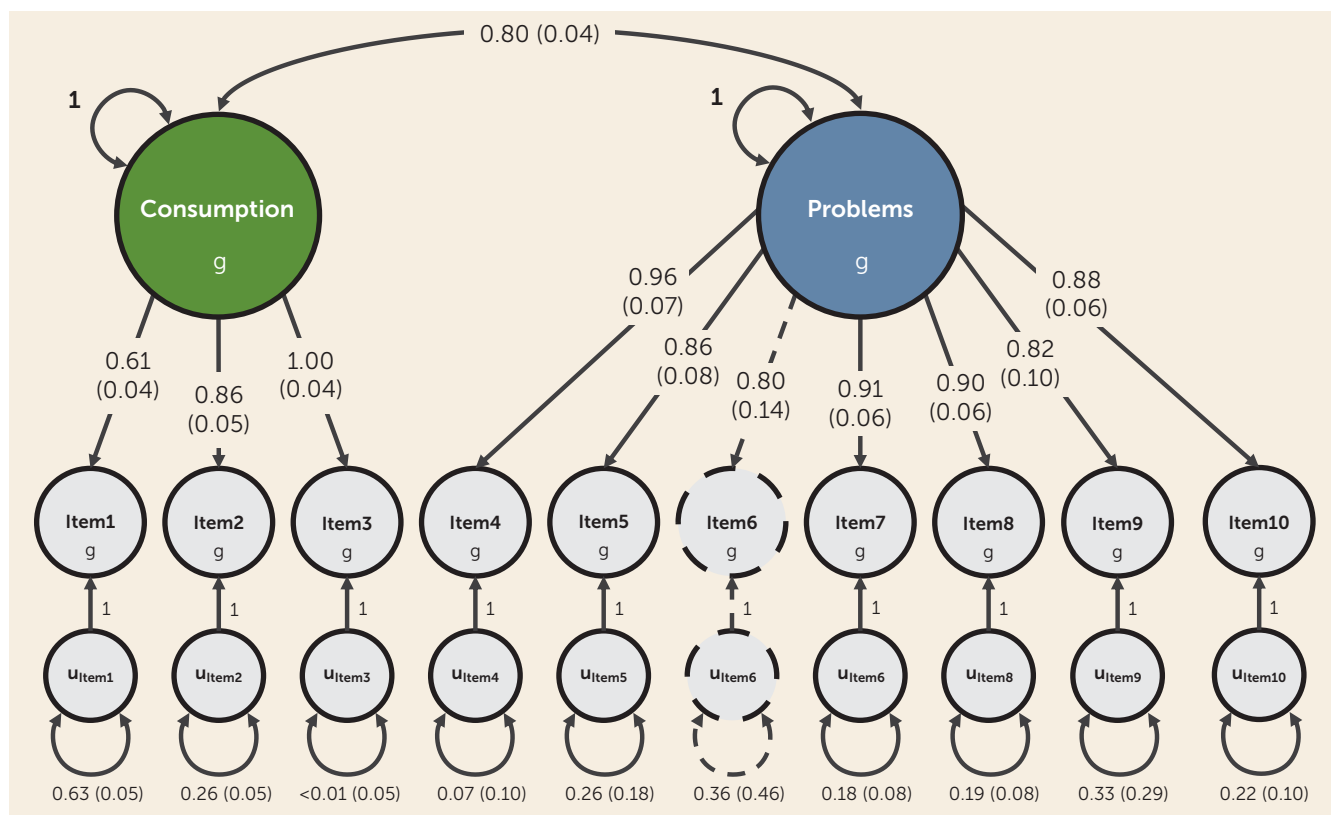
moderately larger than the phenotypic correlations (mean absolute difference=0.198), an effect that was driven by stronger genetic correlations among items 4 through 10 (the problematic alcohol use phenotypes). Of note, all AUDIT items exhibited significant SNP heritability, with the exception of item 6 (see Table S5 in the online supplement). We suspect this may be attributable to the low rates of endorsement for the item in all three cohorts (see Table S2 in the online supplement). For this reason, we excluded item 6 from all subsequent analyses, and a factor extension analysis was used to estimate its expected factor loading in the final model.

We found that a correlated factors model provided the best fit to both the genetic and the phenotypic covariance matrices (phenotypic model: $\chi^2=4252.963$, $df=26$, comparative fit index=0.994, standardized root mean square residual=0.041; genetic model: $\chi^2=142.689$, $df=26$, comparative fit index=0.982, standardized root mean square residual=0.067) (Figure 1; see also Tables S6 and S7 in the online supplement). That is, the patterns of genetic and phenotypic correlations among the AUDIT items could both be represented by a factor model with two correlated factors: one that captured the covariance among alcohol consumption items (items 1–3) and one that captured the covariance among alcohol-related problems (items 4–10). These two latent factors were highly correlated with each other, phenotypically ($r_p=0.825$, $SE=0.002$) and genetically ($r_g=0.801$, $SE=0.037$). Nearly all items had large factor loadings across both levels of analyses except item 1, which consistently had markedly smaller factor loadings and larger residual variances.

The two correlated factors model was compared with other solutions. A model with a single common factor provided acceptable fit for the phenotypic ($\chi^2=14,967.064$, $df=27$, comparative fit index=0.978, standardized root mean square residual=0.070) and genetic ($\chi^2=350.785$, $df=27$, comparative fit index=0.949, standardized root mean square residual=0.094) factor analyses, but it did not minimize the standardized difference between the observed and predicted correlations as well as the correlated factors model (see Table S7 in the online supplement). The parallel factor model (i.e., the sum-score model) exhibited poor fit, reflected by the strong, unanimous bias observed in the model-implied correlations (phenotypic model: $\chi^2=43,655.530$, $df=34$, comparative fit index=0.936, standardized root mean square residual=0.143; genetic model: $\chi^2=607.196$, $df=43$, comparative fit index=0.911, standardized root mean square residual=0.470). Accordingly, we identified the correlated factors model as the best fitting and most appropriate model for further genetic analyses.

Latent Variable Approach Characterizes and Ameliorates Bias in GWAS of Alcohol Consumption

By estimating genetic correlations in a genomic structural equation modeling framework, we identified interesting patterns of relationships between 100 exogenous phenotypes (chosen based on previous findings or hypothesized

FIGURE 1. Genetic relationships between AUDIT items^a

^a The figure is a path diagram of the best-fitting genetic confirmatory factor model for the Alcohol Use Disorders Identification Test (AUDIT), as estimated with genomic structural equation modeling. All parameter estimates are standardized, and standard errors are presented in parentheses. The genetic components of items and factors (denoted by *g*) are inferred variables that are represented as circles. Regression relationships between variables are represented as straight one-headed arrows pointing from the independent variable(s) to the dependent variable(s). Covariance relationships are depicted as curved two-headed arrows linking two variables. The variances for factors are represented as a two-headed arrow connecting the variable to itself, as are the residual variances for individual items (denoted by *u*). As item 6 was included via factor extension, its parameter estimates are illustrated using dashed lines.

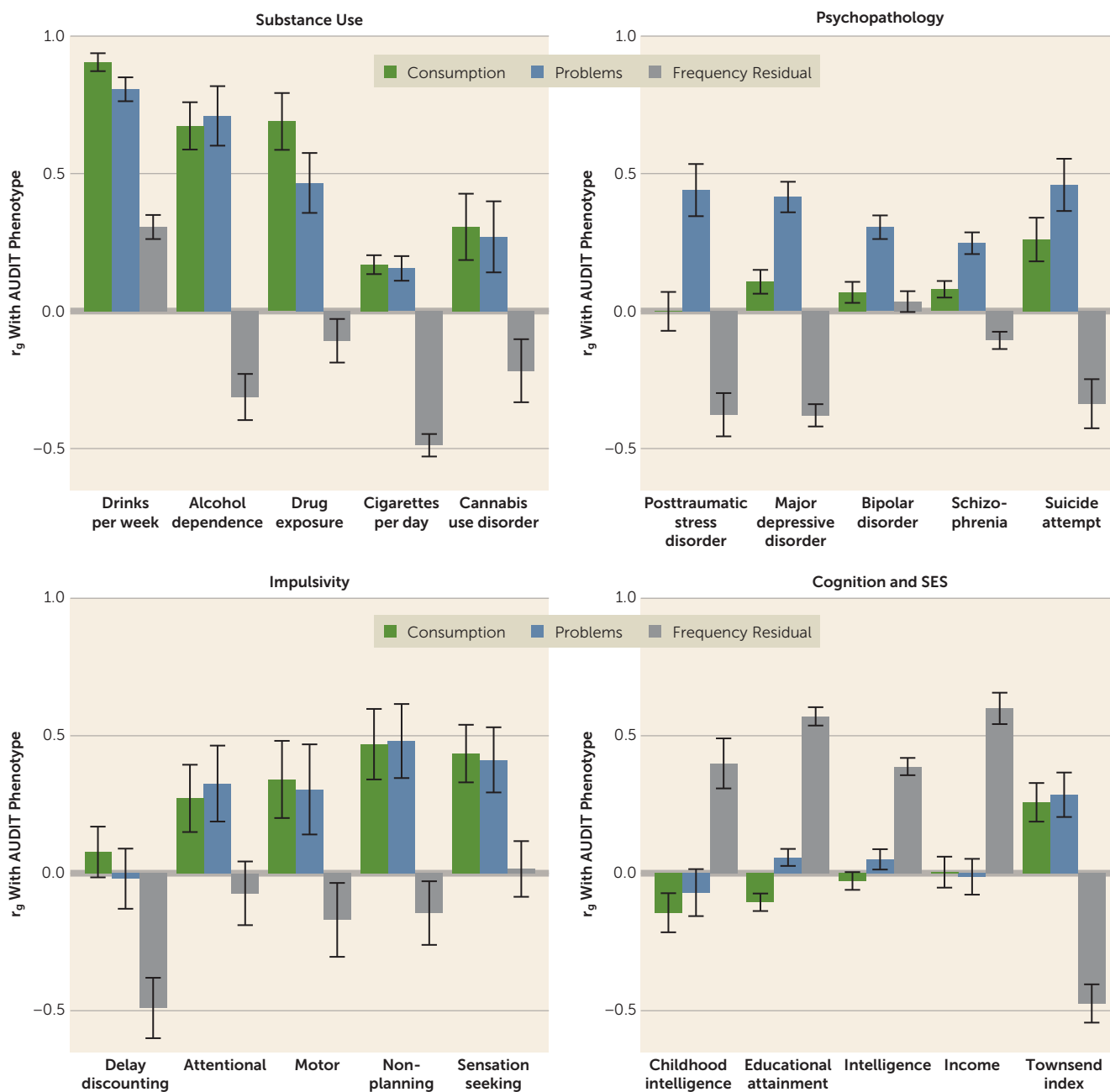
relationships) and the consumption and problems latent genetic factors. We also examined correlations with the residual genetic variance in item 1 (i.e., the genetic variance in item 1 that is unrelated to other AUDIT items; henceforth the “frequency residual”). Results are reported in Table S8 in the online supplement.

For the consumption and problems factors, we found that their patterns of genetic correlation with other phenotypes were much more similar than previously reported for AUDIT-C and AUDIT-P (4). Both the consumption and problems factors showed strong positive genetic correlations with alcohol dependence. The consumption and problems factors were also positively related to other measures of substance use (e.g., cannabis use disorder, impulsivity). Furthermore, the previous positive associations that we observed between AUDIT and indices of socioeconomic status (e.g., educational attainment) were now attenuated.

We did still observe that, compared with the consumption factor, the problems factor was more strongly related to psychopathology (e.g., posttraumatic stress disorder,

depression, bipolar disorder, schizophrenia). We also identified novel divergent associations with pain phenotypes, malnutrition, and measures of social satisfaction (e.g., the problems factor showing genetic overlap with these conditions), suggesting that, as we anticipated, the genetic contributions to alcohol consumption and misuse reflect both complementary and distinct genetic factors.

Finally, the frequency residual was negatively associated with alcohol dependence (Figure 2). We also found positive genetic correlations between the frequency residual and socioeconomic outcomes, including educational attainment, household income, and intelligence. Furthermore, we observed consistently negative genetic correlations between the frequency residual and other psychiatric and substance use disorders, such as major depressive disorder and cannabis use disorder. This result suggests that many of the puzzling genetic correlations previously reported for alcohol consumption were driven by variance related to socially stratified differences in behavior rather than variance related to the alcohol phenotypes of clinical interest.

FIGURE 2. Genetic correlations between latent AUDIT phenotypes and other complex traits^a

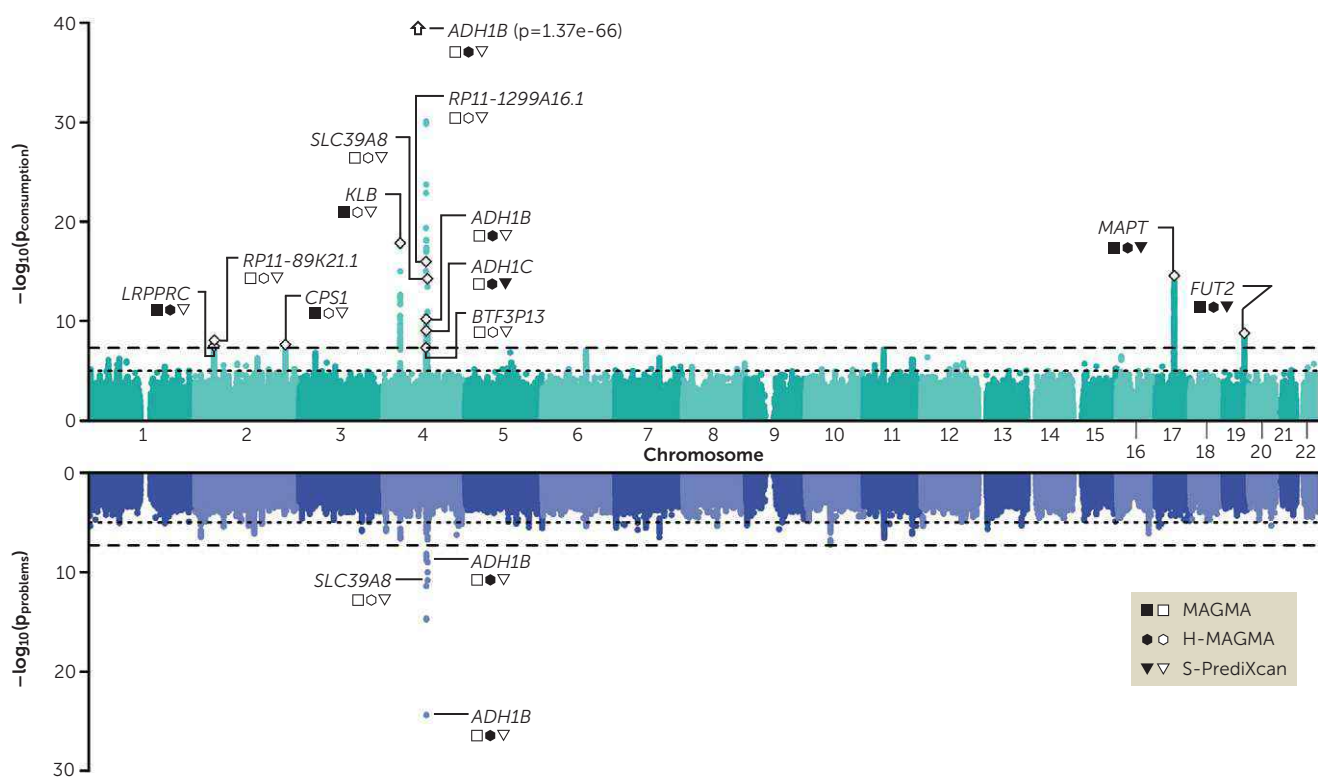
^a The bar charts show the genetic correlation (r_g) results for three Alcohol Use Disorders Identification Test (AUDIT) phenotypes: consumption, problems, and frequency residual. Point estimates and corresponding standard errors are displayed for selected phenotypes related to substance use, psychopathology, impulsivity, cognition, and socioeconomic factors. Full results are reported in Table S8 in the online supplement. SES=socioeconomic status.

Multivariate GWAS Confirms a Distinct Genetic Basis Between Alcohol Consumption and Misuse

The results of our multivariate GWAS for the consumption and problems factors are presented in Figure 3. We identified eight independent loci that were associated with the consumption factor (see Table S9 in the online supplement). For the problems factor, we replicated two loci on chromosome 4, located in the ethanol metabolizing gene

ADH1B (see Table S10 in the online supplement). The signal associated with the latent factors is convergent with that of the sum scores, with a few exceptions (see section 6.1.1 and Tables S11 and S12 in the online supplement).

Some loci included genes that were only associated with the consumption factor (see Table S31 in the online supplement). For example, *KLB*, *RCF1*, and the *MAPT/CRHR1* region, which were previously associated with

FIGURE 3. Multivariate genome-wide association analyses for the latent genetic factors^a

^a The figure is a Miami plot for the two latent genetic factors: consumption (top) and problems (bottom). Independent lead single-nucleotide polymorphisms (SNPs) are labeled with a white diamond. For each lead SNP, the nearest gene is labeled. Additional symbols convey findings from additional biological annotation; filled symbols indicate that the gene was identified in the corresponding pipeline, and empty symbols indicate that the gene was not. The y-axis refers to the significance on a $-\log_{10}$ scale, the x-axis refers to chromosomal position, the horizontal dotted line marks suggestive significance ($p=1 \times 10^{-5}$), and the horizontal dashed line denotes genome-wide significance ($p=5 \times 10^{-8}$).

alcohol consumption behaviors (3–5, 31), were only associated with the consumption factor. We also identified other novel candidate genes for alcohol consumption, such as *CPS1*, which has previously been associated with metabolic conditions (see Table S13 in the online supplement).

We performed in silico gene-based and transcriptome-based analyses (see Tables S15–S30 in the online supplement), which revealed both convergent and divergent associations for the consumption and problems factors (see Table S31 in the online supplement). For example, both factors robustly implicated ethanol metabolizing genes (*ADH1B*, *ADH1C*) and dopamine transmission (*DRD2*, involved in mediating the rewarding effects of drugs [32]), as well as pleiotropic genes previously implicated in anthropometric and metabolic traits (e.g., *CELFI* [5, 33]), and intelligence (e.g., *MTCH2* [34], *FAM180B/NDUFS3* [35]).

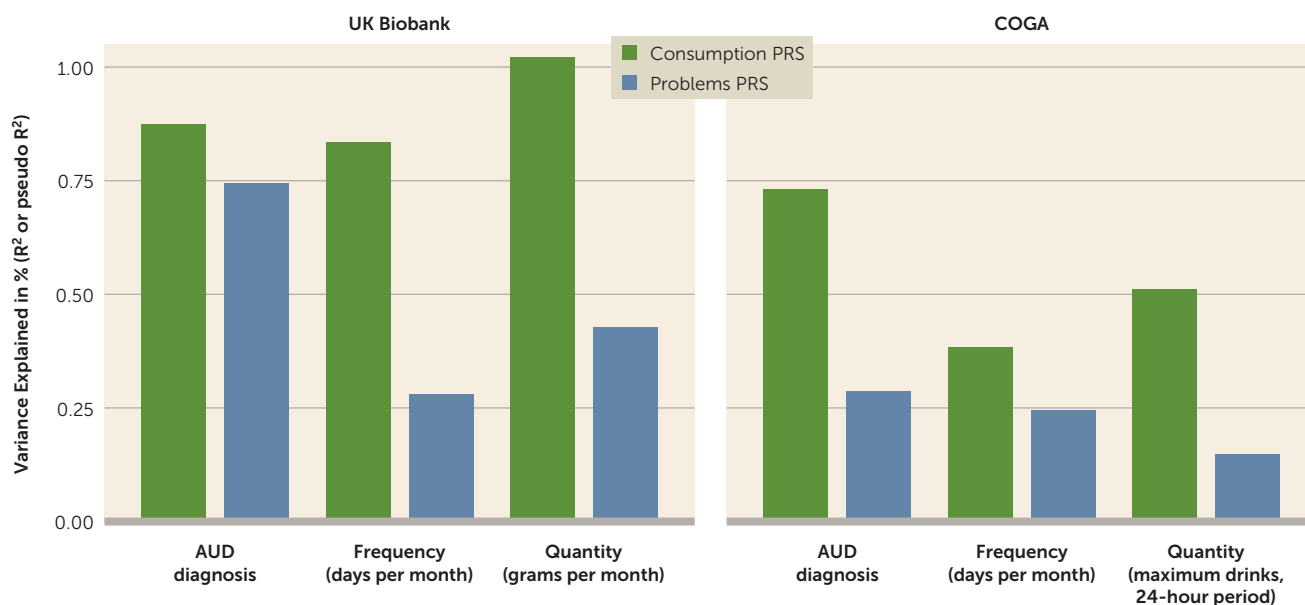
Lastly, gene-set analyses revealed that genes more closely linked to cellular responses to alcohol drinking (e.g., cellular response to retinoic acid) were associated with the consumption factor (see Table S17 in the online supplement), while the gene sets related to postsynaptic modulation of

chemical synaptic transmission were associated with the problems factor (see Table S18 in the online supplement).

Polygenic Risk Analyses

UK Biobank. In UK Biobank, we found that both the consumption and problems PRSs were robustly associated with drinking frequency, drinking quantity, and lifetime AUD (Figure 4). However, the consumption PRS outperformed (i.e., explained more variance) the problems PRS for alcohol consumption phenotypes (see Table S32 in the online supplement). When the latent-factor PRSs and sum-score PRSs for the same construct were both included in the multiple regression model (e.g., the consumption and AUDIT-C PRSs), the consumption PRS outperformed the AUDIT-C PRS in predicting AUD diagnosis and drinking quantity (but not frequency), while the AUDIT-P PRS outperformed the problems PRS across all three phenotypes (see Table S33 in the online supplement).

COGA. In COGA, PRS results aligned with those observed in UK Biobank, with a few exceptions. When both the consumption and problems PRSs were included in the same

FIGURE 4. Associations between consumption and problems PRSs and selected alcohol-related phenotypes^a

^a The bar charts show the variance explained, in percent, by the consumption and problems polygenic risk scores (PRSs) for various clinical and quantitative measures of alcohol use. Values correspond to the proportion of variance explained by the outcome (R^2 or pseudo R^2 , depending on the use of linear or logistic regression; see section 7 in the online supplement for details). Results for the independent UK Biobank subsample and the independent Collaborative Study on the Genetics of Alcoholism (COGA) cohort are presented. Note that the COGA models are not directly comparable to the UK Biobank models, as mixed-effect models were used in COGA. Please also note that because the R^2 for each PRS is calculated from a single PRS model, the values are not independent (because of shared variance between PRSs). Complete results are available in Tables S32–S35 in the online supplement. AUD=alcohol use disorder.

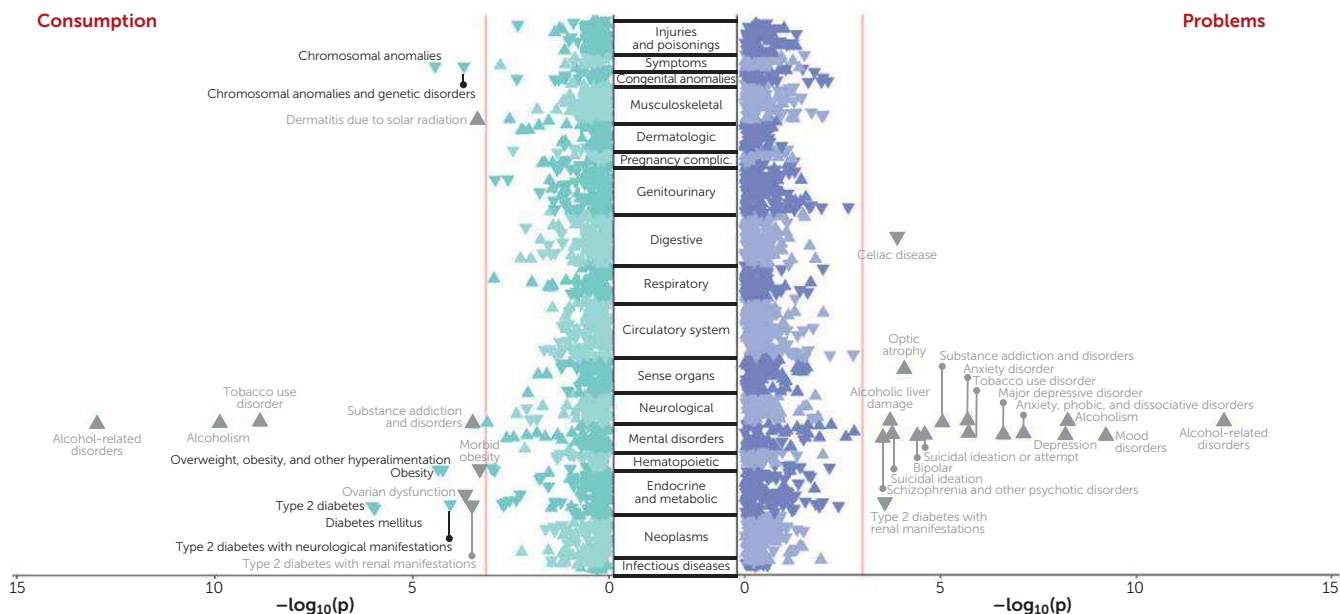
model, only the consumption PRS showed significant associations with drinks per week, maximum number of drinks per 24-hour period, and AUD (see Table S34 in the online supplement). As observed in UK Biobank, when latent-factor PRSs and sum-score PRSs for the same construct were both included in the multiple regression model, the consumption PRS outperformed the AUDIT-C PRS, and the AUDIT-P PRS outperformed the problems PRS (see Table S35 in the online supplement). Interestingly, in those models, we found that the strongest associations were between the consumption PRS and AUD and between the AUDIT-P PRS and AUD.

BioVU. We performed two independent PheWASs of the consumption and problems PRSs to identify whether these two variables would show different patterns of genetic associations with medical outcomes. Of 1,335 phenotypes, 15 were FDR-significantly associated with the consumption PRS (Figure 5; see also Table S36 in the online supplement) and 17 with the problems PRS (see Table S37 in the online supplement). Both PRSs were significantly associated with AUD and other tobacco and substance use disorders. Replicating our previous results for AUDIT-C and AUDIT-P, we observed paradoxical negative associations between the consumption PRS and metabolic conditions, including diabetes mellitus and obesity phenotypes, whereas the problems PRS was primarily positively associated with other psychiatric

disorders, including depression, anxiety disorder, bipolar disorder, schizophrenia, and suicidal ideation or attempt. Intriguingly, the problems PRS was also negatively associated with type 2 diabetes with renal manifestations. Most of the associations did not persist after correcting for AUD, although the direction of effects remained consistent (see Tables S38 and S39 in the online supplement).

DISCUSSION

In this study, we performed the first item-level and the largest GWAS of AUDIT to date ($N=160,824$), and we used genomic structural equation modeling to elucidate the genetic etiology of alcohol consumption and problematic alcohol use. By conducting phenotypic and genetic factor analyses 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 across both levels of analysis. Moreover, by applying empirically derived weights to the AUDIT items in a genomic structural equation modeling framework, we demonstrated that our method can ameliorate confounding biases that have complicated previous work with consumption phenotypes (in particular, the bias present in item 1). Notably, both the consumption and problems factors share a strong positive genetic correlation with alcohol dependence (both r_g values ~ 0.7), and

FIGURE 5. Results of a phenome-wide association study of polygenic risk scores for the consumption and problems PRSs against 1,338 diseases^a

^a Data on diseases were drawn from those available in the biobank from Vanderbilt University Medical Center, BioVU. Phenome-wide association study of both the consumption and problems polygenic risk scores (PRSs) revealed positive genetic associations with alcohol use disorders. The problems PRS was positively genetically associated with multiple psychiatric conditions, whereas the consumption PRS was counterintuitively negatively associated with metabolic conditions. Notably, most of the associations disappear after adjustment for alcohol use disorder diagnosis (nonsignificant associations are highlighted in gray).

we show, for the first time, that the polygenic signal of the consumption factor is strongly associated with several AUD phenotypes in three independent cohorts. Finally, the results of our bioinformatic analyses further illustrate that the consumption and problems factors 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 in several ways, as we delineate below.

First, we built on recent investigations of the genetic etiology of AUD and related traits by analyzing each of the 10 unique items that comprise AUDIT. At this higher resolution, we were able to identify sources of genetic heterogeneity among the items, such as the consistently weaker genetic correlations between frequency of alcohol consumption (item 1) and other drinking patterns (items 2–3) and AUD symptoms (items 4–10). Our item-level approach also allowed us to empirically model the genetic relationships between AUDIT items, providing the first empirical evidence of a correlated, two-factor structure for AUD symptoms at the genetic level. In doing so, we also generated empirically derived weights to determine how individual items contribute to aggregate measures of alcohol consumption and problematic use. This is an important advance from most quantitative or dimensional genetic studies of AUD (and other forms of psychopathology), which often use composite score measures that lack statistical justification.

Second, and perhaps most importantly, we found that the consumption factor was a good genetic proxy of AUD when appropriate weights were applied to the individual items using genomic structural equation modeling. This is a striking change from previous investigations into the divergent genetic bases of alcohol consumption and problematic use, including our own prior analyses of AUDIT. GWASs of alcohol consumption phenotypes have consistently reported low to moderate overlap with AUD, which has surprised many researchers (2–5), and even paradoxical negative associations with a variety of diseases and disorders. Our multivariate approach has ameliorated these issues, producing an aggregate measure of alcohol consumption that is more consistent with the known patterns of alcohol phenotype associations established in the literature, such as a strong genetic correlation with alcohol dependence. Furthermore, we used genetic correlation analyses to characterize the residual genetic variance in frequency of consumption (frequency residual) that is unrelated to other AUDIT items. These analyses revealed that the frequency residual had consistently positive associations with measures of socioeconomic status and consistently negative associations with measures of substance use and psychopathology. Indeed, these genetic correlations are very similar to those observed in GWASs of AUDIT-C (4, 5) and other GWASs of alcohol consumption (3, 4), suggesting that single-item frequency-based measures of alcohol consumption may be particularly susceptible to confounding and/or selection bias. For

example, Marees et al. (36) reported that greater frequency of alcohol consumption was associated with higher socioeconomic status and lower risk of other psychiatric and substance use disorders in UK Biobank. In population-based cohorts with a “healthy volunteer” bias, such as the UK Biobank, the relationship between frequency of alcohol consumption and aspects of physical and mental health may not be fully generalizable (37). This degree of bias, we speculate, will likely vary from population to population.

Third, we confirmed that the genetic contributions to alcohol consumption are partially distinct from those pertaining to problematic consequences of alcohol use. In silico analyses revealed the value of dissecting the two phenotypes, as gene- and transcriptome-based analyses identified partially divergent biological mechanisms for the consumption and problems factors. For example, the corticotropin receptor gene (*CRHRI*), which has been associated with alcohol use in animals and humans (38, 39), was associated with consumption only. As a result, we are now beginning to uncover genetic signals for aspects of alcohol involvement that have the potential to be further analyzed at the molecular, cellular, and circuit levels in cellular and animal model systems.

Fourth, we found that the consumption PRS was strongly associated with AUD even in higher-risk cohorts like COGA. This demonstrates the important downstream effects of allowing items to have different weights in phenotype construction. Whereas our current and previous PRSs for AUDIT-C have been disproportionately influenced by a single item (frequency of consumption) (40), our consumption PRS was composed of the genetic effects shared among all consumption-focused items. The consumption and problems PRSs were both strongly associated with AUD in UK Biobank, even when both scores were entered in the same model. In COGA, both the consumption and problems PRSs were associated with AUD, but the consumption PRS was more strongly associated than the problems PRS. The increased influence of binge drinking (item 3), which had a large factor loading on the consumption factor, may be partially responsible for these stronger associations in a high-risk sample. However, it is perhaps more likely that these differences might be simply explained by differences in item endorsement and thus predictive power of the discovery GWASs (e.g., the consumption factor had a greater mean chi-square than the problems factor).

Finally, our comprehensive PheWAS analyses have linked different facets of AUD liability (via the latent factor-based consumption and problems PRSs) to a myriad of health-related outcomes in a large, independent biobank. We found that the consumption PRS was consistently negatively associated with a broad range of metabolic and congenital conditions. While it is possible that there is still residual bias in the discovery GWAS, it is important to note that this pattern of paradoxical associations with consumption is not observed in the genetic correlation analyses. Thus, it is possible that these negative associations are illustrative of selection

bias or other confounding in BioVU (41), where patients with certain conditions may elect not to drink because of unmeasured factors (e.g., family history, medical advice, contraindications for prescribed medications). Mirroring the genetic correlation results, we also found that the problems PRS was uniquely associated with numerous psychiatric disorders that are commonly reported to co-occur with AUD. However, we determined that the associations between problems PRS and mental health did not persist in the absence of the clinical manifestation of AUD. These findings suggest that the associations with mental health are not a result of horizontal pleiotropy. Instead, they may be a consequence of AUD, be correlated with other risk factors for AUD (along with and/or aside from genetic risk), or be 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.

These findings should be interpreted in light of several limitations. AUDIT is a self-report measure that can be influenced by misreporting, and it only captures alcohol use in the past year, so it can be influenced by longitudinal changes in drinking that may be a consequence of, for example, other illnesses (42). People who stopped drinking or who never drank may represent genetically distinct groups; in our data set, 4,511 individuals were never drinkers, and 4,290 were previous drinkers. While our approach has substantially reduced bias in AUDIT without excluding any individuals from discovery, future studies might consider employing multiple techniques (e.g., separate never drinkers from former drinkers) to further alleviate potential biases associated with frequency of alcohol use in population-based cohorts. Additionally, while the AUDIT PRSs tended to perform similarly in UK Biobank and COGA, the portability of PRSs can be influenced by demographic characteristics such as socioeconomic status, age, and sex (43). It remains to be determined how generalizable the genetics of AUDIT are across different populations, especially in samples of different ancestries (as we included only individuals of European ancestry in the present study) or cultures (e.g., United Kingdom versus United States). A similar point also applies to sex-stratified samples, considering that AUDIT scores differ in men and women. Finally, it is important to note that the problems PRS exhibited weaker associations with AUD and other alcohol phenotypes in comparison to its AUDIT-P counterpart. Although the two predictors generally had similar effects in single-PRS models, the problems PRS was rendered redundant in the cross-method analyses when both of the highly correlated AUDIT-P and problems PRSs (e.g., $r=0.84$ in UK Biobank) were included in the regression models. However, we caution against the interpretation that the univariate GWAS approach is preferable. The multivariate GWAS function of genomic structural equation modeling is not only more flexible than traditional univariate GWAS, but its results may be more robust to confounding, as the software automatically applies a correction for population stratification (9). Furthermore, genomic

structural equation modeling is better suited to investigate nuanced genetic influences, including the possibility of identifying SNPs with heterogeneous effects across symptoms or items.

Analyzing alternative phenotypes as a complementary approach to studying clinically defined AUD, and psychiatric disorders in general, has generated considerable interest in recent years (44). Collectively, our work demonstrates how AUDIT can inexpensively facilitate such efforts. Here, we have shown that, after correcting for some potential biases, item- or symptom-level analyses can help unpack the genetic etiology of AUD by breaking down genetic influences into specific and shared components; notably, this is possible only because we can contrast our results against gold-standard, clinically ascertained AUD GWAS data sets. While composite scores have shown some utility in previous genetic association studies, such studies often rely on strong assumptions that the scale is unidimensional and that each item is equally informative of the construct being measured. In this study, we have shown that the latter assumption is false for AUDIT. In particular, a large proportion of the genetic variance of item 1 appears to be uninformative about a broader consumption construct, as it is related to socially stratified differences in behavior rather than the alcohol phenotypes of clinical interest. Moreover, although we found a notable degree of unidimensionality among the AUDIT items, our results demonstrate that the consumption and problems factors remain distinct in their associations with health.

AUTHOR AND ARTICLE INFORMATION

Department of Psychology, University of Texas at Austin (Mallard, Grotzinger, Harden); Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam (Savage); Department of Psychiatry, Washington University School of Medicine, Saint Louis (Johnson, Anokhin, Agrawal); Department of Psychiatry (Huang, Jennings, Palmer, Sanchez-Roige) and Institute for Genomic Medicine (Palmer), University of California San Diego, La Jolla; Virginia Institute for Psychiatric and Behavioral Genetics, Richmond (Edwards); Department of Biological Psychiatry, Vrije Universiteit Amsterdam (Hottenga, Nivard, de Geus, Boomsma); Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, Tenn. (Gustavson, Davis, Sanchez-Roige); Department of Psychology, Virginia Commonwealth University, Richmond (Dick); Department of Biochemistry and Molecular Biology (Edenberg) and Department of Medical and Molecular Genetics (Lai), Indiana University School of Medicine, Indianapolis; Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City (Kramer); Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, N.Y. (Meyers, Pandey); Department of Psychiatry and Behavioral Sciences and Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tenn. (Davis); Division of Psychiatry, University of Edinburgh, Edinburgh, U.K. (Clarke).

Send correspondence to Dr. Sanchez-Roige (sanchezroige@ucsd.edu).

Data availability: The GWAS summary statistics for each latent AUDIT factor and the sum-score counterparts (AUDIT-C and AUDIT-P) will be made available on the PGC website.

This research was conducted using the UK Biobank Resource under application numbers 11425 and 16406. The AUDIT data collection in

ALSPAC was funded by NIH grant AA018333. 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). For the Netherlands Twin Register, funding was obtained from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMW) grants 904-61-090, 985-10-002, 912-10-020, 904-61-193,480-04-004, 463-06-001, 451-04-034, 400-05-717, Addiction-31160008, 016-115-035, 481-08-011, 400-07-080, 056-32-010, Middelgroot-911-09-032, OCW_NWO Gravity program 024.001.003, NWO-Groot 480-15-001/674, Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL, 184.021.007 and 184.033.111), X-Omics 184-034-019; Spinozapremie (NWO 56-464-14192), KNAW Academy Professor Award (PAH/6635) and University Research Fellow grant (URF) to Dr. Boomsma, Amsterdam Public Health research institute (former EMGO+), Neuroscience Amsterdam research institute (former NCA), the European Community's Fifth and Seventh Framework Program (FP5-LIFE QUALITY-CT-2002-2006, FP7-HEALTH-F4-2007-2013, grant 01254: GenomEUtwin, grant 01413: ENGAGE and grant 602768: ACTION), the European Research Council (ERC Starting 284167, ERC Consolidator 771057, ERC Advanced 230374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), NIH (grants R01D0042157-01A1, R01MH58799-03, MH081802, DA018673, R01 DK092127-04, Grand Opportunity grants 1RC2 MH089951, and 1RC2 MH089995), and the Avera Institute for Human Genetics, Sioux Falls, S.D. Part of the genotyping and analyses was funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health. Computing was supported by NWO through grant 2018/EW/00408559, BiG Grid, the Dutch e-Science Grid, and SURFsara.

PRS analyses using UK Biobank data were carried out on the Genetic Cluster Computer hosted by the Dutch National computing and Networking Services SURFsara. The data sets used for the PheWAS analyses described were obtained from Vanderbilt University Medical Center's BioVU, which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH-funded Shared Instrumentation Grant S10RR025141 and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, and R01HD074711; and additional funding sources listed at <https://vict.vumc.org/biovu-funding>.

Dr. Johnson was supported by funding from NIAAA grant F32AA027435. Ms. Huang, Ms. Jennings, Dr. Palmer, and Dr. Sanchez-Roige were supported by funds from the California Tobacco-Related Disease Research Program (grants 28IR-0070 and T29KT0526). Dr. Edwards was supported by NIAAA grant R01AA027522. Dr. Anokhin was supported by funding from NIH grants K02 DA32573, MH109532, and U10AA008401. Dr. Edenberg and Dr. Kramer were supported by NIAAA grant U10AA008401. Dr. Nivard is supported by NIMH grant R01MH120219, ZonMW grants 849200011 and 531003014 from the Netherlands Organization for Health Research and Development, and a VENI grant awarded by NWO (VI.Veni.191G.030), and he is a Jacobs Foundation Fellow. Dr. Davis obtained support from NIMH grants 1R01MH113362, 1R01MH118233, and 1R56MH120736. Dr. Sanchez-Roige was supported by a NARSAD Young Investigator Award from the Brain and Behavior Foundation (grant 27676).

The Collaborative Study on the Genetics of Alcoholism (COGA): principal investigators, B. Porjesz, V. Hesselbrock, T. Foroud; scientific director, A. Agrawal; translational director, D. Dick. The study includes 11 different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, T. Foroud, J. Nurnberger Jr., Y. Liu); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz, J. Meyers, C. Kamarajan, A. Pandey); 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, R. Hart); Children's Hospital of Philadelphia, University of Pennsylvania (L. Al-masy); Virginia Commonwealth University (D. Dick, J. Salvatore); Icahn School of Medicine at Mount Sinai (A. Goate, M. Kapoor, P. Slesinger); and Howard University (D. Scott). Other COGA collaborators include L. Bauer (University of Connecticut); L. Wetherill, X. Xuei, D. Lai, S. O'Connor, M. Plawecki, S. Lourens (Indiana University); L. Acion (University of Iowa); G. Chan (University of Iowa; University of Connecticut); D.B. Chorlian, J. Zhang, S. Kinreich, G. Pandey (SUNY Downstate); M. Chao (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, S. Saccone (Washington University); F. Aliev, P. Barr (Virginia Commonwealth University). H. Chin and A. Parsian are the NIAAA staff collaborators. The authors also owe a debt of gratitude to Henri Begleiter and Theodore Reich, founding principal investigator and co-principal investigator of COGA, and other past organizers of COGA, including Ting-Kai Li, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions.

The authors report no financial relationships with commercial interests.

Received September 21, 2020; revisions received January 13 and February 16, 2021; accepted February 19, 2021; published online May 14, 2021.

REFERENCES

- Sanchez-Roige S, Palmer AA, Clarke T-K: Recent efforts to dissect the genetic basis of alcohol use and abuse. *Biol Psychiatry* 2020; 87:609–618
- Walters RK, Polimanti R, Johnson EC, et al: Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* 2018; 21:1656–1669
- Liu M, Jiang Y, Wedow R, et al: Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019; 51:237–244
- Kranzler HR, Zhou H, Kember RL, et al: Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun* 2019; 10:1499
- Zhou H, Sealock JM, Sanchez-Roige S, et al: Meta-analysis of problematic alcohol use in 435,563 individuals identifies 29 risk variants and yields insights into biology, pleiotropy, and causality. *Nat Neurosci* 2020; 23:809–818
- Sanchez-Roige S, Palmer AA, Fontanillas P, et al: Genome-wide association study meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) in two population-based cohorts. *Am J Psychiatry* 2019; 176:107–118
- Saunders JB, Aasland OG, Babor TF, et al: Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons With Harmful Alcohol Consumption—II. *Addiction* 1993; 88:791–804
- Litten RZ, Ryan ML, Falk DE, et al: Heterogeneity of alcohol use disorder: understanding mechanisms to advance personalized treatment. *Alcohol Clin Exp Res* 2015; 39:579–584
- Grotzinger AD, Rhemtulla M, de Vlaming R, et al: Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* 2019; 3:513–525
- Boyd A, Golding J, Macleod J, et al: Cohort profile: the “children of the 90s”: the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013; 42:111–127
- Boomsma DI, Vink JM, van Beijsterveldt TCEM, et al: Netherlands Twin Register: a focus on longitudinal research. *Twin Res* 2002; 5:401–406
- Bycroft C, Freeman C, Petkova D, et al: The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; 562:203–209
- Loh P-R, Tucker G, Bulik-Sullivan BK, et al: Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015; 47:284–290
- Jiang L, Zheng Z, Qi T, et al: A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* 2019; 51:1749–1755
- Chang CC, Chow CC, Tellier LC, et al: Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4:7
- Mallard TT, Linnér RK, Grotzinger AD, et al: Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. *BioRxiv*, September 8, 2020 (doi: <https://doi.org/10.1101/603134>)
- Willer CJ, Li Y, Abecasis GR: METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26:2190–2191
- Rossee Y: lavaan: an R package for structural equation modeling. *J Stat Softw* 2012; 48:1–36
- Bulik-Sullivan BK, Loh P-R, Finucane HK, et al: LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; 47:291–295
- Linnér RK, Mallard TT, Barr PB, et al: Multivariate genomic analysis of 1.5 million people identifies genes related to addiction, antisocial behavior, and health. *Nat Neurosci* (in press)
- Cross-Disorder Group of the Psychiatric Genomics Consortium: Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* 2019; 179:1469–1482.e11
- Watanabe K, Taskesen E, van Bochoven A, et al: Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; 8:1826
- de Leeuw CA, Mooij JM, Heskes T, et al: MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput Biol* 2015; 11:e1004219
- Sey NYA, Hu B, Mah W, et al: A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci* 2020; 23:583–593
- Barbeira AN, Dickinson SP, Bonazzola R, et al: Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018; 9:1825
- Ge T, Chen C-Y, Ni Y, et al: Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019; 10:1776
- The Collaborative Study on the Genetics of Alcoholism. *Alcohol Health Res World* 1995; 19:228–236
- Messer LC, Laraia BA, Kaufman JS, et al: The development of a standardized neighborhood deprivation index. *J Urban Health* 2006; 83:1041–1062
- Dennis J, Sealock J, Levinson RT, et al: Genetic risk for major depressive disorder and loneliness in sex-specific associations with coronary artery disease. *Mol Psychiatry* 2019 (<https://doi.org/10.1038/s41380-019-0614-y>)
- Carroll RJ, Bastarache L, Denny JC: R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* 2014; 30:2375–2376
- Evangelou E, Gao H, Chu C, et al: New alcohol-related genes suggest shared genetic mechanisms with neuropsychiatric disorders. *Nat Hum Behav* 2019; 3:950–961
- Volkow ND, Morales M: The brain on drugs: from reward to addiction. *Cell* 2015; 162:712–725
- Hinney A, Albayrak O, Antel J, et al: Genetic variation at the CELF1 (CUGBP, elav-like family member 1 gene) locus is genome-wide associated with Alzheimer's disease and obesity. *Am J Med Genet B Neuropsychiatr Genet* 2014; 165B:283–293
- Davies G, Marioni RE, Liewald DC, et al: Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry* 2016; 21: 758–767
- Savage JE, Jansen PR, Stringer S, et al: Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 2018; 50:912–919

36. Marees AT, Smit DJA, Ong J-S, et al: Potential influence of socioeconomic status on genetic correlations between alcohol consumption measures and mental health. *Psychol Med* 2020; 50:484–498
37. Fry A, Littlejohns TJ, Sudlow C, et al: Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; 186:1026–1034
38. Gelernter J, Sun N, Polimanti R, et al: Genome-wide association study of maximum habitual alcohol intake in >140,000 US European and African American veterans yields novel risk loci. *Biol Psychiatry* 2019; 86:365–376
39. Zorrilla EP, Logrip ML, Koob GF: Corticotropin releasing factor: a key role in the neurobiology of addiction. *Front Neuroendocrinol* 2014; 35:234–244
40. Johnson EC, Sanchez-Roige S, Acion L, et al: Polygenic contributions to alcohol use and alcohol use disorders across population-based and clinically ascertained samples. *Psychol Med* (Online ahead of print, January 20, 2020)
41. Munafò MR, Tilling K, Taylor AE, et al: Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2018; 47:226–235
42. Xue A, Jiang L, Zhu Z, et al: Genome-wide analyses of behavioural traits are subject to bias by misreports and longitudinal changes. *Nat Commun* 2021; 12:20211
43. Mostafavi H, Harpak A, Agarwal I, et al: Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* 2020; 9:e48376
44. Sanchez-Roige S, Palmer AA: Emerging phenotyping strategies will advance our understanding of psychiatric genetics. *Nat Neurosci* 2020; 23:475–480