The Genetic Relationship Between Alcohol Consumption and Aspects of Problem Drinking in an Ascertained Sample


Background: Genomewide association studies (GWAS) have begun to identify loci related to alcohol consumption, but little is known about whether this genetic propensity overlaps with specific indices of problem drinking in ascertained samples.

Methods: In 6,731 European Americans who had been exposed to alcohol, we examined whether polygenic risk scores (PRS) from a GWAS of weekly alcohol consumption in the UK Biobank predicted variance in 6 alcohol-related phenotypes: alcohol use, maximum drinks within 24 hours (MAXD), total score on the Self-Rating of the Effects of Ethanol Questionnaire (SRE-T), DSM-IV alcohol dependence (DSM4AD), DSM-5 alcohol use disorder symptom counts (DSM5AUDSX), and reduction/cessation of problematic drinking. We also examined the extent to which a single nucleotide polymorphism (rs1229984) in ADH1B, which is strongly associated with both alcohol consumption and dependence, contributed to the polygenic association with these phenotypes and whether PRS interacted with sex, age, or family history of alcoholism to predict alcohol-related outcomes. We performed mixed-effect regression analyses, with family membership and recruitment site included as random effects, as well as survival modeling of age of onset of DSM4AD.

Results: PRS for alcohol consumption significantly predicted variance in 5 of the 6 outcomes: alcohol use (Δmarginal $R^2 = 1.39\%$, Δ area under the curve [AUC] = 0.011), DSM4AD (Δmarginal $R^2 = 0.56\%$; ΔAUC = 0.003), DSM5AUDSX (Δmarginal $R^2 = 0.49\%$), MAXD (Δmarginal $R^2 = 0.31\%$), and SRE-T (Δmarginal $R^2 = 0.22\%$). PRS were also associated with onset of DSM4AD (hazard ratio = 1.11, $p = 2.08e-5$). The inclusion of rs1229984 attenuated the effects of the alcohol consumption PRS, particularly for DSM4AD and DSM5AUDSX, but the PRS continued to exert an independent effect for all 5 alcohol measures (Δmarginal $R^2$ after controlling for ADH1B = 0.14 to 1.22%). Interactions between PRS and sex, age, or family history were nonsignificant.

Conclusions: Genetic propensity for typical alcohol consumption was associated with alcohol use and was also associated with 4 of the additional 5 outcomes, though the variance explained in this sample was modest. Future GWAS that focus on the multifaceted nature of AUD, which goes beyond consumption, might reveal additional information regarding the polygenic underpinnings of problem drinking.

Key Words: Alcohol Dependence, Alcohol Consumption, Polygenic Risk, ADH1B.

Alcohol consumption ranges from infrequent intake of alcoholic beverages, through more frequent and heavy episodic drinking (e.g., bingeing), to drinking at problematic levels that are potentially indicative of alcohol use disorders (AUDs). While there is considerable debate surrounding the putative benefits of moderate...
drinking (Stockwell et al., 2016; Wood et al., 2018), excessive alcohol consumption has been unequivocally identified as one of the top 10 contributors to worldwide morbidity and mortality (World Health Organization, 2014). Excessive alcohol consumption can escalate to AUD diagnosis when accompanied by other psychological, behavioral, and physical symptoms, such as loss of control over drinking, tolerance, withdrawal, and persistent drinking despite health problems (Wise and Koob, 2014). It is estimated that AUD-attributable costs approximate 1% of the gross domestic product of developed nations (Rehm et al., 2009).

Both alcohol consumption and AUD (as well as other indices of problematic drinking, e.g., the problem subscale of the Alcohol Use Disorders Identification Test [AUDIT-P]; Sanchez-Roige et al., 2017) are heritable. Twin studies suggest that additive genetic factors contribute to 40 to 70% of the variance in alcohol consumption and AUD (Heath and Martin, 1994; Prescott and Kendler, 1999; Verhulst et al., 2000). The degree to which there is overlap in the genetic factors contributing to alcohol consumption versus AUD is less clear from twin data (Agrawal et al., 2011; Dick et al., 2011; Grant et al., 2009; Kendler et al., 2010), but recent genome-wide association studies (GWAS) have estimated this overlap using linkage disequilibrium score regression (LDSC; Bulik-Sullivan et al., 2015). Based on GWAS of both alcohol consumption (N = 112, 117; Clarke et al., 2017) and DSM-IV alcohol dependence (DSM4AD; N = 14,904 cases and 37,944 controls; Walters et al., 2018), common variants explain, in aggregate, between 9 and 13% of the variance in these phenotypes (i.e., single nucleotide polymorphisms [SNPs] heritability) and the estimated genetic correlation between them ranges from r_g = 0.37 to 0.70 (Walters et al., 2018).

As the correlation between consumption measures and dependence is derived from composite phenotypes (e.g., alcohol dependence diagnosis using structured interviews or clinician reports, consumption via self-reports) and in heterogeneous samples, further work is needed to pinpoint which specific aspects of drinking (e.g., lifetime alcohol use vs. alcohol dependence diagnosis) are most genetically correlated with self-reported levels of alcohol consumption. To study particular features of problem drinking, ascertained samples that are enriched for liability to alcohol dependence may provide increased power through larger sample sizes as well as detailed phenotyping. However, it is likely that several factors might moderate the extent to which genetic liability for alcohol consumption in a specific discovery population, such as the older volunteer cohort that comprises the UK Biobank, is associated with problem drinking in an independent sample. One possibility is that genetic overlap might be more pronounced in a similar age-group as that comprising the discovery GWAS, possibly due to cohort-specific genetic and environmental influences. Other moderators might include sex, cultural effects, and access to alcohol, as well as family history of alcohol problems. For instance, family history has been shown in other studies to accentuate the impact of genetic vulnerability on psychiatrically relevant outcomes (e.g., Agerbo et al., 2015).

Beyond genetic correlations from LDSC, polygenic risk scores (PRS) offer an alternative approach for demonstrating genetic overlap between 2 traits (e.g., alcohol consumption and AUD). PRS represent the additive effects of independent SNPs that are weighted by their effect sizes from a “discovery” GWAS (International Schizophrenia Consortium, 2009). With this approach, every individual in the independent “target” sample is assigned a score that indexes their estimated genetic propensity to the behavior studied in the discovery GWAS. The phenotype of interest in the target sample is then regressed on the polygenic score, and the strength of this association is assessed using R^2 or other measures of predictor efficacy (e.g., area under the curve [AUC]). Although the PRS incorporates additional SNPs beyond those meeting the stringent genomewide significance threshold, PRS typically explain a very small percentage (usually <10%) of the variance in the target sample phenotype (Dudbridge, 2013; Wray et al., 2014). How much variance PRS explain is dependent upon the SNP heritability of the phenotype, the size of both the discovery GWAS and the target sample, the selection thresholds for SNPs included in the PRS, and the methods used for weighting the effect sizes (Dudbridge, 2013). The extent to which a PRS derived from a GWAS of alcohol consumption is associated with aspects of problem drinking is, therefore, an estimate of their genetic commonality, although causal processes can also be represented (Swerdlow et al., 2016).

To date, the most robust genetic signals identified for both alcohol consumption and dependence have been within ADH1B (across multiple ancestries; Bierut et al., 2012; Clarke et al., 2017; Kranzler et al., 2019; Luczak et al., 2006; Sanchez-Roige et al., 2018; Walters et al., 2018). Recent GWAS of alcohol consumption (Jorgenson et al., 2017; Liu et al., 2019; Schumann et al., 2011, 2016) have implicated loci in AUTS2, KLB, GCKR, and other genes; however, evidence for their involvement in the genetics of alcohol dependence remains limited (Sanchez-Roige et al., 2018), although a recent preprint implicated GCKR in a large GWAS of AUD (Kranzler et al., 2019). The protective allele of rs129984, the missense SNP within ADH1B that affects the conversion of ethanol (EtOH) to acetaldehyde, exerts one of the largest single-variant effects on a polygenic trait, with up to a 3-fold decrease in risk for alcohol dependence (Edenberg, 2007; Edenberg and McClintick, 2018). Thus, studies that examine the polygenic overlap between consumption and dependence should account for the role of ADH1B in aggregated genomic propensity.

In this study, we use a large (N = 6,731), deeply phenotyped target sample of European Americans who reported ever drinking alcohol (and enriched for those with alcohol dependence) and a very large (N = 112,117) discovery GWAS in a European-only volunteer cohort (the UK Biobank) to examine the genetic overlap between alcohol consumption and nonproblem as well as problematic drinking.
behavior. Our target sample consists of individuals from the Collaborative Study on the Genetics of Alcoholism (COGA) who have been assessed 1 or more times using an instrument specifically designed to evaluate risk for substance use and common psychiatric disorders. We utilized a PRS approach to examine the polygenic overlap between self-reported alcohol consumption in the UK Biobank volunteer cohort and several aspects of drinking in COGA. We aimed to leverage the unique strengths of both samples in this study to ask: (i) whether a polygenic score for alcohol consumption (average intake per week) from a population-based cohort predicts a range of drinking milestones in a sample that is enriched for familial risk for alcohol problems, (ii) whether any association between the PRS and drinking milestones is moderated by age, sex, or family history of AUDs, and (iii) whether the PRS is associated with aspects of drinking above and beyond the effect of the strongest signal for AUD, rs1229984 in ADH1B.

MATERIALS AND METHODS

Sample

The COGA currently consists of 12,145 individuals with GWAS data. The goal of COGA is to elucidate the genetic underpinnings of AUDs and problem drinking across the lifespan. The study is described in detail elsewhere (Begleiter et al., 1995; Bucholz et al., 2017). Briefly, probands were identified through primarily inpatient alcohol treatment programs at 7 U.S. sites. Probands and their family members were invited to participate if they had a sufficiently large family (usually >3 sibs with parents available) with 2 or more members in the COGA catchment area. Control families (2 parents and 3 or more offspring over the age of 14) were also selected from a variety of sources (e.g., dental clinics, driver license registries). The Institutional Review Boards at all sites approved this study, and written consent was obtained from all participants. As the PRS were derived from a sample of Europeans, only individuals identified as part of the European American subsample of COGA, determined using genomic data, were included in the analyses reported here (N = 7,645). To avoid confounding by individuals at high genetic risk who elected not to drink due to personal (e.g., religious, cultural, health) reasons, we also excluded individuals reporting no lifetime use of alcohol (N = 336). We also excluded those aged 12 to 19 (N = 578) as they may not be past the period of maximal risk for onset of alcohol use and problems, yielding a final analysis N = 6,731.

Genotyping

Genotyping for the COGA European American participants was performed using the Illumina 1M, Illumina OmniExpress, and Illumina 2.5M (Illumina, San Diego, CA), and Smokescreen (BioRealm, Walnut, CA) arrays. Array type was included as a covariate in all analyses. A pruned set of 47,000 variants that were genotyped on all platforms, had minor allele frequencies (MAFs) >10% in the combined samples, Hardy–Weinberg equilibrium (HWE) p-values > 0.001, and missing rates <2%, and were not in linkage disequilibrium (LD, defined as R^2 < 0.5) were used to assess reported pedigree structure using identity-by-descent calculations in PLINK (https://www.cog-genomics.org/plink/1.9/general_usage#cite) (Purcell et al., 2007). Family structures were altered as needed, and SNP genotypes were tested for Mendelian inconsistencies with the revised family structure (O’Connell and Weeks, 1998). Genotype inconsistencies were set to missing. Imputation was to 1,000 Genomes (European (EUR) and African ancestries (AFR), Phase 3, b37, October 2014; build hg19) using SHAPEIT2 (https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html#home) (Delaneau et al., 2012) and then Minimac3 (Das et al., 2016). Only nonpalindromic variants with missing rates <5%, minor allele frequency (MAF) >3%, and HWE p-values > 0.0001 were used for imputation. Imputed SNPs with information scores <0.30 or individual genotype probability scores <0.90 were excluded. For the final dataset for PRS construction, palindromic SNPs (A/T or C/G), monomorphic SNPs, and SNPs that did not pass HWE (p < 1e−6), and SNPs with a MAF less than 0.005 were excluded. In total, 6,881,872 SNPs passed quality control and data cleaning thresholds and were available for analysis. Genotypes that did not pass quality control prior to and upon imputation were excluded from PRS construction.

Measures

All COGA participants were interviewed at least once using a version of the Semi-Structured Assessment for the Genetics of Alcoholism (Bucholz et al., 1994; Hesselbrock et al., 1999). For these analyses, an individual’s interview with the highest lifetime report for each measure was used. Measures included the following:

Alcohol Use. Drinking at least 1 drink a month for 6 or more consecutive months at any point in the lifetime (yes/no).

Maximum Drinks (MAXD). The maximum number of drinks ever consumed in a single 24-hour period (range = 1 to 100; values >100 (N = 32) were constrained to 100). To correct for skew, this variable was transformed by taking the natural logarithm of (MAXD).

Total Score for the Self-Rating of the Effects of EtOH (SRE-T). The SRE-T score was derived from a self-report instrument administered to a smaller subset of COGA participants (N = 4,296) and assesses the participant’s response to 4 items that address the number of drinks required for the participant to feel the effects of alcohol, feel dizzy, or begin to slur in speech, to stumble, or to fall asleep. Scores across 3 time points (first 5 times the participant used alcohol, last
3 months, and period of heaviest drinking) are averaged to create the total (SRE-T) score (Schuckit et al., 1997), win-
sorized at ±2 standard deviations from the mean, and fur-
ther transformed using the square root of the value to correct for skew.

**DSM4AD.** We elected to use DSM-IV (American Psychi-
tric Association, 2000) lifetime criteria for alcohol depen-
dence, the endorsement of 3 or more of 7 dependence criteria
that clustered within a single 12-month period, because it
represents a more severe form of the diagnosis than the
DSM-5. **Age of onset of alcohol dependence** (DSM-IV) was
used in survival analyses, with those who had not met criteria
for DSM4AD censored at their age at the last interview.

**DSM-5 Alcohol Use Disorder Symptom Count
(DSM5AUDSX).** We elected to use DSM-5 criteria for
the quantitative index of AUD severity as it captures addi-
tional aspects of the disorder beyond DSM-IV dependence,
thus providing a broader scale of liability. The lifetime
endorsement of the 11 DSM-5 (American Psychiatric
Association, 2013) criteria for AUD, summed at the interview
with maximum number of symptoms, was used.

**Reduction/Cessation of Problematic Drinking.** An unor-
dered categorical measure was created to represent 3 groups:
(i) those who did not meet criteria for DSM-5 AUD during
their lifetime; (ii) individuals who had a lifetime history of
DSM-5 AUD and were current problematic drinkers due to
an active AUD diagnosis in the past 12 months or were
high-risk drinkers (defined as men: ≥5 drinks/d or ≥15 drinks
in 1 week; women: ≥4 drinks/d or ≥8 in 1 week; National
Institute on Alcohol Abuse and Alcoholism, 2004); and (iii)
those who had a lifetime history of AUD but had reduced/ ceased their drinking and either did not report any AUD cri-
teria (except craving), or were not high-risk drinkers, or were
abstinent from alcohol, all in the past 12 months (McCutch-
eon et al., 2017; Schuckit et al., 2018). A comparison of
Group A against either Group B or Group C contrasts presence or absence of a lifetime diagnosis of DSM-5 AUD,
while the comparison of Group B and Group C stratifies those with a lifetime diagnosis into high-risk drinkers, includ-
ing those with active AUD, and low-risk drinkers who may also be in abstinence remission.

**Covariates** included sex, participant’s age at their last interview, birth cohort (dummy variables representing birth years prior to 1930, 1930 to 1949, 1950 to 1969, and 1970 and after), the first 3 ancestral principal components, array
type, recruitment site, and the total number of interviews with the participant.

**Family History.** A binary measure representing whether at least one of the biological parents of the respondent had a
history of DSM-5 AUD was used. Various sources of information
were used to derive family history (e.g., parental interview, respondent report; see Bucholz et al., 2017;
McCutcheon et al., 2017, for details). Of the analytic sample, 1,656 individuals did not have a report on family history,
and 485 individuals had a missing report on 1 parent while
the other parent was confirmed to be family history negative. Thus, family history was not included as a covariate in all
analyses, and for analyses including family history (e.g.,
PRS*family history), the sample size was smaller
\(N = 5,075\).

**Statistical Analysis**

**Construction of PRS.** Effect sizes and effect alleles were
derived from a GWAS of \(N = 112,117\) unrelated European-
ancestry individuals in the UK Biobank (Clarke et al., 2017).
Participants were asked about their current drinking status
(never, previous, current, prefer not to say) and their average
weekly and monthly consumption of a variety of alcoholic
beverages (e.g., red wine, white wine, beer, spirits). An overall
measure of average alcohol intake per week was derived
from these measures. Age and weight were then regressed
onto alcohol consumption in units per week in males and
females separately, and the residuals from these regressions
were then pooled (males + females) to form the alcohol con-
sumption phenotype of interest. A GWAS was conducted
with 12,489,782 quality-controlled SNPs and UK Biobank
(UK) assessment center, genotyping batch, and 15 principal
components included as additional covariates. In COGA,
after removing palindromic/ambiguous SNPs from the sum-
mary statistics, PRS were coded for every individual by multiply-
ing an individual’s number of effect alleles at a
particular SNP by that SNP’s effect size (beta) from the dis-
covery GWAS, then averaging across SNPs to create 1 score per person. Clumping was done using the European subset
of the 1,000 Genomes Phase 3 sample (1000 Genomes Pro-
ject Consortium, 2015) as an external LD reference panel,
using a 500 kb physical distance and an LD threshold of
\(r^2 \geq 0.25\). Scores representing effect sizes with increasingly lenient thresholds of statistical significance in the discovery
GWAS were constructed (\(p_T < 0.0001\), \(p_T < 0.001\),
\(p_T < 0.01\), \(p_T < 0.05\), \(p_T < 0.10\), \(p_T < 0.20\), \(p_T < 0.30\),
\(p_T < 0.40\), \(p_T < 0.50\)). Scores were standardized before statisti-
cal analysis.

**Data Analysis.** All analyses were conducted in R (R Core
Team, 2017). First, mixed-effect regressions were used to
examine whether alcohol consumption PRS predicted (i) alcohol use, (ii) MAXD, (iii) DSM5AUDSX, (iv) DSM4AD,
(v) SRE-T, and (vi) cessation/reduction of problem drinking,
and to determine which PRS threshold (i.e., \(p_T\)) was most
predictive of the drinking measure based on the \(p\)-value and
\(R^2\). For the cessation/reduction measure, groups were con-
trasted with each other using binary comparisons (e.g.,
Group B vs. Group C). All regressions controlled for the
covariates mentioned above as fixed effects, while the family
identifier and recruitment site were included in the models as
random effects (family nested within site). To assess model fit
and the relative amount of variance explained by the PRS, we used the “MuMIn” (https://cran.r-project.org/web/packages/MuMIn/index.html) package in R to calculate both marginal and conditional $R^2$ for each mixed model (Barton, 2012; Nakagawa and Schielzeth, 2012). We use the marginal $R^2$ to select the most predictive $p_T$. The proportion of variance attributable to the PRS ($\Delta$ marginal $R^2$) was estimated as the difference between the marginal $R^2$ (i.e., $\Delta$ marginal) of the model that included covariates and the PRS and the model with covariates alone (i.e., marginal $R^2$[full model] − marginal $R^2$[model without PRS]). Further, Cox proportional hazards survival analyses for onset of DSM4AD were conducted using the survival and survminer packages in R (https://rpkgs.datanovia.com/survminer/index.html) (Kassambara et al., 2017; Therneau and Lumley, 2015) with the same covariates as above. Violations of the proportional hazards assumption for the PRS were tested using scaled Schoenfeld residuals. For graphical depiction of cumulative survival curves, quartiles of PRS were computed using scaled Schoenfeld residuals. For graphical depiction of proportional hazards assumption for the PRS were tested (Veff = 4.82). Thus, for the first set of analyses, where PRS (including 9 p-value thresholds) predicted measures of drinking, we corrected for 45 tests (9 PRS thresholds for each of the 5 independent phenotypic tests), for a Bonferroni-corrected $p$-value of 0.05/45 = 1.1e-03. For follow-up analyses (e.g., adjustment for rs1229984), where only a single PRS at one p-value threshold was used, a stringent correction of 0.05/5 = 0.01 was implemented.

## RESULTS

Characteristics of the COGA sample ($N = 6,731$) are provided in Table 1. As expected, nearly 90% of the sample reported a lifetime history of alcohol use (at least once a month for at least 6 consecutive months). The prevalence of problem drinking (including DSM4AD and DSM5AUD)

<table>
<thead>
<tr>
<th>Table 1. Characteristics of 6,731 Participants of European Descent More Than 20 Years Old in the COGA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Overall</strong> (n = 6,731)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Female (%)</strong></td>
</tr>
<tr>
<td>Born before 1990                                                                                       17.6%</td>
</tr>
<tr>
<td>Born 1930 to 1949                                                                                       41.7%</td>
</tr>
<tr>
<td>Born in 1950 to 1969                                                                                     36.1%</td>
</tr>
<tr>
<td>Born in 1970-now                                                                                       4.6%</td>
</tr>
<tr>
<td>Age at last interview (mean [SD])                                                                        38.697 (14.189)</td>
</tr>
<tr>
<td>Family history positive*(%) (N = 5,075)                                                                  71.1%</td>
</tr>
<tr>
<td>Number of interviews (mean [SD])                                                                       2.246 (1.528)</td>
</tr>
<tr>
<td>Alcohol use                                                                                             89.9%</td>
</tr>
<tr>
<td>DSM-IV alcohol dependence (DSM4AD)                                                                      33.8%</td>
</tr>
<tr>
<td>Age at onset of alcohol dependence (mean [SD])                                                          33.3 (15.1)</td>
</tr>
<tr>
<td>Alcohol dependent with active AUD or high-risk drinking                                                  63.9%</td>
</tr>
<tr>
<td>Reduction/cessation* (N = 6,718)                                                                         33.4%</td>
</tr>
<tr>
<td>Active DSM-5 AUD or high-risk* drinkers                                                                  16.0%</td>
</tr>
<tr>
<td>Nonactive DSM-5 AUD who are currently abstinent or currently low-risk drinkers                          50.5%</td>
</tr>
<tr>
<td>No lifetime DSM-5 AUD diagnosis—and low-risk drinkers                                                    50.5%</td>
</tr>
<tr>
<td>Maximum drinks in 24 hrs (MAXD; mean [SD])                                                             18.9 (16.4)</td>
</tr>
<tr>
<td>Self-Rating of the Effects of Ethanol (SRE-T) score                                                     2.21 (0.61)</td>
</tr>
<tr>
<td>DSM-5 Alcohol use disorder symptom count (DSM5AUDSX; mean [SD])                                        3.07 (3.47)</td>
</tr>
</tbody>
</table>

*Family history positive: either or both parents have history of lifetime AUD.
*Reflects past-12-month report at final interview.
*High risk defined as ≥5 drinks/d or ≥15 drinks in 1 week for men and ≥4 drinks/d or ≥8 in 1 week for women.
was relatively high as well. For instance, the lifetime prevalence of DSM4AD was 33.8%, while 19.0, 11.1, and 22.5% endorsed 2 to 3, 4 to 5, and 6 or more DSM5AUD lifetime criteria, respectively.

As shown in Table 2, the alcohol consumption PRS predicted alcohol use, DSM4AD, and DSM5AUDSX at all $p_T$, even after correction for multiple testing. It predicted the greatest variance in alcohol use ($\Delta$ marginal $R^2 = 0.0139$; $p < 0.2$), followed by DSM4AD ($\Delta$ marginal $R^2 = 0.0056$; $p < 0.01$). DSM5AUDSX was also significantly associated with the PRS ($\Delta$ marginal $R^2 = 0.0049$; $p_T < 0.5$). MAXD was associated with the PRS at thresholds $p_T < 0.5$ to 0.001 (maximum $\Delta$ marginal $R^2 = 0.0031$), while SRE-T scores were only associated with the alcohol consumption PRS at $p_T < 0.1$ ($\Delta$ marginal $R^2 = 0.0022$). The results pattern for reduction/cessation of problem drinking suggests that the PRS were only associated with the difference between those with no lifetime DSM-5 AUD (Group A) versus the group with active DSM-5 AUD and/or high-risk drinking (Group B) ($p_T < 0.10$: odds ratio [OR] = 1.21, 95% confidence interval (CI) = 1.14 to 1.27, $p = 1.29e-8$). The PRS was not significantly associated with the comparisons between those with no lifetime DSM5AUD versus lifetime DSM5AUD but now being low-risk/abstinent (Group C; OR = 1.09, 95% CI = 1.01 to 1.18, $p = 0.028$) or between those with active DSM5AUD/high-risk drinking and those who were low-risk/abstinent (OR = 0.94, 95% CI = 0.86 to 1.02, $p = 0.105$). Because of

<table>
<thead>
<tr>
<th>N SNPs</th>
<th>Alcohol use</th>
<th>MAXD</th>
<th>SRE-T</th>
<th>DSM-5 AUD symptom count (DSM5AUDSX)</th>
<th>DSM-IV alcohol dependence (DSM4AD)</th>
<th>Reduction/cessation of problem drinkinga (<em>no lifetime AUD vs. active AUD/high-risk drinking only</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T &lt; 0.5$</td>
<td>313,399 1.11</td>
<td>0.26</td>
<td>0.18</td>
<td>0.49 (p = 4.54e-10)</td>
<td>0.57</td>
<td>0.69</td>
</tr>
<tr>
<td>$p_T &lt; 0.4$</td>
<td>268,741 1.13</td>
<td>0.25</td>
<td>0.18</td>
<td>0.47</td>
<td>0.52</td>
<td>0.61</td>
</tr>
<tr>
<td>$p_T &lt; 0.3$</td>
<td>217,799 1.28</td>
<td>0.25</td>
<td>0.18</td>
<td>0.46</td>
<td>0.47</td>
<td>0.67</td>
</tr>
<tr>
<td>$p_T &lt; 0.2$</td>
<td>159,282 1.39 (p = 3.22e-8)</td>
<td>0.26</td>
<td>0.17</td>
<td>0.43</td>
<td>0.49</td>
<td>0.76</td>
</tr>
<tr>
<td>$p_T &lt; 0.1$</td>
<td>91,805 1.27</td>
<td>0.31 (p = 1.29e-7)</td>
<td>0.22 (p = 7.01e-4)</td>
<td>0.44</td>
<td>0.59</td>
<td>0.83 (p = 1029e-8)</td>
</tr>
<tr>
<td>$p_T &lt; 0.05$</td>
<td>51,951 1.35</td>
<td>0.27</td>
<td>0.18</td>
<td>0.36</td>
<td>0.48</td>
<td>0.72</td>
</tr>
<tr>
<td>$p_T &lt; 0.01$</td>
<td>13,626 1.04</td>
<td>0.29</td>
<td>0.16</td>
<td>0.37</td>
<td>0.56 (p = 1.55e-7)</td>
<td>0.45</td>
</tr>
<tr>
<td>$p_T &lt; 0.001$</td>
<td>2,048 0.48</td>
<td>0.18</td>
<td>0.17</td>
<td>0.22</td>
<td>0.36</td>
<td>0.32</td>
</tr>
<tr>
<td>$p_T &lt; 0.0001$</td>
<td>349 0.58</td>
<td>0.08</td>
<td>0.10</td>
<td>0.14</td>
<td>0.29</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Percent variance explained was determined via $\Delta$ marginal $R^2$. The most predictive threshold (based on $p$-value) for each outcome is bolded. Alcohol use was defined as having had at least 1 drink per month for at least 6 consecutive months. The SRE-T is the SRE score averaged across 3 time points (first 5 times the participant used alcohol, last 3 months, and period of heaviest drinking). There were 3 pairwise comparisons for the reduction/cessation outcome: comparing (i) those with no lifetime AUD, (ii) those with lifetime AUD and current symptoms or no symptoms but high-risk drinking, and (iii) those who stopped drinking altogether (were abstinent for the past 12 months) or who were low-risk drinking; only the comparison between those with no lifetime AUD and those with lifetime AUD and/or high-risk drinking is shown here. AUD, alcohol use disorder; COGA, Collaborative Study on the Genetics of Alcoholism; MAXD, maximum drinks; PRS, polygenic risk scores; SNPs, single nucleotide polymorphisms.

*Denotes significance at Bonferroni-corrected $p$-value of 0.05/45 = 1.e-3.

Table 3. Association Between PRS Derived for Alcohol Consumption and Problem Drinking Indices, Adjusting for rs1229984 (ADH1B)

<table>
<thead>
<tr>
<th>DSM-IV alcohol dependence (DSM4AD)</th>
<th>DSM-5 AUD symptom count (DSM5AUDSX)</th>
<th>Alcohol use</th>
<th>SRE-T</th>
<th>MAXD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most predictive PRS</td>
<td>$p_T &lt; 0.01$</td>
<td>$p_T &lt; 0.5$</td>
<td>$p_T &lt; 0.2$</td>
<td>$p_T &lt; 0.1$</td>
</tr>
<tr>
<td>$\Delta$ marginal $R^2$ for PRS (%)</td>
<td>0.562</td>
<td>0.490</td>
<td>1.389</td>
<td>0.217</td>
</tr>
<tr>
<td>$\Delta$ marginal $R^2$ for ADH1B (%)</td>
<td>1.19</td>
<td>0.527</td>
<td>0.549</td>
<td>0.403</td>
</tr>
<tr>
<td>Beta (SE) for ADH1B</td>
<td>$-0.951$ (0.159),</td>
<td>$-0.967$ (0.165),</td>
<td>$-0.674$ (0.159),</td>
<td>$-0.165$ (0.035),</td>
</tr>
<tr>
<td>$p = 2.25e{-9}$</td>
<td>$p = 5.31e{-9}$</td>
<td>$p = 2.15e{-5}$</td>
<td>$p = 2.65e{-6}$</td>
<td>$p = 3.10e{-7}$</td>
</tr>
<tr>
<td>$\Delta$ marginal $R^2$ for ADH1B, adjusting for PRS (%)</td>
<td>1.13</td>
<td>0.486</td>
<td>0.489</td>
<td>0.380</td>
</tr>
<tr>
<td>$\Delta$ marginal $R^2$ for PRS, adjusting for ADH1B (%)</td>
<td>0.420</td>
<td>0.395</td>
<td>1.22</td>
<td>0.141</td>
</tr>
<tr>
<td>Beta (SE) for PRS, after adjusting for ADH1B</td>
<td>0.144 (0.032),</td>
<td>0.221 (0.040),</td>
<td>0.235 (0.046),</td>
<td>0.023 (0.008),</td>
</tr>
<tr>
<td>$p = 6.31e{-6}$</td>
<td>$p = 2.65e{-8}$</td>
<td>$p = 3.90e{-7}$</td>
<td>$p = 0.006$</td>
<td>$p = 1.81e{-6}$</td>
</tr>
</tbody>
</table>

AUD, alcohol use disorder; MAXD, maximum drinks; PRS, polygenic risk scores.
this limited association with the PRS, the reduction/cessation measure was not carried forward into the conditional (controlling for rs1229984) follow-up analysis. DSM4AD and DSM5AUDSX were strongly correlated with maximum cigarettes smoked \( (p < 0.001) \), but alcohol consumption PRS were not predictive of maximum cigarettes smoked \( (N = 5,436, p = 0.67) \).

Even though PRS \( p_T < 0.01 \) was associated with DSM4AD at \( x < 1.1e-3 \), the difference in AUC attributable to the addition of the PRS was modest \( (\text{AUC-covariates} = 0.710; \text{AUC-covariates + PRS = 0.713}; \text{equality test } \chi^2 = 6.62, p = 0.01) \). Results for alcohol use were far stronger: The AUC in a model where PRS alone predicted alcohol use was 0.57. Even in comparison to the model with covariates alone \( (\text{AUC} = 0.688) \), addition of PRS resulted in a statistically significant increase in prediction of alcohol use \( (\text{AUC} = 0.70; \text{equality test } \chi^2 = 9.36, p = 0.0022) \). Youden’s \( J \) was 0.11 (sensitivity = 0.49, specificity = 0.62), and a PRS cut-point of 0.0524 was determined to maximize classification, but at a weak AUC of 0.56.

The correlations between the PRS and rs1229984 ranged from \(-0.15 \) to \(-0.05 \). As shown in Table 3, the variance predicted by the PRS was partially independent of the inclusion of rs1229984 in \( ADH1B \). There was a significant association between rs1229984 genotype and alcohol use, DSM4AD, DSM5AUDSX, MAXD, and SRE-T with rs1229984 predicting 0.34 to 1.19% of the variance in these drinking measures. After including rs1229984 genotype as a covariate, the alcohol consumption PRS continued to predict 0.14 to 1.22% of the variance across the drinking measures. In particular, rs1229984 genotype had the least influence on alcohol use, for which the PRS predicted more variance than the \( ADH1B \) genotype \( (\text{Δmarginal } R^2 \text{ for ADH1B conditional on PRS = 0.49}; \text{Δmarginal } R^2 \text{ for PRS conditional on ADH1B = 1.22%).}

Within the survival analysis framework, alcohol consumption PRS \( (p_T < 0.01) \) were associated with a 1.11 [95\% CI = 1.06 to 1.16, \( p = 2.08e-5 \)] hazards of onset of DSM4AD, even after adjustment for covariates. There was no evidence for violation of the proportional hazards assumption. As shown in Fig. 1 (panel A), individuals with scores in the lowest quartile were least likely to have met criteria for DSM4AD, with survival probabilities being higher particularly in those aged 35 and older. Adjustment for rs1229984, which was strongly associated with decreased risk of onset of DSM4AD \( (\text{hazard ratio } [HR] = 0.48, 95\% \text{ CI} = 0.37 \text{ to } 0.63, p = 7.85e-08) \), had little influence on the effect of the overall PRS \( (HR = 1.09, 95\% \text{ CI} = 1.04 \text{ to } 1.14, p = 2.86e-4) \).

As expected, sex was strongly associated with all aspects of problem drinking, including DSM4AD and DSM5AUDSX \( (p < 2e-16; \text{higher likelihood of problem drinking in males}) \). Family history was positively associated with all drinking measures except for alcohol use, Age at last interview was associated with DSM4AD, DSM5AUDSX, and the reduction/cessation comparison of those with no lifetime AUD.
related to typical consumption of various types of alcoholic beverages (e.g., beer, wine, liquor). Such assessments are amenable to collection in very large samples such as the UK Biobank, and meta-analyses of GWAS data for such simple alcohol measures have succeeded in locus discovery (Clarke et al., 2017; Jorgenson et al., 2017; Liu et al., 2019; Schumann et al., 2016). It is unclear whether all loci identified using these indices of typical drinking will overlap with the genetic variants that contribute to other drinking milestones and features, including alcohol dependence. Not only does alcohol dependence involve high levels of alcohol consumption, but it also includes a significant loss of control over drinking, drinking to ameliorate negative mood states, impairment of interpersonal and vocational functioning, and a general transition from impulsive to compulsive use (Koob and Kreek, 2007). One prior GWAS noted that the genetic correlation between alcohol consumption and dependence was modest and variable (Walters et al., 2018). Another recent study that draws on the UK Biobank data suggests that indices of recent problem drinking (e.g., failed to do what was expected because of drinking) were genetically correlated with psychopathology, while indices of more moderate alcohol consumption were not (and in some instances were negatively correlated) (Sanchez-Roige et al., 2018). The present study provides evidence that the aggregated genetic effects related to typical alcohol consumption in an older volunteer cohort do predict significant (albeit very modest) proportions of variance in various indices of drinking, including problem drinking, over and above the effects of rs1229984. However, in the current analyses the strongest associations occurred with a broad measure of alcohol use. This is not entirely surprising, since the alcohol consumption PRS is based on typical weekly intake, which itself is partly derived from frequency of drinking.

Five findings are noteworthy. First, alcohol consumption PRS predicted a very modest proportion of variance (<0.6%) in DSM-IV alcohol dependence (DSM4AD) and DSM-5 AUD symptomatology (DSM5AUDSX). This estimate is relatively consistent with the extent to which PRS tend to explain cross-trait variance. For instance, a recent study examined the extent to which alcohol consumption PRS (derived from a smaller meta-analysis; Schumann et al., 2016) predicted recent drinking (alcohol consumption and AUDIT scores) in a sample of 5,456 participants aged 18 to 83 years (Mies et al., 2018); at the most predictive thresholds, alcohol consumption PRS predicted 0.11% of the variance in recent drinking. A similar analysis of the Avon Longitudinal Study of Parents and Children found that an 89-SNP risk score (derived from literature searches) predicted up to 0.66% of the variance in typical alcohol consumption (Taylor et al., 2016). Despite this consistency, the AUC estimate suggests that the consumption PRS minimally (but significantly) improves classification for DSM4AD. These findings point to the extremely high polygenicity underlying alcohol intake and problem drinking such that even the aggregated effects of SNPs explain only modest
proportions of variance. These findings also suggest that there may be only modest genetic overlap between the range of normal alcohol consumption and problematic drinking phenotypes.

Second, despite being a quantitative index of alcohol consumption, maximum number of drinks consumed in a single 24-hour period (MAXD) was less effectively predicted by the alcohol consumption PRS ($R^2 \leq 0.31\%$). MAXD is genetically correlated with problem drinking (Agrawal et al., 2009; Grant et al., 2009) and correlates well with quantitative indices of tolerance (Kendler et al., 2012; Schuckit et al., 2008). The lower prediction might be related to MAXD being potentially influenced by a single episode of heavy drinking and, thus, not indicative of either typical or problem drinking. Likewise, total scores on the subjective ratings of EtOH (SRE-T) were not predicted by alcohol consumption PRS. SRE-T is an index of alcohol sensitivity and has primarily been studied as a predisposing factor for and predictor of later problem drinking (Schuckit and Smith, 2013). Individuals with higher SRE-T scores demonstrate lower level of response to alcohol, potentially via dampening of neural and physiological pathways in response to alcohol (Schuckit, 2018; Schuckit et al., 2008). There are currently no published GWAS of SRE-T, although a meta-analysis of a related but etiologically distinct trait representing subjective ratings of EtOH during the first 5 times that alcohol was consumed (SRE-5) did not find loci previously associated with alcohol consumption or dependence to be related to it (Edwards et al., 2018). Thus, even though SRE is heritable (Kalu et al., 2012) and related to alcohol consumption, its genetic underpinnings may be quite different from those related to alcohol intake.

A third notable observation is that the alcohol consumption PRS explained the most variance for alcohol use ($R^2 \leq 1.39\%$), which represents drinking at least once a month for 6 consecutive months at some point during the lifetime and is ubiquitously endorsed in this sample (89.9%); this is likely also common in the UK Biobank sample. In fact, the addition of the PRS significantly improved the AUC for alcohol use, compared to a model of covariates only. It is likely that alcohol consumption in the UK Biobank represents normative patterns of drinking and does not sufficiently index problem drinking. Thus, in COGA, PRS derived from such an index of normative alcohol consumption was most closely related to an equally heterogeneous measure of drinking that included normative (and problem) drinkers. We were unable to examine typical alcohol consumption as it was not assessed in COGA in a similar manner as UK Biobank. Nonetheless, our results suggest that in this high-risk sample, PRS from large-scale GWAS of alcohol consumption were more strongly associated with propensity to alcohol use than with measures of problem drinking or disorder.

Fourth, we found little evidence for shared genetic overlap between weekly alcohol consumption and reduction/cessation of problematic drinking. Prior studies have identified candidate variants related to treatment-dependent remission (e.g., Karpyak et al., 2014), and emerging evidence suggests high sibling concordance for abstinence remission (McCutcheon et al., 2017), suggesting heritable influences on the transition from active AUD to low-risk drinking and abstinence. However, while PRS distinguished those without a lifetime history of DSM-5 AUD from those with an active diagnosis or high-risk drinking, they were unrelated to low-risk drinking or abstinence in those with a lifetime history of DSM-5 AUD. This could be partially due to a lack of power: Only 16% of the individuals were in the category of low-risk drinking and abstinence. This finding could also suggest that even though there is polygenic overlap between alcohol consumption and severity of AUD (e.g., DSM5AUDSX), and individuals who successfully reduce their problem drinking might have a less severe form of the disorder, the genetic propensity that extends beyond severity and specifically predicts cessation is distinct. One might speculate that genetic influences on remission relate less to genetic liability to alcohol intake and more to aspects of socioeconomic status (e.g., Trim et al., 2013), personality characteristics, and other comorbid psychiatric disorders (Lopez-Quintero et al., 2011). Future studies should further explore the heritability of cessation and its coheritability with other drinking measures.

Fifth, the addition of rs1229984 in the models attenuated the effect of the PRS, although only modestly for alcohol use. While rs1229984 exerts one of the strongest effect sizes observed for psychiatric and behavioral traits (Edenberg and McClintick, 2018), other loci related to alcohol dependence are expected to have noticeably more modest effects (Walters et al., 2018). In our analyses, the conditional effect of rs1229984 (i.e., $R^2$ for rs1229984 when including PRS in the model) was greater for DSM4AUD and DSM5AUDSX (explaining 0.49 to 1.13% of the variance) than the conditional effect of the PRS after adjusting for rs1229984 (0.40 to 0.42% of variance explained). In fact, the addition of PRS to the model had negligible impact on the variance in DSM4AUD and DSM5AUDSX already explained by rs1229984, highlighting this variant’s robust effect on problem drinking. In contrast, rs1229984 did not substantially reduce the variance in alcohol use associated with the PRS; rather, the alcohol consumption PRS explained a greater percent of the variance in alcohol use (1.22%) than rs1229984 genotype (0.49%). This suggests that, despite being an aggregate polygenic index of variants with small individual effect sizes, the alcohol consumption PRS are superior predictors of alcohol use compared to rs1229984 alone.

One prior study suggested the polygenic prediction is maximized when the discovery and target cohorts have similar demographic characteristics and ascertainment strategies (Savage et al., 2018). Thus, it is also possible that the variance associated with the PRS is greater in the subset of COGA that is demographically matched to the UK Biobank participants (e.g., older age). However, age did not appear to
moderate the association between the PRS and alcohol-related outcomes in the current study, nor was there support for absence of proportionality of hazards in the survival model, suggesting homogeneity of effects across age. Relatedly, even though family history was a strong predictor of problem drinking, we found no evidence that it moderated the effect of the PRS in our sample. This might be due to the partially high-risk nature of the COGA sample. Despite differences between the characteristics of the discovery sample and COGA, we detected appreciable variance in alcohol use.

Some limitations of this sample are also noteworthy. First, we restricted our analyses here to individuals of European American descent because the discovery GWAS was limited to individuals of European ancestry, and cross-ancestral predictions have several limitations (Martin et al., 2017). Second, despite being fairly large and predominantly ascertained for families with many alcohol-dependent individuals, the COGA sample size may still be underpowered to detect the very modest percentages of variance explained by common SNPs for alcohol-related traits (Bogdan et al., 2018; Clarke et al., 2017; Gratten et al., 2014; Sanchez-Roige et al., 2017; Walters et al., 2018). In particular, although certain types of interaction (e.g., crossover effects) do not necessitate significant main effects, we were likely underpowered to test such interactions with PRS where a main effect of PRS was undetectable.

While large-scale population-based studies, such as the UK Biobank, are valuable in their potential to identify multiple risk loci for easily collectible phenotypes, such as alcohol consumption, and contribute to the development of well-powered PRS, they may not adequately capture the genetic factors that contribute to problem drinking, which itself is a multifaceted phenotype. Until there are similarly large-scale GWAS conducted in samples ascertained for problem drinking, the amount of genetic overlap between alcohol consumption and problematic drinking behaviors remains unclear.

ACKNOWLEDGMENTS

The COGA (principal investigators: B. Porjesz, V. Hesselbrock, H. Edenberg, and 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, and 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, and S. Lourens (Indiana University); G. Chan (University of Iowa; University of Connecticut); J. Meyers, D. Chorlian, C. Kamarajan, A. Pandey, and J. Zhang (SUNY Downstate); J.-C. Wang, M. Kapoor, and S. Bertelsen (Icahn School of Medicine at Mount Sinai); A. Anokhin, V. McCutcheon, and S. Sacco (Washington University); J. Salvatore, F. Aliev, and 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 was supported by National Institutes of Health (NIH) Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse. This research was also supported by MH109532 (ECJ, AA, HJE) and F32AA027435 (ECJ).

CONFLICT OF INTEREST

J.I.N. is an investigator for Janssen and Assurex.

REFERENCES


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.


