Published in final edited form as:
Hum Genet. 2013 October ; 132(10): 1141–1151. doi:10.1007/s00439-013-1318-z.

A meta-analysis of two genome-wide association studies to identify novel loci for maximum number of alcoholic drinks

Manav Kapoor[1], **Jen-Chyong Wang**[1], **Leah Wetherill**[2], **Nhung Le**[1], **Sarah Bertelsen**[1], **Anthony L Hinrichs**[1], **John Budde**[1], **Arpana Agrawal**[1], **Kathleen Bucholz**[1], **Danielle Dick**[3], **Oscar Harari**[1], **Victor Hesselbrock**[4], **John Kramer**[5], **John I Nurnberger Jr**[2], **John Rice**[1], **Nancy Saccone**[1], **Marc Schuckit**[6], **Jay Tischfield**[7], **Bernice Porjesz**[8], **Howard J Edenberg**[2], **Laura Bierut**[1], **Tatiana Foroud**[2], and **Alison Goate**[1,*]

Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63110

[1]Washington University School of Medicine

[2]Indiana University School of Medicine

[3]Virginia Institute of Psychiatric and Behavioral Genetics

[4]University of Connecticut Health Center

[5]University of Iowa Carver College of Medicine

[6]University of California San Diego

[7]Rutgers University

[8]SUNY Health Sciences Center

Abstract

Maximum number of alcoholic drinks consumed in a 24-h period (maxdrinks) is a heritable (> 50%) trait and is strongly correlated with vulnerability to excessive alcohol consumption and subsequent alcohol dependence (AD). Several genome-wide association studies (GWAS) have studied alcohol dependence, but few have concentrated on excessive alcohol consumption. We performed two GWAS using maxdrinks as an excessive alcohol consumption phenotype: one in 118 extended families (N=2322) selected from the Collaborative Study on the Genetics of Alcoholism (COGA), and the other in a case-control sample (N=2593) derived from the Study of Addiction: Genes and Environment (SAGE). The strongest association in the COGA families was detected with rs9523562 ($p = 2.1\times10^{-6}$) located in an intergenic region on chromosome 13q31.1; the strongest association in the SAGE dataset was with rs67666182 ($p = 7.1\times10^{-7}$), located in an intergenic region on chromosome 8. We also performed a meta-analysis with these two GWAS and demonstrated evidence of association in both datasets for the *LMO1* ($p = 7.2\times10^{-7}$) and *PLCL1* genes ($p = 4.1\times10^{-6}$) with increased maxdrinks. A variant in *AUTS2* and variants in *INADL, C15orf32* and *HIP1* that were associated with measures of alcohol consumption in a meta-analysis of GWAS studies and a GWAS of alcohol consumption factor score also showed nominal association in the current meta-analysis. The present study has identified several loci that warrant further examination in independent samples. Among the top SNPs in each of the dataset ($p\leq10^{-4}$) far more showed the same direction of effect in the other dataset than would be expected by chance ($p = 2\times10^{-3}$, 3×10^{-6}), suggesting that there are true signals among these top SNPs, even though no SNP reached genome-wide levels of significance.

*Corresponding Author: Alison Goate, D.Phil., Samuel & Mae S. Ludwig Professor of Genetics in Psychiatry, Professor of Genetics, Professor of Neurology, Dept. of Psychiatry, B8134, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110.

Keywords

Alcohol consumption; maximum number of alcoholic drinks; GWAS; COGA; SAGE

Introduction

Excessive alcohol consumption is a modifiable behavioral risk factor for morbidity and one of the leading causes of death in the United States (Mokdad et al. 2004). Every year nearly 85,000 (3.5%) people die due to short and long-term consequences of alcohol use (Mokdad et al. 2004). Consumption of alcohol in large quantities is capable of damaging every organ and system of the body (Caan and De Belleroche 2002) and is associated with increased risk of alcohol dependence (AD), cancer, alcoholic liver disease, cardiovascular disease, chronic pancreatitis and malabsorption (Muller et al. 1985; Testino 2008). The International Agency for Research on Cancer and the U.S. Department of health and Human Services has listed alcohol as a known carcinogen. Epidemiological studies indicate that moderate alcohol consumption by younger women could increase the risk of breast cancer later in life (Allen et al. 2009; Brooks and Zakhari 2013; Chen et al. 2011).

Several genome-wide association analysis (GWAS) have been carried out to find genes that contribute to risk for AD (Bierut et al. 2010; Edenberg et al. 2010; Kendler et al. 2011; Treutlein et al. 2009; Wang et al. 2012a), and alcohol consumption (Baik et al. 2011; Heath et al. 2011; Schumann et al. 2011). Recently, Schumann and colleagues (Schumann et al. 2011) performed a meta-analysis using 12 population-based GWAS and reported genome-wide association of a SNP in the autism susceptibility candidate 2 gene (*AUTS2*) with alcohol consumption (grams per day per kilogram body weight). In another GWAS, Baik and colleagues (Baik et al. 2011) identified twelve SNPs on chromosome 12 that were strongly associated with alcohol consumption (grams per day) in Korean men. A GWAS of a quantitative factor score created from indices of excessive alcohol consumption that included maximum number of alcoholic drinks consumed in 24 hours (maxdrinks), reported many variants approaching the genome-wide threshold for suggestive significance ($p < 5 \times 10^{-5}$) (Heath et al. 2011). Bierut and coworkers (2012) used maxdrinks as a standalone measure of alcohol consumption and identified a SNP (rs1229984) in the *ADH1B* gene strongly associated with lower consumption of alcohol in populations of European and African ancestry. This SNP has been demonstrated to be functional and to have a strong impact on risk for alcoholism in all populations (Hurley and Edenberg 2012).

The present study took advantage of both family and case-control study designs in two large complementary and well-characterized European American (EA) cohorts assessed using the Semi-Structured Assessment for the Genetics of Alcoholism, which includes the “maxdrinks” measure of alcohol consumption (maxdrinks). Maxdrinks is heritable and strongly correlated with excessive alcohol consumption and alcohol dependence (AD); (Grant et al. 2009; Kendler et al. 2010; Saccone et al. 2000). GWAS was performed in each cohort and then results were combined in a subsequent meta-analysis. We found suggestive evidence of association with several novel loci and supportive evidence of association with several previously reported loci.

Materials and Methods

Subjects

The data presented here were generated through analyses using two datasets.

The Collaborative Study on the Genetics of Alcoholism (COGA) sample— Following the approval of institutional review boards at all participating institutions, AD probands > 18 years old were recruited through alcohol treatment programs and administered the adult version of a validated poly-diagnostic instrument, the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al. 1994). Relatives of the probands and comparison families below the age of 18 years were administered an adolescent version of the SSAGA. The COGA sample used here consisted of 2322 subjects of European descent from 118 extended families with a measure of maxdrinks. The characteristics of the study participants are listed in Table 1. Further details about this dataset, are described elsewhere (Wang et al. 2012a).

The Study of Addiction: Genetics and Environment (SAGE) sample—The Institutional Review Board at each contributing institution reviewed and approved the protocols for genetic studies under which all subjects were recruited. The Study of Addiction: Genetics and Environment (SAGE) is funded as part of the Gene Environment Association Studies (GENEVA) initiative supported by the National Human Genome Research Institute. The subjects were selected from three large, complementary datasets: COGA, Family Study of Cocaine Dependence (FSCD), and Collaborative Genetic Study of Nicotine Dependence (COGEND). In the present study, we removed 129 individuals from the SAGE study who were also part of the 118 extended families in COGA. A total of 2593 subjects of European descent with the maxdrinks measure (Table 1) were used for the association analysis. Further details of the SAGE sample have been described previously (Bierut et al. 2010).

Phenotype

The measure for lifetime maximum number of alcoholic drinks consumed in 24 hour period (maxdrinks) in both the COGA and SAGE datasets was derived from the SSAGA interview (Bucholz et al. 1994). The specific question asked to assess maxdrinks was “What is the largest number of drinks you have ever had in a 24-hour period”? A standard drink of alcohol was defined as a 4 oz. glass of wine, a 12 oz. bottle of beer, or a 1.5 oz. shot glass of 80 proof liquor. The average maxdrinks in the COGA sample and the SAGE samples were similar, with average maxdrinks=25 and maxdrinks=13 in men and women, respectively (Table 1). Phenotypic distributions for both the COGA and SAGE datasets are presented in Supplementary figure 1. Extreme values of maxdrinks (greater than 100) were set at 100. Individuals who did not consume alcohol were classified as unknown and were removed from subsequent analyses. Prior to association analysis, maxdrinks in both datasets were natural log-transformed to acquire a normal distribution (Supplementary figure 1).

Genotyping

COGA sample—Genotyping was performed at the Genome Technology Access Center at Washington University School of Medicine in St. Louis (http://gtac.wustl.edu/) using the Illumina Human OmniExpress array 12.VI (Illumina, San Diego, CA, USA). A total of 707,557 autosomal SNPs passed quality control (Wang et al. 2012a). SNPs with a minor allele frequency (MAF) below 5% (n=115,872) were excluded from further analysis. One hundred eighteen families, who self-reported European descent were included in the dataset. We utilized EIGENSTRAT (Price et al. 2006) along with HapMap European reference samples to confirm European ancestry.

SAGE sample—DNA samples were genotyped on the Illumina Human 1 M beadchip (Illumina Inc., San Diego, CA, USA) by the Center for Inherited Diseases Research (CIDR) at Johns Hopkins University. A total of 948,658 SNPs passed data-cleaning procedures and further within sample filtering for autosomal and X-chromosome markers yielded 948,142

markers. HapMap genotyping controls, duplicates, related subjects and outliers were removed from the sample set (Bierut et al. 2010). A total of 771,842 SNPs on the autosomal chromosomes with MAF greater than equal to 5% were used for further analysis. The software package EIGENSTRAT (Price et al. 2006) was used to calculate principal components reflecting continuous variation in allele frequencies representing ancestral differences. First principal component (PC1) distinguished African-American participants from European-American (EA) participants. The EA subset of the SAGE sample was used for the final association analysis.

Imputation

Because different genotyping arrays were used in the COGA and SAGE studies, there were only 483,037 common SNPs across both datasets after quality control. In order to perform a meta-analysis, we imputed the SNPs for both datasets.

COGA dataset—We used the program BEAGLE version 3.3.1 (Browning and Browning 2007) to impute SNPs on autosomes and X chromosome that were not genotyped on the Illumina Omni Express array. Since our sample was European American, we used as a reference set the genotypic data from the EUR in the August 2010 release of the 1000 Genomes Project, provided with the Beagle 3.3.1 release. SNPs with a correlation between the best-guess genotype and allele dosage greater than 0.3 (r^2>0.3) were used in the analyses. For individual-level genotype data, we retained genotypes having a probability ≥80% (from the gprob metric in Beagle); all other genotypes were set to missing. We converted genotypic probability data into most-likely genotypes. This helped us in detecting genotypic errors in families. We used these most-likely genotypes in further cleaning and analysis. The imputed SNPs were cleaned using the same methods as genotyped SNPs. Mendelian errors in the imputed data were removed. After a careful quality review, a total of 4,058,415 SNPs were used for association analysis. To account for uncertainty, we used the mean of the distribution of imputed genotypes, which corresponds to an expected allelic or genotypic count (dosage) for each individual.

SAGE dataset—Imputation for the SAGE data was performed using BEAGLE version 3.3.1, (Browning and Browning 2007) for all autosomes and the X chromosome, with the December 2010 released genotypes and haplotype reference panels selected from the 1000 Genomes Project (www.1000genomes.org). EUR analysis panel was used to impute the EA samples. In SAGE we did not have the advantage of family based cleaning; therefore, we based our analysis on 4,058,415 common SNPs that met imputation quality control criteria across both SAGE and COGA datasets.

Data Analysis

COGA dataset—We first tested for the effect of covariates on maxdrinks. As expected, gender was a highly significant predictor of maxdrinks (P=0.0001). Because the COGA dataset consists of large multi-generation families we used birth cohort to account for the secular effects on maxdrinks across generations. We divided subjects into 4 cohorts based on their year of birth (< 1930< 1930 – 1949 – 1950 – 1969, and ≥1970). The effect of age at interview was not independent of birth cohort, therefore it was not used as a covariate in the final association analysis. Quantitative trait association analysis was performed using a mixed linear model (MLM) implemented in the GWAF (Chen and Yang 2010) package of R, using gender and birth cohort as covariates and natural log transformed maxdrinks as the trait. The correlated nature of the COGA sample was accounted for using random effects correlated according to the degree of relatedness within a family using a kinship matrix.

SAGE dataset—As described previously (Bierut et al. 2010), covariates included sex, age, sample origin (COGA, COGEND and FSCD), cocaine dependence and nicotine dependence. The relationship between age at interview and maxdrinks was not linear in the SAGE dataset. Therefore we avoided using age as a continuous covariate in the analysis. We coded age at interview into four dummy categorical variables (defined, using quartiles, as 34 years and younger, 35–39 years, 40–44 years, and 45 years and older) and included them as covariates. Nicotine and cocaine dependence were included as covariates in SAGE because the COGEND and FSCD datasets were ascertained on nicotine dependence and cocaine dependence respectively. Association of natural log transformed maxdrinks was examined by linear regression using PLINK (Purcell et al. 2007).

Meta-analysis

We used METAL software (Willer et al. 2010) to combine the p values and effect size estimates from the SAGE and COGA association results. METAL selects the reference allele for each marker and calculates the z-statistic characterizing the evidence for association. The z-statistic summarizes the magnitude and direction of the effect relative to the reference allele. Overall statistic and p-values were then calculated from the weighted sum of the individual statistics.

Conditional Analysis

The minor allele of rs1229984 in *ADH1B* gene was previously shown to be protective against alcohol dependence in Asian samples, in which this allele is common (Chen et al. 1999; Edenberg 2007; Thomasson et al. 1991; Wang et al. 2012b). We have shown that the minor allele is also protective for excessive alcohol consumption in European Americans and African Americans (Bierut et al. 2012). We therefore performed a conditional analysis and subsequent meta-analysis in the SAGE and COGA datasets, including rs1229984 genotype as a covariate. Because this SNP has a low minor allele frequency and is hard to impute, this SNP was directly genotyped in the SAGE and COGA datasets (Bierut et al. 2012).

Power of the study

We conducted power calculations under an additive genetic model, at alpha = 5×10^{-8}, assuming the trait SNP and marker are in complete linkage disequilibrium with specified D′, for a range of assumed effect sizes starting with the strongest effect identified in the present study, using the Genetic Power Calculator (Purcell et al. 2003).

Gene based association analysis

We used the program, VEGAS (Liu et al. 2010) to perform gene-based tests for association on the results from the meta-analysis of COGA and SAGE. This program annotates SNPs to corresponding genes, produces a gene-based test statistic, and then uses simulation to calculate an empirical gene-based p-value.

Genome-wide complex trait analysis (GCTA)

The GCTA program as described by Yang and colleagues (Yang et al. 2011) was used to estimate the proportion of phenotypic variance for maxdrinks explained by the genome-wide SNPs genotyped on the Illumina 1M chip in the SAGE dataset. This program fits a linear model to the phenotype data and uses a restricted maximum likelihood to estimate the variance explained by the SNPs.

Results

COGA—No SNP, either genotyped or imputed, passed our pre-specified threshold of genome-wide significance of $p<5\times10^{-8}$ for association with maxdrinks (Supplementary figure 2). A total of 479 SNPs showed association at $p < 1 \times 10^{-4}$ (Supplementary table 1), which is higher than the number of SNPs (406) (p = 0.015) expected by chance. The strongest association with maxdrinks was detected with an imputed SNP, rs9523562 (beta = −0.13, $p=2.1 \times 10^{-6}$), on chromosome 13q31.1. Several genotyped and imputed SNPs that are highly correlated with rs9523562 (R^2=0.8, D′=0.95) also showed association with maxdrinks ($5.0 \times 10^{-5} \leq p \leq 2.1 \times 10^{-6}$) (Supplementary table 1). An intronic SNP rs2283970 in *CACNG2* (calcium channel, voltage-dependent, gamma subunit 2) also showed suggestive association (beta = −0.13, $p=3.9 \times 10^{-6}$).

SAGE—In the SAGE dataset, a total of 646 SNPs showed association with maxdrinks at $p < 1 \times 10^{-4}$; however, none of these variants met criteria for genome-wide significance (Supplementary table 2, supplementary figure 3). The observed number of SNPs is almost 1.5 times more than the number of SNPs expected to show suggestive association at this level by chance ($p < 1 \times 10^{-4}$). rs67666182 near *SNX16* (sorting nexin 16) on chromosome 8 showed suggestive association ($9.9 \times 10^{-5} < p < 7.1 \times 10^{-7}$). Other SNPs in high linkage disequilibrium with rs67666182 (r^2=0.8 D′=0.95) also provided evidence of association (Supplementary table 2). Several SNPs in cytochrome P450, family 4, subfamily F, polypeptide 8 (*CYP4F8),* muskelin 1, intracellular mediator containing kelch motifs *(MKLN1),* LIM domain only 1 *(LMO1),* phenylalanyl-tRNA synthetase 2, mitochondria *(FARS2)* and SH3-domain GRB2-like 3 (*SH3GL3)* were also associated with maxdrinks ($p < 1.0 \times 10^{-4}$).

Meta-analysis using SAGE and COGA dataset

There was limited replication and overlap for the top signals identified in either the COGA or the SAGE dataset (Supplementary Tables 1 & 2). The association of SNPs ($p < 1 \times 10^{-4}$) near the neural cell adhesion molecule (*NCAM2*) and M-phase phosphoprotein (*MPHOSPH6*) genes observed in the COGA study showed nominal association in the SAGE study ($0.04 \leq p \leq 0.05$). The association of SNPs near the LIM domain only (*LMO1*), Toll like receptors (*TLR1*, *TLR10*) and Ataxin 2 binding protein 1 (*A2BP1*) genes identified in SAGE were also weakly ($0.03 \leq p \leq 0.06$) associated with maxdrinks in COGA. However, the number of SNPs with $p < 1 \times 10^{-4}$ in both COGA and SAGE with the same direction of effect in the other study was significantly higher than expected by chance: 270/479 of the COGA SNPs (P=0.003) and 380/646 of the SAGE SNPs ($p=4\times10^{-6}$).

Table 2 presents the results from the meta-analysis of COGA and SAGE samples. An intronic variant rs4758317 in *LMO1* showed evidence for association ($p=7.2\times10^{-7}$) with evidence coming from both SAGE (beta = 0.08, $p=6.0\times10^{-6}$) and COGA (beta = 0.05, $p=2.9\times10^{-2}$) (Table 2, supplementary figure 4). Similarly, we observed a suggestive association ($p= 4.1\times10^{-6}$) of maxdrinks with an intronic variant in *PLCL1* (rs67031482), driven by both COGA (beta = −0.07, $p=1.1 \times 10^{-2}$) and SAGE (beta = −0.07, $p=1.2\times10^{-4}$) datasets. In populations of European descent, this SNP is in a region of linkage disequilibrium of ~ 200 kb on chromosome 2q that spans the entire *PLCL1* gene (Supplementary figure 5). Table 2 presents the results from the meta-analysis of COGA and SAGE samples. For meta-analysis we had > 90% power to detect a genome-wide association given a SNP that explains 1% of the variance for maxdrinks.

Conditional analysis

The SAGE EA dataset used in the present study was part of the meta-analysis of rs1229984 reported by our group (Bierut et al, 2011). In the current EA subset of SAGE we validated the reported association of rs1229984 with maxdrinks ($p = 4.32 \times 10^{-7}$). Although the signal was not genome-wide significant, it explained a large proportion (1%) of variance for maxdrinks in the SAGE EAs. This SNP was also nominally associated with maxdrinks in the COGA family dataset ($p = 0.012$) with protective effect (−0.19). The meta-analysis of these two datasets for rs1229984 resulted in genome-wide significant association with maxdrinks ($p = 2.04 \times 10^{-8}$). To test the possible moderating effect of rs1229984 on the relationship between maxdrinks and other genetic variants, we further analyzed the COGA and SAGE GWAS including rs1229984 as a covariate. Rs4758317 in *LMO1* remained the strongest signal with a slight improvement in effect (beta value changed from −0.078 to −0.080) and p value (changed from 7.2×10^{-7} to 5.1×10^{-7}). The meta-analysis resulted in a nominal improvement in the evidence for association of some other SNPs as well, but no new suggestive association results emerged through this analysis (supplementary table 4).

Estimation of variance explained by genome-wide SNPs

We estimated the proportion of variance for the maxdrinks explained by 771,842 SNPs genotyped on 2593 unrelated subjects from the SAGE dataset using linear model analysis as implemented in GCTA program. The results showed that nearly 40% of the variance for maxdrinks can be explained by considering all SNPs simultaneously.

Gene-based association analysis

Several genes that showed association in single SNP meta-analysis were also found to be significant in gene-based association tests (Supplementary table 3). The test confirmed that multiple SNPs in a gene support the evidence of single SNP association analysis. Exocyst complex component 5 (*EXOC5)* on chromosome 14 and solute carrier family 26 member 4 (*SLC26A4)* on chromosome 7 were among the top genes identified in both gene-based and single SNP association analysis ($p < 5 \times 10^{-5}$). *LMO1* ($p = 0.01$) and *PLCL1* ($p = 4.13 \times 10^{-4}$) also showed nominal significance in gene-based analysis.

Discussion

In the present study, we identified several SNPs showing suggestive association with maxdrinks ($p < 5.0 \times 10^{-5}$) in the COGA or the SAGE datasets, but no SNP met the threshold for genome-wide significance. The strongest signals from the COGA GWAS were near a non-coding RNA (LincRNA) on chromosome 13q31.1. These SNPs were not significant in SAGE ($0.5 \leq p \leq 0.91$), but showed the same direction of effect. LincRNAs are usually associated with open chromatin signatures such as histone modification sites. There is some emerging evidence suggesting that lincRNAs regulate gene expression both during normal development and under pathological conditions, including neuropsychiatric disorders (Dudley et al. 2011; Mattick 2009). Out of 479 SNPs ($p < 1 \times 10^{-4}$) identified in the COGA dataset, SNPs near *NCAM2* and *MPHOSPH6* showed nominal replication in SAGE ($p < 0.05$), while many other SNPs showed the same direction of effect. In the SAGE GWAS the strongest signal, rs67666182, along with highly correlated SNPs ($R^2 > 0.8$) near *FABP5* showed no replication in COGA, while SNPs in *LMO1*, and near *TLR1*, *TLR10* and *A2BP1* showed nominal replication. Because a large excess of SNPs that were nominally significant ($p<10^{-4}$) in either COGA or SAGE showed the same direction of effect in the other study, this suggests that within these SNPs there is a reproducible signal of variants that affect maxdrinks.

In meta-analysis, the strongest association ($p = 7.2 \times 10^{-7}$) was detected with rs4758317, an intronic SNP in *LMO1*, a cysteine rich, two LIM domain transcription regulator. This gene is expressed in the brain and involved in gene regulation within neural lineage cells potentially by direct DNA binding or by binding to other transcription factors. Recently researchers have showed that members of the *LMO* gene family are involved in regulation of behavioral responses to ethanol in Drosophila melanogaster and the mouse (Lasek et al. 2011). In flies, decreased expression of Drosophila Lim-only (dLmo) was associated with increased sensitivity to ethanol-induced sedation, whereas increased expression of dLmo led to increased resistance to ethanol-induced sedation. A similar increase in ethanol sensitivity was observed in $Lmo3^{-/-}$ mice. Injecting lentivirus expressing either shLmo3.8 or shScr in single- cell embryos knocked down the Lmo3 expression in mice. In addition to an effect on ethanol sedation, reducing Lmo3 levels also correlated with decreased ethanol consumption, suggesting that Lmo3 may play a role in alcohol preference in mammals. In another study, Wang and colleagues used quantitative PCR to measure *LMO1* expression in a set of 23 tumors. They reported that rs110419, a variant in *LMO1,* is associated with *LMO1* mRNA expression ($p = 0.01$) (Wang et al. 2011). This SNP showed nominal association with maxdrinks ($p = 1.29 \times 10^{-4}$) and is also in strong LD ($R^2 = 0.93$, $D' = 0.96$) with the strongest SNP rs4758317 ($p = 7.2 \times 10^{-7}$) in the present study. *LMO1* was also nominally significant in gene-based association analysis performed using VEGAS ($p = 0.01$).

Another genetic locus of interest is a 200 kb region, where 93 SNPs spanning phospholipase C-like 1 (*PLCL1)* on the long arm of chromosome 2, showed evidence of association in both COGA ($7.21 \times 10^{-3} \leq p \leq 2.71 \times 10^{-2}$) and SAGE ($2.80 \times 10^{-3} \leq p \leq 1.22 \times 10^{-4}$) subjects (Supplement figure 5). This protein was first identified as a novel inositol 1,4,5-triphosphate binding protein. It has a number of binding partners including GABA(A) receptor associated protein and beta subunits of GABA(A) receptors. *PLCL1* is part of biological processes such as intracellular signal transduction, lipid metabolism and behavior (KEGG http://www.genome.jp/kegg). We used online function prediction programs like HaploReg (Ward and Kellis 2012), Polyphen (Adzhubei et al. 2010) and SCAN (http://www.scandb.org) to predict the function of variants within the *PLCL1* locus. The strongest signal in this gene is an intronic SNP rs67031482 ($p=4.7 \times 10^{-6}$), which is in LD ($R^2 = 0.72$, $D' = 0.92$) with rs1064213 ($p = 1.1.\times10^{-4}$), a missense variant that leads to substitution of a conserved residue valine to isoleucine at position 667 in the PI_PLC Y-box region. This conserved region has been shown to be important for the catalytic activity of the protein (Jiang et al. 1994). Interestingly, in GWAS of Molecular Genetics of Schizophrenia control sample, researchers reported a variant rs10180112 in *PLCL1* which showed moderate evidence of association ($p=5.22 \times 10^{-4}$) with alcohol dependence symptom counts in African Americans (Kendler et al. 2011). This variant showed nominal association with maxdrinks ($p= 1.9 \times 10^{-3}$) in COGA families, but no association was detected in the SAGE sample. The SNP rs10180112 is in low LD ($R^2 = 0.036$, $D' = 0.51$) with rs67031482, the most strongly associated SNP with maxdrinks in the current datasets. This same SNP, rs10180112, is a putative expression quantitative trait locus (eQTL) in monocytes (Zeller et al. 2010), as reported on an eQTL browser (http://eqtl.uchicago.edu/). The variant rs1579695 ($p=1.12 \times 10^{-5}$) identified in the present study was also found to be a potential cis eQTL in monocytes as shown in the data of Zeller and colleagues (Zeller et al. 2010). This gene also showed nominal significance with maxdrinks in gene-based association analysis performed using VEGAS ($p = 4.13 \times 10^{-4}$). Variants identified in the present study remained significant even after controlling for *ADH1B* SNP, which shows that these genes have independent effects on alcohol consumption.

A major problem in combining alcohol consumption data across studies has been the variability in the way these data have been defined and measured (Agrawal et al. 2012). A strength of the present study is that the meta-analysis incorporated two datasets in which the

same assessment tool and quantitative measure of excessive alcohol consumption was employed. Despite the same assessment tool, a potential weakness of our study is the differences in ascertainment between the two studies: multiply affected families versus a case control design. These two designs may enrich for different risk factors. In particular, the large families in COGA may be enriched for rare highly penetrant variants, for which we likely have low power to replicate in a case control design.

It is perhaps not surprising that the present study was unable to identify genome-wide signals for this quantitative trait. The genome-wide significance level of 5×10^{-8} is conservative for large EA families such as those in COGA, where the number of independent tests was small due to extended linkage disequilibrium (LD). Our power calculation showed that in the combined COGA and SAGE sample we had good power (~90%) to detect SNPs explaining approximately 1% of the variance.

Several SNPs identified as candidates in a previously reported meta-analysis and GWAS of alcohol consumption also showed nominal evidence of replication in the present study. Statistics for the SNPs that showed the strongest association with alcohol consumption in previous studies were extracted from the present meta-analysis. SNPs in autism susceptibility candidate 2 *(AUTS2)* (Schumann et al. 2011), inaD-like (*INADL),* chromosome 15 open reading frame 32 *(C15orf32),* and huntingtin interacting protein (*HIP1)* (Heath et al. 2011) were nominally associated with maxdrinks ($p < 0.05$) in the present dataset.

The present study has identified several novel candidate genes that may influence alcohol consumption, and suggests that other SNPs among the top hits ($p<10^{-4}$) may also be true causal factors. Using the GCTA program, we found that nearly 40% of the variance for maxdrinks can be explained by considering all SNPs simultaneously on the Illumina 1M chip. The strongest SNPs in the present study explain a very small proportion of this variance. This shows that part of the missing heritability for maxdrinks is tagged by SNPs on the 1M chip, but due to stringent p value thresholds for GWAS, we likely ignored some real signals, underscoring the need to find effective ways to extract meaningful genetic data from the noise. Another concern is that variants like rs1229984, which explain a relatively large fraction of the variance, are not present on most genotyping chips because of their relatively low frequency in European populations (MAF=3%). There may be other rare and novel variants like this that influence maxdrinks, which are not tagged by common SNPs and are difficult to impute. To identify these variants, whole genome, exome or targeted sequencing in large numbers of people will be needed. Further biological affirmation and replication in independent datasets will be required to confirm the role of the genes identified in this study in alcohol consumption traits such as maxdrinks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Collaborative Study on the Genetics of Alcoholism (COGA): COGA, Principal Investigators B Porjesz, V Hesselbrock, H Edenberg, L Bierut includes ten different centers: University of Connecticut (V Hesselbrock); Indiana University (HJ Edenberg, J Nurnberger Jr, T Foroud); University of Iowa (S Kuperman, J Kramer); SUNY Downstate (B Porjesz); Washington University in Saint Louis (L Bierut, A Goate, J Rice, K Bucholz); University of California at San Diego (M Schuckit); Rutgers University (J Tischfield); Southwest Foundation (L Almasy), Howard University (R Taylor) and Virginia Commonwealth University (D Dick). A Parsian and M Reilly are the NIAAA Staff Collaborators. We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and CoPI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, currently a consultant with COGA, P Michael Conneally, Raymond Crowe and Wendy Reich, for their

critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA).

The Study of Addiction: Genetics and Environment (SAGE): Funding support for SAGE was provided through the NIH Genes, Environment and Health Initiative (GEI) (U01 HG004422). SAGE is one of the GWAS funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of data sets and samples was provided by COGA (U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392) and the Family Study of Cocaine Dependence (FSCD; R01 DA013423, R01 DA019963). Genotyping at the Johns Hopkins University Center for Inherited Disease Research was supported by the NIH GEI (U01HG004438) Grant, NIAAA, NIDA and the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease'

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4): 248–249. nmeth0410-248 [pii]. 10.1038/nmeth0410-248 [PubMed: 20354512]

Agrawal A, Freedman ND, Cheng YC, Lin P, Shaffer JR, Sun Q, Taylor K, Yaspan B, Cole JW, Cornelis MC, DeSensi RS, Fitzpatrick A, Heiss G, Kang JH, O'Connell J, Bennett S, Bookman E, Bucholz KK, Caporaso N, Crout R, Dick DM, Edenberg HJ, Goate A, Hesselbrock V, Kittner S, Kramer J, Nurnberger JI Jr, Qi L, Rice JP, Schuckit M, van Dam RM, Boerwinkle E, Hu F, Levy S, Marazita M, Mitchell BD, Pasquale LR, Bierut LJ. Measuring alcohol consumption for genomic meta-analyses of alcohol intake: opportunities and challenges. Am J Clin Nutr. 2012; 95(3):539–547.10.3945/ajcn.111.015545 [PubMed: 22301922]

Allen NE, Beral V, Casabonne D, Kan SW, Reeves GK, Brown A, Green J. Moderate alcohol intake and cancer incidence in women. J Natl Cancer Inst. 2009; 101(5):296–305.10.1093/jnci/djn514 [PubMed: 19244173]

Baik I, Cho NH, Kim SH, Han BG, Shin C. Genome-wide association studies identify genetic loci related to alcohol consumption in Korean men. Am J Clin Nutr. 2011; 93(4):809–816.10.3945/ajcn.110.001776 [PubMed: 21270382]

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman S, Lynskey M, Mann K, Neuman RJ, Nothen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP. A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A. 2010; 107(11): 5082–5087. 0911109107 [pii]. 10.1073/pnas.0911109107 [PubMed: 20202923]

Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, Agrawal A, Bucholz KK, Grucza R, Hesselbrock V, Kramer J, Kuperman S, Nurnberger J, Porjesz B, Saccone NL, Schuckit M, Tischfield J, Wang JC, Foroud T, Rice JP, Edenberg HJ. ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Mol Psychiatry. 2012; 17(4):445–450. mp2011124 [pii]. 10.1038/mp.2011.124 [PubMed: 21968928]

Brooks PJ, Zakhari S. Moderate alcohol consumption and breast cancer in women: from epidemiology to mechanisms and interventions. Alcohol Clin Exp Res. 2013; 37(1):23–30.10.1111/j.1530-0277.2012.01888.x [PubMed: 23072454]

Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007; 81(5):1084–1097. S0002-9297(07)63882-8 [pii]. 10.1086/521987 [PubMed: 17924348]

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol. 1994; 55 (2):149–158. [PubMed: 8189735]

Caan, W.; De Belleroche, J. Drink, Drugs and Dependence: From Science to Clinical Practice. Taylor & Francis Group; 2002.

Chen CC, Lu RB, Chen YC, Wang MF, Chang YC, Li TK, Yin SJ. Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. Am J Hum Genet. 1999; 65(3):795–807.10.1086/302540 [PubMed: 10441588]

Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. Bioinformatics. 2010; 26(4):580–581. btp710 [pii]. 10.1093/bioinformatics/btp710 [PubMed: 20040588]

Chen WY, Rosner B, Hankinson SE, Colditz GA, Willett WC. Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. JAMA: the journal of the American Medical Association. 2011; 306(17):1884–1890.10.1001/jama.2011.1590 [PubMed: 22045766]

Dudley KJ, Li X, Kobor MS, Kippin TE, Bredy TW. Epigenetic mechanisms mediating vulnerability and resilience to psychiatric disorders. Neurosci Biobehav Rev. 2011; 35(7):1544–1551. S0149-7634(11)00008-X [pii]. 10.1016/j.neubiorev.2010.12.016 [PubMed: 21251925]

Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. Alcohol research & health: the journal of the National Institute on Alcohol Abuse and Alcoholism. 2007; 30 (1):5–13. [PubMed: 17718394]

Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, Bierut LJ, Bucholz KK, Goate A, Aliev F, Dick D, Hesselbrock V, Hinrichs A, Kramer J, Kuperman S, Nurnberger JI Jr, Rice JP, Schuckit MA, Taylor R, Todd Webb B, Tischfield JA, Porjesz B, Foroud T. Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol Clin Exp Res. 2010; 34(5):840–852. ACER1156 [pii]. 10.1111/j.1530-0277.2010.01156.x [PubMed: 20201924]

Grant JD, Agrawal A, Bucholz KK, Madden PA, Pergadia ML, Nelson EC, Lynskey MT, Todd RD, Todorov AA, Hansell NK, Whitfield JB, Martin NG, Heath AC. Alcohol consumption indices of genetic risk for alcohol dependence. Biol Psychiatry. 2009; 66(8):795–800.10.1016/j.biopsych.2009.05.018 [PubMed: 19576574]

Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, McEvoy BP, Schrage AJ, Grant JD, Chou YL, Zhu R, Henders AK, Medland SE, Gordon SD, Nelson EC, Agrawal A, Nyholt DR, Bucholz KK, Madden PA, Montgomery GW. A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. Biol Psychiatry. 2011; 70(6):513–518. S0006-3223(11)00210-1 [pii]. 10.1016/j.biopsych.2011.02.028 [PubMed: 21529783]

Hurley TD, Edenberg HJ. Genes encoding enzymes involved in ethanol metabolism. Alcohol research: current reviews. 2012; 34 (3):339–344. [PubMed: 23134050]

Jiang H, Wu D, Simon MI. Activation of phospholipase C beta 4 by heterotrimeric GTP-binding proteins. The Journal of biological chemistry. 1994; 269 (10):7593–7596. [PubMed: 8125982]

Kendler KS, Kalsi G, Holmans PA, Sanders AR, Aggen SH, Dick DM, Aliev F, Shi J, Levinson DF, Gejman PV. Genomewide association analysis of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. Alcohol Clin Exp Res. 2011; 35(5):963–975.10.1111/j.1530-0277.2010.01427.x [PubMed: 21314694]

Kendler KS, Myers J, Dick D, Prescott CA. The relationship between genetic influences on alcohol dependence and on patterns of alcohol consumption. Alcohol Clin Exp Res. 2010; 34(6):1058–1065.10.1111/j.1530-0277.2010.01181.x [PubMed: 20374207]

Lasek AW, Giorgetti F, Berger KH, Tayor S, Heberlein U. Lmo genes regulate behavioral responses to ethanol in Drosophila melanogaster and the mouse. Alcohol Clin Exp Res. 2011; 35(9):1600–1606.10.1111/j.1530-0277.2011.01506.x [PubMed: 21599714]

Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010; 87(1):139–145. S0002-9297(10)00312-5 [pii]. 10.1016/j.ajhg.2010.06.009 [PubMed: 20598278]

Mattick JS. The genetic signatures of noncoding RNAs. PLoS Genet. 2009; 5(4):e1000459.10.1371/journal.pgen.1000459 [PubMed: 19390609]

Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA: the journal of the American Medical Association. 2004; 291(10):1238–1245.10.1001/jama.291.10.1238 [PubMed: 15010446]

Muller D, Koch RD, von Specht H, Volker W, Munch EM. Neurophysiologic findings in chronic alcohol abuse. Psychiatrie, Neurologie, und medizinische Psychologie. 1985; 37 (3):129–132.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38(8): 904–909.10.1038/ng1847 [PubMed: 16862161]

Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003; 19 (1):149–150. [PubMed: 12499305]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–575. S0002-9297(07)61352-4 [pii]. 10.1086/519795 [PubMed: 17701901]

Saccone NL, Kwon JM, Corbett J, Goate A, Rochberg N, Edenberg HJ, Foroud T, Li TK, Begleiter H, Reich T, Rice JP. A genome screen of maximum number of drinks as an alcoholism phenotype. Am J Med Genet. 2000; 96 (5):632–637. [PubMed: 11054770]

Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, Desrivieres S, Aliev FA, Khan AA, Amin N, Aulchenko YS, Bakalkin G, Bakker SJ, Balkau B, Beulens JW, Bilbao A, de Boer RA, Beury D, Bots ML, Breetvelt EJ, Cauchi S, Cavalcanti-Proenca C, Chambers JC, Clarke TK, Dahmen N, de Geus EJ, Dick D, Ducci F, Easton A, Edenberg HJ, Esko T, Fernandez-Medarde A, Foroud T, Freimer NB, Girault JA, Grobbee DE, Guarrera S, Gudbjartsson DF, Hartikainen AL, Heath AC, Hesselbrock V, Hofman A, Hottenga JJ, Isohanni MK, Kaprio J, Khaw KT, Kuehnel B, Laitinen J, Lobbens S, Luan J, Mangino M, Maroteaux M, Matullo G, McCarthy MI, Mueller C, Navis G, Numans ME, Nunez A, Nyholt DR, Onland-Moret CN, Oostra BA, O'Reilly PF, Palkovits M, Penninx BW, Polidoro S, Pouta A, Prokopenko I, Ricceri F, Santos E, Smit JH, Soranzo N, Song K, Sovio U, Stumvoll M, Surakk I, Thorgeirsson TE, Thorsteinsdottir U, Troakes C, Tyrfingsson T, Tonjes A, Uiterwaal CS, Uitterlinden AG, van der Harst P, van der Schouw YT, Staehlin O, Vogelzangs N, Vollenweider P, Waeber G, Wareham NJ, Waterworth DM, Whitfield JB, Wichmann EH, Willemsen G, Witteman JC, Yuan X, Zhai G, Zhao JH, Zhang W, Martin NG, Metspalu A, Doering A, Scott J, Spector TD, Loos RJ, Boomsma DI, Mooser V, Peltonen L, Stefansson K, van Duijn CM, Vineis P, Sommer WH, Kooner JS, Spanagel R, Heberlein UA, Jarvelin MR, Elliott P. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. Proc Natl Acad Sci U S A. 2011; 108(17):7119–7124.10.1073/pnas.1017288108 [PubMed: 21471458]

Testino G. Alcoholic diseases in hepato-gastroenterology: a point of view. Hepato-gastroenterology. 2008; 55 (82–83):371–377. [PubMed: 18613369]

Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li TK, Wang SP, Lin YT, Lu RB, Yin SJ. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. Am J Hum Genet. 1991; 48 (4):677–681. [PubMed: 2014795]

Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, Ludwig KU, Frank J, Wichmann HE, Schreiber S, Dragano N, Sommer WH, Leonardi-Essmann F, Lourdusamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nothen MM, Kiefer F, Spanagel R, Mann K, Rietschel M. Genome-wide association study of alcohol dependence. Arch Gen Psychiatry. 2009; 66(7):773–784. 66/7/773 [pii]. 10.1001/archgenpsychiatry.2009.83 [PubMed: 19581569]

Wang JC, Foroud T, Hinrichs AL, Le NX, Bertelsen S, Budde JP, Harari O, Koller DL, Wetherill L, Agrawal A, Almasy L, Brooks AI, Bucholz K, Dick D, Hesselbrock V, Johnson EO, Kang S, Kapoor M, Kramer J, Kuperman S, Madden PA, Manz N, Martin NG, McClintick JN, Montgomery GW, Nurnberger JI Jr, Rangaswamy M, Rice J, Schuckit M, Tischfield JA, Whitfield JB, Xuei X, Porjesz B, Heath AC, Edenberg HJ, Bierut LJ, Goate AM. A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53. Mol Psychiatry. 2012a mp2012143 [pii]. 10.1038/mp.2012.143

Wang JC, Kapoor M, Goate AM. The genetics of substance dependence. Annu Rev Genomics Hum Genet. 2012b; 13:241–261.10.1146/annurev-genom-090711-163844 [PubMed: 22703173]

Wang K, Diskin SJ, Zhang H, Attiyeh EF, Winter C, Hou C, Schnepp RW, Diamond M, Bosse K, Mayes PA, Glessner J, Kim C, Frackelton E, Garris M, Wang Q, Glaberson W, Chiavacci R,

Nguyen L, Jagannathan J, Saeki N, Sasaki H, Grant SF, Iolascon A, Mosse YP, Cole KA, Li H, Devoto M, McGrady PW, London WB, Capasso M, Rahman N, Hakonarson H, Maris JM. Integrative genomics identifies LMO1 as a neuroblastoma oncogene. Nature. 2011; 469(7329): 216–220. nature09609 [pii]. 10.1038/nature09609 [PubMed: 21124317]

Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40(Database issue):D930–934. gkr917 [pii]. 10.1093/nar/gkr917 [PubMed: 22064851]

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26(17):2190–2191. btq340 [pii]. 10.1093/bioinformatics/btq340 [PubMed: 20616382]

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88(1):76–82. S0002-9297(10)00598-7 [pii]. 10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]

Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, Maouche S, Germain M, Lackner K, Rossmann H, Eleftheriadis M, Sinning CR, Schnabel RB, Lubos E, Mennerich D, Rust W, Perret C, Proust C, Nicaud V, Loscalzo J, Hubner N, Tregouet D, Munzel T, Ziegler A, Tiret L, Blankenberg S, Cambien F. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One. 2010; 5(5):e10693.10.1371/journal.pone.0010693 [PubMed: 20502693]

Table 1

Characteristics of COGA and SAGE subjects

	COGA (118 families)		SAGE EA (N = 2593)[a]	
	Male	**Female**	**Male**	**Female**
N	1101	1221	1153	1440
Age at Interview (Mean + s.d.) yrs	34.95 ± 15.47	36.47 ± 15.78	39.41 ± 9.72	38.66 ± 8.44
maxdrinks (Mean + s.d.) max number of drinks in 24 hrs	23.72 ± 18.09[b]	12.52 ± 12.68[b]	25.48 ± 19.79[c]	13.10 ± 14.23[c]

[a]The demographic data reported for SAGE is for 2593 subjects after removing 129 subjects which overlapped with COGA

[b]two-tailed p for t-test <0.001

[c]two-tailed p for t-test < 0.001

Table 2

Summary of results for meta-analysis of COGA and SAGE datasets ($p \leq 1.00 \times 10^{-5}$)

SNP	CHR[a]	BP[b]	Gene	A1	A2	MAF[c]	COGA p	SAGE p	Meta p
rs1229984[D]	4	100239319	*ADH1B*	T	C	0.03	1.20E-02	4.32E-07	2.04E-08
rs4758317	11	8250811	*LMO1*	C	A	0.42	2.85E-02[G]	5.97E-06	7.20E-07
rs59677118	9	14441677	*NA*	A	G	0.06	7.83E-03	4.70E-05	1.16E-06
rs55731057	4	38798935	*NA*	T	C	0.25	6.69E-02	1.17E-05	3.31E-06
rs67031482	2	198911166	*PLCL1*	C	T	0.48	1.06E-02	1.22E-04	4.07E-06
rs5743604	4	38801285	*TLR1*	G	A	0.24	6.27E-02	1.63E-05	4.09E-06
rs1353899	3	177228979	*NA*	G	T	0.22	4.22E-02[G]	2.45E-05	4.21E-06
rs745899	2	198908040	*PLCL1*	A	T	0.48	1.06E-02	1.34E-04	4.24E-06
rs10497813	2	198914072	*PLCL1*	G	T	0.48	1.04E-02	1.42E-04[G]	4.36E-06
rs2117339	2	198915734	*PLCL1*	C	T	0.48	1.02E-02	1.41E-04	4.40E-06
rs7144649	14	57822216	*NA*	G	A	0.23	1.15E-01[G]	5.76E-06[G]	4.40E-06
rs12329164	2	198905224	*PLCL1*	C	G	0.48	1.06E-02	1.39E-04	4.49E-06
rs10206714	2	198906481	*PLCL1*	C	T	0.48	1.06E-02	1.37E-04	4.50E-06
rs6732340	2	198912952	*PLCL1*	C	G	0.48	1.09E-02	1.41E-04	4.63E-06
rs1392219	3	177222429	*NA*	T	C	0.22	3.68E-02	3.72E-05[G]	5.00E-06
rs59972978	20	56058355	*NA*	T	A	0.20	1.32E-03	1.02E-03	5.21E-06
rs2196174	2	198905172	*PLCL1*	A	G	0.48	1.06E-02	1.73E-04	5.49E-06
rs4833095	4	38799710	*TLR1*	C	T	0.26	8.11E-02[G]	1.50E-05[G]	5.55E-06
rs1371664	2	198904465	*PLCL1*	T	A	0.48	1.06E-02	1.80E-04	5.61E-06
rs9712275	2	198907143	*PLCL1*	C	T	0.48	1.42E-02	1.31E-04	5.77E-06
rs11690205	2	198933948	*PLCL1*	C	T	0.48	8.26E-03	2.44E-04	5.87E-06
rs10497811	2	198902909	*PLCL1*	A	C	0.48	1.06E-02	1.80E-04	5.94E-06
rs7587251	2	198930197	*PLCL1*	T	G	0.48	8.26E-03	2.41E-04	6.06E-06
rs962210	2	198902230	*PLCL1*	A	T	0.48	1.09E-02	1.83E-04	6.11E-06
rs35062652	2	198927927	*PLCL1*	C	T	0.47	9.29E-03	2.29E-04	6.13E-06

Hum Genet. Author manuscript; available in PMC 2014 October 01.

SNP	CHR[a]	BP[b]	Gene	A1	A2	MAF[c]	COGA p	SAGE p	Meta p
rs11690149	2	198933804	*PLCL1*	C	T	0.48	8.26E-03	2.49E-04	6.18E-06
rs16985179	22	28029538	*NA*	T	C	0.11	2.57E-04[G]	3.37E-03[G]	6.23E-06
rs17616434	4	38812876	*NA*	C	T	0.26	7.17E-02[G]	2.17E-05[G]	6.71E-06
chr2:198911426	2	198911426	*PLCL1*	G	A	0.48	1.06E-02	2.13E-04	6.73E-06
rs36061340	8	101807230	*NA*	T	C	0.05	4.43E-02	5.00E-05	6.81E-06
rs7572733	2	198929806	*PLCL1*	C	T	0.48	7.23E-03	3.25E-04[G]	6.97E-06
rs2196175	2	198905073	*PLCL1*	T	A	0.48	1.06E-02	2.18E-04	6.98E-06
rs7553212	1	216738788	*ESRRG*	G	A	0.32	7.25E-02[G]	2.32E-05[G]	7.14E-06
rs1583792	2	198900288	*PLCL1*	C	T	0.48	1.43E-02	1.70E-04	7.26E-06
rs1866666	2	198940607	*PLCL1*	T	C	0.48	1.05E-02	2.47E-04	7.79E-06
rs6834581	4	38788234	*NA*	C	T	0.25	3.79E-02	6.67E-05	8.18E-06
rs9973400	2	198941578	*PLCL1*	T	C	0.48	9.95E-03	2.82E-04	8.29E-06
chr7:20089751	7	20089751	*NA*	A	G	0.09	5.93E-03	4.71E-04	8.54E-06
rs62202398	20	46801224	*NA*	A	G	0.06	3.81E-01	1.69E-06	8.62E-06
rs11851015	14	57669533	*EXOC5*	G	A	0.15	3.01E-03[G]	8.55E-04	8.64E-06
rs4543123	4	38792524	*NA*	G	A	0.24	3.34E-02	8.26E-05[G]	8.75E-06
rs2188561	7	107336058	*SLC26A4*	A	C	0.21	1.47E-01	9.17E-06	9.23E-06
chr14:101189019	14	101189019	*NA*	T	C	0.16	1.83E-02	1.79E-04	9.61E-06

[a] Chromosome;

[b] Chromosomal position (base pairs) based on human genome build 19, dbSNP 137

[c] Minor allele frequency estimated on founders of COGA dataset;

[D] *ADH1B* variant with MAF < 0.05 genotyped in COGA and SAGE datasets;

[G] genotyped SNPs

Hum Genet. Author manuscript; available in PMC 2014 October 01.