Genome-wide association studies identify genetic loci related to alcohol consumption in Korean men

Ikyung Baik, Nam H Cho, Seong Hwan Kim, Bok-Ghee Han, and Chol Shin

ABSTRACT

Background: Genome-wide association (GWA) studies regarding the quantitative trait of alcohol consumption are limited.

Objective: The objective of the study was to explore genetic loci associated with the amount of alcohol consumed.

Design: We conducted a GWA study with discovery data on single nucleotide polymorphisms (SNPs) for 1721 Korean male drinkers aged 40–69 y who were included in an urban population–based cohort. Another sample that comprised 1113 male drinkers who were from an independent cohort enrolled in a rural area served as a resource for replication. At baseline (18 June 2001 through 29 January 2003), members of both cohorts provided information on average daily alcohol consumptions, and their DNA samples were collected for genotyping.

Results: We tested 315,914 SNPs of discovery data by using multivariate linear regression analysis adjusted for age and smoking, and 12 SNPs on chromosome 12q24 had genome-wide significant associations with alcohol consumption; adjusted $P$ values by using Bonferroni correction were $1.6 \times 10^{-3}$ through $5.8 \times 10^{-46}$. We observed most SNPs in intronic regions and showed that the genes that harbor SNPs were C12orf51, CCDC63, MYL2, OAS3, CUX2, and RPH3A. In particular, signals in or near C12orf51, CCDC63, and MYL2 were successfully replicated in the test for 317,951 SNPs; rs2074356 in C12orf51 was in high linkage disequilibrium with SNPs in ALDH2, but other SNPs were not.

Conclusions: In a GWA study, we identified loci and alleles highly associated with alcohol consumption. The findings suggest the need for further investigations on the genetic propensity for drinking excessive amounts of alcohol. Am J Clin Nutr doi: 10.3945/ajcn.110.001776.

INTRODUCTION

Heavier alcohol consumption is associated with increased risk of developing alcohol dependence (AD) in a dose-response relation (1). Although the quantity of drinking is not a diagnostic element of AD (2), heavy drinking of $\geq 3$ drinks/d significantly increased AD risk in men (1). Because heritability estimates of AD of $\approx 50\%$ appear to be nontrivial (3), intensive investigations on genetic loci for alcohol-related traits have been carried out (4, 5) and only a few studies considered the amount of alcohol consumed as a quantitative phenotype (6–8). In linkage-analysis studies, loci related to average daily alcohol consumption were shown on chromosomes 9, 15, and 16, and loci for maximum alcohol consumption were observed on chromosomes 4 and 9 (6–8). Accumulated data from genome-wide linkage studies for AD agree on chromosome 4 (9–12) but have suggested different regions, such as chromosomes 1, 2, 3, 7, and 12, where genomic markers of AD are located (9, 11, 13, 14). Candidate-gene investigations have also provided supportive evidence for chromosome 4 in the association with alcohol consumption and dependence and showed polymorphic positions of the alcohol dehydrogenase gene cluster (15, 16) and the $\gamma$-aminobutyric acid A receptor gene cluster (17, 18). However, regarding other suggested loci, particular genes located on each chromosome are still largely unidentified. Taken together, previous findings from linkage-mapping and candidate-gene approaches suggest that multiple genes may play a role in determining the propensity of alcohol consumption and dependence albeit providing inconsistent and limited information. A genome-wide association (GWA) approach is an emerging methodology that enables the identification of genetic loci and variants associated with complex traits and diseases (19). A few GWA studies have been conducted for AD in populations with European ancestry (20–22), but a GWA study for alcohol consumption has not been reported. In a German male population that included AD cases and control subjects, an association with genome-wide significance was shown for 2 single nucleotide polymorphisms (SNPs) of chromosome 2q35 where the gene that encodes peroxisomal trans-2-enoyl-coenzyme A reductase (PECR) is located (20). Other studies reported less-substantial associations and suspected regions on chromosome 11 (21, 22). Thus, further data from GWA investigations, in particular by using alternative phenotypes relevant for AD, are still called for (4).

1 From the Department of Foods and Nutrition, College of Natural Sciences, Kookmin University, Seoul, Republic of Korea (IB); the Department of Preventive Medicine, Ajou University School of Medicine, Suwon, Republic of Korea (NH); the Department of Internal Medicine, Korea University Ansan Hospital, Ansan, Republic of Korea (SHK and CS); the Center for Genome Science, National Institutes of Health, Seoul, Republic of Korea (B-GH); and the Institute of Human Genomic Study, Korea University Ansan Hospital, Ansan, Republic of Korea (CS).

2 Supported by a National Research Foundation of Korea Grant funded by the Korean Government (2009-0070038) and a grant from the Korea Centers for Disease Control and Prevention (budgets 2001-347-6111-221, 2002-347-6111-221 and the Korean Genome Analysis Project 4845-301).

3 Address correspondence to C Shin, Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Korea University Ansan Hospital, 516 Gojandong, Danwongu, Ansan 425-707, Republic of Korea. E-mail: chol-shin@korea.ac.kr.

Received August 4, 2010. Accepted for publication December 23, 2010.
doi: 10.3945/ajcn.110.001776.
Therefore, to explore genetic loci associated with alcohol consumption, we conducted a GWA study in 1721 Korean male drinkers from an urban population–based cohort and analyzed data for 1113 male drinkers from another independent cohort that included a rural population that served as a resource for replication. For this study, we used the average daily alcohol consumption as a quantitative trait for all drinkers and additionally used the scores on the Alcohol Use Disorder Identification Test (AUDIT), which were available for 469 drinkers.

SUBJECTS AND METHODS

Study population

Two independent prospective cohort studies began in 2001 and were embedded within the Korean Genome Epidemiology Study with the goal of conducting GWA studies. To effectively achieve the goal, a multidisciplinary research consortium was developed under the Korea Association Resource project. Detailed information on the study design and procedures for each study were available in previous reports (23, 24). Each study comprised a population-based sample of male and female Koreans aged 40–69 y and from the same ethnic background, but cohort members were enrolled at the following 2 different sites: Ansan, which is an urban community with a population of 555,000, and Ansung, which is a rural community including 133,000 residents, on the basis of the 2000 census. To enroll members for each cohort, the most efficient method was used on the basis of knowledge about characteristics of each community. For enrollment at the Ansan site, 10,957 eligible subjects were identified by telephone contact on the basis of a 2-stage cluster-sampling method with the information of a governing district called Dong and demographic characteristics. Similarly, Ansung members were recruited from 5 of 11 governing districts called Myon by using a cluster-sampling method, and as a result, 7192 eligible subjects were identified by mail or telephone contact and a door-to-door visit. For the baseline health examination from 18 June 2001 to 29 January 2003, 5020 participants (2523 men and 2497 women) from Ansan and 5018 (2239 men and 2779 women) from Ansung visited the Korea University Ansan Hospital and the Ajou University Medical Center, respectively. All participated in a comprehensive health examination, interviews, and collection procedures of biospecimens for assays at each study site. In particular, participants completed an interview-based questionnaire for demographic information, medical histories, health conditions, family disease history, lifestyles including smoking and alcohol drinking, and dietary intakes. All of these procedures were administered by health professionals who were trained to follow a standardized protocol. Cohort members were followed up biennially with a scheduled site visit for similar interviews and health examinations. At each visit, participants signed an informed consent form, which was approved by the Human Subjects Review Committee at the Korea University Ansan Hospital or the Ajou University Medical Center.

In this study, we only analyzed GWA data for male drinkers because most women in the 2 cohorts were abstainers. We denoted GWA data of male drinkers from the cohort study conducted at the Korea University Ansan Hospital as discovery data and GWA data of male drinkers from the cohort study at the Ajou University Medical Center as replication data.

Alcohol consumption and AUDIT

The questionnaire information on alcohol consumption status is available elsewhere (25). Briefly, participants were asked whether they ever consumed alcoholic beverage in their lifetime and whether there was a time in their life when they regularly consumed at least one drink of any alcoholic beverage in every month. Individuals who provided positive answers to both questions were classified as drinkers. Participants were also asked to complete a table that inquired about the information on drinking amounts and patterns in the past 30 d. According to prevailing alcoholic beverages in Korea, the information on the average frequency of drinking occasions, amounts of drinks consumed for a typical occasion, and the volume of one standard drink for each type of alcoholic beverage was collected. Thus, the total daily alcohol consumption (g alcohol/d) was calculated by summing up the beverage-specific alcohol amount consumed. The average alcohol consumption per day that was calculated with the baseline information was strongly correlated with serum concentrations of 

\[ r = 0.49, P < 0.001 \]  

(Spearman’s). On the basis of similar procedures, alcohol consumption was reassessed and calculated every 2-y follow-up period. During the first half period of the fourth follow-up (May 2009 through January 2010), AUDIT, which was developed by the World Health Organization, was conducted in the cohort study of the Korea University Ansan Hospital, and its total scores were strongly correlated with the amount of alcohol consumed in 469 male drinkers (Spearman r = 0.76, P < 0.001). Within AUDIT, participants were asked a question on the alcohol flush reaction of whether or not his face turns red when taking as little as one drink of an alcoholic beverage.

Genotype data and inclusion criteria

A previous report provided detailed information on the preparation of genomic samples, genotyping method, and quality control for GWA data (23). Genomic DNA was collected from blood samples, and 500 ng DNA was processed according to the Affymetrix Genome-wide Human SNP array 5.0 (Affymetrix Inc, Santa Clara, CA). Primarily, DNA samples from 10,004 individuals were obtained from the 2 cohorts. After excluding 11 contaminated samples, genotype data were available for a total of 9993 individuals. Further exclusion was made on the basis of the following criteria: call rates <96%, sex discrepancy compared with questionnaire data from the cohort studies, cancer cases that were identified when biospecimens were collected at baseline, or average pairwise identity-by-state values >0.8 as estimated values from first-degree relatives of Korean sib-pair samples. Remaining data included 8842 men and women who were alcohol drinkers or abstainers (23). Genotype calls were made for 500,568 SNPs by using Bayesian robust linear modeling by using the Mahalanobis distance genotyping algorithm. After discriminating SNPs on the standard protocol for quality control, 352,228 SNPs remained and were used for principal component analysis to confirm the applicability of Japanese in Tokyo and Han Chinese in Beijing (JPT/CHB) HapMap reference (http://hapmap.ncbi.nlm.nih.gov/). With the use of such a reference, SNP imputation was completed to generate imputed GWA data (23).
On the basis of the original GWA data that included 352,228 SNPs, we included only male drinkers who provided baseline information for average daily alcohol consumption, and thus, 1721 men and 1113 men entered discovery data and replication data, respectively. After the evaluation of SNP call rates (>95%), the Hardy-Weinberg equilibrium test \((P > 0.0001)\), and minor allele frequency (>5%), discovery genotyping data included 315,914 SNPs for GWA analysis. With the use of the same criteria, replication data included 317,951 SNPs for analysis. In 1721 male drinkers, 27% of them completed AUDIT, and their 316,706 SNP data were also analyzed.

### Statistical analyses

Descriptive statistics on the baseline characteristics of study populations were calculated with discovery and replication data. To explore genetic loci associated with alcohol consumption, we conducted multivariate linear regression analysis with PLINK software (version 1.06; http://pngu.mgh.harvard.edu/purcell/plink/) (26). For such an analysis, we used the average daily alcohol consumption transformed by the natural logarithm after comparison of normal quantile-quantile (Q-Q) plots of the residuals with various transformations on the variable. Baseline information of age and smoking status was collected from the data set of each cohort study and was taken into account in the analysis. Because smoking status is considered to be closely linked with alcohol drinking, overadjustment might be a concern if smoking status is used as a covariate in analyses. However, because analyses with only age produced similar results, we have presented data adjusted for age and smoking status. Additive, dominant, and recessive models were applied to the analysis. For multiple testing corrections, we used Bonferroni adjustment. On the basis of association results from an additive model, we constructed Q-Q plots for discovery and replication data by comparing the distribution of observed \(P\) values for given SNPs with a theoretical distribution. To depict the plots, we used the R statistical software package (version 2.9.2; http://cran.us.r-project.org/). In addition, we examined the linkage disequilibrium (LD) with \(D^{'},\) between all pairs of biallelic SNP loci and obtained haplotypes with Haplovew software (version 4.1; http://www.broad.mit.edu/mpg/haplovew/). With the use of expectation-maximization algorithm as implemented in this program, haplotypes were reconstructed for the set of all SNPs with a \(P\) value of the Hardy-Weinberg equilibrium test set at 0.05. The amount of alcohol consumed was compared according to the categories of SNPs variants by using the analysis of covariance procedure with the Scheffe post hoc test. To conduct such an analysis, the SAS program (SAS 9.1.3; SAS Institute, Cary, NC) was used; all tests were 2-tailed, and statistical significance was set at 0.05. In the analysis for AUDIT scores, we classified drinkers into 2 groups as \(<8\) and \(\geq8\) on the basis of total scores and performed multivariate logistic regression analysis with PLINK software (version 1.06, http://pngu.mgh.harvard.edu/purcell/plink/) (26) to obtain an odds ratio and its 95% CI.

### RESULTS

The mean daily consumption of alcohol was \(\approx30\) g/d and the proportion of heavy drinkers who consumed more than 80 g/d was \(\approx7\%\) in both data sets (Table 1).

![Figure 1](image.png)

**FIGURE 1.** Manhattan plots of discovery genome-wide association data for the association with natural log-transformed average alcohol consumption per day among 1721 Korean male drinkers. The y axis shows the \(-\log_{10}\) unadjusted \(P\) values, and the squares indicate positions along the chromosome (Chr.). The solid line indicates the \(P\) value of \(1 \times 10^{-6}\).
significance level were shown on chromosome 12q24. The range of P values for significant SNPs was 1.0 × 10⁻⁷ to 3.7 × 10⁻¹⁸, whereas P values for SNPs in other chromosomes were unable to reach statistical significance.

Detailed GWA findings from additive-model analyses for natural log-transformed alcohol consumption by using discovery and replication data are shown in Table 2. Genes that harbored significant SNPs were chromosome 12 open reading frame 51 (C12orf51), coiled-coil domain containing 63 (CCDC63), ventricular regulatory myosin light chain (MYL2), 2'-5'-oligoadenylate synthetase 3 (OAS3), cut-like homebox 2 (CUX2), and rabphilin 3A homolog (RPH3A). Most SNPs were located in the introns of each gene; only rs2072134 was within the 3' untranslated region of OAS3. Among 12 genome-wide significant signals shown in discovery data, 6 SNPs were replicated, and carriers of major alleles of the SNPs were more likely to be heavy drinkers than were carriers of minor alleles. Highly significant signals were rs2074356 and rs11066280 in or near the C12orf51 gene and rs12229654 located near MYL2; in pooled data, the corresponding adjusted P values were 9.5 × 10⁻⁹, 9.3 × 10⁻⁵, and 3.8 × 10⁻³. The CCDC63 gene included 3 significant SNPs (ie, rs10849915, rs11065756, and rs2238149), and 2 of them were replicated with genome-wide significance.

In the MYL2 gene, rs3782889 was significantly associated with alcohol consumption, and this association was replicated; adjusted P values were 9.5 × 10⁻²⁰ in discovery data, 3.0 × 10⁻¹⁶ in replication data, and 1.5 × 10⁻²² in pooled data (Table 2). Similar findings were observed in results from dominant-model analyses (see supplemental Table S1 under “Supplemental data” in the online issue), whereas only rs2078521 of CUX2 was significantly associated with alcohol consumption in recessive-model analyses with discovery data.

We compared the average amount of alcohol consumed per day according to homozygous and heterozygous variants at given SNPs, which were significantly associated with alcohol consumption in discovery data (Table 3). In drinkers with common homozygous types, means of daily alcohol consumption were >30 g, which approximately corresponded to 3 drinks. In discovery data, men who had common homozygous types were more likely to be heavy drinkers than were men with heterozygote or variant homozygous types (P < 0.05; Scheffe test). Similar findings, except for the results seen with SNPs in CCDC63, CUX2, and RPH3A, were observed in replication data (Table 3).

Suggestive SNPs from additive-model analyses with pooled data were listed (see supplemental Table S2 under “Supplemental data” in the online issue); as genes of interest, there were the cholinergic receptor/muscarinic 3 gene (CHRM3) on chromosome 1, the hydroxytryptamine (serotonin) receptor 4 gene (HTR4), and the phosphodiesterase 4D/cyclicAMP-specific (PDE4D) gene on chromosome 5, and the neuronal periodic acid Schiff (PAS) domain protein 3 (NPAS3) gene on chromosome 14, which have been implicated in genomic investigations for drug addiction (27, 28). In the list, most promising SNPs with their allelic contributions were rs2072133 of OAS3, rs2999918, rs1016078, rs16941759, and rs1016079 of transmembrane protein 116 (TMEM116), rs4766660, rs10850088, rs4767017, and rs2384069 of RPH3A, rs11066211 of C12orf51, rs2078851 of CUX2, rs933296 of MYL2, rs2238153 of ataxin 2 (ATXN2), rs12321677 and rs9971746 of a disintegrin and metalloproteinase (ADAM)
metallopeptidase domain 1 (ADAM1, pseudogene), and rs12423041 and rs16941724 of mitogen-activated protein kinase-activated protein kinase 5 (MAPKAPK5) (see supplemental Table S2 under “Supplemental data” in the online issue).

With the use of discovery and imputed data, we examined LD and focused on the significant loci in this study. We observed that rs2074356 and rs11066280 in or near C12orf51 were in high LD, and SNPs in C12orf51, TMEM116, and MAPKAPK5, which are in the vicinity to the aldehyde dehydrogenase gene (ALDH2), were also in high LD (Figure 2). In imputed data, several intronic SNPs in ALDH2, such as rs2238151, rs11069628, rs4648328, rs4646777, rs10744777, rs4646778, rs2158029, rs16941669, rs11066029, and rs7296651, were observed to be in high LD (D’ > 0.99) with rs2074356 in C12orf51 (see supplemental Figure S2 under “Supplemental data” in the online issue). A previous study in European young adults demonstrated that intronic SNPs such as rs2238151 and rs4648328 in ALDH2 are in substantial LD, and their variants are associated with delayed alcohol metabolism (29). In contrast, SNPs in the CCDC63 gene and the MYL2 gene are a fair distance away from the region that contains the ALDH2 and C12orf51 genes. Among significant SNPs in CCDC63 and MYL2, rs10849915, rs11065756, and rs3782889 are in high LD (D’ > 0.99) (see supplemental Figures S2 and S3 under “Supplemental data” in the online issue).

We performed analyses for AUDIT scores in 469 drinkers who were included in discovery data. The odds ratio of having total scores ≥8, which indicated alcohol problems, was 0.15 for carriers with a heterozygote type of rs11066280 compared with subjects who had a common homozygote type. A similar odds ratio was observed for rs11066280 located near C12orf51. These 2 SNPs in or near C12orf51 were significantly associated with AUDIT scores even after multiple testing correction (Table 4). Furthermore, rs11066280 was highly associated with the alcohol flush reaction, which was asked of 469 drinkers. Carriers with a heterozygote type of rs11066280 had

### Table 3

Comparison of average daily alcohol consumption (in g) according to genotypes of significant single nucleotide polymorphisms (SNPs)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>M/m alleles</th>
<th>Study data</th>
<th>Common homozygote</th>
<th>Heterozygote</th>
<th>Variant homozygote</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2074356</td>
<td>C12orf51</td>
<td>T/C</td>
<td>Discovery data</td>
<td>32.1 (29.6, 34.5)</td>
<td>15.9 (9.9, 21.2)</td>
<td>6.3 (47.9, 60.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>33.0 (29.6, 36.3)</td>
<td>15.7 (7.8, 23.6)</td>
<td>0.89 (102.0, 103.7)</td>
</tr>
<tr>
<td>rs11066280</td>
<td>C12orf51</td>
<td>A/T</td>
<td>Discovery data</td>
<td>32.4 (30.4, 34.4)</td>
<td>15.6 (11.0, 20.2)</td>
<td>4.9 (43.8, 53.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>33.4 (29.9, 36.8)</td>
<td>17.2 (10.0, 24.3)</td>
<td>22.4 (16.4, 61.3)</td>
</tr>
<tr>
<td>rs1229654</td>
<td>MYL2</td>
<td>T/G</td>
<td>Discovery data</td>
<td>32.8 (30.7, 34.8)</td>
<td>17.4 (13.3, 21.6)</td>
<td>16.0 (8.4, 40.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.6 (29.1, 31.4)</td>
<td>18.9 (13.5, 24.4)</td>
<td>6.7 (40.6, 53.9)</td>
</tr>
<tr>
<td>rs10849915</td>
<td>CCDC63</td>
<td>A/G</td>
<td>Discovery data</td>
<td>32.3 (29.7, 35.0)</td>
<td>21.0 (16.2, 25.8)</td>
<td>14.6 (3.3, 32.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.8 (28.1, 35.4)</td>
<td>26.3 (20.0, 32.5)</td>
<td>25.1 (1.2, 49.1)</td>
</tr>
<tr>
<td>rs11065756</td>
<td>CCDC63</td>
<td>G/A</td>
<td>Discovery data</td>
<td>32.3 (29.7, 34.9)</td>
<td>20.6 (15.8, 25.3)</td>
<td>13.6 (6.1, 33.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.9 (28.3, 35.6)</td>
<td>25.8 (19.6, 32.1)</td>
<td>25.1 (1.2, 49.0)</td>
</tr>
<tr>
<td>rs3782889</td>
<td>MYL2</td>
<td>T/C</td>
<td>Discovery data</td>
<td>32.1 (30.0, 34.3)</td>
<td>23.2 (19.4, 27.0)</td>
<td>18.9 (4.5, 33.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.7 (29.6, 33.9)</td>
<td>22.7 (18.8, 26.6)</td>
<td>18.8 (3.6, 34.0)</td>
</tr>
<tr>
<td>rs2238149</td>
<td>CCDC63</td>
<td>A/G</td>
<td>Discovery data</td>
<td>32.2 (29.6, 34.8)</td>
<td>19.8 (14.9, 24.7)</td>
<td>22.0 (1.4, 42.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.1 (27.5, 34.7)</td>
<td>26.9 (20.4, 33.4)</td>
<td>40.3 (16.4, 64.3)</td>
</tr>
<tr>
<td>rs2072134</td>
<td>OAS3</td>
<td>G/A</td>
<td>Discovery data</td>
<td>30.7 (28.2, 33.2)</td>
<td>21.6 (15.6, 27.6)</td>
<td>22.0 (6.6, 50.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.8 (28.4, 35.2)</td>
<td>20.6 (12.5, 28.8)</td>
<td>36.8 (10.4, 103.2)</td>
</tr>
<tr>
<td>rs11066453</td>
<td>OAS3</td>
<td>A/G</td>
<td>Discovery data</td>
<td>31.2 (29.1, 33.2)</td>
<td>21.2 (16.4, 26.1)</td>
<td>32.9 (8.3, 57.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.9 (28.5, 35.5)</td>
<td>25.4 (19.1, 31.7)</td>
<td>25.1 (1.4, 48.9)</td>
</tr>
<tr>
<td>rs11065783</td>
<td>MYL2</td>
<td>A/G</td>
<td>Discovery data</td>
<td>31.0 (28.1, 34.0)</td>
<td>27.7 (23.8, 31.6)</td>
<td>18.7 (7.7, 29.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>32.9 (29.0, 36.9)</td>
<td>27.1 (21.9, 32.4)</td>
<td>16.2 (0.3, 32.0)</td>
</tr>
<tr>
<td>rs4766553</td>
<td>CUX2</td>
<td>A/C</td>
<td>Discovery data</td>
<td>32.2 (28.1, 36.3)</td>
<td>30.1 (26.7, 33.4)</td>
<td>23.5 (18.6, 28.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>31.0 (25.1, 36.9)</td>
<td>30.8 (26.4, 35.2)</td>
<td>28.0 (21.1, 34.9)</td>
</tr>
<tr>
<td>rs7958347</td>
<td>RPH3A</td>
<td>T/C</td>
<td>Discovery data</td>
<td>31.7 (28.4, 35.1)</td>
<td>30.4 (27.7, 33.0)</td>
<td>25.3 (21.2, 29.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication data</td>
<td>30.7 (25.9, 35.4)</td>
<td>30.8 (26.2, 35.4)</td>
<td>26.8 (17.5, 36.2)</td>
</tr>
</tbody>
</table>

1 All values are means (95% CIs) and were obtained from a general linear model (ANCOVA) that contained age and smoking status (never smoker, former smoker, or current smoker of ≤10, 10.1–20, or >20 cigarettes/d). M/m, major and minor. Across groups, values in the same row with different superscript letters were significantly different, P < 0.05 (Scheffe test).
a 13.4-times higher odds (95% CI: 7.6–23.2) of an alcohol flush reaction than carriers with a common homozygous type (data available from authors on request). Carriers with a heterozygous type of rs12229654 in MYL2, compared with carriers who had a common homozygous type, also showed a lower odds ratio (0.23; 95% CI: 0.13, 0.40) of having higher AUDIT scores, but the association of this particular signal became insignificant after Bonferroni adjustment.

DISCUSSION

The current GWA study showed loci and variants associated with the amount of alcohol consumed. Significant GWA signals were observed in or near several genes on chromosome 12q24, such as C12orf51, CCDC63, MYL2, OAS3, CUX2, and RPH3A, in discovery data that included 1721 male drinkers. In particular, SNPs in or near C12orf51, CCDC63, and MYL2 were successfully replicated in 1113 male drinkers. Carriers with minor alleles of significant signals showed reduced alcohol consumption. Furthermore, additional analyses in 456 drinkers from discovery data revealed that signals in or near C12orf51 were significantly associated with AUDIT scores. In LD analyses with imputed data, SNPs in C12orf51 and ALDH2 were in high LD, whereas those in CCDC63 and MYL2 that shared substantial LD were in a different LD block far from the region of C12orf51 and ALDH2.

The main strengths of our investigation were the use of GWA data from population-based cohorts, the collection of information on alcohol consumption by using a structured questionnaire and interviews, and the investigation in drinkers from a general

![FIGURE 2. Association analysis of single nucleotide polymorphisms (SNPs) in C12orf51, CCDC63, MYL2, and OAS3 on chromosome 12q24 with the use of discovery genome-wide association (GWA) data. (A) P values for the association testing of discovery GWA data with natural log-transformed alcohol consumption. The association was drawn from linear regression analysis adjusted for age and smoking status on the basis of the additive model. The y axis shows the \(-\log_{10}\) unadjusted P values, and the x axis shows the physical position of chromosome 12q24. Red diamonds indicate highly significant SNPs. (B) Gene locations for strong signals that correspond to (A) and (C). (C) A linkage disequilibrium (LD) map that shows the associations of 2 SNPs; red dots indicate highly significant SNPs, and blocks show strong LD signals.](image-url)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Cytologic location</th>
<th>Position</th>
<th>Nearest gene</th>
<th>M/m alleles</th>
<th>MAF</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11066280</td>
<td>12q24.13</td>
<td>111302166</td>
<td>C12orf51</td>
<td>A/T</td>
<td>0.17</td>
<td>0.18 (0.11, 0.31)</td>
<td>5.03×10^{-05}</td>
</tr>
<tr>
<td>rs2074356</td>
<td>12q24.13</td>
<td>111129784</td>
<td>C12orf51</td>
<td>C/T</td>
<td>0.15</td>
<td>0.15 (0.08, 0.28)</td>
<td>1.17×10^{-03}</td>
</tr>
</tbody>
</table>

1 M/m, major and minor; MAF, minor allele frequency; OR, odds ratio. According to total scores of AUDIT, 267 drinkers (AUDIT scores ≥8) and 202 drinkers (AUDIT scores <8) were grouped for analysis. ORs (95% CIs) were obtained from the multivariate logistic regression model that contained age and smoking status (never smoker, former smoker, or current smoker of ≤10, 10.1–20, or >20 cigarettes/d). Adjusted P values were obtained by using Bonferroni correction.

TABLE 4

Single nucleotide polymorphisms (SNPs) that showed the smallest P values in the association with higher scores of the Alcohol Use Disorder Identification Test (AUDIT) in 469 male drinkers from discovery data

BAIK ET AL
population. To the best of our knowledge, this study provided the first data on the GWA with alcohol consumption that was analyzed as a quantitative trait. In associations with alcohol-related variables, no GWA study has been reported in Asians. Some limitations should be considered when interpreting our results. First, we were unable to assess AD defined by the diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (2). The 2 cohort studies included in this investigation did not consider evaluations for AD at baseline, but investigators in one of the studies involved the AUDIT during the follow-up period. Because our follow-up data showed a high correlation between the amount of alcohol consumed and AUDIT scores, we reasonably regarded heavy drinking as potential AD. Second, although previous studies considered non-synonymous SNPs of ALDH2, such as rs671, which is a critical region associated with alcohol drinking behavior (30–32), we were unable to examine the association between alcohol consumption and such specific locus. Nonetheless, the LD findings as well as the genome associations with the alcohol flush reaction appeared to indicate that signals in or near C12orf51 are highly likely to originate from the region of the ALDH2 gene. As a third limitation, our study included only a small sample size of male drinkers and was unable to analyze GWA data for female drinkers because the majority of women abstained from or rarely drank alcohol. In addition, our study populations were Koreans and, thus, the generalization of our findings for other ethnicities is limited. The exact functional roles of the suggested loci, in particular the CCDC63 and MYL2 region, in alcohol drinking remain unclear at this point, and thus further investigations on these genes are warranted.

Some studies reported that CCDC63, MYL2, OAS3, CUX2, and RPH3A were associated with drug dependence and affective disorders (33–35). Because alcoholism is considered a polygenic trait and a multifactorial disease with psychological manifestations, genes involved in neurophysiologic mechanisms are expected to play an important role in alcohol consumption. The LD findings on CCDC63, MYL2, and loci around ALDH2 support that SNPs in CCDC63 and MYL2 may be associated with alcohol drinking independent of the effects of ALDH2.

Accumulated data have shown a strong association between the ALDH2 gene and alcohol drinking behavior mainly in Asians (30–32). The ALDH2 gene encodes a mitochondrial aldehyde dehydrogenase, which is a hepatic enzyme that takes part in the oxidation of acetaldehyde to acetate after ethanol is metabolized into acetaldehyde with alcohol dehydrogenase. It plays a part in the oxidation of acetaldehyde to acetate after ethanol is metabolized into acetaldehyde with alcohol dehydrogenase. It has been reported that ALDH2*1 and ALDH2*2, as major polymorphisms of ALDH2, differently influence enzyme activity (4, 16), and the ALDH2*2 variant dominantly contributed to a reduction of enzyme activity and ineffective clearance of toxic acetaldehyde, which resulted in aversive physiologic responses such as facial flushing, hypotension, palpitations, tachycardia, and vomiting (4, 16). Such unpleasant reactions have been implicated in alcohol drinking behavior because individuals with a homogenous genotype of ALDH2*2 are unlikely to be frequent consumers of alcohol or alcoholic (31, 32). A meta-analysis study showed that men with the ALDH2*2 variant had a 0.12-fold odds of alcoholism than were men with ALDH2*1 (32).

In conclusion, the current GWA study in male Korean drinkers from population-based cohorts showed loci and SNPs variants that were highly associated with alcohol consumption and suggested that causal genetic regions associated with alcohol consumption may reside on chromosome 1q24, at least in East Asians. It needs to be determined whether CCDC63 and MYL2 are novel genes that affect alcohol consumption. Further GWA investigations on excessive alcohol consumption are warranted to enrich information for a genomic approach to develop strategies for early prevention against AD.

The authors’ responsibilities were as follows—IB: provided study inference, conducted statistical analyses, and wrote the manuscript; NH, SHK, and BGH: oversaw the conduct of the study; CS: designed, initiated, and oversaw the conduct of the study and had primary responsibility for the final content of the manuscript; and all authors: read and approved the final manuscript. None of the authors declared a conflict of interest.

REFERENCES

17. Covault J, Gelemter J, Hesselbrock V, Nellissey M, Kranzler HR. Allelic and haplootypic association of GABRA2 with alcohol