Molecular Psychiatry (2013) 18, 1218–1224

www.nature.com/mp

ORIGINAL ARTICLE

A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies *C15orf53*

J-C Wang[1,15], T Foroud[2,15], AL Hinrichs[1,15], NXH Le[1], S Bertelsen[1], JP Budde[1], O Harari[1], DL Koller[2], L Wetherill[2], A Agrawal[1], L Almasy[3], AI Brooks[4], K Bucholz[1], D Dick[5], V Hesselbrock[6], EO Johnson[7], S Kang[8], M Kapoor[1], J Kramer[9], S Kuperman[10], PAF Madden[1], N Manz[8], NG Martin[11], JN McClintick[12], GW Montgomery[11], JI Nurnberger Jr[13], M Rangaswamy[8], J Rice[1], M Schuckit[14], JA Tischfield[3], JB Whitfield[11], X Xuei[12], B Porjesz[8], AC Heath[1], HJ Edenberg[2,12], LJ Bierut[1] and AM Goate[1]

Several studies have identified genes associated with alcohol-use disorders (AUDs), but the variation in each of these genes explains only a small portion of the genetic vulnerability. The goal of the present study was to perform a genome-wide association study (GWAS) in extended families from the Collaborative Study on the Genetics of Alcoholism to identify novel genes affecting risk for alcohol dependence (AD). To maximize the power of the extended family design, we used a quantitative endophenotype, measured in all individuals: number of alcohol-dependence symptoms endorsed (symptom count (SC)). Secondary analyses were performed to determine if the single nucleotide polymorphisms (SNPs) associated with SC were also associated with the dichotomous phenotype, DSM-IV AD. This family-based GWAS identified SNPs in *C15orf53* that are strongly associated with DSM-IV alcohol-dependence symptom counts ($P = 4.5 \times 10^{-8}$, inflation-corrected $P = 9.4 \times 10^{-7}$). Results with DSM-IV AD in the regions of interest support our findings with SC, although the associations were less significant. Attempted replications of the most promising association results were conducted in two independent samples: nonoverlapping subjects from the Study of Addiction: Genes and Environment (SAGE) and the Australian Twin Family Study of AUDs (OZALC). Nominal association of *C15orf53* with SC was observed in SAGE. The variant that showed strongest association with SC, rs12912251 and its highly correlated variants ($D' = 1$, $r^2 \geqslant 0.95$), have previously been associated with risk for bipolar disorder.

Molecular Psychiatry (2013) **18,** 1218–1224; doi:10.1038/mp.2012.143; published online 23 October 2012

Keywords: *C15orf53*; DSM-IV alcohol-dependence symptoms; family-based GWAS; quantitative traits

INTRODUCTION

Alcohol-use disorders (AUDs) are among the most common and costly public health problems throughout the world.[1] Family and twin studies have provided evidence for a genetic predisposition toward AUDs,[2,3] with genetic factors accounting for approximately 40–60% of the total variance in risk for alcohol dependence (AD).[3–8]

A variety of study designs have been employed to identify genes influencing the vulnerability to AD. Genome-wide association studies (GWAS) are a potentially more comprehensive way to study a complex, common disease like AD where we have little knowledge of disease pathophysiology. Several GWAS have sought to identify variants associated with the risk for AD using case–control designs, including treatment seeking subjects with AD,[9] individuals selected from densely affected families with AD,[10] a case–control series drawn from treatment and community-based samples from several diseases,[11] subjects ascertained from large unselected sibships and individuals selected for heavier alcohol use.[12] GWAS using quantitative traits derived from alcohol consumption and AD symptomatology have also been examined in a population-based sample[13] and an Australian sample of related individuals.[12] Results thus far have identified interesting candidate genes for AD, although the overlap of the top genetic signals across studies has been limited.

The Collaborative Study of the Genetics of Alcoholism (COGA) provides another opportunity to examine genes associated with problematic alcohol use. COGA is a multisite, longitudinal study established to identify vulnerability genes for AD by recruiting multiplex alcohol-dependent families, as well as representative families from the community.[14–18] In the current analysis, we performed a family-based GWAS in large multigenerational families severely affected by AD. These families likely represent a subgroup enriched for AD susceptibility alleles.

The power to identify genes contributing to the risk for disease may be increased through the analysis of quantitative endophenotypes highly correlated with that disorder but measurable in all

[1]Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO, USA; [2]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; [3]Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA; [4]Department of Genetics, Rutgers University, Piscataway, NJ, USA; [5]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; [6]Department of Psychiatry, University of Connecticut Health Center, Farmington, CT, USA; [7]Division of Health, Social and Economic Research, Research Triangle Institute International, Research Triangle Park, NC, USA; [8]Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY, USA; [9]Department of Psychiatry, University of Iowa College of Medicine, Iowa City, IA, USA; [10]Division of Child Psychiatry, University of Iowa Hospitals, Iowa City, IA, USA; [11]Queensland Institute of Medical Research, Queensland, Australia; [12]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA; [13]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA and [14]San Diego Department of Psychiatry, University of California, La Jolla, CA, USA. Correspondence: Dr AM Goate, Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid Ave, Saint Louis, MO 63110, USA.
E-mail: goatea@psychiatry.wustl.edu
[15]These authors contributed equally to this work.
Received 29 December 2011; revised 26 July 2012; accepted 4 September 2012; published online 23 October 2012

individuals. Rather than focus on the presence or absence of AD, we used the number of AD symptoms as our primary phenotype. Some research has indicated that AD may be better captured with a symptom count (SC) rather than with a dichotomous diagnosis.[19–21] Evidence from twin studies has shown that two quantitative measures, dependence symptoms and alcohol consumption, are highly correlated with AD and index closely to the risk for AD.[22,23] SCs can be computed for any drinker, including older adolescents who are just beginning to use alcohol but may not fulfill the criteria for AD, thus allowing us to use more of our sample in the analysis and increase the power to detect association. As most other studies on alcoholism have used a dichotomous diagnosis, DSM-IV AD, we analyzed the regions of interest identified in the SC analysis to evaluate if similar findings emerge.

MATERIALS AND METHODS

COGA study subjects

Following the approval of institutional review boards at all participating institutions, AD probands were recruited through alcohol treatment programs and administered a validated poly-diagnostic instrument, the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), to assess AD.[14,16,18,19,24] Individuals below the age of 18 years were administered an adolescent version of the SSAGA. The same assessment approach, used for all probands and their relatives, was repeated at several year intervals for a large number of individuals.

The goal of this study was to identify genetic variants associated with alcohol-related phenotypes. All families were reviewed and the genetically most informative subset of COGA families that could be used for analyses of a variety of alcohol-related phenotypes, including DSM-IV SC, were selected for a family-based GWAS. Prioritization in selecting subjects for analysis was the basis of a higher number of AD family members, the number of relatives who supplied DNA, as well as the number of family members with another key COGA phenotype, electrophysiology measures. To reduce the heterogeneity in the sample, only families that were primarily of Caucasian descent were selected for genotyping, yielding a total of 2322 subjects from 118 extended families who were genotyped. The resulting data set includes multigenerational families affected by AUDs with an average of 20 subjects per family (Supplementary Figure S1). After genotype's quality control and cleaning, correcting pedigree inconsistencies, and processing the phenotypes (see below), 2010 genotyped individuals were included in the subsequent analyses. Full details on the genotype cleaning are included in the supplementary information.

Phenotype

We computed the SC using the seven lifetime diagnostic criteria for DSM-IV AD. This measure, with a value ranging from 0–7, was available on all individuals with an adolescent or adult SSAGA who had reported ever-consuming alcohol. When longitudinal data were available, we used the maximum number of symptoms endorsed at any interview. Individuals who were younger than 15 years at the most recent interview were excluded from the analysis because SC in this population is likely to be nonrepresentative of adult cohorts. Of the 2010 subjects $\geqslant$15 years who drank, 622 did not report any of the 7 symptoms of AD, and 765 had 3 or more such symptoms. The distribution of SC is shown in Figure 1.

DSM-IV AD was used as a secondary phenotype, to enable comparison with other studies in the literature. Given the wide age range of the subjects included in the analysis and the fact that many had not passed through the age of risk for an alcohol-use disorder, the following algorithm was developed to reach the final diagnosis after considering all the evaluations. Individuals aged 15 years or older who met DSM-IV criteria at any evaluation were classified as alcohol dependent. Individuals aged 23 years and older who drank but did not meet criteria for AD on any adult SSAGA were classified as unaffected. Individuals who did not consume alcohol, were under the age of 15 years at all evaluations, or were aged 15–22 years and did not meet the DSM-IV criteria, were classified as unknown and were removed from subsequent analyses—this avoids ascribing an unaffected status to an adolescent or young adult who may be at high genetic risk but not yet past the peak period of vulnerability. There were 684 subjects in the 118 extended families who met the criteria for AD using either an adult or adolescent SSAGA (Supplementary Table S1). Among 1638 remaining individuals, 964 were classified as unaffected. The average number of individuals in a family diagnosed with AD was 5.9 (Supplementary Figure S2).

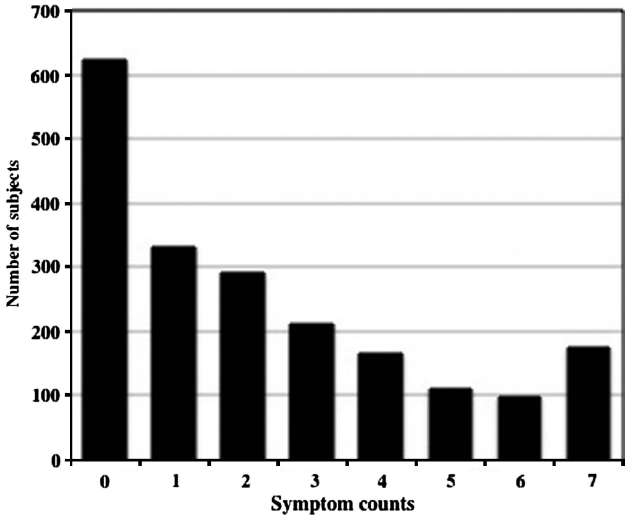

Figure 1. Distribution of DSM_IV alcohol-dependence symptom counts in the genotyped GWAS sample.

Genotype

Genotyping was performed at the Genome Technology Access Center at Washington University School of Medicine in Saint Louis (http://gtac.wustl.edu/) using the Illumina Human OmniExpress array 12.VI (Illumina, San Diego, CA, USA). We also included in the analysis, genotypes for subjects ($n = 275$) from these 118 families who were genotyped in a previous case–control GWAS using the Illumina 1M array.[10] For quality control purposes, 51 of the 275 subjects were genotyped again on the Illumina Human OmniExpress array. Imputed data were obtained using the program BEAGLE.[25] A detailed description of imputation and subsequent data cleaning is included in the supplementary information.

Statistical analysis

A total of 707 557 autosomal single-nucleotide polymorphisms (SNPs) passed quality control. Given their limited power to detect the association, SNPs with a minor allele frequency below 5% ($n = 115\,872$) were excluded from further analysis. Thus the association analysis was performed with the remaining 591 685 SNPs, giving a Bonferroni-corrected threshold for genome-wide significance of $P = 8.45 \times 10^{-8}$.

We first tested the effect of covariates on our phenotypes, SC and DSM-IV AD. As expected, gender was a highly significant predictor of SC and DSM-IV AD, and was included as a covariate in all analyses. We identified cohort effects and therefore divided subjects into four cohorts the basis of their year of birth ($<$1930, 1930–1949, 1950–1969 and $\geqslant$1970). For SC, the age-squared parameter was still be significant after cohort effect was included, and the final model therefore included gender, age, age-squared and cohort. For DSM-IV AD, the age-squared parameter was not a significant factor after considering cohort and was omitted. The first principal component from the EIGENSTRAT analysis (pc1), although not statistically significant, was still included in all analyses to reduce the risk of false-positive associations owing to population stratification.

In this sample, the SC phenotype best fit a negative binomial distribution, which was identified by applying PROC COUNTREG and PROC SGPLOT in SAS (http://support.sas.com/rnd/app/da/glimmix.html). By specifying a negative binomial distribution and a logarithmic link function, we parametrically modeled the observed trait distribution and included relevant covariates described above. Association with SC was analyzed for each SNP using a dose–effect model (number of minor alleles present in each individual), as implemented in PROC GLIMMIX from SAS. To control for relatedness, the test was placed in a general linear mixed model framework[26] using an independent working correlation matrix where each family was a separate cluster.

Inflation of *P* values was revealed by preliminary examination of the quantile–quantile (Q–Q) plot (Supplementary Figure S3). The genomic inflation factor (GIF), calculated by computing the median of the χ^2 statistics divided by the median of the central χ^2 distribution with df $=1$, was 1.25. To control for this inflation, we used the Genomic Control[27] method with a λ value of 1.25. In particular, we recomputed the level of association with each marker by dividing the observed χ^2 by the inflation factor λ value of 1.25. We verified that these new, inflation-corrected *P*-values had a GIF of 1, indicating no further inflation.

The analyses of AD were conducted using the GWAF, an R package for genome-wide association analyses with family data.[28] A logistic regression model was employed with gender, age and cohort included as covariates, and a log additive model for each SNP was tested for association. The generalized estimating equation (GEE) framework was used to control for relatedness. No inflation of *P*-values was observed ($\lambda = 1.05$).

Replication samples

The Study of Addiction: Genetics and Environment (SAGE) sample. The SAGE sample is a case–control series selected from three large, complementary data sets: COGA, Family Study of Cocaine Dependence and Collaborative Genetics Study of Nicotine Dependence. After removing 129 individuals in SAGE who were also part of the 118 extended families in the primary analysis, data from 2647 subjects of European descent were used to replicate promising associations ($P < 0.0001$) identified in the COGA sample. Detailed characteristics of this sample and the genotyping platform were described in Bierut *et al.*[11] Imputed dosage data were obtained using the same method, as described in the supplementary information. The distribution of SC was similar to that of the COGA sample. We used PROC GLIMMIX in SAS to test the association of individual SNPs with SC, including age, age-squared, gender, nicotine dependence, cocaine dependence and pc1 as covariates. We used the GEE framework described above to analyze the association with AD.

The Australian Twin-Family Study of AUD (OZALC) Sample. The twins in this study were initially ascertained through the Australian Twin Registry, followed by cascading recruitment of non-twin siblings, parents, adult offspring and spouses.[12] Data from 6166 subjects of European descent were used for replication analysis with the SC. Detailed characteristics of this sample and genotyping platform were described in Heath *et al.*[12] Imputed dosage data were obtained using MACH (http://www.sph.umich.edu/csg/abecasis/MACH). The association of individual SNPs with SC was performed using PROC GLIMMIX from SAS. Age and gender were included as covariates for the association analysis. The GEE model described above was used for the association analysis with AD.

RESULTS

Association with DSM-IV symptom counts

Results of the entire genome are summarized in the Manhattan plot (Figure 2). After correcting for inflated λ, 72 SNPs of 591 685 genotyped autosomal SNPs tested showed the evidence of association with SC with inflation-corrected *P*-values $< 10^{-4}$ (Supplementary Table S2). None of these 72 SNPs reached genome-wide significance. Among these top signals, we identified 7 chromosomal loci containing 3 or more SNPs within 50 kb of each other that show association with SC (Table 1).

The strongest association was detected with rs12903120 ($P = 5.45 \times 10^{-8}$, inflation-corrected $P = 1.09 \times 10^{-6}$) in an uncharacterized gene, *C15orf53* on chromosome 15q14. Two other genotyped SNPs, rs12916379 and rs2132157, that are highly correlated with rs12903120 ($D' = 1$, $r^2 = 1$ in HapMap EUR reference sample) also showed strong association with SC (with inflation-corrected $P = 2.79 \times 10^{-6}$ and $P = 3.02 \times 10^{-6}$, respectively) (Supplementary Table S3). Rs12916379 is located in the 3′ untranslated region of *C15orf53* (Figure 3). Using imputed genotypes, 15 additional SNPs in *C15orf53* gene region also showed suggestive evidence of association (inflation-corrected $P \leqslant 1.0 \times 10^{-5}$) with SC (Supplementary Table S3). Three of these 15 SNPs have stronger association (inflation-corrected $9.4 \times 10^{-7} \leqslant P \leqslant 9.7 \times 10^{-7}$) with SC than the genotyped SNP rs12903120. A non-synonymous coding SNP in *C15orf53* that is highly correlated with rs12903120 (rs7165988; $r^2 = 0.95$, $D' = 1$ in HapMap EUR reference sample) is associated with SC at an inflation-corrected $P = 1.7 \times 10^{-6}$.

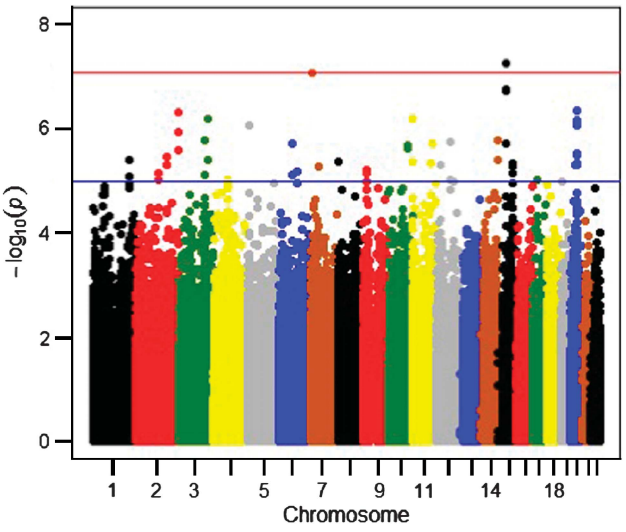

Figure 2. Manhattan plot of genome-wide association results for DSM-IV alcohol-dependence SC using negative binomial analysis. –log10 values shown here were not corrected for inflation factor.

Association with DSM-IV alcohol dependence in the regions of interest

We tested whether SNPs in the seven chromosomal regions associated with SC were also associated with DSM-IV AD. Our association analysis using a GEE model found none of the genotyped SNPs in these regions attained genome-wide significance with DSM-IV AD. By comparing the associations for SC with DSM-IV AD, we observed a less significant association with the dichotomous diagnostic trait than the association with SC (Table 1). However, the effect sizes between the two phenotypes were highly correlated (Figure 4).

Replication of association in SAGE and OZALC studies

Seventy-two SNPs that associated with SC (with inflation-corrected $P < 10^{-4}$) in the COGA families were tested in SAGE. On the basis of a prior hypothesis (our initial results from GWAS) for each SNP regarding the direction of effect, we found 5/72 SNPs showing nominal association with the same direction of effect for SC ($0.007 \leqslant P \leqslant 0.05$). However, none of these SNPs are significant after correcting for multiple testing. We then tested the replication of the 18 SNPs in *C15orf53* that showed suggestive evidence of association with SC in COGA sample (inflation-corrected $P = 1.0 \times 10^{-5}$). Using the imputed data, eight of these 18 SNPs showed nominal association with SC in SAGE sample ($P \leqslant 0.03$) (Supplementary Table S3). These 18 SNPs all lie within a single LD bin ($r^2 \geqslant 0.8$, $D' \geqslant 0.9$), so the data shown in Supplementary Table S3 reflects a single statistical test.

Sixty-nine of the 72 SNPs associated with SC (with inflation-corrected $P < 10^{-4}$) in the COGA families were tested in OZALC. None of these SNPs showed association with SC in this sample (Supplementary Table S2). Association of SNPs in the *C15orf53* gene region (rs2132157, rs12916379 and rs12903120, with $P = 0.03$, $P = 0.03$ and $P = 0.05$, respectively) was observed with DSM-IV AD but not with SC (data not shown).

DISCUSSION

We conducted a family-based GWAS and identified genome-wide significant association ($P = 5.4 \times 10^{-8}$) between SC and SNPs in

Table 1. Chromosome regions containing three or more SNPs within 50 kb of each other that are associated with symptom count at inflation corrected $P<0.0001$ and the comparison of association between alcohol-dependence symptom count and DSM-IV alcohol dependence in these regions

Chromosome location	*SNP*	*Position (hg19)*	*Gene/ transcript*	*Gene/ transcript position*	*Alcohol-dependence symptom count*			*DSM-IV alcohol dependence*	
					Effect	P-*value*	*Corrected* P-*value*	*Effect*	P-*value*
1q32.3	rs612414	212602176	NENF	212606229---212619721	−0.15	1.07E−05	7.85E−05	−0.26	2.25E-03
	rs583058	212610755	NENF		−0.13	1.42E−05	9.92E−05	−0.23	9.44E-04
	rs4804	212619339	NENF		−0.14	8.36E−06	6.44E−05	−0.24	7.15E-04
	rs483954	212620214	NENF		−0.15	3.90E−06	3.47E−05	−0.27	6.30E-04
2q37.3	rs896543	237509207	CXCR7	237478380–237490997	−0.22	4.75E−07	6.29E−06	−0.56	4.35E-06
	rs6431476	237517937	CXCR7		−0.20	1.14E−06	1.28E−05	−0.51	3.92E-06
	rs7594454	237537935	CXCR7		−0.18	2.60E−06	2.49E−05	−0.44	3.05E-05
3q24	rs7431637	143049769	SLC9A9	142984064–143567373	−0.14	7.58E−06	5.95E−05	−0.31	3.60E-04
	rs10446322	143068250	SLC9A9		−0.15	1.73E−06	1.80E−05	−0.33	5.77E-05
	rs868702	143085345	SLC9A9		0.15	1.73E−06	1.79E−05	0.34	2.58E-05
4q21.21	rs12513014	81061422	intergenic		0.16	1.38E−05	9.70E−05	0.35	1.02E-03
	rs13102102	81073672	intergenic		0.14	1.11E−05	8.12E−05	0.35	1.37E-05
	rs13138779	81087073	intergenic		0.21	9.75E−06	7.30E−05	0.49	7.69E-04
9p22.2	rs10963462	18130036	intergenic		−0.18	6.22E−06	5.07E−05	−0.40	1.18E-05
	rs763976	18134914	intergenic		−0.17	1.00E−05	7.46E−05	−0.40	7.40E-06
	rs12006002	18166899	intergenic		0.15	7.24E−06	5.73E−05	0.45	4.87E-07
15q14	rs7168475	38960882	C15orf53	38988799–38992239	−0.18	1.98E−06	2.00E−05	−0.39	4.66E–04
	rs12903120	38988097	C15orf53		−0.18	5.45E−08	1.09E−06	−0.38	7.62E-06
	rs12916379	38991520	C15orf53		−0.17	1.74E−07	2.79E−06	−0.36	4.06E-05
	rs2132157	38992547	C15orf53		−0.17	1.92E−07	3.02E−06	−0.36	4.35E-05
15q24.2	rs2029519	75415962	intergenic		0.14	1.07E−05	7.87E−05	0.29	1.60E-03
	rs4479194	75422131	intergenic		0.15	5.27E−06	4.43E−05	0.30	1.02E-03
	rs7172677	75424593	intergenic		0.15	7.15E−06	5.68E−05	0.32	5.35E-04
20q11.22	rs6060124	33536897	GSS	33516236–33543601	0.16	4.41E−07	5.92E−06	0.33	2.03E-04
	rs6088664	33551100	MYH7B	33543704–33590240	−0.15	3.23E−06	2.98E−05	−0.27	1.91E-03
	rs6579204	33553677	MYH7B		0.15	4.54E−06	3.93E−05	0.29	1.25E-03

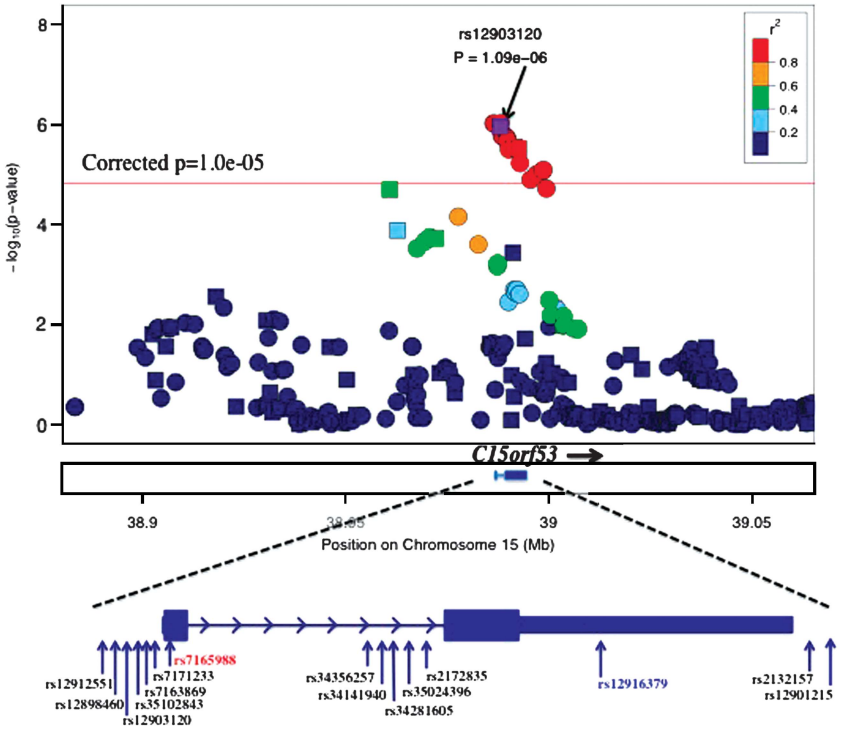

Figure 3. Plot of chromosome 15q14 association with DSM_IV alcohol-dependence SC. Squares represent genotyped SNPs; circles represent imputed SNPs. SNP rs7165988 (in red) is a non-synonymous coding variant. SNP rs12916379 (in blue) is in 3′ untranslated region of *C15orf53*.

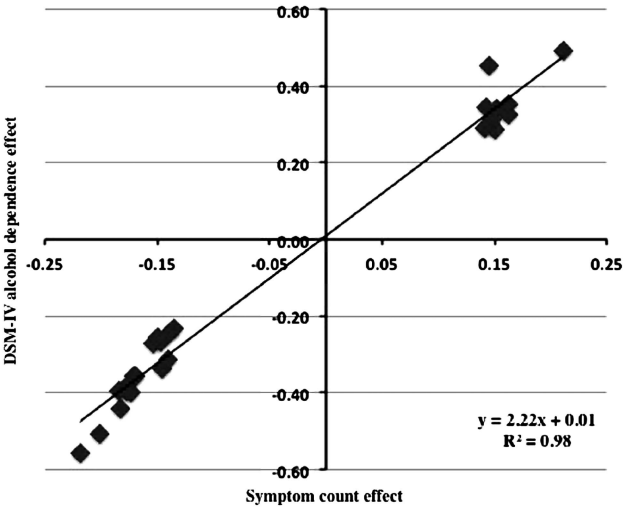

Figure 4. Correlation between symptom-count effect and DSM-IV alcohol-dependence effect in COGA sample.

chromosome 15q14. However, the Q–Q plot for the COGA family-based sample suggests GIF of 1.25. One possible explanation is the presence of polygenic inheritance.[29] The SC trait is likely due to many loci of small effect. Because of this and the high LD in the dense map, it is possible that there are relatively few genomic regions free of loci with small effect moderately inflating GIF. A second possibility is that the inflation may be due to the presence of very large families in our sample. As our association test controls for relatedness by treating each family as a separate cluster, this may be insufficient in families where distantly related individuals should not be clustered into the same class. After correcting for the GIF of 1.25 using the genomic controls method, none of the SNPs associated with SC reached genome-wide significance.

Most GWAS studies do not employ family-based data; thus, there are inherent challenges in estimating the power of the sample to detect particular effect sizes. However, the implementation of a correlation matrix subdivided by family clusters to control for relatedness among the families pares SC analysis to an association test similar to a case–control study, as do the GEE used to analyze AD. Both methods provide similar power estimates based on simulation studies.[30,31] For this sample, we estimated power using Quanto.[32] These extended pedigrees have 70% power to detect an effect size of 1.1 for SC when the minor allele frequency is between 0.10 and 0.30.

The strongest association (inflation-corrected $9.4 \times 10^{-7} \leqslant P \leqslant 9.8 \times 10^{-6}$) was detected with a group of 18 highly correlated variants ($r^2 \geqslant 0.95$) within and flanking *C15orf53*. Interestingly, recent GWAS data has reported consistent evidence, showing that variation in a region close to *C15orf53* influences susceptibility for bipolar disorder.[33–36] SNPs that influenced bipolar disorder susceptibility (rs12912251, rs2172835 and rs12899449) also show strong association with SC (inflation corrected $P = 9.4 \times 10^{-7}$, $P = 3.1 \times 10^{-6}$, and $P = 1.26 \times 10^{-5}$, respectively) in COGA families severely affected by AD. The alleles that are associated with reduced risk for BD are also associated with lower dependence SCs. We did not detect association between SCs and other variants reported by the Psychiatric GWAS Consortium Bipolar Disorder Working Group[35] to be strongly associated with bipolar disorder. This suggests a specific effect of this gene on risk for bipolar disorder and AD rather than a more general shared underlying liability to both disorders. Studies of psychiatric disorders have shown that bipolar disorder and alcoholism commonly co-occur,[37–39] and that individuals with bipolar disorder have a greater likelihood of AUDs than the general population.[40] Approximately 46% of subjects with bipolar disorder type I have AUDs, while 39% of subjects with a less severe form of mania, bipolar disorder II, have an AUDs.[37–39] The association of *C15orf53* with SC detected in this study suggests the possibility of a specific genetic link between bipolar disorder and alcoholism.

Two of the imputed SNPs showing strong association with SC are located within the exons of *C15orf53* (Figure 3). *C15orf53* encodes an uncharacterized protein of 179 amino acids with homology to uncharacterized proteins in other species, including chimpanzee, gibbon and orangutan. Rs7165988 in exon 1 results in a non-synonymous coding change of valine to leucine at codon 3. This substitution is predicted to be possibly damaging by POLYPHEN.[41] Rs12916379 is located within the 3′ UTR of *C15orf53* and could influence transcript stability. Our preliminary data showed that *C15orf53* mRNA expression is detectable in nine brain regions tested, although the expression level is low. The method of this assay is included in the supplementary information.

There are five additional candidate genes located in chromosomal regions showing association with SC (Table 1). Among these, *SLC9A9* encodes a sodium and hydrogen exchanger in chromosome 3q24, and is of particular interest as it has been linked to tobacco smoking initiation, a behavior highly comorbid with alcohol use.[42] Studies have repeatedly shown evidence of association between *SLC9A9* and attention-deficit hyperactivity disorder.[43–45] Because inattention is a predictor of smoking initiation,[46] and attention-deficit hyperactivity disorder and smoking are co-transmitted through families more often than expected by chance,[47] it is possible that *SLC9A9* influences both smoking and inattention. The association of *SLC9A9* with SC identified in this study suggests a genetic connection between attention-deficit hyperactivity disorder, smoking and alcoholism. Our preliminary data showed that *SLC9A9* mRNA expression is detectable in human frontal cortex. Total mRNA expression in brain tissues derived from alcoholics is 1.09 fold higher than in the brain tissues derived from nonalcoholic subjects. Another gene of interest, *CXCR7* in chromosome 2q37.3, encodes C-X-C chemokine receptor type 7, a member of the G-protein-coupled receptor family. Our preliminary assay with human frontal cortices showed *CXCR7* total-mRNA expression, which is 1.33 fold higher in alcoholics, compared with the expression level in nonalcoholic subjects. The method of these assays is included in the supplementary information.

To date, four previous case–control GWAS analyses of AUDs have not provided evidence of association that reached genome-wide significance with AD or quantitative traits.[9–12] In this study, our approach to identifying genetic risk factors for alcohol problems focused primarily on a quantitative measure of alcohol SC, rather than an AUD diagnosis. Our SC measure was on the basis of seven DSM-IV AD criteria and deliberately excluded the 4 criteria associated with DSM-IV alcohol abuse. We crafted this measure to allow the most straightforward comparisons between findings for SC and findings for diagnosis of dependence. Dimensional dependence measures such as SC are more powerful than dichotomized phenotypic measures (that is, DSM-IV AD) for detecting risk factors, especially in samples containing adolescent subjects. In the present study, we compared chromosomal regions that showed strong associations with both SC and DSM-IV AD, and consistently observed a stronger relationship for genetic variants with SC than with DSM-IV AD (Table 1, Figure 4); the effect sizes and directions for both phenotypes were relatively consistent.

Although we are encouraged by our findings, we recognize that there are limitations to our study. Among our top candidate genes, we only detected nominal association for *C15orf53* and SC in SAGE. Several reasons could possibly explain this limited replication. First, the power to replicate findings of small effect across studies in samples of the size used in this study is low. Second, in contrast to the COGA sample, neither the SAGE nor the

OZALC samples were ascertained from large families severely affected by AUDs. It is possible that severely affected families have a concentration of genetic variants that influence risk for alcoholism and that may have less effect on alcoholism in the general population. A coordinated evaluation including many more families severely affected by alcoholism is necessary to confirm our findings.

In summary, our family-based GWAS identified SNPs in the gene *C15orf53* that showed suggestive evidence of association with DSM-IV alcohol-dependence SC. Interestingly, SNPs in this gene have previously been associated with risk for bipolar disorder in other GWAS and suggest there may be some common genetic factors contributing to both disorders.

CONFLICT OF INTEREST

Doctors LJ Bierut, AM Goate, AJ Hinrichs, J Rice and JC Wang are listed as inventors on the patent 'Markers for Addiction' (US 20070258898) covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. The remaining authors declare no conflict of interest.

ACKNOWLEDGEMENTS

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 Co-PI 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'

The Australian Twin-family Study of Alcohol-Use Disorder (OZALC) Sample: The OZALC study was supported by National Institutes of Health Grants AA07535, AA07728, AA13320, AA13321, AA14041, AA11998, AA17688, DA012854 and DA019951; by Grants from the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485 and 552498); by Grants from the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016 and DP0343921); and by the 5th Framework Programme (FP-5) GenomEUtwin Project (QLG2-CT-2002-01254). Genotyping at Center for Inherited Disease Research was supported by a Grant to the late Richard Todd, MD, PhD., former Principal Investigator of Grant AA13320. We acknowledge the contribution of Anjali Henders and Yi-Ling for their technical assistance.

REFERENCES

1 World Health Organization. *Global strategy to reduce the harmful use of alcohol*. WHO Press: Geneva, Switzerland, 2010.
2 Goodwin DW, Schulsinger F, Moller N, Hermansen L, Winokur G, Guze SB *et al.* Drinking problems in adopted and nonadopted sons of alcoholics. *Arch Gen Psychiatry* 1974; **31**: 164–169.
3 Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ *et al.* Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychol Med* 1997; **27**: 1381–1396.
4 Dick DM, Latendresse SJ, Lansford JE, Budde JP, Goate A, Dodge KA *et al.* Role of GABRA2 in trajectories of externalizing behavior across development and evidence of moderation by parental monitoring. *Arch Gen Psychiatry* 2009; **66**: 649–657.
5 Kendler KS, Neale MC, Heath AC, Kessler RC, Eaves LJ. A twin-family study of alcoholism in women. *Am J Psychiatry* 1994; **151**: 707–715.
6 Prescott CA, Kendler KS. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *Am J Psychiatry* 1999; **156**: 34–40.
7 Prescott CA, Sullivan PF, Kuo PH, Webb BT, Vittum J, Patterson DG *et al.* Genomewide linkage study in the Irish affected sib pair study of alcohol dependence: evidence for a susceptibility region for symptoms of alcohol dependence on chromosome 4. *Mol Psychiatry* 2006; **11**: 603–611.
8 Schuckit MA, Edenberg HJ, Kalmijn J, Flury L, Smith TL, Reich T *et al.* A genome-wide search for genes that relate to a low level of response to alcohol. *Alcohol Clin Exp Res* 2001; **25**: 323–329.
9 Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P *et al.* Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry* 2009; **66**: 773–784.
10 Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L *et al.* Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol Clin Exp Res* 2010; **34**: 840–852.
11 Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E *et al.* A genome-wide association study of alcohol dependence. *Proceedings of the National Academy of Sciences of the United States of America* 2010; **107**: 5082–5087.
12 Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA *et al.* A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. *Biol Psychiatry* 2011; **70**: 513–518.
13 Kendler KS, Kalsi G, Holmans PA, Sanders AR, Aggen SH, Dick DM *et al.* Genomewide association analysis of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. *Alcohol Clin Exp Res* 2011; **35**: 963–975.
14 Begleiter H, Porjesz B, Wang W. Event-related brain potentials differentiate priming and recognition to familiar and unfamiliar faces. *Electroencephalogr Clin Neurophysiol* 1995; **94**: 41–49.
15 Edenberg HJ. The collaborative study on the genetics of alcoholism: an update. *Alcohol Res Health* 2002; **26**: 214–218.
16 Foroud T, Edenberg HJ, Goate A, Rice J, Flury L, Koller DL *et al.* Alcoholism susceptibility loci: confirmation studies in a replicate sample and further mapping. *Alcohol Clin Exp Res* 2000; **24**: 933–945.
17 Nurnberger Jr JI, Wiegand R, Bucholz K, O'Connor S, Meyer ET, Reich T *et al.* A family study of alcohol dependence: coaggregation of multiple disorders in relatives of alcohol-dependent probands. *Arch Gen Psychiatry* 2004; **61**: 1246–1256.
18 Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P *et al.* Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet* 1998; **81**: 207–215.
19 Bucholz KK, Heath AC, Reich T, Hesselbrock VM, Kramer JR, Nurnberger Jr JI *et al.* Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multicenter family study of alcoholism. *Alcohol Clin Exp Res* 1996; **20**: 1462–1471.
20 Hasin DS, Liu X, Alderson D, Grant BF. DSM-IV alcohol dependence: a categorical or dimensional phenotype? *Psychol Med* 2006; **36**: 1695–1705.
21 Heath AC, Martin NG. Genetic influences on alcohol consumption patterns and problem drinking: results from the Australian NH&MRC twin panel follow-up survey. *Ann NY Acad Sci* 1994; **708**: 72–85.
22 Grant JD, Agrawal A, Bucholz KK, Madden PA, Pergadia ML, Nelson EC *et al.* Alcohol consumption indices of genetic risk for alcohol dependence. *Biol Psychiatry* 2009; **66**: 795–800.
23 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**: 1058–1065.
24 Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. *Addiction* 1999; **94**: 1361–1370.
25 Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009; **84**: 210–223.
26 Hadfield JD, Nakagawa S. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J Evol Biol* 2010; **23**: 494–508.
27 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.
28 Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* 2010; **26**: 580–581.
29 Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ *et al.* GIANT Consortium. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* 2011; **19**: 807–812.

30 Litaker M, Ferris D. A simulation study to evaluate ANOVA and GEE for Comparing Correlated Proportions with Missing Values. Proceedings of the 12th Annual Conference of the SouthEast SAS Users Group, Nashville TN, October 31–November 2, 2004.
31 Xue X, Gange SJ, Zhong Y, Burk RD, Minkoff H, Massad LS *et al.* Marginal and mixed-effects models in the analysis of human papillomavirus natural history data. *Cancer Epidemiol Biomarkers Prev* 2010; **19**: 159–169.
32 Gauderman WMJ. QUANTO 1.2. A computer program for power and sample size calculations for genetic-epidemiology studies. http://hydra.usc.edu/gxe, 2006.
33 Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L *et al.* Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 2008; **40**: 1056–1058.
34 Liu Y, Blackwood DH, Caesar S, de Geus EJ, Farmer A, Ferreira MA *et al.* Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder. *Mol Psychiatry* 2011; **16**: 2–4.
35 Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N *et al.* Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43**: 977–983.
36 Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W *et al.* Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry* 2009; **14**: 755–763.
37 Kessler RC, Nelson CB, McGonagle KA, Liu J, Swartz M, Blazer DG *et al.* Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. *Br J psychiatry Suppl* 1996; **30**: 17–30.
38 Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL *et al.* Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA*, 1990; 264: 2511–2518.
39 Sonne SC, Brady KT. Bipolar disorder and alcoholism. NIAAA publication, November 2002. pubs.niaaa.nih.gov/publications/arh26-2/103-108.htm.
40 Helzer JE, Pryzbeck TR. The co-occurrence of alcoholism with other psychiatric disorders in the general population and its impact on treatment. *J Stud Alcohol* 1988; **49**: 219–224.
41 Vasily Ramensky PB, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; **30**: 3894–3900.
42 Vink JM, Smit AB, de Geus EJ, Sullivan P, Willemsen G, Hottenga JJ *et al.* Genome-wide association study of smoking initiation and current smoking. *Am J Hum Genet* 2009; **84**: 367–379.
43 Franke B, Vasquez AA, Johansson S, Hoogman M, Romanos J, Boreatti-Hümmer A *et al.* Multicenter analysis of the SLC6A3/DAT1 VNTR haplotype in persistent ADHD suggests differential involvement of the gene in childhood and persistent ADHD. *Neuropsychopharmacology* 2010; **35**: 656–664.
44 Markunas CA, Quinn KS, Collins AL, Garrett ME, Lachiewicz AM, Sommer JL *et al.* Genetic variants in SLC9A9 are associated with measures of attention-deficit/hyperactivity disorder symptoms in families. *Psychiatr Genet* 2010; **20**: 73–81.
45 Mick E, Todorov A, Smalley S, Hu X, Loo S, Todd RD *et al.* Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 898–905 e3.
46 Barman SK, Pulkkinen L, Kaprio J, Rose RJ. Inattentiveness, parental smoking and adolescent smoking initiation. *Addiction* 2004; **99**: 1049–1061.
47 Monuteaux MC, Faraone SV, Hammerness P, Wilens TE, Fraire M, Biederman J *et al.* The familial association between cigarette smoking and ADHD: a study of clinically referred girls with and without ADHD, and their families. *Nicotine Tob Res* 2008; **10**: 1549–1558.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)