Molecular Psychiatry (2012) 17, 445–450

www.nature.com/mp

ORIGINAL ARTICLE

ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry

LJ Bierut[1], AM Goate[1], N Breslau[2], EO Johnson[3], S Bertelsen[1], L Fox[1], A Agrawal[1], KK Bucholz[1], R Grucza[1], V Hesselbrock[4], J Kramer[5], S Kuperman[6], J Nurnberger[7], B Porjesz[8], NL Saccone[9], M Schuckit[10], J Tischfield[11], JC Wang[1], T Foroud[12], JP Rice[1] and HJ Edenberg[12,13]

[1]*Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA;* [2]*Department of Epidemiology, Michigan State University, East Lansing, MI, USA;* [3]*Division of Health, Social and Economic Research, Research Triangle Institute International, Research Triangle Park, NC, USA;* [4]*Department of Psychiatry, University of Connecticut, Farmington, CT, USA;* [5]*Department of Psychiatry, University of Iowa, Iowa City, IA, USA;* [6]*Division of Child Psychiatry, University of Iowa Hospitals, Iowa City, IA, USA;* [7]*Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA;* [8]*Department of Psychiatry, State University of New York, Brooklyn, NY, USA;* [9]*Department of Genetics, Washington University School of Medicine, St Louis, MO, USA;* [10]*Department of Psychiatry, University of California-San Diego, La Jolla, CA, USA;* [11]*Department of Genetics, Rutgers University, Piscataway, NJ, USA;* [12]*Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA and* [13]*Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA*

A coding variant in alcohol dehydrogenase 1B (*ADH1B*) (rs1229984) that leads to the replacement of Arg48 with His48 is common in Asian populations and reduces their risk for alcoholism, but because of very low allele frequencies the effects in European or African populations have been difficult to detect. We genotyped and analyzed this variant in three large European and African-American case–control studies in which alcohol dependence was defined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria, and demonstrated a strong protective effect of the His48 variant (odds ratio (OR) 0.34, 95% confidence interval (CI) 0.24, 0.48) on alcohol dependence, with genome-wide significance (6.6×10^{-10}). The hypothesized mechanism of action involves an increased aversive reaction to alcohol; in keeping with this hypothesis, the same allele is strongly associated with a lower maximum number of drinks in a 24-hour period (lifetime), with $P = 3 \times 10^{-13}$. We also tested the effects of this allele on the development of alcoholism in adolescents and young adults, and demonstrated a significantly protective effect. This variant has the strongest effect on risk for alcohol dependence compared with any other tested variant in European populations.
Molecular Psychiatry (2012) **17,** 445–450; doi:10.1038/mp.2011.124; published online 4 October 2011

Keywords: *ADH1B*; alcohol dehydrogenase; alcohol dependence; association study; genetics; protective allele

Introduction

Alcohol dehydrogenases catalyze the first step in the oxidative metabolism of ethanol (beverage alcohol). The Arg48His coding single-nucleotide polymorphism (SNP) rs1229984 in *ADH1B*, encoding the β-subunit of alcohol dehydrogenase, dramatically affects enzyme activity; enzymes with His48 oxidize ethanol approximately 70- to 80-fold faster than those with Arg48.[1] In East Asian populations, the His48 allele is common and lowers the risk for alcoholism, with an odds ratio (OR) of approximately 0.2 for a single allele.[1–4] Although it does not trigger a full flushing reaction, the protective mechanism is hypothesized to involve the aversive effects of transiently elevated acetaldehyde, which is conceptually similar to the protection accorded by the severe flushing experienced by individuals with the inactive variant of aldehyde dehydrogenase 2 (*ALDH2*).[1]

Recently, a meta-analysis of published studies on different populations, using varying definitions of alcohol dependence, abuse or alcohol-related diseases, reported a very robust association with rs1229984 in Asian populations[5] (47 studies, P value $= 7 \times 10^{-42}$). However, the results for alcohol dependence in European subjects were modest (19 studies, P value $= 0.002$), and African-American subjects were not studied. This His48 allele has a

Correspondence: Dr LJ Bierut, Department of Psychiatry, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, MO 63110, USA.
E-mail: laura@wustl.edu
Received 3 June 2011; revised 28 July 2011; accepted 23 August 2011; published online 4 October 2011

low frequency in populations of European or African descent.

In contrast to the modest findings of rs1229984 with alcohol dependence in subjects of European descent, Arg48His has been strongly implicated in cancers of the upper aerodigestive tract, specifically esophageal cancer, with the protective influence of His48 being most pronounced in drinkers ($N = 15\,841$, OR = 0.64 (95% confidence interval (CI) 0.59, 0.71), $P = 1 \times 10^{-20}$).[6,7] No evidence of association was seen in the relatively modest African-American population with cancer ($N = 1079$, OR = 0.99 (95% CI 0.46, 2.13), $P = 0.98$).

The purpose of this study was two-fold. We examined the association of rs1229984 in three large case–control studies of subjects of European and African descent in which alcohol dependence was directly diagnosed to test whether this SNP shows protective effects on alcohol dependence and alcohol consumption in non-Asian samples. We then confirmed the protective effect of the His48 allele in an adolescent and young adult population that is in the critical period of risk for the development of alcohol dependence.

Subjects and methods

Samples for this meta-analysis were drawn from three existing studies, the Collaborative Study on the Genetics of Alcoholism (COGA), the Collaborative Genetic Study of Nicotine Dependence (COGEND) and the Family Study of Cocaine Dependence (FSCD), each of which is described below. For the purpose of this study, all case subjects met a lifetime diagnosis for alcohol dependence based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV).[8] Control subjects had at least one drink of alcohol, but never met criteria for DSM-IV alcohol dependence. The characteristics of adult subjects from COGA, COGEND and FSCD are presented in Table 1.

Collaborative Study on the Genetics of Alcoholism

COGA was initiated in 1989, with funding from the National Institute on Alcohol Abuse and Alcoholism (NIAAA).[9,10] Subjects were recruited from seven sites across the United States. Alcohol-dependent probands were systematically recruited from treatment facilities, and relatives were interviewed. Comparison

Table 1 Characteristics of adult subjects

	Study		
	COGA N = *1800*	*COGEND* N = *2666*	*FSCD* N = *1166*
Sex, N (%)			
Male	971 (53.9)	1,018 (38.2)	579 (49.7)
Female	829 (46.1)	1,648 (61.8)	587 (50.3)
Age, years			
Mean ± s.d.	43.1 ± 11.4	36.5 ± 5.5	37.0 ± 8.8
Range	18.0–79.0	25.0–47.0	18.0–60.0
Race, N (%)			
European American	1353 (75.2)	1981 (74.3)	558 (47.9)
African American	447 (24.8)	685 (25.7)	608 (52.1)
DSM-IV alcohol dependence			
Diagnosis, *N* (%)	1139 (63.3)	612 (23.0)	547 (46.9)
FTND nicotine dependence			
Never smokers, *N* (%)	509 (30.3)	0	505 (43.3)
Nicotine dependent, *N* (%)	803 (47.9)	1407 (52.8)	445 (38.2)
Not nicotine dependent, *N* (%)	366 (21.8)	1259 (47.2)	216 (18.5)
DSM-IV cocaine dependence			
Diagnosis, *N* (%)	481 (26.7)	219 (8.2)	551 (47.3)
rs1229984[a]			
AA	5 (0/5)	6 (0/6)	1 (0/1)
AG	96 (38/58)	162 (21/141)	51 (20/31)
GG	1699 (1101/598)	2498 (591/1907)	1114 (527/587)

Abbreviations: COGA, Collaborative Study on the Genetics of Alcoholism; COGEND, Collaborative Genetic Study of Nicotine Dependence; FSCD, Family Study of Cocaine Dependence; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; FTND, Fagerström Test for Nicotine Dependence.
[a]Total *N* (case *n*/control *n*).

families were drawn from the same communities. A case–control sample of biologically unrelated individuals was drawn from the sample.[11] All cases met DSM-IV criteria for alcohol dependence. Control subjects were defined as individuals who consumed alcohol but did not meet any definition of alcohol dependence or alcohol abuse, or abuse or dependence on other drugs (except nicotine). For this study 1800 COGA case–control subjects were studied, including 1139 DSM-IV alcohol-dependent cases and 661 non-alcohol-dependent controls.

Collaborative Genetic Study of Nicotine Dependence
COGEND was initiated in 2001 as a three-part program project grant, funded through the National Cancer Institute.[12] Nicotine-dependent and non-dependent smoking subjects were recruited from communities in Detroit and St Louis, using a Health Maintenance Organization in Detroit, and the Missouri Family Registry in St Louis maintained at Washington University for research purposes. Nicotine-dependent subjects were current smokers with a Fagerström Test for Nicotine Dependence (FTND)[13] score of $\geqslant 4$. Non-nicotine-dependent subjects were individuals who smoked at least 100 cigarettes in their lifetime, who never had any symptoms of dependence (lifetime FTND = 0). For this study 2666 COGEND subjects were used, including 612 DSM-IV alcohol-dependent cases and 2054 non-alcohol-dependent controls. This sample includes 1610 subjects who were previously examined by Sherva *et al*.[14]

Family Study of Cocaine Dependence
FSCD was initiated in 2000 as a case–control study, funded through the National Institute on Drug Abuse.[15] Cocaine-dependent individuals were systematically recruited from chemical dependency treatment facilities in the greater St Louis metropolitan area. Community-based comparison subjects with no substance dependence were recruited through the Missouri Family Registry and matched by age, race, gender and residential zip code. For this study 1166 FSCD subjects were used, including 547 DSM-IV alcohol-dependent cases and 619 non-alcohol-dependent comparison subjects.

Adolescent and young adult study as part of COGA
The adolescent and young adults study is a sample of subjects from the COGA study who had at least one assessment between the years 1989–2008 when the subjects were between the ages of 12 and 25 years (Table 2). Subjects were recruited from families affected with alcoholism (as either a first- or a second-degree relative), and from community-based comparison families. For this analysis, 2039 adolescents and young adults who ever had a drink of alcohol were studied. All subjects were of European descent. This sample is independent of the COGA case–control sample.

Table 2 Characteristics of adolescent and young adult sample

	Adolescents and young adults, N = *2039*
Sex, N (%)	
Male	983 (48.2)
Female	1056 (51.8)
Age at first interview, years	
Mean ± sd	15.9 ± 5.0
Range	6–25
DSM-IV alcohol dependence	
Diagnosis, *N* (%)	302 (14.8)

Phenotyping for all studies
Alcohol-related phenotypes were derived using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) designed for genetic studies of alcoholism.[16–18] The SSAGA comprehensively assesses alcohol, nicotine and drug use, and has high reliability and validity.[16–18] Alcohol and drug dependence were defined using DSM-IV criteria,[8] and nicotine dependence was characterized using the FTND.[13]

In the case–control samples, two phenotypes were selected for study: lifetime diagnosis of DSM-IV alcohol dependence, and the maximum number of alcoholic drinks consumed in a 24-hour period. A standard drink of alcohol was defined as a glass of wine, a bottle of beer or a shot glass of hard liquor.

For the adolescent and young adult study, subjects $\geqslant 18$ years were assessed with the SSAGA. Adolescent subjects, 17 years of age or younger, were assessed with the childhood version of the SSAGA (C-SSAGA). DSM-IV-defined alcohol dependence was the outcome under study.

Genotyping of rs1229984
DNA was extracted from blood samples or transformed lymphoblastoid cell lines.

COGA sample and FSCD sample. We used the PrimerPicker software (http://www.kbioscience.co.uk/) to design the assay, and followed the protocol described in the KASPar SNP Genotyping System Manual (http://www.kbioscience.co.uk), with detection performed using the ABI 7900 HT Fast Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). A subset of samples (69%) was also genotyped using the Sequenom MassArray system (Sequenom, San Diego, CA, USA). Samples with discrepant results (0.3%) were discarded.

COGEND sample. A large component of this study was previously genotyped at this variant as part of a previous genotyping project.[19] To complete the genotyping in the sample, we used Sequenom MassArray, homogenous MassEXTEND (hME) or iPLEX assays for genotyping, using Sequenom SpectroTYPER software v3.4.[20]

COGA adolescent and young adult sample. rs1229984 was included in a set of 384 SNPs genotyped using the Illumina GoldenGate custom-designed platform at the Microarray Core of the Genome Sequencing Center at Washington University. We used the default algorithm in BeadStudio (Illumina; San Diego, CA, USA) to cluster genotypes.

Statistical analysis

Each study was analyzed separately. All studies have previously undergone extensive genotyping, and population structure has been reviewed with Eigenstrat.[21] Self-reported race ('White' or 'Black') corresponded to genetically determined groups.

Because of the low minor allele frequency for rs1229984 and the small number of subjects homozygous for the minor allele (less than 0.5% in each study), the G/A heterozygous subjects and A/A homozygous subjects were collapsed into a single group for analysis and compared with G/G homozygous subjects.

In the adult case–control sample, alcohol dependence was analyzed using logistic regression to estimate ORs for the variant rs1229984. Maximum drinks consumed in a 24-hour period were examined by linear regression of the log-transformed number of drinks. Prior to log transformation, the few extreme values of maximum drinks in 24-hour (greater than 100) were set at 100. Each study was examined to determine the characteristics that predicted alcohol dependence and maximum drinks consumed in a 24-hour period. Age, gender, nicotine dependence and cocaine dependence were significantly associated with alcohol dependence and maximum drinks consumed in a 24-hour period. Primary analyses were adjusted for sex, age (in quartiles), and nicotine and cocaine dependence.

To determine the sensitivity of analyses to the inclusion of covariates, analyses were repeated using SNP only as a predictor variable, and then with SNP, age (quartiles) and gender as predictor variables. In addition, analyses including principal components adjusting for population structure were performed. We combined results from the three studies to calculate a pooled estimate and meta-analysis *P* value using a fixed-effects model for subjects of European and African descent.

In the adolescent and young adult sample, survival analysis was used to examine alcohol dependence. Only subjects who reported having at least one drink of alcohol were included. A total of 302 adolescents and young adults had developed DSM-IV alcohol dependence. Those subjects who had not developed alcohol dependence were designated as censored.

Results and discussion

The frequency of alcohol dependence and comorbid nicotine and cocaine dependence is consistent with the study design and ascertainment protocols for each study (Table 1). rs1229984 is strongly associated with alcohol dependence in each dataset (Table 3). See Supplementary Table 1 for the results for the full models. In the meta-analysis across the three adult studies, which included 2298 alcohol-dependent cases and 3334 non-dependent controls, the allele encoding His48 is significantly associated with a reduced risk for alcohol dependence, with an OR of 0.34 (95% CI 0.24, 0.48) and *P* value of 6.6×10^{-10}. The magnitude of effect is similar in both European-American and African-American populations. Analyses were repeated using principal components, to ensure that the association is not an artifact of population stratification, and results are essentially unchanged (Supplementary Table 2). The findings are robust not only with respect to ethnicity but also with respect to the major covariates (age, sex, nicotine and cocaine dependence), as removing or limiting covariates did not markedly affect the results (Supplementary Tables 3 and 4). These results are consistent with the recent meta-analysis of Asian populations (47 studies, AG (Arg/His) and AA (His/His) versus GG (Arg/Arg), OR = 0.24 (95% CI 0.19, 0.31), *P* value = 2×10^{-31}).[5]

These results are further supported by analyses of an independent sample of adolescents and young adults (aged 12–25 years) of European descent. Using survival analysis, the His48 allele is associated with a 0.51 reduced hazard of developing alcohol dependence ($N = 2039$; hazard ratio = 0.51 (95% CI 0.30, 0.88) and *P* value = 0.015) (Figure 1).

The His48 allele is also associated with lower alcohol consumption, as measured by the subjects' lifetime maximum alcohol consumption in a 24-hour period ($\beta = -0.28$ (95% CI −0.35, −0.20), *P* value = 3.24×10^{-13}) (Table 3). A similar association is seen in both European-American and African-American subjects. These findings are consistent with, but exceed in magnitude, those from a prior report of association between His48 and maximum drinks consumed in a 24-hour period.[22]

These results provide strong evidence that the His48 allele in β-ADH is strongly associated with a reduced likelihood of a lifetime diagnosis of alcohol dependence and a lower maximum number of drinks consumed in a 24-hour period in individuals of both European and African ancestry. The effect of this association is seen early in the course of illness, and predicts the development of alcohol dependence in adolescents and young adults who are at the beginning of their drinking career. The direction of these effects is similar to that found in Asians, among whom the allele frequency is much higher.

Because rs1229984 is not on most genome-wide association study arrays, is relatively uncommon (< 5% minor allele frequency) in populations of European and African descent, and the accuracy of imputation is poor, the association with alcohol dependence has not been detected in several recent genome-wide association studies,[11,23,24] including studies from which these subjects were drawn.[11,24] For example, this SNP was imputed from the GWAS

Table 3 Association of rs1229984 with DSM-IV alcohol dependence and maximum drinks consumed in a 24-hour period

		Outcome					
		DSM-IV alcohol dependence[a]			*Maximum drinks in a 24-hour period*[b]		
Ethnicity	*MAF*	N	*OR (95% CI)*	P	N	β *(95% CI)*	P
COGA							
EA	0.03	1250	0.25 (0.12, 0.48)	5.97×10^{-5}	1246	−0.42 (−0.58, −0.25)	4.61×10^{-7}
AA	0.01	428	0.37 (0.06, 1.89)	2.48×10^{-1}	425	−0.31 (−0.74, 0.12)	1.57×10^{-1}
COGEND							
EA	0.04	1981	0.38 (0.23, 0.63)	2.19×10^{-4}	1981	−0.22 (−0.31, −0.12)	9.15×10^{-6}
AA	0.01	685	0.47 (0.10, 2.23)	3.42×10^{-1}	685	−0.002 (−0.41, 0.41)	9.92×10^{-1}
FSCD							
EA	0.03	558	0.29 (0.10, 0.85)	2.32×10^{-2}	550	−0.24 (−0.50, 0.01)	6.19×10^{-2}
AA	0.02	608	0.40 (0.13, 1.28)	1.24×10^{-1}	588	−0.70 (−1.11, −0.28)	9.97×10^{-4}
Meta-analysis							
EA		**3789**	**0.32** (0.22, 0.47)	$\mathbf{6.03 \times 10^{-9}}$	**3777**	**−0.27** (−0.35, −0.19)	$\mathbf{1.33 \times 10^{-11}}$
AA		1721	0.41 (0.18, 0.93)	3.21×10^{-2}	1698	−0.34 (−0.58, −0.10)	6.00×10^{-3}
Combined		**5510**	**0.34** (0.24, 0.48)	$\mathbf{6.57 \times 10^{-10}}$	**5475**	**−0.28** (−0.35, −.020)	$\mathbf{3.24 \times 10^{-13}}$

Abbreviations: AA, African American; CI, confidence interval; EA, European American; MAF, minor allele frequency; OR, odds ratio.
[a]Logistic regression adjusted for sex, age (in quartiles), and nicotine and cocaine dependence.
[b]Linear regression of the log-transformed number of drinks adjusted for sex, age (in quartiles), and nicotine and cocaine dependence.
Bold values indicate genome-wide significant results.

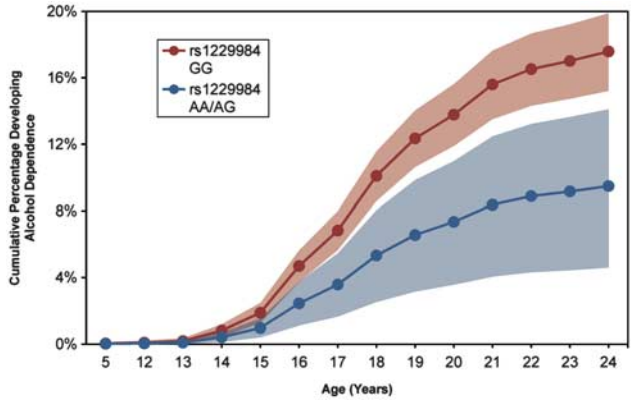

Figure 1 Survival estimates from proportional hazards regression model adjusted for gender and age. Shaded areas represent the 95% confidence intervals for each estimate.

data using BEAGLE.[25] Though over 98% of the genotyped GG samples are correctly assigned, only 68% of the AG heterozygote and AA homozygote genotypes are accurately imputed. As a result, testing this association required direct genotyping of this SNP.

Although the population-attributable risk of His48 differs because of the varying allele frequencies between ethnicities, at an individual level the effect of this *ADH1B* variant on the amount of alcohol consumed and on the risk of developing alcohol dependence is similar regardless of ancestry. This polymorphism is an exception to the generally small effect sizes associated with other variants that affect complex traits, including alcohol dependence; rs1229984 has a substantial effect on the risk of developing alcohol dependence, reducing it by two-thirds.

Conflict of interest

Dr Bierut, Dr Rice, Dr Goate and Dr Wang are inventors on the patent 'Markers for Addiction' (US 20070258898) covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. Dr NL Saccone is the spouse of Dr S Saccone, who is listed as an inventor on the patent. Dr Bierut served as a consultant for Pfizer Inc. in 2008. Dr Breslau, Dr Johnson, Ms Bertelsen, Mr Fox, Dr Agrawal, Dr Bucholz, Dr Grucza, Dr Hesselbrock, Dr Kramer, Dr Nurnberger, Dr Porjesz, Dr Schuckit, Dr Tischfield, Dr Foroud and Dr Edenberg declare no conflict of interest.

Acknowledgments

Support for the Collaborative Study on the Genetics of Alcoholism (COGA) was provided by the National Institute on Alcohol Abuse and Alcoholism, and the National Institute on Drug Abuse (U10 AA008401). Support for the Collaborative Genetic Study of Nicotine Dependence (COGEND) was provided by the National Cancer Institute (P01 CA089392). Sup-

port for the Family Study of Cocaine Dependence (FSCD) was provided by the National Institute on Drug Abuse (R01 DA013423, R01 DA019963). This work was also supported by NIDA Grant K02 DA021237 to LJB. A detailed list of acknowledgments and funding sources can be found in the Supplementary Information online. We thank Sherri Fisher for her organization and assistance in the development of this manuscript.

References

1 Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health* 2007; **30**: 5–13.
2 Chen CC, Lu RB, Chen YC, Wang MF, Chang YC, Li TK *et al.* Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *Am J Hum Genet* 1999; **65**: 795–807.
3 Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li TK *et al.* Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *Am J Hum Genet* 1991; **48**: 677–681.
4 Whitfield JB. Alcohol dehydrogenase and alcohol dependence: variation in genotype-associated risk between populations. *Am J Hum Genet* 2002; **71**: 1247–1250; author reply 1250–1241.
5 Li D, Zhao H, Gelernter J. Strong association of the alcohol dehydrogenase 1B Gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. *Biol Psychiatry* 2011; **70**: 504–512.
6 Ding JH, Li SP, Cao HX, Wu JZ, Gao CM, Liu YT *et al.* Alcohol dehydrogenase-2 and aldehyde dehydrogenase-2 genotypes, alcohol drinking and the risk for esophageal cancer in a Chinese population. *J Hum Genet* 2010; **55**: 97–102.
7 McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang SC, Byrnes G *et al.* A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet* 2011; **7**: e1001333.
8 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association: Washington, DC, 1994.
9 Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li T-K, Schuckit MA *et al.* The Collaborative Study on the Genetics of Alcoholism. *Alcohol Health Res World* 1995; **19**: 228–236.
10 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.
11 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.
12 Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF *et al.* Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007; **16**: 24–35.
13 Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* 1991; **86**: 1119–1127.
14 Sherva R, Rice JP, Neuman RJ, Rochberg N, Saccone NL, Bierut LJ. Associations and interactions between SNPs in the alcohol metabolizing genes and alcoholism phenotypes in European Americans. *Alcohol Clin Exp Res* 2009; **33**: 848–857.
15 Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB. Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend* 2008; **95**: 14–22.
16 Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger Jr JI *et al.* 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**: 149–158.
17 Bucholz KK, Hesselbrock VM, Shayka JJ, Nurnberger Jr JI, Schuckit MA, Schmidt I *et al.* Reliability of individual diagnostic criterion items for psychoactive substance dependence and the impact on diagnosis. *J Stud Alcohol* 1995; **56**: 500–505.
18 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.
19 Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 2007; **16**: 36–49.
20 Wang JC, Grucza R, Cruchaga C, Hinrichs AL, Bertelsen S, Budde JP *et al.* Genetic variation in the CHRNA5 gene affects mRNA levels and is associated with risk for alcohol dependence. *Mol Psychiatry* 2009; **14**: 501–510.
21 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**: 904–909.
22 Macgregor S, Lind PA, Bucholz KK, Hansell NK, Madden PA, Richter MM *et al.* Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum Mol Genet* 2009; **18**: 580–593.
23 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.
24 Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E *et al.* A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* 2010; **107**: 5082–5087.
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.

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