Multiple distinct CHRNB3–CHRNA6 variants are genetic risk factors for nicotine dependence in African Americans and European Americans

Robert C. Culverhouse[1,2], Eric O. Johnson[3], Naomi Breslau[4], Dorothy K. Hatsukami[5], Brooke Sadler[6,7], Andrew I. Brooks[8], Victor M. Hesselbrock[9], Marc A. Schuckit[10], Jay A. Tischfield[8], Alison M. Goate[6,11], Nancy L. Saccone[2,12] & Laura J. Bierut[6]

Department of Medicine, Washington University School of Medicine, Saint Louis, MO, USA,[1] Division of Biostatistics, Washington University School of Medicine, Saint Louis, MO, USA,[2] Behavioral Health Epidemiology Program, Research Triangle Institute International, Research Triangle Park, NC, USA,[3] Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA,[4] Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA,[5] Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO, USA,[6] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA,[7] Department of Genetics, Rutgers University, Piscataway, NJ, USA,[8] Department of Psychiatry, University of Connecticut Health Center, Farmington, CT, USA,[9] Department of Psychiatry, University of California, San Diego, San Diego, CA, USA,[10] Hope Center for Neurological Disorders, Department of Neurology, Washington University School of Medicine, Saint Louis, MO, USA[11] and Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA[12]

ABSTRACT

Aims Studies have shown association between common variants in the α6–β3 nicotinic receptor subunit gene cluster and nicotine dependence in European ancestry populations. We investigate whether this generalizes to African Americans, whether the association is specific to nicotine dependence and whether this region contains additional genetic contributors to nicotine dependence. **Design** We examined consistency of association across studies and race between the α6β3 nicotinic receptor subunit locus and nicotine, alcohol, marijuana and cocaine dependence in three independent studies. **Setting** United States of America. **Participants** European Americans and African Americans from three case–control studies of substance dependence. **Measurements** Subjects were evaluated using the Semi-Structured Assessment for the Genetics of Alcoholism. Nicotine dependence was determined using the Fagerström Test for Nicotine Dependence. **Findings** The single nucleotide polymorphism rs13273442 was associated significantly with nicotine dependence across all three studies in both ancestry groups [odds ratio (OR) = 0.75, $P = 5.8 \times 10^{-4}$ European Americans; OR = 0.80, $P = 0.05$ African Americans]. No other substance dependence was associated consistently with this variant in either group. Another SNP in the region, rs4952, remains modestly associated with nicotine dependence in the combined data after conditioning on rs13273442. **Conclusions** The common variant rs13273442 in the CHRNB3–CHNRA6 region is associated significantly with nicotine dependence in European Americans and African Americans across studies recruited for nicotine, alcohol and cocaine dependence. Although these data are modestly powered for other substances, our results provide no evidence that correlates of rs13273442 represent a general substance dependence liability. Additional variants probably account for some of the association of this region to nicotine dependence.

Keywords African Americans, chromosome 8, European Americans, genetic risk, rs13273442, rs1451240, rs4952, rs6474412, substance dependence.

INTRODUCTION

Although the genetics of substance dependence are complex, recent studies have successfully identified several genes that contribute to the development of nico-tine dependence. The region of chromosome 8p11 that includes the α6–β3 nicotinic receptor subunit gene cluster has been associated with smoking behavior [1–7]. A set of common, highly correlated variants ($r^2 = 1.0$) tagged by rs13273442, rs6474412 and rs1451240 is

associated with smoking behaviors at genome-wide significance in European ancestry populations: for the quantitative smoking phenotype 'cigarettes per day' ($P = 1.3 \times 10^{-8}$ using rs6474412 [1]) and for dependence using a case–control phenotype based on the Fagerström Test for Nicotine Dependence [8] ($P = 2.4 \times 10^{-8}$ using rs1451240 [7]). When robust evidence is discovered that a genetic variant contributes to dependence on a particular substance, three questions logically arise: (i) is the identified association robust across different ancestral populations; (ii) is the association specific to a single substance, or does it represent a general substance dependence risk; and (iii) are there additional, statistically independent genetic associations in the region?

The goal of this study is to explore these three questions. To examine consistency of the genetic association across populations, we meta-analyzed three independent studies of substance dependence, which collectively include 5171 subjects of European American and African American descent. Comparing results in both populations is essential for determining whether a finding can be generalized and may help to identify contributors to health disparities between African Americans and European Americans [9]. For example, although African Americans smoke fewer cigarettes than European Americans, they have a higher incidence of lung cancer (76 versus 70 per 100 000) [10]. This health disparity underscores the need for studies to identify genetic factors contributing to nicotine dependence in African Americans. To test whether the genetic association is specific to nicotine dependence or whether rs13273442 tags a non-specific genetic liability to substance dependence, we utilized the fact that each study comprehensively assessed nicotine, alcohol, marijuana and cocaine dependence. Whether or not the variation in this correlated cluster is specific to nicotine dependence is a key factor for improved understanding of the underlying biology of dependence. Finally, to examine the question of additional genetic associations to nicotine in this region, we performed analyses of rs4952 conditioned on the rs13273442 genotype. We targeted rs4952 as a potential second contributor to nicotine dependence in this region because it has been reported previously as being associated with nicotine dependence [6], and it is only modestly correlated with rs13273442 ($r^2 = 0.103$, D′ = 1.0 in European Americans, $r^2 = 0.005$, D′ = 1.0 in African Americans). Knowledge of multiple signals in this region will help to determine the focus of future genetic research.

METHODS

Data

Subjects were recruited by three independent studies of addiction: the Collaborative Genetic Study of Nicotine Dependence (COGEND), the Collaborative Study on the Genetics of Alcoholism (COGA) and the Family Study of Cocaine Dependence (FSCD). The Institutional Review Board at each contributing institution reviewed and approved the protocols for genetic studies under which all subjects were recruited. Subjects provided written informed consent. All subjects included in these analyses passed high quality control procedures for phenotypes and genotypes.

Collaborative Genetic Study of Nicotine Dependence (COGEND)

COGEND was designed as a community-based study of nicotine dependence. Subjects were recruited from Detroit, MI and St Louis, MO. More than 53 000 subjects were screened by telephone, more than 2800 were personally interviewed and more than 2700 donated blood samples for genetic studies [5,6]. To be recruited, subjects had to meet one of two ascertainment criteria. Nicotine-dependent case subjects were current smokers who met criteria for nicotine dependence defined as having a current FTND score of 4 or more [5]. Non-dependent control subjects smoked at least 100 cigarettes in their life-time, with a life-time maximum FTND score of 0 or 1. As part of the study design strategy, controls were oversampled to have a life-time maximum FTND score of 0, with 99.2% (910 of 917) of the European controls having a score of 0 and 82.1% (193 of 235) of the African American controls having a score of 0. There were no inclusionary or exclusionary criteria regarding alcohol or drug dependence. The present study includes 2597 unrelated, genotyped subjects from COGEND.

Collaborative Study on the Genetics of Alcoholism (COGA)

An alcohol-dependent case and non-alcohol-dependent control series of unrelated individuals was selected from more than 10 000 subjects who participated in the genetic arm of COGA. COGA systematically recruited families with multiple members affected with alcohol dependence and community-based comparison families from participating centers across the United States. From this larger sample, a subset of unrelated alcohol-dependent individuals was selected for genotyping [11]. Control subjects, who used alcohol but never met criteria for alcohol or drug dependence, were selected for genotyping. Because this study focused on alcohol dependence, nicotine dependence was not an exclusionary criterion for the control subjects. The present study includes 1403 unrelated, genotyped subjects from COGA.

Family Study of Cocaine Dependence (FSCD)

FSCD was designed as a genetic and environmental study of drug dependence with a focus on cocaine dependence.

Cocaine-dependent subjects were recruited systematically from chemical dependency treatment units in the greater St Louis, MO, metropolitan area [12]. Community-based control subjects were identified through the Missouri Family Registry and matched by age, race, gender and residential zip code. The community-based control subjects had used alcohol, but never met criteria for any alcohol, nicotine or other drug dependence. The present study includes 1171 unrelated genotyped subjects from FSCD.

Assessment

All three studies used an assessment of substance dependence based on the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) [13]. This shared methodology of interview administration, question format and queried domains allowed harmonization of phenotypic data across all studies. The FTND was used to evaluate life-time history of nicotine dependence, with a score of 4 or more defining 'dependence' for this study. Life-time history of dependence on alcohol, cocaine and marijuana was determined according to DSM-IV criteria. Although opioid and sedative dependence were assessed, these diagnoses were not included in this analysis because of the small number of subjects dependent on these substances and reduced power to detect association.

Ancestry was assessed by self-report and confirmed for all participants through a principal component analysis of ancestry informative markers using EIGENSTRAT software [14].

Demographics

The characteristics of the study participants are listed in Table 1. The sample comprises 68% European American and 32% African American subjects. Comorbid substance dependence is common, with more than half the subjects dependent on any given substance being dependent on at least one other substance as well. The extent of comorbidity in these data is illustrated in Fig. 1. Demographics broken down by ethnic subgroups can be found in Supporting information, Table S1. Rates of substance use and rates of dependence among those who have used are shown in Supporting information, Table S2. We define use of cigarettes as having smoked at least 100 cigarettes life-time (a common definition). For the other substances, use is defined as having tried it at least once. Each study's demographic characteristics are consistent with its recruitment design.

Genotyping and data cleaning

All genotyping was performed by the Center for Inherited Disease Research (CIDR). The COGA and FSCD samples were genotyped using Illumina Human1Mv1_C BeadChips. COGEND subjects were genotyped using two genome-wide single nucleotide polymorphism (SNP) arrays: Illumina Human1Mv1_C and Illumina HumanOmni2.5M. Extensive cleaning was undertaken to ensure high-quality genotyping by examining batch effects, call rates and other quality metrics [11].

Our analyses of consistency of association across populations and specificity of substance focused on correlates ($r^2 = 1.0$ in the European population) of the

Table 1 Characteristics of the sample.

Characteristic	COGEND	COGA	FSCD	Total
Sample size (*n*)	2575	1347	1170	5092
Sex, *n* (%)				
Males	978 (38.0%)	713 (52.9%)	581 (49.7%)	2319 (44.8%)
Females	1597 (62.0%)	634 (47.1%)	589 (50.3%)	2852 (55.2%)
Age, years				
Mean ± SD	36.6 ± 5.5	42.9 ± 10.3	37.0 ± 8.8	38.4 ± 8.3
Range	23–47	18–77	18–60	18–77
Ancestry, *n* (%)				
European American	1899 (73.7%)	974 (72.3%)	558 (47.7%)	3498 (67.6%)
African American	676 (26.3%)	373 (27.7%)	612 (52.3%)	1673 (32.4%)
Diagnoses, *n* (%)				
Nicotine dependence	1423 (55.3%)	634 (47.1%)	448 (38.3%)	2505 (49.2%)
Alcohol dependence	574 (22.3%)	855 (63.5%)	548 (46.8%)	1977 (38.8%)
Cocaine dependence	209 (8.1%)	384 (28.5%)	555 (47.4%)	1148 (22.6%)
Marijuana dependence	295 (11.5%)	289 (21.5%)	300 (25.6%)	884 (17.4%)

COGEND = Collaborative Genetic Study of Nicotine Dependence; COGA = Collaborative Study on the Genetics of Alcoholism; FSCD = Family Study of Cocaine Dependence; SD = standard deviation.
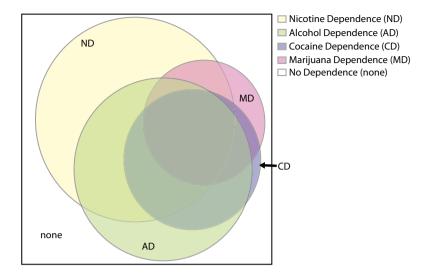
Figure 1 Comorbidity in the combined Collaborative Genetic Study of Nicotine Dependence (COGEND), Collaborative Study on the Genetics of Alcoholism (COGA) and Family Study of Cocaine Dependence (FSCD) data. Areas of the circles are proportional to the fraction of the sample dependent on each substance. Alcohol dependence (38.8% of the subjects), cocaine dependence (22.6%) and marijuana dependence (17.4%) were defined by DSM-IV criteria. Nicotine dependence (49.2%) was defined as Fagerström Test for Nicotine Dependence (FTND) score ≥ 4. Overlap represents comorbidity. For instance, in these data most of the subjects who were dependent on cocaine were also dependent on alcohol. In this sample, 34.4% of the individuals were not dependent on any of the four substances

common chromosome 8p11 variants rs6474412 and rs1451240 reported previously as having GWAS-significant associations to nicotine dependence. The SNP Annotation and Proxy Search (SNAP) tool [15], using the 1000 Genomes CEU sample, identified 18 additional SNPs that are highly correlated ($r^2 = 1.0$) with these two SNPs in subjects of European ancestry. Only one SNP genotyped in our data, rs13273442, is highly correlated in both Europeans ($r^2 = 1.0$) and the African Yoruba (YRI) populations ($r^2 \geq 0.89$) to the two SNPs rs6474412 and rs1451240. The A allele displays a frequency of 0.25 in the European CEU reference sample and a frequency of 0.64 in the African YRI reference sample. Supporting information Table S3 provides information about the correlations in YRI for the 20 SNPs that are correlated perfectly in the CEU reference.

Our examination of whether variants in addition to rs13273442 may contribute to the association between the *CHRNB3–CHRNA6* region and nicotine dependence focused on the nearby variant rs4952. The T allele of rs4952 is uncommon in both European Americans (3.0%) and African Americans (0.8%). The variant rs4952 is in linkage disequilibrium with rs13273442, but only modestly correlated with it ($r^2 = 0.103$, D′ = 1.0 in European Americans, $r^2 = 0.005$, D′ = 1.0 in African Americans).

Statistical analyses

The primary analysis approach for all three arms of our investigation was a meta-analysis of odds ratios (ORs)

using a consistently coded reference allele across the three studies in our largest ethnic sample, the subjects of European ancestry. This strategy maximized power to detect robust association while recognizing the potential impact of differences in ascertainment protocols and environmental differences among the studies. Analysis of the African American samples explored the consistency of signals in a distinct ancestral group. Finally, we conducted meta-analyses of the six ancestry × study strata to evaluate evidence for or against a consistent association across ancestral groups.

All computations of ORs and standard errors for the substudies were performed using SAS software [16]. In each analysis, covariates representing sex and age [using quartiles, defined by 34 years and younger (reference), 35–39 years, 40–44 years and 45 years and older] were included as categorical variables. Because use of a substance is required to develop dependence, binary covariates representing use of each of the four substances were also included. Meta-analyses were performed using the R package r-meta. None of the corresponding tests for heterogeneity based on Woolf's test were significant (range $P = 0.13–P = 0.95$). For this reason, reported *P*-values for the meta-analyses are based on fixed-effects models.

Primary genetic association analyses

To account for high rates of comorbid substance dependence in these subjects, we chose a method that could simultaneously model these disorders and their associa-

tion with genotype [17]. We utilized a logistic regression method in which genotype (coded 0, 1 or 2 to represent the number of coded risk alleles) is expressed as the dependent outcome variable:

$$\log\left(\frac{P_1 + P_2}{1 - P_1 - P_2}\right) = \alpha_1 + \sum_{i=1}^{k} \beta_i D_i$$
$$+ \text{demographic covariates.}$$

In this model, P_1 and P_2 represent an individual's probability of carrying one or two copies of the risk allele, respectively, and the D_i represents diagnoses for dependence on the substances evaluated in this study: nicotine, alcohol, marijuana and cocaine. This model makes a 'proportional odds' assumption which, in this case, is equivalent to assuming an additive genetic model. The demographic covariates used were sex and age quartiles. This model was then extended to include covariates for 'use' of each substance as well as dependence. Inclusion of 'use' covariates creates a more complex model, but has the advantage of distinguishing between individuals who have used a substance, but not become dependent, from those who cannot be dependent because they have never used the substance. These primary models, which controlled for demographics and substance use covariates within each study × ancestry stratum, were then meta-analyzed. The stratified results, along with the meta-analysis results, were the primary tools we used to investigate the cross-population consistency of results and the question of whether rs13273442 represented a general liability to dependence on any substance, or if it appeared to be specific to nicotine dependence.

To investigate the question of whether rs4952 reflected additional association between this region and nicotine dependence, we used a traditional logistic regression model. Nicotine dependence was the dependent variable, and rs4952 was the predictor of interest. The variant rs13273442 was included as a conditioning covariate, and variables representing dependence on alcohol, cocaine and marijuana were included to account for comorbidity, along with sex and age quartiles.

RESULTS

Association of rs13273442 in European and African ancestry samples

The variant rs13273442 is associated with nicotine dependence in the European Ancestry group (Table 2). The point estimates of the ORs display a consistent direction of effect for the A allele across all three studies, resulting in a highly significant association ($P = 5.8 \times 10^{-4}$) between the variant and nicotine dependence after adjusting for comorbid alcohol,

Table 2 Association between rs13273442 genotype and dependence for multiple substances in each of the individual subsamples together with meta-analysis within ancestry group and across all strata (fixed-effects meta-analysis model).

Ancestry	Study	n	Nicotine OR	(CI)	P	Alcohol OR	(CI)	P	Cocaine OR	(CI)	P	Marijuana OR	(CI)	P
EA	COGEND	1899	0.79	(0.65–0.96)	0.02	1.03	(0.82–1.29)	0.82	1.06	(0.70–1.62)	0.77	0.79	(0.55–1.12)	0.18
EA	COGA	974	0.56	(0.39–0.80)	1.4e-3	1.31	(0.88–1.96)	0.18	1.18	(0.78–1.79)	0.44	0.85	(0.58–1.26)	0.42
EA	FSCD	558	0.99	(0.59–1.72)	0.99	0.82	(0.49–1.37)	0.45	1.45	(0.74–2.84)	0.27	1.08	(0.66–1.76)	0.77
EA	Meta-analysis	3431	0.75	(0.64–0.88)	5.8e-4	1.05	(0.87–1.27)	0.59	1.17	(0.89–1.53)	0.26	0.87	(0.69–1.09)	0.22
AA	COGEND	676	0.97	(0.71–1.32)	0.84	1.19	(0.80–1.76)	0.39	0.93	(0.51–1.68)	0.80	1.31	(0.86–1.97)	0.20
AA	COGA	373	0.58	(0.36–0.96)	0.03	1.66	(0.86–3.18)	0.13	1.19	(0.64–2.19)	0.58	1.00	(0.61–1.65)	0.99
AA	FSCD	612	0.69	(0.45–1.08)	0.11	0.92	(0.61–1.37)	0.67	1.21	(0.62–2.38)	0.58	1.19	(0.78–1.80)	0.42
AA	Meta-analysis	1661	0.80	(0.64–1.00)	0.05	1.12	(0.87–1.45)	0.37	1.09	(0.76–1.57)	0.64	1.18	(0.91–1.52)	0.20
EA + AA	Meta-analysis	5092	0.77	(0.67–0.88)	8.6e-5	1.08	(0.93–1.25)	0.34	1.14	(0.92–1.42)	0.24	1.00	(0.84–1.18)	0.97

COGEND = Collaborative Genetic Study of Nicotine Dependence; COGA = Collaborative Study on the Genetics of Alcoholism; FSCD = Family Study of Cocaine Dependence; EA = European ancestry; AA = African American. Results for dependence variables from analyses predicting genotypes using all substance dependencies, sex and age quartile as predictors. Model: genotype = sex + age quartile + nicotine dependence + alcohol dependence + marijuana dependence + cocaine dependence + smoking exposure (yes/no at least 100 cigarettes/life-time) + ever drank alcohol (yes/no) + ever used cocaine (yes/no) + ever smoked marijuana (yes/no). Minimum P-value range for tests of heterogeneity: (0.14–0.95). CI = confidence interval; OR = odds ratio.

marijuana and cocaine dependence. The meta-analyzed European ancestry samples result in an estimated OR = 0.75 [95% confidence interval (CI) = 0.64–0.88].

In the African American sample (Table 2), the results for rs13273442 are consistent with those found in the European ancestry sample. After including all four substance dependence diagnoses in the model, the meta-analysis results for nicotine dependence indicate a protective effect for the A allele seen in the point estimates in each of the substudies. Meta-analyzing the three African American samples results in a nominally significant association ($P = 0.05$) and an OR estimate (OR = 0.80; 95% CI = 0.64–1.00) very similar to the one estimated from the European samples.

Meta-analyzing the six race × study strata (Table 2) tightened the confidence interval for the estimated nicotine dependence OR of 0.77, corresponding to an improvement in the *P*-value to 8.6×10^{-5}.

Although no significant heterogeneity was detected for any of the meta-analyses, for completeness we include the results of meta-analysis using random-effects models in Supporting information, Table S4. Results from traditional logistic regression (predicting dependence status using the variant as a predictor) were consistent with our primary analyses and are included in Supporting information, Table S5.

Association of rs13273442 is specific for nicotine dependence

None of the other substance dependence covariates displayed a statistically significant association with rs13273442. Alcohol dependence is common in the European American subjects (37.8%, $n = 1298$), and hence we expect this meta-analysis to be well powered to detect association to this substance. The European ancestry meta-analysis-estimated OR for alcohol, adjusted for all other substances, is a modest 1.05 ($P = 0.59$), and the ORs point estimates in the different studies range from protective (0.82 for FSCD) to risk (1.31 for COGA). Similarly, there is no clear genetic contribution to cocaine or marijuana dependence (OR = 1.17 and 0.87, respectively), but the power to detect an association for these substances is reduced due to the smaller number of dependence diagnoses [16.8% ($n = 578$) and 15.8% ($n = 541$), respectively].

Similar findings resulted from the African American analyses. Alcohol dependence does not demonstrate a statistically significant independent association in these data (OR = 1.12, $P = 0.37$). Neither cocaine nor marijuana dependence are associated significantly with rs13273442, although the power to detect association is reduced due to the smaller number of individuals dependent on these substances [34.3% ($n = 570$) and 20.7%

($n = 343$), respectively]. Furthermore, the direction of effect is not consistent across studies or populations.

By meta-analyzing the six race × study strata (Table 2), the differences between nicotine and the other substances became even more apparent. In contrast to the nicotine dependence meta-analysis *P*-value (8.6×10^{-5}), the meta-analysis *P*-values for alcohol, cocaine and marijuana are 0.34, 0.24 and 0.97, respectively, again demonstrating no statistical evidence for association between variation in rs13273442 and dependence on these substances independent of the association with nicotine dependence.

Results of testing for an additional genetic contributor beyond rs13273442

Because of the low allele frequency of rs4952, this association was modeled as homozygous for the common allele (CC) versus all other genotypes (CT and TT). Results from our analyses of rs4952, controlling for rs13273442 by including it as a covariate in the model, are listed in Table 3. The power is decreased for this analysis compared to our primary analysis of rs13273442 due to the low frequency of the T allele of rs4952. Nonetheless, in both ancestry groups the summary estimate of effect is strongly protective (OR = 0.75 for European Americans, OR = 0.59 for African Americans), although the confidence intervals are wide. Meta-analysis across the ancestral populations yields a statistically significant OR between the common CC genotype and carriers of the T allele (OR = 0.72, $P = 0.02$).

DISCUSSION

When a genetic risk factor for the development of substance dependence is identified, several questions arise. These coordinated meta-analyses of data from three studies that focused on nicotine, alcohol and cocaine dependence increase our understanding of the role that variants in the chromosome 8p11 region, which includes the α6–β3 nicotinic receptor subunit gene cluster, play in the development of substance dependence.

First, this study extends the finding of genetic association with nicotine dependence into an African American population. Thus far, the majority of genetic studies, including those of substance dependence, have been conducted using European ancestry samples [18]. Our findings indicate that the A allele of rs13273442 is associated with protection from nicotine dependence in African American subjects as well as European American subjects. This allele, which is common in both populations, displays a consistent protective effect across all three of our African American samples and represents a similar strong decrease in risk for each copy of the A allele

Table 3 Association between rs4952 and nicotine dependence, conditioned on rs13273442 (fixed-effects meta-analysis model).

| Ancestry | Study | n | Nicotine (fixed-effects) | | | Het P |
			OR	(CI)	P	
EA	COGEND	1899	0.82	(0.56–1.21)	0.32	
EA	COGA	974	0.44	(0.22–0.86)	0.02	
EA	FSCD	558	1.39	(0.47–4.12)	0.55	
EA	Meta-analysis	3431	0.75	(0.54–1.03)	0.08	0.15
AA	COGEND	676	0.43	(0.16–1.17)	0.10	
AA	COGA	373	1.38	(0.19–10.14)	0.75	
AA	FSCD	612	1.81	(0.19–17.19)	0.61	
AA	Meta-analysis	1661	0.59	(0.28–1.17)	0.13	0.33
EA + AA	Meta-analysis	5092	0.72	(0.53–0.96)	0.02	0.26

Het P = P-value for a test of heterogeneity among the meta-analysis strata. Model: nicotine dependence = rs4952 + rs13273442 + sex + age quartile + alcohol dependence + cocaine dependence + marijuana dependence. COGEND = Collaborative Genetic Study of Nicotine Dependence; COGA = Collaborative Study on the Genetics of Alcoholism; FSCD = Family Study of Cocaine Dependence; EA = European ancestry; AA = African American; CI = confidence interval; OR = odds ratio.

carried in both ancestral populations (OR = 0.80 for African Americans; OR = 0.75 for European Americans). Meta-analyzing the European Ancestry and African American data sets strengthened the statistical evidence for association, resulting in a P-value of 8.6×10^{-5} and further increasing confidence in the validity of this genetic finding. This conclusion is supported by a recent study of nicotine dependence that focused on this region [19] and reported a significant association for rs6474412, a correlate of rs13273442 ($r^2 \geq 0.89$), in three populations: European American, African American and Asian.

Next, these analyses demonstrate that the association between a common variant in the chromosome 8p11 region is specific to nicotine dependence and does not represent a general substance dependence liability locus. These meta-analyses found that the nicotine dependence result is robust, even after adjusting for comorbidity with other substances. In contrast, none of the other three dependence diagnoses (alcohol, cocaine, marijuana) displayed a statistically significant association in the meta-analysis. The lack of association results must be interpreted with knowledge of the limitations of power. The sample sizes for the individual strata were modest. For the modest effect sizes expected for complex phenotypes such as alcohol and other drug dependence, large samples are needed to confidently rule out an association. Thus, while the evidence points to this locus having a specific risk effect associated with nicotine dependence, more modest genetic risks for alcohol, marijuana or cocaine may be present, but beyond the limits of detection in this study.

We also examined this locus to determine if the association signal was represented by one variant or if there were potentially multiple distinct variants associated with nicotine dependence. The variant rs4952 was the

focus of this investigation, and we chose it because of its low correlation with rs13273442 in both ancestral populations ($r^2 = 0.10$, in European Americans, $r^2 < 0.01$ in African Americans) and a previous report of association to nicotine dependence [6]. We recognized that the power of this analysis would be decreased due to the low frequency of the T allele for rs4952 [European Americans (3.0%) and African Americans (0.8%)]. In spite of the modest power, tests of association between rs4952 and nicotine dependence conditioned on the common variant rs13273442 resulted in point estimates indicating a protective effect for the minor allele in both groups (OR = 0.75 for European Americans, OR = 0.59 for African Americans). Meta-analysis across the two ancestral groups tightened the confidence interval enough that the result is statistically significant (OR = 0.72; P = 0.02). We thus believe that multiple variants in this region contribute to the genetic risk for nicotine dependence.

In conclusion, substance dependence has a major impact on public health, and smoking is a leading modifiable contributor to death world-wide, killing more than 5 million people annually [20]. Common variation in the chromosome 8p11 region, which contains the α6 and β3 nicotinic receptor subunit genes, contributes to heaviness of smoking in European Americans and is associated with nicotine dependence in African Americans. General liability and substance-specific genetic risk variants are predicted to contribute to the development of dependence. However, our data suggest that the genetic variants tagged by rs13273442 in this region most likely represent a risk factor that is specific to nicotine dependence. This variant does not appear to be associated with a general liability to alcohol or other drug dependence, although the power is reduced to confidently rule out a risk for other substance dependence. Finally, our findings

suggest that multiple variants contribute to the association between this region and nicotine dependence. This work represents important next steps to improve our understanding of the genetic factors that underlie addiction, which we hope will provide insights into how to reduce nicotine dependence and improve smoking cessation.

Declaration of interests

Drs Bierut and Goate are listed as inventors on Issued US Patent 8,080,371, 'Markers for Addiction' covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. Dr NL Saccone is the spouse of Dr SF Saccone, who is also listed as an inventor on the above patent. In 2013, Dr Marc Schuckit gave a presentation on the genetics of alcohol to Anheuser-Busch-InBev executives at Yale University. He received the cost of his travel and an honorarium.

Acknowledgements

References

1. Thorgeirsson T. E., Gudbjartsson D. F., Surakka I., Vink J. M., Amin N., Geller F. *et al.* Sequence variants at CHRNB3–CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* 2010; **42**: 448–53.
2. Hoft N. R., Corley R. P., McQueen M. B., Schlaepfer I. R., Huizinga D., Ehringer M. A. *et al.* Genetic association of the CHRNA6 and CHRNB3 genes with tobacco dependence in a nationally representative sample. *Neuropsychopharmacology* 2009; **34**: 698–706.
3. Zeiger J. S., Haberstick B. C., Schlaepfer I., Collins A. C., Corley R. P., Crowley T. J. *et al.* The neuronal nicotinic receptor subunit genes (CHRNA6 and CHRNB3) are associated with subjective responses to tobacco. *Hum Mol Genet* 2008; **17**: 724–34.
4. Ehringer M. A., McQueen M. B., Hoft N. R., Saccone N. L., Stitzel J. A., Wang J. C. *et al.* Association of CHRN genes with 'dizziness' to tobacco. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B**: 600–9.
5. Bierut L. J., Madden P. A., Breslau N., Johnson E. O., Hatsukami D., Pomerleau O. F. *et al.* Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007; **16**: 24–35.
6. Saccone S. F., Hinrichs A. L., Saccone N. L., Chase G. A., Konvicka K., Madden P. A. *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.
7. Rice J. P., Hartz S. M., Agrawal A., Almasy L., Bennett S., Breslau N. *et al.* CHRNB3 is more strongly associated with Fagerström test for cigarette dependence-based nicotine dependence than cigarettes per day: phenotype definition changes genome-wide association studies results. *Addiction* 2012; **107**: 2019–28.
8. Heatherton T. F., Kozlowski L. T., Frecker R. C., Fagerström K. O. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *Br J Addict* 1991; **86**: 1119–27.
9. Centers for Disease Control and Prevention (CDC). CDC Health Disparities and Inequalities Report—United States, 2011. *Morb Mortal Wkly Rep* 2011; **60**: 1–113.
10. Centers for Disease Control and Prevention (CDC). Racial/ethnic disparities and geographic differences in lung cancer incidence—38 States and the District of Columbia, 1998–2006. *Morb Mortal Wkly Rep* 2010; **59**: 1434–8.
11. Bierut L. J., Agrawal A., Bucholz K. K., Doheny K. F., Laurie C., Pugh E. *et al.* A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* 2010; **107**: 5082–7.
12. Bierut L. J., Strickland J. R., Thompson J. R., Afful S. E., Cottler L. B. Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend* 2008; **95**: 14–22.
13. Bucholz K. K., Cadoret R., Cloninger C. R., Dinwiddie S. H., Hesselbrock V. M., Nurnberger J. I. Jr. *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**: 149–58.
14. Price A. L., Patterson N. J., Plenge R. M. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–9.
15. Johnson A. D., Handsaker R. E., Pulit S. L., Nizzari M. M., O'Donnell C. J., de Bakker P. I. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; **24**: 2938–9.
16. SAS/STAT (2002–2003). Cary, NC: SAS Institute, Inc.
17. Grucza R. A., Wang J. C., Stitzel J. A., Hinrichs A. L., Saccone S. F., Saccone N. L. *et al.* A risk allele for nicotine dependence in CHRNA5 is a protective allele for cocaine dependence. *Biol Psychiatry* 2008; **64**: 922–9.
18. Hartz S. M., Johnson E. O., Saccone N. L., Hatsukami D., Breslau N., Bierut L. J. *et al.* Inclusion of African Americans in genetic studies: what is the barrier? *Am J Epidemiol* 2011; **174**: 336–44.
19. Cui W. Y., Wang S., Yang J., Yi S. G., Yoon D., Kim Y. J. *et al.* Significant association of CHRNB3 variants with nicotine dependence in multiple ethnic populations. *Mol Psychiatry* 2013; **18**: 1149–51.
20. World Health Organization (WHO). *Tobacco fact sheet*. Geneva: WHO; 2011.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1 Demographics for the samples by ancestry group.

Table S2 Use prevalence and dependence prevalence in those who have used the substance.

Table S3 Correlation substructure in Africans for single nucleotide polymorphisms (SNPs) correlated perfectly ($R^2 = 1.0$) with rs13273442 in Europeans.

Table S4 Association between rs13273442 genotype and dependence for multiple substances in each of the individual subsamples together with meta-analysis within ancestry group and across all strata (random effects meta-analysis model).

Table S5 Association between rs13273442 genotype and dependence for four substances based on logistic regression with substance-specific dependence as the dependent variable (separate analyses for each substance) (fixed-effect meta-analysis model).