Addiction
RESEARCH REPORT
doi:10.1111/j.1360-0443.2012.03922.x

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

John P. Rice[1], Sarah M. Hartz[1], Arpana Agrawal[1], Laura Almasy[2], Siiri Bennett[3], Naomi Breslau[4], Kathleen K. Bucholz[1], Kimberly F. Doheny[5], Howard J. Edenberg[6], Alison M. Goate[1], Victor Hesselbrock[7], William B. Howells[1], Eric O. Johnson[8], John Kramer[9], Robert F. Krueger[16], Samuel Kuperman[10], Cathy Laurie[3], Teri A. Manolio[11], Rosalind J. Neuman[1], John I. Nurnberger Jr[12], Bernice Porjesz[13], Elizabeth Pugh[5], Erin M. Ramos[11], Nancy Saccone[14], Scott Saccone[1], Marc Schuckit[15] & Laura J. Bierut[1], as part of the GENEVA Consortium

Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA,[1] Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA,[2] Department of Biostatistics, University of Washington, Seattle, WA, USA,[3] Department of Epidemiology, Michigan State University, East Lansing, MI, USA,[4] Center for Inherited Disease Research, Johns Hopkins School of Medicine, Baltimore, MD, USA,[5] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA,[6] Department of Psychiatry, University of Connecticut, Farmington, CT, USA,[7] Division of Health, Social and Economic Research, Research Triangle Institute International, Research Triangle Park, NC, USA,[8] Department of Psychiatry, University of Iowa College of Medicine, Iowa City, IA, USA,[9] Division of Child Psychiatry, University of Iowa Hospitals, Iowa City, IA, USA,[10] National Human Genome Research Institute, Bethesda, MD, USA,[11] Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA,[12] Department of Psychiatry and Behavioral Sciences, State University of New York, Brooklyn, NY, USA,[13] Department of Genetics, Washington University School of Medicine, St Louis, MO, USA,[14] Department of Psychiatry, University of California–San Diego, La Jolla, CA, USA[15]and Department of Psychology, University of Minnesota, Minneapolis, MN, USA[16]

ABSTRACT

Aims Nicotine dependence is a highly heritable disorder associated with severe medical morbidity and mortality. Recent meta-analyses have found novel genetic loci associated with cigarettes per day (CPD), a proxy for nicotine dependence. The aim of this paper is to evaluate the importance of phenotype definition (i.e. CPD versus Fagerström Test for Cigarette Dependence (FTCD) score as a measure of nicotine dependence) on genome-wide association studies of nicotine dependence. **Design** Genome-wide association study. **Setting** Community sample. **Participants** A total of 3365 subjects who had smoked at least one cigarette were selected from the Study of Addiction: Genetics and Environment (SAGE). Of the participants, 2267 were European Americans, 999 were African Americans. **Measurements** Nicotine dependence defined by FTCD score ≥4, CPD. **Findings** The genetic locus most strongly associated with nicotine dependence was rs1451240 on chromosome 8 in the region of *CHRNB3* [odds ratio (OR) = 0.65, $P = 2.4 \times 10^{-8}$]. This association was further strengthened in a meta-analysis with a previously published data set (combined $P = 6.7 \times 10^{-16}$, total $n = 4200$). When CPD was used as an alternate phenotype, the association no longer reached genome-wide significance ($\beta = -0.08$, $P = 0.0004$). **Conclusions** Daily cigarette consumption and the Fagerstrom Test for Cigarette Dependence show different associations with polymorphisms in genetic loci.

Keywords CHRNB3, genome-wide association studies, meta-analysis, phenotype definitions.

Correspondence to: John P. Rice, Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63110, USA.
E-mail: john@zork.wustl.edu
Submitted 12 September 2011; initial review completed 7 November 2011; final version accepted 18 April 2012

INTRODUCTION

Tobacco use is one of the leading causes of mortality world-wide. Because most regular smoking occurs in the context of nicotine dependence, nicotine dependence is frequently the focus of studies on tobacco use [1]. Among current smokers, approximately 60% are nicotine-dependent based on the Fagerström Test for Cigarette Dependence (FTCD), a well-established scale for assessing nicotine dependence [2]. Evidence for genetic factors contributing to the risk of smoking behaviors and nicotine dependence is seen in the clustering of heavy smoking and nicotine dependence in families and the similarity of smoking behaviors in genetically identical twins [3,4].

Numerous studies have found an association between nicotine dependence and single nucleotide polymorphisms (SNPs) in the α5 nicotinic receptor gene, *CHRNA5* [5–12]. To maximize the power to detect genetic variants associated with smoking quantity (as an alternative to nicotine dependence), three large research consortia performed genome-wide association meta-analyses of cigarettes per day (CPD) in a combined sample of more than 75 000 subjects [13–15]. The strongest association was the variant in *CHRNA5*, with a combined *P*-value less than 1×10^{-70}. However, several other variants were discovered with genome-wide significance: variants near the nicotinic receptor subunit genes *CHRNB3* and *CHRNA6* on chromosome 8 (rs6474412, $P = 1.4 \times 10^{-8}$, a region associated previously with other nicotine phenotypes [12,16–19]), variants near the nicotine metabolizing enzyme genes *CYP2A6* and *CYP2B6* on chromosome 19 (rs4105144, $P = 2.2 \times 10^{-12}$) and variants in a non-coding region located on chromosome 10q23 (rs1329650, $P = 5.7 \times 10^{-10}$).

Because genome-wide association studies (GWAS) have stringent *P*-value requirements, the issue of statistical power is highly relevant. Although most attempts at maximizing power in GWAS focus on increasing sample size, as in the above meta-analyses, power can also be increased by reducing phenotypic variance by increasing either precision of measurement or phenotypic homogeneity of the subjects.

Although CPD is the most common phenotypic measurement of smoking behavior, there is strong epidemiological evidence that the number of cigarettes smoked per day varies across cultures and ethnicities. For example, African Americans smoke fewer cigarettes than do European Americans [20]. However, the FTCD score, ranging from 0 to 10 where CPD can account for a maximum of four levels, defined as 1–10, 11–20, 21–30 or 31 or more, appears to be an invariant measure of nicotine dependence across ethnicities [20]. Therefore, we hypothesized that a GWAS with FTCD rather than CPD may have increased power to detect variants associated with nicotine dependence, especially in a multi-ethnic sample.

To clarify the relationship between FTCD-based nicotine dependence and CPD in the context of a GWAS, we defined FTCD-based nicotine-dependent cases and non-nicotine-dependent controls from the Study of Addiction: Genetics and Environment (SAGE), a multi-ethnic, case–control sample selected for alcohol dependence [21]. By including a diverse set of study participants, we have the opportunity to extend our investigation beyond the previous studies in European Americans, and specifically address the role that phenotype definition plays in genome-wide association studies.

METHODS

Data

This analysis uses a subset of subjects who have ever smoked from the SAGE, part of the Gene Environment Association Studies (GENEVA) program of the National Institutes of Health (NIH) Genes, Environment, and Health Initiative [22]. For the overall SAGE project, unrelated alcohol-dependent cases ($n = 1897$) and non-alcohol-dependent control subjects ($n = 1937$) were selected from three large, complementary data sets: 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). Characteristics of the individual data sets are given in the supporting information (Table S3). The Institutional Review Board at each contributing institution reviewed and approved the protocols for genetic studies of substance dependence under which all subjects were recruited. Subjects provided informed consent for genetic studies and agreed to allow their genetic and phenotypic information to be shared with qualified investigators through NIH repositories. For each of the three studies, we describe the sampling schemes used to recruit subjects and select for genotyping.

COGEND

COGEND was designed as a community based case–control family study of nicotine dependence. The COGEND ascertainment protocol identified current smokers with nicotine dependence defined by an FTCD score ≥ 4 (maximum score of 10); non-nicotine-dependent subjects who had smoked at least 100 cigarettes and had a life-time FTCD score of zero were also recruited. All subjects were ascertained from Detroit and St Louis. Approximately 53 000 subjects were screened by telephone, 2800 were interviewed personally and nearly 2700 donated blood samples for genetic studies. The COGEND study contributed 275 nicotine-dependent

cases and 1082 non-nicotine-dependent smoking controls to these nicotine dependence genetic analyses. Of these, 125 nicotine-dependent cases and 706 non-nicotine-dependent controls overlap with the samples used in Bierut *et al*. and Saccone *et al*. [5,12].

COGA

A case–control series of unrelated individuals was selected from more than 8000 subjects who participated in the genetic arm of COGA. COGA systematically recruited subjects from participating centers in Hartford, Connecticut; Indianapolis, Indiana; Iowa City, Iowa; New York City, New York; San Diego, California; St Louis, Missouri; and Washington, DC. For SAGE, cases met life-time criteria for DSM-IV alcohol dependence; the majority of cases were recruited from alcoholism treatment centers. Control subjects, biologically unrelated to cases, were individuals who consumed alcohol but never experienced any significant alcohol- or drug-related problems, as reported on the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA). The COGA study contributed 649 nicotine-dependent cases and 553 non-nicotine-dependent smoking control subjects to the subsequent nicotine dependence genetic analyses.

FSCD

Cocaine-dependent cases were recruited systematically from chemical dependency treatment units in the greater St Louis metropolitan area. Community-based control subjects were identified through the Missouri Family Registry and matched by age, race, gender and residential zip code. Controls were biologically unrelated individuals from the same communities who consumed alcohol, but had no life-time history of dependence on any substance. FSCD contributed 370 nicotine-dependent cases and 436 non-nicotine-dependent smoking control subjects.

Nicotine dependence and smoking phenotypes

We used several approaches to define the most appropriate phenotype for the genetic association analysis of smoking. First, to enhance sample homogeneity, we eliminated 143 individuals who had substance dependence diagnoses other than nicotine dependence, alcohol dependence, alcohol and cocaine dependence or other substance abuse (except nicotine dependence). These 143 subjects are labeled as 'Other' in the SAGE files available through the database of Genotypes and Phenotypes (dbGaP; accession number phs000092.v1.p1).

We included all subjects who had ever smoked a cigarette. Cases were defined with an FTCD score of 4 or more and controls had an FTCD score ≤ 3, based on the dichotomous nicotine dependence phenotype used for COGEND. For our case definition, we also included 100 individuals who smoked on average more than a pack a day, but had a missing FTCD score. This is consistent with previous research that found that most individuals who smoke more than 20 cigarettes a day have an FTCD score of 4 or greater [8]. The final sample for association testing contains 1294 nicotine-dependent cases and 2071 non-dependent controls who have smoked at least one cigarette (Table 1).

A relatively large proportion of individuals in this sample have a diagnosis of alcohol dependence and/or cocaine dependence because the COGA and FSCD studies were designed to examine these disorders. This reflects the elevated rates of nicotine dependence in individuals with comorbid substance dependence conditions.

CPD is an alternative phenotype for smoking behavior that has been studied in previous GWAS. To evaluate CPD, a 4-point ordinal scale was created: at most 10 cigarettes daily, 11–20 cigarettes daily, 21–30 cigarettes daily and more than 30 cigarettes daily. This phenotype has been used in other studies [9]. We used this to examine further our top GWAS finding.

Genotyping and data cleaning

As part of GENEVA, DNA samples were genotyped on the Illumina Human 1M-Duo beadchip by the Center for Inherited Disease Research (CIDR) at Johns Hopkins University.

The Illumina 1M–Duo array has a total of 1 072 820 probes, of which 23 812 are 'intensity-only', leaving 1 049 008 probes as SNP assays. These SNP assays demonstrate excellent data quality—95% of SNPs have a missing call rate <1.4% and the median of the missing call rate is 0.05%. A thorough data cleaning procedure was applied to ensure the highest possible data quality, including the use of HapMap controls, detection of gender and chromosomal anomalies, hidden relatedness, population structure, missing call rates, batch effects, Mendelian error detection, duplication error detection and Hardy–Weinberg equilibrium [22]. Of the 1 049 008 SNPs, 948 658 SNPs passed data cleaning procedures. Further details are provided in the comprehensive data cleaning report posted on the GENEVA website (http://www.genevastudy.org/docs/GENEVA_Alcohol_QC_report_8Oct2008.pdf).

Population stratification

The composition of the samples in terms of self-identified ethnicity was 2267 European Americans (self-reported 'white'), 99 Hispanic Americans and 999 African Americans (self-reported 'black'). Subjects who self-identified as both African American and Hispanic were labeled as African Americans.

Table 1 Characteristics of the sample.

	Nicotine-dependent smokers			*Controls*		
	EA n = 828	*AA n = 466*	*Total n = 1294*	*EA n = 1538*	*AA n = 533*	*Total n = 2071*
No comorbid diagnosis	9%	6%	8%	73%	21%	71%
Alcohol dependence	91%	94%	41%	27%	35%	30%
Cocaine dependence	43%	66%	51%	8%	20%	11%
Cigarettes per day						
0–10	4%	24%	11%	88%	90%	89%
11–20	39%	53%	44%	11%	9%	10%
21–30	25%	9%	19%	1%	0%	1%
≥31	32%	14%	25%	0%	0%	0%
Age (years)						
Mean age	40	40	40	38	40	39
<35	29%	18%	25%	35%	20%	31%
35–39	23%	27%	24%	23%	25%	24%
40–44	23%	32%	26%	25%	32%	27%
≥45	25%	23%	24%	16%	23%	18%
Male	60%	55%	58%	36%	47%	39%
Female	40%	45%	42%	63%	53%	61%
Income <$20 000	15%	37%	21%	3%	18%	5%
No high school degree	14%	33%	23%	3%	13%	7%

EA: European Americans; AA: African Americans.

We used the software package EIGENSTRAT [23] to calculate principal components reflecting continuous variation in allele frequencies, representing ancestral differences in subjects. Two principal components were identified. The first distinguished African American from European American participants, and the second distinguished Hispanic from non-Hispanic subjects. These scores were included to control for effects of population stratification. In addition, we used self-reported ethnicity (European American, African American or Hispanic) as a categorical variable and compared results with those using the first two principal components.

Statistical analyses

Two genome-wide association analyses were conducted in PLINK [24]. The first used logistic regression with nicotine dependence as the dependent variable, and the second used linear regression with CPD as the dependent variable. Genotypes were coded log-additively (0, 1, 2 copies of the minor allele). Covariates representing sex, age [defined, using quartiles, as 3 indicator variables representing 34 years and younger (reference), 35–39 years, 40–44 years and 45 years and older], self-reported ethnicity and alcohol and cocaine dependence (the diagnoses used to ascertain subjects for the original COGA and FSCD studies) were included.

The QQ-Plot of the association between nicotine dependence (FTCD ≥ 4) and the 948 658 SNPs may be seen in Fig. S1 of the supporting information. The lambda value is 1.02, reflecting adequate control of population stratification using self-reported ethnicity. The Manhattan plots for the two analyses are shown in Fig. S2 of the supporting information.

To evaluate the robustness of the findings, we analyzed the association between rs1451240 and several smoking phenotypes (CPD and FTCD, coded as a continuous variable in linear regression, dichotomous variable in logistic regression and ordinal variable in ordered logistic regression). In addition, we modified the inclusion criteria to include only individuals who smoked regularly to determine whether or not this changed the results.

We performed a meta-analysis of the independent subset of this study using Metal [25,26]. The primary sample we used comprised the 2590 subjects who were not included in the previous GWAS of nicotine dependence [5]. These results were combined with the published odds ratio (OR) and corresponding statistics from the association study in 1610 subjects from Saccone *et al.* [27].

We graphically evaluated linkage disequilibrium and GWAS *P*-values using the software WGAviewer (http://people.genome.duke.edu/~dg48/WGAViewer/) [28].

Population-based analysis of smoking phenotypes

To clarify discrepancies between results obtained using FTCD-based nicotine dependence and results obtained

with CPD, we compared FTCD with CPD in a population-based sample. As part of recruitment to the COGEND study, subjects were selected randomly from the St Louis region using the Missouri Family Registry, sent a letter, and telephoned [3,12]. To evaluate the relationship between the phenotypes of CPD and the FTCD score, of the 28 658 subjects who completed the telephone screening, we selected the 14 343 subjects who had smoked at least 100 cigarettes in their life-time and calculated the correlation between FTCD and CPD. We also calculated the polychoric correlations between the FTCD score components and CPD.

RESULTS

The region most strongly associated with nicotine dependence in this study is represented by 14 SNPs in a 40-kb region on chromosome 8 (Table 2), with a single bin reaching genome-wide significance ($P < 5 \times 10^{-8}$). The most significant SNP is rs1451240 (OR = 0.65, $P = 2.4 \times 10^{-8}$), and this SNP tags a bin on chromosome 8 including part of *CHRNB3* in both African Americans and European Americans (supporting information, Figs S2 and S3). Of interest, the SNP most strongly associated with nicotine dependence in previous studies, rs16969968 in *CHRNA5* on chromosome 15, had an OR consistent with published studies (OR = 1.31), but the *P*-value was not genome-wide significant ($P = 6.2 \times 10^{-4}$). In contrast with the results from nicotine dependence, the GWAS using CPD as the dependent variable did not find any SNP to be significantly associated at a genome-wide level (Fig. 1).

The *CHRNB3* region of chromosome 8 in general, and this signal in particular, has been associated previously with CPD, but of the many genome-wide association studies using CPD as the primary phenotype [7,9,15,29], the only previously published genome-wide significant association with this region has been in a large meta-analysis including over 75 000 subjects [15]. Specifically, rs1451240 has an r2 of 1.0 (based on 1000 genome pilot 1 data, CEU) with the two chromosome 8 SNPs published in the meta-analysis. To clarify the difference between our study of only 3365 subjects and the large meta-analysis, we examined the effects of phenotype definition, ethnicity and comorbid diagnoses on the association with this SNP.

There are two primary differences between the FTCD definition of nicotine dependence and CPD: (i) nicotine dependence is based on a 10-point FTCD scale computed from six items including CPD, and (ii) nicotine dependence is a dichotomous variable whereas CPD is an ordinal variable. Therefore, we created four phenotypes for evaluation: dichotomous nicotine dependence, dichotomous CPD, ordinal FTCD score and ordinal CPD. Using each of

Table 2 The top 10 associated SNPs from the GWAS analysis of nicotine dependence.

SNP	*Chr*	*Position (Ensembl 56)*	*Locus and context*	*Test allele*	*EA cases*	*EA cntrls*	*EA OR (95% CI)*	*aa cases*	*AA cntrls*	*AA OR (95% CI)*	*Hisp cases*	*Hisp cntrls*	*Full data set adjusted OR (95% CI)*	*P*
rs1451240	8	42,546,711	CHRNB3, Intergenic	A	0.18	0.23	0.65 (0.53, 0.79)	0.62	0.68	0.64 (0.50, 0.81)	0.25	0.31	0.65 (0.56, 0.76)	2.44E-08
rs4736835	8	42,547,033	CHRNB3, Intergenic	T	0.19	0.23	0.65 (0.53, 0.80)	0.62	0.68	0.64 (0.51, 0.82)	0.25	0.31	0.65 (0.56, 0.76)	2.99E-08
rs10958725	8	42,524,584	CHRNB3, Intergenic	T	0.18	0.23	0.65 (0.53, 0.80)	0.66	0.71	0.62 (0.48, 0.79)	0.25	0.31	0.65 (0.56, 0.76)	3.06E-08
rs6474413	8	42,551,064	CHRNB3, Upstream	C	0.19	0.23	0.65 (0.53, 0.79)	0.66	0.71	0.63 (0.50, 0.81)	0.27	0.31	0.65 (0.56, 0.76)	3.62E-08
rs1955185	8	42,549,647	CHRNB3, Upstream	G	0.19	0.23	0.65 (0.53, 0.80)	0.66	0.71	0.64 (0.50, 0.82)	0.27	0.31	0.65 (0.56, 0.76)	4.64E-08
rs4950	8	42,552,633	CHRNB3, 5prime Utr	C	0.19	0.23	0.64 (0.52, 0.79)	0.7	0.75	0.65 (0.50, 0.85)	0.28	0.33	0.65 (0.56, 0.76)	9.50E-08
rs7004381	8	42,551,161	CHRNB3, Upstream	A	0.19	0.23	0.65 (0.53, 0.80)	0.57	0.62	0.68 (0.54, 0.86)	0.25	0.31	0.67 (0.57, 0.77)	9.93E-08
rs13280604	8	42,559,586	CHRNB3, Intronic	G	0.19	0.23	0.65 (0.53, 0.80)	0.7	0.74	0.65 (0.50, 0.84)	0.28	0.33	0.66 (0.56, 0.77)	1.04E-07
rs10958726	8	42,535,909	CHRNB3, Intergenic	G	0.18	0.23	0.65 (0.53, 0.80)	0.57	0.63	0.68 (0.54, 0.86)	0.25	0.3	0.67 (0.57, 0.77)	1.15E-07
rs13273442	8	42,544,017	CHRNB3, Intergenic	A	0.19	0.23	0.65 (0.53, 0.80)	0.57	0.62	0.69 (0.55, 0.87)	0.25	0.31	0.67 (0.58, 0.78)	1.38E-07

SNP: single nucleotide polymorphism; OR: odds ratio; CI: confidence interval; EA: European Americans; AA: African Americans; Hisp: Hispanic; GWAS: genome-wide association studies.

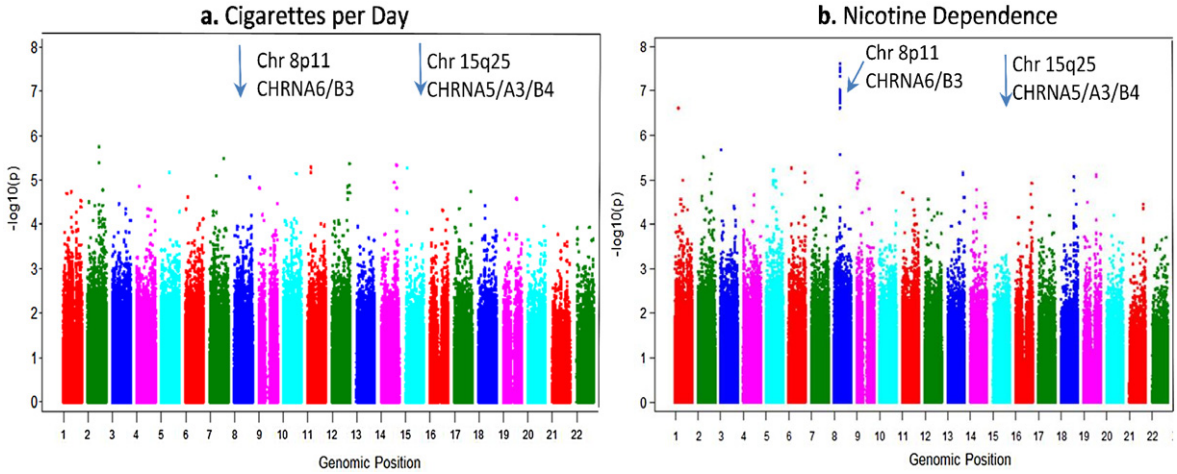

Figure 1 Manhattan plot of *P*-values from multivariate logistic regression models testing for association with cigarettes per day (a) and nicotine dependence (b)

Table 3 Stratified analysis of association between smoking behavior and the *CHRNB3* region on chromosome 8 tagged by rs1451240.

	Association between smoking & rs1451240							
		Dichotomous traits				*Ordinal traits*		
		FTCD ≤ 3 versus FTCD ≥ 4		*CPD ≤ 20 versus CPD > 20*		*FTCD*		*CPD*	
	n	*OR*	*P*	*OR*	*P*	*β*	*P*	*β*	*P*
Full sample	3365	0.65	2.E-08	0.73	0.0004	−0.5	0.0004	−0.08	0.0004
Stratified analyses									
Male	1555	0.69	0.0004	0.75	0.01	−0.65	0.002	−0.1	0.006
Female	1810	0.60	0.00002	0.72	0.02	−0.35	0.04	−0.06	0.04
Interaction *P*-value		0.17		0.09		0.32		0.0001	
European Americans	2366	0.64	1.E-05	0.65	8.E-05	−0.54	0.001	−0.1	0.0005
African Americans	999	0.64	3.E-04	0.91	6.E-01	−0.37	0.08	−0.03	0.36
Interaction *P*-value		0.8		0.11		0.55		0.16	
Age quartile 1 (<34)	972	0.79	0.1	0.92	0.7	−0.22	0.4	−0.02	0.6
Age quartile 2 (35–39)	809	0.53	6.E-06	0.91	0.6	−0.73	0.005	−0.06	0.2
Age quartile 3 (40–44)	899	0.57	0.0005	0.72	0.06	−0.43	0.06	−0.1	0.04
Age quartile 4 (≥45)	685	0.71	0.03	0.51	0.0001	−0.52	0.14	−0.16	0.004
Interaction *P*-value		0.001		0.004		0.7		0.004	
No comorbid substance dependence	1564	0.68	0.03	0.59	0.09	−0.1	0.6	−0.02	0.25
Alcohol dependence	916	0.63	0.0001	0.62	0.0005	−1	0.0006	−0.2	0.0004
Alcohol and cocaine dependence	885	0.67	0.0020	0.9	0.38	−0.6	0.01	−0.04	0.4
Interaction *P*-value		0.41		0.3		0.8		3.71E-08	

Interaction *P*-values correspond to the test that the single nucleotide polymorphism (SNP)*strata analysis in the full sample is statistically significant, i.e. the association between SNP and phenotype differs across strata (d.f. = number of strata minus 1). FTCD: Fagerström Test for Cigarette Dependence; CPD: cigarettes per day; OR: odds ratio.

these phenotypes, we looked at the association with rs1451240 in multiple stratifications of the data: gender, ethnicity, age and comorbid substance dependence (Table 3).

A substantial decrease in power is noted both with the conversion of nicotine dependence to a dichotomous CPD phenotype and with the conversion of nicotine dependence to an ordinal FTCD variable. This loss of power is consistent across strata: within nearly every stratum of gender, ethnicity, age and comorbid diagnosis, the strongest association with rs1451240 is seen in the nicotine dependence phenotype. Indeed, tests of proportional odds

for both CPD and FTCD scores indicate that there is a threshold effect ($P < 0.0001$ in both cases). Further, varying the definitions of cases and controls does not seem to impact the results (supporting information, Tables S1 and S2).

There does not appear to be an effect of gender or comorbid substance dependence on the strength of the association. However, the strength of this difference varies across ethnicity and age. Specifically, despite the fact that the FTCD-based definition of nicotine dependence appears to have an equivalent relationship to rs1451240 across ethnicities, the relationship between this SNP and CPD is much weaker in African Americans ($\beta = -0.03$, $P = 0.35$) than in European Americans ($\beta = -0.11$, $P = 0.0005$). This supports theories that nicotine dependence in African Americans is not captured fully by CPD, probably related to observations that nicotine-dependent African Americans smoke fewer CPD than do European Americans [20]. The equivalence of these ORs across ethnicities is striking, especially given that the allele frequencies differ widely in the two groups. For example, the 'A' allele of rs1451240 has frequencies of approximately 25% in European American controls and 70% in African American controls. The phenomenon of similar ORs across ethnicities despite different allele frequencies is considered further evidence of a true biological association [30].

To clarify the relationship between CPD and FTCD-based nicotine dependence, we evaluated the correlation between total FTCD score and CPD. In a population sample of subjects from Missouri who had smoked at least 100 cigarettes, the correlation between FTCD and CPD was 0.81 in European Americans ($n = 11\,312$) and 0.71 in African Americans ($n = 3031$). The FTCD items related to early-morning smoking had the lowest tetrachoric correlations with CPD, both in European Americans (0.35) and African Americans (0.32): item 3: 'which cigarette would you hate most to give up (first morning cigarette)?' and item 5: 'do you smoke more frequently during the first hours after waking than during the rest of the day?'. In a population sample of Missouri smokers, the correlation between CPD and early-morning smoking was 0.69 in EA ($n = 11\,286$) and 0.58 in AA ($n = 3028$). FTCD has been described previously as a two-dimensional phenotype characterized by (i) CPD and (ii) early-morning smoking [31]. The association between early-morning smoking and rs1451240 is given in the supporting information (Table S2). These results suggest that studies using CPD as a phenotype may be missing this important component of nicotine dependence. Furthermore, this discrepancy appears to be of particular relevance in populations of African descent.

Our analysis of the association between nicotine dependence and the SNP rs1451240 was combined into a meta-analysis with an independent study of nicotine dependence [12]. The previously published study had some subjects who overlapped with the current study. Using the published OR from this study, and eliminating the overlapping subjects from our current study, we computed a meta-analysis *P*-value of 6.7×10^{-16} ($n = 4200$ subjects); further evidence that this association is, indeed, real.

DISCUSSION

We compared two GWAS of smoking behavior to evaluate the importance of phenotype in GWAS. We found SNPs on chromosome 8 in the region of *CHRNB3* that reached genome-wide significance in their association with nicotine dependence, but did not reach genome-wide significance in the GWAS using CPD as a dependent variable. Interestingly, the association was stronger in our combined sample of 4200 ($P = 6.7 \times 10^{-16}$) than in the meta-analysis of CPD with a combined sample of more than 75 000 subjects ($P = 1.3 \times 10^{-8}$) [15]. We attribute this discrepancy to the use of an FTCD-based definition of nicotine dependence rather than CPD.

It is important to note that although the correlation between FTCD and CPD is relatively high, the slight change of phenotype from FTCD-based nicotine dependence to CPD changes the results of the study. This has implications in other fields of medicine, implying that a small change in phenotype may expose previously undiscovered variants, and these variants may have specific roles in distinguishing differences between the two phenotypes. Rather than focusing only on increasing the sample size via meta-analyses, this study shows that samples with precise phenotypes may find previously undiscovered variants by conducting association studies using secondary phenotypes.

We specifically examined the relationship between FTCD and CPD to clarify the discrepancy between our FTCD-based results and CPD-based results. Specifically, FTCD includes measures for early-morning smoking that are not well captured by CPD. The difference between these phenotypes may be explained by the contrast in African Americans: although the odds ratios for the SNP using the FTCD-based definition of nicotine dependence are identical in European Americans and African Americans, the effect size for the regression onto CPD is subjectively smaller in African Americans than in European Americans (although not statistically significantly smaller). The inconsistent measurement of CPD compared to FTCD has been described previously in the literature [20].

It is interesting to note that the FTCD phenotype is strongest as a dichotomous variable, and the highly significant test for proportional odds indicates that the

relationship between nicotine dependence and this region is a threshold phenomenon. This suggests that the relationship between CHRNB3 and smoking behavior may be more related to specific component of nicotine dependence rather than smoking quantity.

A second characteristic of our data set that differs from previously published studies is the enrichment of our sample for substance dependence. Although we did not see a statistical interaction between comorbid diagnosis and the genetic association, our sample was ascertained primarily for substance dependence (alcohol and cocaine). Of interest, the relationship between this region and alcohol dependence has been noted in the literature [21,32,33]. This highlights the complex relationship between comorbid substance use disorders and genetic susceptibilities. Further, although this analysis shows a GWAS-significant association with FTCD-based nicotine dependence that was also seen in a large meta-analysis using CPD as the primary phenotype, it would be interesting to examine the association with this variant in other data sets that have measured FTCD.

Our study highlights a variant associated with nicotine dependence that is more strongly associated with an FTCD-based definition of nicotine dependence than the more common phenotype of CPD. This serves as a striking example of how small changes in phenotype can expose new genetic variants associated with disease.

Declarations of interest

Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative (GEI) (U01 HG004422). SAGE is one of the genome-wide association studies 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 the Collaborative Study on the Genetics of Alcoholism (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). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease' (HHSN268200782096C). The Collaborative Genetic Study of Nicotine Dependence (COGEND) project is a collaborative research group and part of the NIDA Genetics Consortium. Subject collection was supported by NIH grant P01 CA089392 (L.J. Bierut) from the National Cancer Institute. Phenotypic and genotypic data are stored in the NIDA Center for Genetic Studies (NCGS) at http://zork.wustl.edu/ under NIDA Contract HHSN271200477451C (J. Tischfield and J. Rice). Lead investigators directing data collection are L.J. Bierut, N. Breslau, D. Hatsukami, and E.O. Johnson. The authors thank Heidi Kromrei and Tracey Richmond for their assistance in data collection. Additional funding was provided by NIH grants UL1RR024992, K02DA021237, UL1RR024992, KL2RR024994 and T32MH014677. John P. Rice, Scott Saccone, Alison Goate and Laura J. Bierut are listed as inventors on Issued U.S. Patent 8,080,371, 'Markers for Addiction' covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Laura J. Bierut acted as a consultant for Pfizer in 2008.

Acknowledgements

The authors thank the following people from Washington University who contributed without compensation: Sherri Fisher for her assistance in data collection and editing of the manuscript, and Hilary Davidson for her assistance with the manuscript.

References

1. Fiore M. C., Jaen C. R., Baker T. B., Bailey W. C., Benowitz N. L., Curry S. J. *et al*. Treating tobacco use and dependence: 2008 update. Rockville, MD: US Department of Health and Human Services; 2008.
2. Heatherton T. F., Kozlowski L. T., Frecker R. C., Fagerstrom K. O. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* 1991; **86**: 1119–27.
3. Bierut L. J., Dinwiddie S. H., Begleiter H., Crowe R. R., Hesselbrock V., Nurnberger J. I. Jr *et al*. Familial transmission of substance dependence: alcohol, marijuana, cocaine, and habitual smoking: a report from the Collaborative Study on the Genetics of Alcoholism. *Arch Gen Psychiatry* 1998; **55**: 982–8.
4. Li M. D. The genetics of nicotine dependence. *Curr Psychiatry Rep* 2006; **8**: 158–64.
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. Amos C. I., Wu X., Broderick P., Gorlov I. P., Gu J., Eisen T. *et al*. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008; **40**: 616–22.
7. Berrettini W., Yuan X., Tozzi F., Song K., Francks C., Chilcoat H. *et al*. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry* 2008; **13**: 368–73.
8. Stevens V. L., Bierut L. J., Talbot J. T., Wang J. C., Sun J., Hinrichs A. L. *et al*. Nicotinic receptor gene variants

influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 3517–25.
9. Thorgeirsson T. E., Geller F., Sulem P., Rafnar T., Wiste A., Magnusson K. P. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008; **452**: 638–42.
10. Weiss R. B., Baker T. B., Cannon D. S., von Niederhausern A., Dunn D. M., Matsunami N. *et al.* A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet* 2008; **4**: e1000125.
11. Saccone N. L., Culverhouse R. C., Schwantes-An T. H., Cannon D. S., Chen X., Cichon S. *et al.* Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet* 2010; **6**: e1001053.
12. 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. [Research Support, NIH, Extramural].
13. Heath A. C., Madden P. A., Bucholz K. K., Dinwiddie S. H., Slutske W. S., Bierut L. J. *et al.* Genetic differences in alcohol sensitivity and the inheritance of alcoholism risk. *Psychol Med* 1999; **29**: 1069–81.
14. Liu J. Z., Tozzi F., Waterworth D. M., Pillai S. G., Muglia P., Middleton L. *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010; **42**: 436–40.
15. 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.
16. 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.
17. Greenbaum L., Kanyas K., Karni O., Merbl Y., Olender T., Horowitz A. *et al.* Why do young women smoke? I. Direct and interactive effects of environment, psychological characteristics and nicotinic cholinergic receptor genes. *Mol Psychiatry* 2006; **11**: 312–22, 223. [Research Support, non-US Government].
18. Hoft N. R., Corley R. P., McQueen M. B., Schlaepfer I. R., Huizinga D., Ehringer M. A. Genetic association of the CHRNA6 and CHRNB3 genes with tobacco dependence in a nationally representative sample. *Neuropsychopharmacology* 2009; **34**: 698–706. [Research Support, NIH, Extramural].
19. 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.
20. Johnson E. O., Morgan-Lopez A. A., Breslau N., Hatsukami D. K., Bierut L. J. Test of measurement invariance of the FTND across demographic groups: assessment, effect size, and prediction of cessation. *Drug Alcohol Depend* 2008; **93**: 260–70.
21. Foroud T., Edenberg H. J., Goate A., Rice J., Flury L., Koller D. L. *et al.* Alcoholism susceptibility loci: confirmation studies in a replicate sample and further mapping. *Alcohol Clin Exp Res* 2000; **24**: 933–45.
22. Laurie C. C., Doheny K. F., Mirel D. B., Pugh E. W., Bierut L. J., Bhangale T. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* 2010; **34**: 591–602.
23. Price A. L., Patterson N. J., Plenge R. M., Weinblatt M. E., Shadick N. A., Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–9.
24. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.
25. Sanna S., Jackson A. U., Nagaraja R., Willer C. J., Chen W. M., Bonnycastle L. L. *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* 2008; **40**: 198–203.
26. Willer C. J., Sanna S., Jackson A. U., Scuteri A., Bonnycastle L. L., Clarke R. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**: 161–9.
27. Saccone N. L., Wang J. C., Breslau N., Johnson E. O., Hatsukami D., Saccone S. F. *et al.* The CHRNA5–CHRNA3–CHRNB4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res* 2009; **69**: 6848–56.
28. Ge D., Zhang K., Need A. C., Martin O., Fellay J., Urban T. J. *et al.* WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* 2008; **18**: 640–3.
29. Spitz M. R., Amos C. I., Dong Q., Lin J., Wu X. The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *J Natl Cancer Inst* 2008; **100**: 1552–6.
30. Ioannidis J. P., Ntzani E. E., Trikalinos T. A. 'Racial' differences in genetic effects for complex diseases. *Nat Genet* 2004; **36**: 1312–8.
31. Richardson C. G., Ratner P. A. A confirmatory factor analysis of the Fagerstrom Test for Nicotine Dependence. *Addict Behav* 2005; **30**: 697–709.
32. Landgren S., Engel J. A., Andersson M. E., Gonzalez-Quintela A., Campos J., Nilsson S. *et al.* Association of nAChR gene haplotypes with heavy alcohol use and body mass. *Brain Res* 2009; **1305**: S72–9. [Research Support, non-US Government].
33. Hoft N. R., Corley R. P., McQueen M. B., Huizinga D., Menard S., Ehringer M. A. SNPs in CHRNA6 and CHRNB3 are associated with alcohol consumption in a nationally representative sample. *Genes Brain Behav* 2009; **8**: 631–7. [Research Support, NIH, Extramural].

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Q–Q plot of *P*-values from multivariate logistic regression models testing for association with nicotine dependence.

Figure S2 Linkage disequilibrium structure and *P*-values in the European American subset of the sample in the CHRNB3 region of chromosome 8, the region of the genome most strongly associated with nicotine

dependence in our study. Figure created by WGAViewer (http://people.genome.duke.edu/~dg48/WGAViewer/) [1].

Figure S3 Linkage disequilibrium structure and *P*-values in the African American subset of the sample in the CHRNB3 region of chromosome 8, the region of the genome most strongly associated with nicotine dependence in our study. Figure created by WGAViewer (http://people.genome.duke.edu/~dg48/WGAViewer/) [1].

Table S1 Association between rs1451240 and smoking phenotypes.

Table S2 Association between rs1451240 and smoking phenotypes.

Table S3 Characteristic of sample, by contributing study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

