

Nicotinic Receptor Gene Variants Influence Susceptibility to Heavy Smoking

Victoria L. Stevens,¹ Laura J. Bierut,² Jeffrey T. Talbot,¹ Jen C. Wang,² Juzhong Sun,¹ Anthony L. Hinrichs,² Michael J. Thun,¹ Alison Goate,² and Eugenia E. Calle¹

¹Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, Georgia and ²Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri

Abstract

Heavy smoking is a strong predictor of nicotine dependence, which is a major impediment to smoking cessation. Although both heavy smoking and nicotine dependence are highly heritable, previous attempts to identify genes influencing these phenotypes have been largely unsuccessful until very recently. We studied 1,452 heavy smokers (defined as smoking at least 30 cigarettes per day for at least 5 years) and 1,395 light smokers (defined as smoking <5 cigarettes per day for at least 1 year) to investigate the association of common variants in nicotinic receptor subunit genes with smoking behavior. Compared with the most common allele, two separate groups of single nucleotide polymorphisms (SNP) in the *CHRNA5-CHRNA3-CHRNA4* gene cluster

were associated with heavy smoking with a very high statistical significance. One group of eight SNPs, which included a nonsynonymous SNP in the *CHRNA5* gene, was in strong linkage disequilibrium and associated with increased risk of heavy smoking. A second group of SNPs not strongly correlated with the first was associated with decreased risk of heavy smoking. Analyses that combined both groups of SNPs found associations with heavy smoking that varied by >2-fold. Our findings identify two loci in the *CHRNA5-CHRNA3-CHRNA4* gene cluster that predict smoking behavior and provide strong evidence for the involvement of the $\alpha 5$ nicotinic receptor in heavy smoking. (Cancer Epidemiol Biomarkers Prev 2008;17(12):3517–25)

Introduction

Smoking is the leading cause of preventable death, resulting in an estimated 438,000 premature deaths annually in the United States (1). Approximately 40% of these deaths are due to cancer, with lung cancer being the major contributor (1). Lung cancer risk increases with higher numbers of cigarettes smoked per day (2). Heavy smoking is also a strong predictor of nicotine dependence (3), which hampers the efforts by many of the 45.1 million smokers in the United States to quit smoking (4). Increased understanding of the genetic mechanisms underlying heavy smoking and nicotine dependence may aid in the development of drug therapies to help smokers quit.

Twin studies have indicated that genes influence individual susceptibility to heavy smoking and other phenotypes related to nicotine dependence (5), although the specific genes have been difficult to identify. However, a recent genome-wide association study (GWAS) involving ~7,500 people from two European populations found a common haplotype in the *CHRNA5-CHRNA3-CHRNA4* gene cluster to be associated with smoking quantity (6). Single nucleotide polymorphisms (SNP) in these nicotinic receptor subunit genes were also associated with nicotine dependence as defined by the

Fagerström Test for Nicotine Dependence (FTND) scores in another GWAS (7) and parallel candidate gene study (8) and with early age of tobacco initiation (9). Recently, three additional GWASs (10–12) reported that SNPs in the *CHRNA5-CHRNA3-CHRNA4* gene cluster are associated with increased risk of lung cancer. One of these studies (10) also found these SNPs to be associated with smoking quantity and nicotine dependence and therefore suggested that the link with lung cancer is primarily mediated through the smoking-related phenotypes. However, the other two studies (11, 12) found little or no evidence for the involvement of smoking behavior in the association and concluded that the genetic variation affected lung cancer directly. Thus, although there is considerable evidence for the importance of common variants in the *CHRNA5-CHRNA3-CHRNA4* gene cluster, it is not clear which phenotypes are directly associated with these SNPs.

As ligand-gated ion channels consisting of five subunits, nicotinic receptors are the primary targets for nicotine that initiate the brain responses to smoking. It is biologically plausible, therefore, that these genes may influence smoking intensity and nicotine dependence. However, nicotine-mediated activation of these receptors *in vitro* has been shown to stimulate cellular proliferation and to inhibit apoptosis of bronchial epithelial cells (13), raising the possibility that nicotine exposure could influence lung cancer directly rather than through the exposure to carcinogenic compounds in cigarette smoke. Thus, biologically plausible mechanisms could explain a direct influence of variation in nicotinic receptors on smoking behaviors and a direct or indirect influence on lung cancer risk.

Received 6/26/08; revised 9/11/08; accepted 9/25/08.

Grant support: The Collaborative Genetic Study of Nicotine Dependence funded by National Cancer Institute grant P01 CA089392 and the Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine.

Requests for reprints: Victoria L. Stevens, Department of Epidemiology and Surveillance Research, American Cancer Society, 250 Williams Street Northwest, Atlanta, GA 30303-1002. Phone: 404-329-5197; Fax: 404-327-6450. E-mail: Victoria.Stevens@cancer.org

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-08-0585

In this study, we investigated the association of SNPs in the *CHRNA5-CHRNA3-CHRNA4* gene cluster with heavy smoking using participants in the American Cancer Society Cancer Prevention Study II (CPS-II) Nutrition Cohort (14). Smoking intensity and duration information collected at three time points over 15 years was used to identify 1,452 heavy smokers and 1,395 light smokers for this study. The association of the *CHRNA5-CHRNA3-CHRNA4* gene cluster with heavy smoking was explored by genotyping 17 SNPs spanning over 200 kb of chromosome 15 that include these genes. In addition, the association of common SNPs in other nicotinic receptor genes, specifically in the *CHRNA3-CHRNA6* gene region, and the *CHRNA4* and *CHRNA2* genes, was investigated to determine the importance of these genes in heavy smoking.

Materials and Methods

Study Population. The smokers used in this study were participants in the American Cancer Society CPS-II Cohort, a prospective study of cancer mortality that began in 1982, and the CPS-II Nutrition Cohort, a prospective study of cancer incidence formed in 1992 using a subset of CPS-II participants. Approximately 1.2 million Americans, who were recruited from all 50 states and had a median age of 57 in 1982, were enrolled in CPS-II. All participants completed a self-administered baseline questionnaire at enrollment, which included questions on lifetime history of smoking behavior, demographic, medical, lifestyle, environmental, and dietary characteristics. In 1992, ~184,000 CPS-II participants were enrolled in the CPS-II Nutrition Cohort and completed a mailed questionnaire that included questions on smoking behavior, demographics, diet, and other lifestyle factors. The recruitment and characteristics of this cohort have been described elsewhere (14). Follow-up questionnaires were sent to all living Nutrition Cohort members in 1997 and every 2 y afterwards to update exposure information and to ascertain newly diagnosed cases of cancer. Incident cases reported via questionnaire response were verified through medical records, linkage with state cancer registries, or death certificates (14).

Between June 1998 and June 2001, blood samples were collected from a subset of CPS-II Nutrition Cohort participants (21,965 women and 17,411 men); fractionated into serum, plasma, buffy coat, and RBCs; and stored in liquid nitrogen vapor phase at -130°C until needed for analysis. Subsequently, CPS-II Nutrition Cohort participants who did not provide a blood sample were invited to donate a buccal cell sample, which was self-collected using the "swish and spit" method as described by Feigelson et al. (15). The sample was mailed to a central biorepository where the cells were collected by centrifugation and placed into long-term storage in liquid nitrogen vapor phase. This effort resulted in the collection of buccal cell samples from 38,180 women and 31,824 men in 2001 and 2002. Thus, we had DNA from buffy coat or buccal cell samples from 109,380 individuals.

Smoking Phenotypes. Information on smoking behavior from questionnaires administered in 1982, 1992, and 1997 was used to identify light and heavy smokers

among participants with a DNA sample. To be considered for inclusion in this study, participants had to have reported smoking more than 100 cigarettes in their lifetime.

Individuals who reported smoking at least 30 cigarettes per day for at least 5 y were defined as heavy smokers. Heavy smoking was a fairly common phenotype among both men and women. We randomly selected 750 heavy smoking men and 750 heavy smoking women for this study (total $N = 1,500$). Twelve percent of these heavy smokers were still actively smoking in 1997.

Light smokers were defined as individuals who reported having smoked for at least 1 y during their lifetime and, in 1982 and 1992, reported always having smoked fewer than 5 cigarettes per day and, in 1997, fewer than 10 cigarettes per day (the option of fewer than 5 cigarettes per day was not provided on this questionnaire). Light smoking was a rare phenotype, especially among men. Among participants with DNA, 461 men and 1,482 women met the criteria for light smoking. To have a total sample of light smokers the same size as the heavy smoking sample, we included all light smoking men and a random sample of 1,039 light smoking women for a total of 1,500 light smokers. Less than 2% of these light smokers were still actively smoking in 1997.

SNP Selection and Genotyping. SNPs chosen for genotyping included those in the neuronal nicotinic acetylcholine receptor (AChR) genes *CHRNA3*, *CHRNA4*, *CHRNA5*, *CHRNA6*, *CHRNA2*, *CHRNA3*, and *CHRNA4* with $P < 0.05$ in a recent study of nicotine dependence (7, 8). Additional haplotype-tagging SNPs were chosen using the method of Gabriel et al. (16) and HapMap data available for Caucasians (Centre d'Etude du Polymorphisme Humain) in July 2006 (data release 21/phase II). SNPs that failed assay design were replaced by SNPs in high linkage disequilibrium (LD) with the problematic SNPs. These selection criteria resulted in a total of 61 SNPs being chosen for genotyping. An additional 200 SNPs were selected and genotyped across the genome

Table 1. Frequencies of selected characteristics of the light and heavy smokers included in this study

Variable	Light smoker ($n = 1,395$)	Heavy smoker ($n = 1,452$)
Gender		
Male	31.2	50.1
Female	68.8	49.9
Age in 97 (mean \pm SD)	67.1 ± 6.0	67.3 ± 5.7
Smoking status in 97		
Current	1.5	11.9
Former	98.5	88.1
Education		
\leq High school graduate	19.4	25.8
Some college	28.5	35.5
College graduate	51.3	38.2
Missing	0.7	0.5
Alcohol use in 97		
Noncurrent	31.4	31.4
<1 drink/week	15.3	11.2
1-6 drinks/week	36.2	25.3
1 drink/day	9.8	17.1
2+ drinks/day	2.2	12.2
Missing	5.2	2.8

NOTE: Values are expressed as percentages unless otherwise noted.

Table 2. Characteristics, minor allele frequencies, and associations with nicotine dependence for the *CHRNA5-CHRNA3-CHRNA4* SNPs genotyped in this study

No.	SNP	Position	Gene	SNP region	Minor allele	Minor allele frequency		<i>P</i> *	Excluded LCa <i>P</i> [†]	OR [‡] (95% CI)
						Cases, <i>n</i> (%)	Controls, <i>n</i> (%)			
1	rs17483686	76520445			T	1,088 (37.6)	899 (32.0)	5.32E-06	1.57E-05	1.30 (1.16-1.46)
2	rs8034191	76593078			C	1,095 (37.8)	884 (31.9)	5.44E-07	1.40E-06	1.34 (1.19-1.50)
3	rs905739	76632165			C	562 (19.5)	700 (25.3)	6.98E-07	3.71E-07	0.72 (0.64-0.82)
4	rs2036527	76638670			T	1,092 (37.8)	874 (31.5)	1.14E-07	4.40E-07	1.36 (1.21-1.53)
5	rs684513	76645455	<i>CHRNA5</i>	Intron 1	G	518 (18.0)	666 (24.1)	1.13E-07	4.70E-08	0.70 (0.61-0.80)
6	rs588765	76652480	<i>CHRNA5</i>	Intron 1	T	1,221 (42.2)	1,165 (42.0)	0.85	0.89	0.99 (0.89-1.10)
7	rs17486278	76654537	<i>CHRNA5</i>	Intron 1	C	1,097 (38.0)	872 (31.6)	7.00E-08	2.13E-07	1.37 (1.22-1.54)
8	rs569207	76660174	<i>CHRNA5</i>	Intron 1	A	563 (19.5)	721 (26.0)	1.40E-08	7.00E-09	0.69 (0.61-0.79)
9	rs637137	76661031	<i>CHRNA5</i>	Intron 2	A	564 (19.5)	725 (26.1)	1.30E-08	6.00E-09	0.69 (0.61-0.79)
10	rs11633585	76661547	<i>CHRNA5</i>	Intron 2	C	148 (5.1)	121 (4.3)	0.28	0.24	1.15 (0.89-1.47)
11	rs16969968	76669980	<i>CHRNA5</i>	Nonsynonymous, N398D	A	1,093 (37.8)	875 (31.4)	6.30E-08	1.36E-07	1.37 (1.22-1.54)
12	rs514743	76671282	<i>CHRNA5</i>	Intron 5	T	1,063 (36.8)	1,029 (37.2)	0.53	0.74	0.96 (0.86-1.08)
13	rs578776	76675455	<i>CHRNA3</i>	3' UTR	T	707 (24.4)	839 (30.3)	1.37E-06	6.15E-07	0.75 (0.66-0.84)
14	rs1051730	76681394	<i>CHRNA3</i>	Synonymous	T	1,073 (37.8)	860 (31.4)	9.30E-08	2.51E-07	1.37 (1.22-1.53)
15	rs3743078	76681814	<i>CHRNA3</i>	Intron 4	G	571 (19.8)	738 (26.6)	5.00E-09	2.00E-09	0.69 (0.60-0.78)
16	rs11637630	76686774	<i>CHRNA3</i>	Intron 4	G	559 (19.4)	724 (26.2)	5.00E-09	2.00E-09	0.68 (0.60-0.78)
17	rs1878399	76699058	<i>CHRNA3</i>	Intron 1	G	1,217 (42.3)	1,166 (42.0)	0.81	0.94	0.99 (0.89-1.10)
18	rs1948	76704454	<i>CHRNA4</i>	3' UTR	T	965 (33.3)	962 (34.7)	0.12	0.20	0.91 (0.82-1.02)
19	rs12914008	76710560	<i>CHRNA4</i>	Nonsynonymous, I91T	A	130 (4.5)	99 (3.6)	0.082	0.076	1.27 (0.97-1.66)
20	rs17487223	76711042	<i>CHRNA4</i>	Intron 2	T	1,151 (39.8)	936 (33.9)	8.07E-07	1.68E-06	1.33 (1.18-1.48)
21	rs950776	76713073	<i>CHRNA4</i>	Intron 2	C	986 (34.1)	959 (34.6)	0.37	0.59	0.95 (0.85-1.06)
22	rs12440014	76713781	<i>CHRNA4</i>	Intron 2	G	586 (20.3)	732 (26.4)	3.45E-07	1.71E-07	0.72 (0.64-0.82)
23	rs11636605	76715933	<i>CHRNA4</i>	Intron 1	A	492 (17.0)	619 (22.4)	1.18E-06	5.02E-07	0.72 (0.63-0.82)
24	rs1316971	76717565	<i>CHRNA4</i>	Intron 1	A	490 (17.0)	611 (22.1)	3.18E-06	8.87E-07	0.73 (0.64-0.83)
25	rs3813567	76721606			C	505 (17.5)	632 (22.8)	1.25E-06	6.89E-07	0.72 (0.63-0.82)
26	rs11633223	76722531			C	1,113 (38.6)	1,064 (38.6)	0.81	0.89	0.99 (0.88-1.10)
27	rs3971872	76729090			T	246 (8.5)	204 (7.4)	0.18	0.185	1.15 (0.94-1.40)
28	rs1996371	76743861			G	1,251 (43.3)	1,075 (38.9)	3.33E-04	3.55E-04	1.22 (1.09-1.36)
29	rs16970006	76757314			C	203 (7.0)	222 (8.0)	0.12	0.13	0.85 (0.69-1.04)

**P* for trend, adjusted for gender and age (per year).

[†]*P* for trend after exclusion of individuals with lung cancer (10 light smokers, 75 heavy smokers), adjusted for gender and age (per year).

[‡]Per allele OR, adjusted for gender and age (per year).

and used for the assessment of population stratification in these samples (17).

SNPs were genotyped using the Illumina GoldenGate custom genotyping method by the Microarray Core of the Genome Sequencing Center at Washington University (Director, Dr. Seth Crosby). For the GoldenGate data, clustering was done using the default algorithm in BeadStudio (Illumina) and manually reviewed by two technicians. Using the filter function in BeadStudio, we removed SNPs with a call frequency of <95% ($n = 3$) and SNPs for which Hardy-Weinberg equilibrium was rejected ($P < 0.05$; $n = 2$). Thus, genotyping results for 56 nicotinic receptor SNPs and 200 population SNPs were used in this study. No genotyping results were obtained for 102 samples, leaving 1,479 heavy smokers and 1,419 light smokers for analyses.

Statistical Analyses. The presence of population stratification among our study subjects was evaluated using a STRUCTURE analysis (18) of 200 SNPs genotyped in our subjects and available for four ethnicities in the HapMap data. Subject ethnicity was plotted in relationship to Central European, Yoruban in Nigeria, and Japanese from Tokyo and Han Chinese in Beijing populations. Individuals found to have <85% European ancestry were excluded from the analyses. This resulted in the exclusion of 51 individuals (27 heavy smokers

and 24 light smokers). Thus, 1,452 heavy smokers and 1,395 light smokers were included in the statistical analyses.

The association with heavy smoking was evaluated assuming an additive genetic model using a per allele odds ratio (OR), which was determined using unconditional logistic regression in which a continuous variable for the number of minor alleles (0, 1, or 2) was entered into the regression model. The associated *P* value is derived from the Wald χ^2 with 1 degree of freedom. LD measurements (D' and r^2) were made using Haploview v4.0.

Gene-gene interactions were evaluated using the likelihood ratio test. All models were adjusted for age (in single year categories) and gender (male or female). Reported *P* values were two sided and were corrected for multiple comparisons using the Bonferroni method where indicated.

Results

The characteristics of the light and heavy smokers included in this study are shown in Table 1. Although there are roughly equal numbers of men and women among the heavy smokers, only 31% of the light smokers are men. These gender differences in smoking prevalence

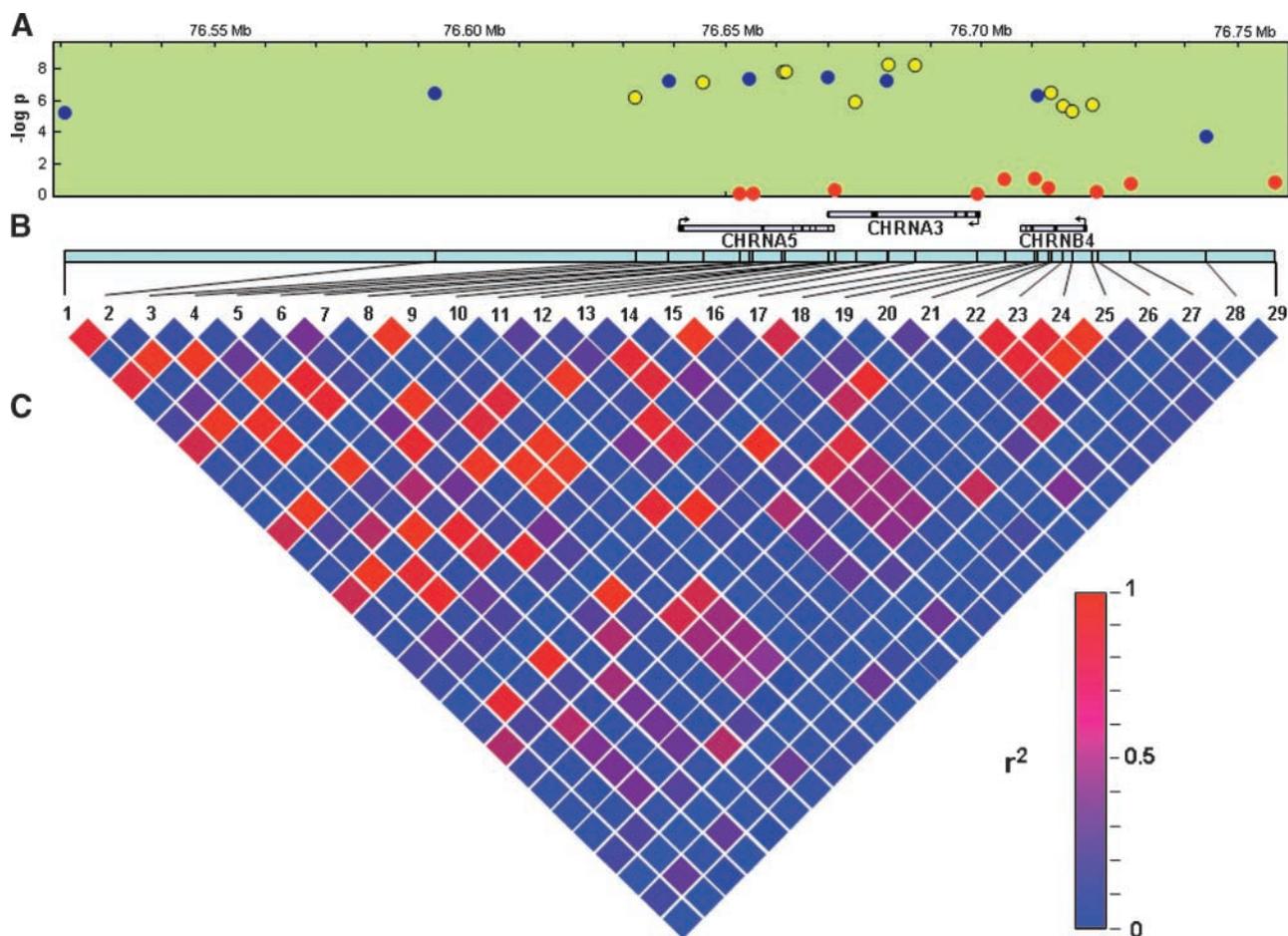


Figure 1. Association of SNPs in the *CHRNA5-CHRNA3-CHRN4* gene cluster with heavy smoking. **A.** The $-\log$ of the P value for the association of each of the SNPs with nicotine dependence (Y axis) as well as their relative positions on chromosome 15 (X axis). *Blue circles*, SNPs associated with increased risk; *yellow circles*, SNPs associated with decreased risk; *red circles*, SNPs not associated with heavy smoking. **B.** The positions of the *CHRNA5*, *CHRNA3*, and *CHRN4* genes are shown in the appropriate place relative to the chromosomal location shown on the X axis on the top of **A**. *Darkened squares*, exons; *light purple*, introns. **C.** LD matrix for the 29 SNPs genotyped in this study. The color of each block indicates the amount of LD using the scale of colors for r^2 from 0 to 1 in the lower right hand corner. The number for each SNP corresponds to that listed in the first column of Table 1.

are consistent with those documented for Americans in the birth cohorts of the study population and are likely due to both cultural and biological factors (19). The heavy smokers were slightly less educated and likely to consume more alcohol than the light smokers. The percentages that were current smokers were low in both groups (1.5% for light smokers and 11.9% for heavy smokers), likely reflecting their advanced age (smoking prevalence of U.S. adults ages ≥ 65 years is $< 9\%$; ref. 20). The characteristics of the individuals excluded from the analysis because of ethnic background or genotyping problems were similar to those included in the study (data not shown).

Nineteen SNPs on chromosome 15 in the region that encompasses the *CHRNA5-CHRNA3-CHRN4* gene cluster were strongly associated with heavy smoking (Table 2; Fig. 1). All of these associations were highly statistically significant even after conservative Bonferroni correction for multiple comparisons (P values ranged 0.019– $2.8E-07$ after correction). In addition, all the asso-

ciations remained significant for both men and women when gender-specific analyses were done (data not shown). These findings replicated the association for nine of the SNPs that had been identified by a GWAS (7) and parallel candidate gene study (8) of nicotine dependence and identified 10 new associated SNPs in the same region.

The minor alleles of eight of the statistically significant SNPs were associated with an increased risk of heavy smoking (Table 2, highlighted with a black background). Seven of these SNPs (all but rs1996371) share a high degree of LD (D' between 0.81 and 1.0, r^2 between 0.61 and 0.99). The rs1996371 SNP is also in LD with the other seven SNPs, although less so (D' between 0.55 and 0.72, r^2 between 0.22 and 0.39). The association between rs1996371 and heavy smoking was completely attenuated [per allele OR, 1.02; 95% confidence interval (95% CI), 0.89–1.18] when any other of the risk SNPs in this group were included in the logistic regression model. Thus, these eight SNPs represent a single

Table 3. The risks of heavy smoking associated with each genotype of the significant SNPs in the A5A3B4 gene cluster are shown

SNP	Position	Genotype	No. cases	No. controls	OR (95% CI)
rs17483686	76520445	AA	545	632	1.00 (reference)
		AT	714	627	1.34 (1.14-1.57)
		TT	187	131	1.65 (1.28-2.14)
					<i>P</i> trend = 5.32E-06
rs8034191	76593078	TT	539	636	1.00 (reference)
		TC	723	618	1.41 (1.20-1.66)
		CC	186	133	1.70 (1.32-2.19)
					<i>P</i> trend = 5.44E-07
rs905739	76632165	TT	938	769	1.00 (reference)
		TC	444	530	0.69 (0.59-0.81)
		CC	59	85	0.58 (0.41-0.83)
					<i>P</i> trend = 6.98E-07
rs2036527	76638670	CC	540	646	1.00 (reference)
		CT	718	612	1.43 (1.22-1.68)
		TT	187	131	1.76 (1.36-2.28)
					<i>P</i> trend = 1.14E-07
rs684513	76645455	CC	965	795	1.00 (reference)
		CG	422	512	0.69 (0.59-0.81)
		GG	48	77	0.51 (0.35-0.75)
					<i>P</i> trend = 1.57E-05
rs17486278	76654537	AA	534	638	1.00 (reference)
		AC	721	608	1.45 (1.23-1.70)
		CC	188	132	1.78 (1.37-2.29)
					<i>P</i> trend = 1.13E-07
rs569207	76660174	GG	940	754	1.00 (reference)
		GA	451	545	0.67 (0.57-0.78)
		AA	56	88	0.52 (0.37-0.75)
					<i>P</i> trend = 1.40E-08
rs637137	76661031	TT	938	755	1.00 (reference)
		TA	452	547	0.67 (0.57-0.78)
		AA	56	89	0.52 (0.37-0.74)
					<i>P</i> trend = 1.30E-08
rs16969968	76669980	GG	539	650	1.00 (reference)
		GA	717	609	1.45 (1.24-1.70)
		AA	188	133	1.77 (1.37-2.29)
					<i>P</i> trend = 6.30E-08
rs578776	76675455	CC	830	670	1.00 (reference)
		CT	525	587	0.72 (0.62-0.85)
		TT	91	126	0.59 (0.44-0.79)
					<i>P</i> trend = 1.37E-06
rs1051730	76681394	CC	536	642	1.00 (reference)
		CT	693	594	1.43 (1.22-1.68)
		TT	190	133	1.78 (1.38-2.30)
					<i>P</i> trend = 9.30E-08
rs3743078	76681814	CC	931	748	1.00 (reference)
		CG	453	544	0.68 (0.58-0.80)
		GG	59	97	0.49 (0.34-0.68)
					<i>P</i> trend = 5.00E-09
rs11637630	76686774	AA	938	749	1.00 (reference)
		AG	447	540	0.67 (0.57-0.79)
		GG	56	92	0.50 (0.35-0.70)
					<i>P</i> trend = 5.00E-09
rs17487223	76711042	CC	505	605	1.00 (reference)
		CT	729	616	1.45 (1.23-1.71)
		TT	211	160	1.64 (1.29-2.09)
					<i>P</i> trend = 8.07E-07
rs12440014	76713781	CC	922	749	1.00 (reference)
		CG	454	542	0.69 (0.59-0.81)
		GG	66	95	0.57 (0.41-0.80)
					<i>P</i> trend = 3.45E-07
rs11636605	76715933	GG	1,007	831	1.00 (reference)
		GA	384	485	0.66 (0.56-0.77)
		AA	54	67	0.67 (0.46-0.98)
					<i>P</i> trend = 1.18E-06
rs1316971	76717565	GG	1,008	837	1.00 (reference)
		GA	380	481	0.65 (0.55-0.77)
		AA	55	65	0.72 (0.49-1.05)
					<i>P</i> trend = 3.18E-06

(Continued on the following page)

Table 3. The risks of heavy smoking associated with each genotype of the significant SNPs in the A5A3B4 gene cluster are shown (Cont'd)

SNP	Position	Genotype	No. cases	No. controls	OR (95% CI)
rs3813567	76721606	TT	990	823	1.00 (reference)
		TC	403	500	0.66 (0.56-0.78)
		CC	51	66	0.66 (0.45-0.96)
					<i>P</i> trend = 1.25E-06
rs1996371	76743861	AA	452	525	1.00 (reference)
		AG	735	637	1.37 (1.15-1.62)
		GG	258	219	1.42 (1.13-1.77)
					<i>P</i> trend = 3.33E-04

NOTE: Associations are adjusted for age and gender.

correlated cluster that is very significantly associated with increased risk of heavy smoking.

The minor alleles of the other 11 SNPs with significant *P* values in the *CHRNA5-CHRNA3-CHRNA4* gene cluster were associated with decreased risk of heavy smoking (Table 2, highlighted by gray background). As with the risk SNPs, these protective SNPs share varying levels of LD (*D'* between 0.64 and 1.0, *r*² between 0.39 and 0.98) and represent a single group of correlated SNPs. Although the LD across this region seems high when measured by *D'*, the correlation between the risk and protective SNPs is low (*r*² ≤ 0.2), indicating that they represent two unrelated associations (see Fig. 1).

The genotype associations for the risk SNPs in the *CHRNA5-CHRNA3-CHRNA4* gene cluster show a dose response [for rs16969968: GA OR, 1.45 (95% CI, 1.24-1.70); AA OR, 1.77 (95% CI, 1.37-2.29); Table 3]. The same is true for the protective SNPs [for rs3743078: CG OR, 0.68 (95% CI, 0.58-0.80); GG OR, 0.49 (95% CI, 0.34-0.68); Table 3]. The highly significant *P* trends for these results support both sets of variants following an additive genetic model. All the findings in Table 3 remain statistically significant after correction for multiple comparisons except for the results for rs1996371.

Because the risk and protective SNPs represent two separate associations, the combined influence of the variants was assessed to determine how heavy smoking was affected by genotype combinations of SNPs from each group. The nonsynonymous SNP rs16969968 was used to represent the risk group and rs3743078, one of the most significantly associated SNPs, was used to represent the protective group. The combination of the

wild-type genotype of rs16969968 (GG) and homozygous variant genotype of rs3743078 (GG) is associated with the lowest risk for heavy smoking (Table 4), whereas the AA/CC genotype combination is associated with the highest risk (Table 4). The odds of heavy smoking relative to light smoking among smokers with the highest risk genotype combination were 2.43 times higher than in those with the lowest risk genotype combination.

Three of the correlated SNPs we found associated with increased risk of heavy smoking (rs803191, rs16969968 and rs1051730) have also been found to be associated with increased risk of lung cancer by three large GWASs (10-12). In our study, 10 light smokers and 75 heavy smokers had been diagnosed with incident lung cancer. To determine if any part of the association we observed between these SNPs and heavy smokers was due to a direct association of the SNPs with lung cancer, we conducted a sensitivity analysis excluding the 85 lung cancer cases. The associations of both the risk and protective SNPs with heavy smoking changed very little (≤0.01 for per-allele and ≤0.02 for genotype ORs). As reflected in the trend *P* values, which are listed in the column labeled excluded LCa *P* in Table 2, all previously significant results remained significant. The *P* values for the risk SNPs (dark rows) were slightly attenuated, whereas those for the protective SNPs (gray rows) were somewhat more significant.

We also assessed the association of the risk and protective SNPs with lung cancer directly while controlling for smoking phenotype. No association was observed for the protective SNPs and lung cancer risk [for rs3743078: CG OR, 0.89 (95% CI, 0.55-1.44); GG OR, 1.31 (95% CI, 0.51-3.36)]. However, the homozygous variant genotype of the risk SNPs was significantly associated with an increased risk of lung cancer [for rs16969968: GA OR, 0.67 (95% CI, 0.41, 1.10); AA OR, 1.80 (95% CI, 1.02, 3.20)].

The associations with heavy smoking for the SNPs in the *CHRNA3-CHRNA6* gene region or in the *CHRNA4* and *CHRNA2* genes are shown in Table 5. Two SNPs (rs7012713, in the *CHRNA3* gene, and rs7828365, located 5' to the *CHRNA6* gene) were significantly associated with heavy smoking [for rs7012713: per allele OR, 1.42 (95% CI, 1.05-1.94), *P* = 0.023; for rs7828365: per allele OR, 0.84 (95% CI, 0.72-0.99), *P* = 0.039]. However, these associations were no longer statistically significant after correction for multiple comparisons. No other statistically significant associations were found for these four genes.

Table 4. Combined influence of a SNP with a minor allele associated with increased risk (rs16969968) and a SNP with a minor allele associated with decreased risk of heavy smoking (rs3743078)

rs16969968	rs3743078		
	CC	CG	GG
GG	258/237	222/311	59/97
GA	1.00 (reference)	0.66 (0.52-0.85)	0.56 (0.39-0.82)
AA	485/376	231/233	0/0
	1.20 (0.96-1.51)	0.96 (0.74-1.24)	—
	188/133	0/0	0/0
	1.36 (1.02-1.81)	—	—

NOTE: The numbers at the top of each cell indicate the number of cases and controls with the specific genotype combination.

Table 5. Characteristics, minor allele frequencies, and associations with heavy smoking for various nicotinic receptor SNPs

Region/SNP	Gene	Position	SNP region	Minor allele	Minor allele frequency		P*	OR [†] (95% CI)
					Cases, n (%)	Controls, n (%)		
B3A6								
rs10958726		42655066		G	615 (21.3)	641 (23.0)	0.19	0.92 (0.81-1.04)
rs13277254		42669139		G	616 (21.3)	645 (23.4)	0.13	0.91 (0.80-1.03)
rs1530847		42667396		C	429 (14.8)	462 (16.7)	0.073	0.88 (0.76-1.01)
rs6474413		42670221		C	613 (21.2)	644 (23.2)	0.13	0.91 (0.80-1.03)
rs7004381		42670318		A	613 (21.2)	638 (23.0)	0.18	0.92 (0.81-1.04)
rs4952	<i>CHRNA3</i>	42706222	Synonymous	T	117 (4.1)	121 (4.3)	0.85	0.97 (0.75-1.27)
rs4953	<i>CHRNA3</i>	42706816	Synonymous	C	118 (4.1)	122 (4.4)	0.83	0.97 (0.75-1.27)
rs7012713	<i>CHRNA3</i>	42711460	Intron	T	103 (3.6)	77 (2.8)	0.023	1.42 (1.05-1.94)
rs6987323		42716389		G	122 (4.2)	122 (4.4)	0.99	1.00 (0.77-1.30)
rs7017612		42718402		C	575 (19.9)	595 (21.4)	0.29	0.93 (0.82-1.06)
rs9298628		42725148		T	566 (19.6)	591 (21.3)	0.21	0.92 (0.81-1.05)
rs2304297	<i>CHRNA6</i>	42727356	3' UTR	C	640 (22.2)	660 (23.9)	0.26	0.93 (0.82-1.05)
rs17621710	<i>CHRNA6</i>	42734717	Intron 2	T	330 (11.4)	319 (11.5)	0.70	0.97 (0.82-1.14)
rs16891604	<i>CHRNA6</i>	42737870	Intron 2	A	154 (5.3)	146 (5.3)	0.82	0.97 (0.77-1.23)
rs7828365		42748471		T	340 (11.8)	389 (14.0)	0.036	0.84 (0.72-0.99)
rs6474417		42748677		A	192 (6.8)	157 (6.0)	0.12	1.19 (0.96-1.48)
rs6474419		42754456		G	88 (3.0)	105 (3.8)	0.16	0.81 (0.60-1.09)
A4								
rs6011770	<i>CHRNA4</i>	61447875	3' UTR	T	84 (2.9)	106 (3.8)	0.068	0.76 (0.57-1.02)
rs2236196	<i>CHRNA4</i>	61448000	3' UTR	G	805 (28.1)	731 (26.5)	0.26	1.07 (0.95-1.21)
rs3787138	<i>CHRNA4</i>	61449668	Intron 5	G	385 (13.4)	326 (11.8)	0.088	1.15 (0.98-1.35)
rs6010918	<i>CHRNA4</i>	61459945	Intron 2	A	154 (5.4)	120 (4.3)	0.12	1.21 (0.95-1.56)
rs2273505	<i>CHRNA4</i>	61461322	Intron 2	T	203 (7.1)	192 (6.9)	0.71	1.04 (0.84-1.28)
B2								
rs4845652		152804829		T	260 (9.0)	276 (9.9)	0.25	0.90 (0.75-1.08)
rs2072659	<i>CHRNA2</i>	152815145	3' UTR	G	305 (10.5)	266 (9.6)	0.28	1.10 (0.92-1.31)
rs3811450		152817656		T	202 (7.0)	205 (7.4)	0.56	0.94 (0.76-1.16)
rs9427092		152820346		C	594 (20.6)	567 (20.4)	0.60	1.04 (0.91-1.18)
rs11264222		152820981		T	819 (28.6)	785 (28.5)	0.89	0.99 (0.88-1.12)

Abbreviation: UTR, untranslated region.

*P for trend, adjusted for gender and age (per year).

†Per allele OR, adjusted for gender and age.

Discussion

We have shown that two distinct variant groups in the *CHRNA5-CHRNA3-CHRNA4* gene cluster are strongly associated with heavy smoking. Of the eight SNPs associated with increased risk of heavy smoking, only rs16969968 alters the coding sequence of these genes. This SNP was associated with nicotine dependence in previous GWAS and candidate gene studies (7, 8) and is the most likely candidate for the polymorphism responsible for increased risk of heavy smoking. The 11 SNPs associated with decreased risk of heavy smoking either are intronic or are outside the translated region of the three genes. Thus, it is not clear whether one of these SNPs or a polymorphism in LD with these variants is the functional variant responsible for this association. Our findings with the risk and protective SNP genotype combinations define a gradient of genetic predisposition to smoking intensity. The odds of heavy smoking behavior were twice as high in individuals with the highest risk compared with the lowest risk genotype combination.

rs16969968 is a nonsynonymous SNP in the *CHRNA5* gene that replaces an aspartic acid with an asparagine at codon 398 of the protein. It is highly conserved across species and the presence of this variant in the $\alpha 5$ receptor subunit has been shown to reduce receptor response to nicotine in *in vitro* studies (21). This subunit is highly expressed in several areas of the brain, including the

ventral tegmental area (22) where nicotine interacts with AChRs. Mice in which $\alpha 5$ expression has been knocked out show greatly reduced sensitivity to nicotine-induced behaviors and seizures (23). However, the consequences of the $\alpha 5$ knockout are not as dramatic as those caused by the loss of other nicotinic receptor subunits (24), resulting in most investigation of AChRs focusing on other subunits. Our convincing findings implicating the $\alpha 5$ receptor subunit and the rs16969968 SNP in smoking behavior argue strongly for investigation into how the reduced activity of this receptor influences downstream consequences of nicotine exposure. Such research may lead to the development of pharmacologic agents that could eliminate the consequences of this amino acid replacement and aid in smoking cessation.

We used heavy and light smoking rather than nicotine dependence because of the questionnaire information collected from the study participants. The number of cigarettes smoked per day is highly heritable (25) and predictive of nicotine dependence (3, 6). To evaluate how our cigarettes per day phenotype compared with that defined by FTND scores, we compared these characteristics in another study in which over 16,000 smokers, ages 25 to 44 years, provided information on both variables. These smokers were participants in the Collaborative Genetic Study of Nicotine Dependence (COGEN; ref. 26). As shown in Fig. 2, over 94% of the respondents who reported currently smoking five or fewer cigarettes per day had a lifetime maximum FTND

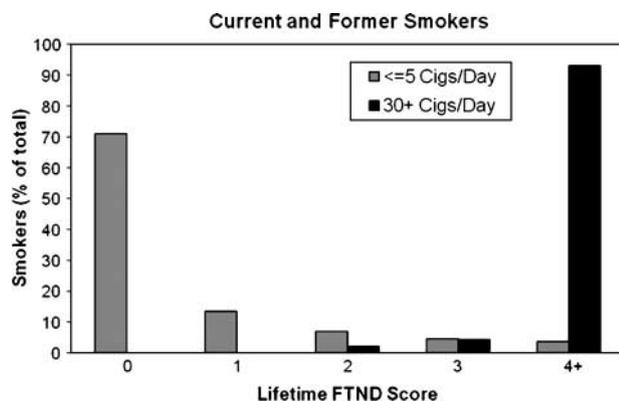


Figure 2. FTND scores for smokers who smoked ≤ 5 cigarettes per day (light gray columns) or 30+ cigarettes per day (black columns). The percentage of the total number of smokers in each group is shown. The data used to create this figure were collected as part of the COGEN study (26).

score of <4 , which is consistent with a classification of nondependent, and 84% had a FTND score of 1 or less. Over 96% of current smokers who reported smoking 30 or more cigarettes per day were nicotine dependent using the FTND criteria of a score of 4 or more. Similar findings were seen in former smokers. This comparison of number of cigarettes per day and FTND score in the COGEN population and a previous analysis of cigarettes per day with DSM-IV–defined nicotine dependence (6) indicate that the majority of our heavy and light smokers would be classified as nicotine dependent and nondependent, respectively.

Several of the SNPs we found associated with heavy smoking in the *CHRNA5-CHRNA3-CHRNA4* gene cluster have been previously associated with this phenotype (6) and with nicotine dependence defined by FTND scores (7, 8). This consistency of findings of the genetic association across the several studies using different measures for smoking behaviors and nicotine dependence shows the robustness of these findings.

In addition to previous associations with nicotine dependence (7, 8) and heavy smoking (6), three of the SNPs in the *CHRNA5-CHRNA3-CHRNA4* gene cluster (rs8034191, rs16969968 and rs1051730) have been found to be associated with lung cancer risk (11, 12). In our analysis, the association of these SNPs with heavy smoking was unchanged when individuals with lung cancer were excluded, suggesting that the significant variants we observed predict for heavy smoking, regardless of whether they predict for lung cancer. Although we did find that the homozygous variant genotype of the risk SNPs was significantly associated with increased risk of lung cancer, this finding could be due to chance; it was based on a small number of lung cancer cases, and the OR was <1.0 in the heterozygous variant.

The association with SNPs that we found to be protective for heavy smoking with lung cancer was only investigated in one of the GWASs for lung cancer. Hung et al. (11) found three of these SNPs (rs637137, rs578776, and rs3743078) to be unrelated to lung cancer risk. The association of these SNPs with heavy smoking in our study was unaffected by the removal of individuals with

lung cancer and no direct association with lung cancer risk was observed. Thus, comparison of these risk and protective SNPs in lung cancer studies may be helpful in defining whether associations with the cancer are direct or are through an association with smoking behavior.

The strengths of our study include the large population size and comprehensive SNP coverage of the genes investigated. In addition, the use of a single racial/ethnic group and population stratification analysis with 200 independent SNPs allows us to eliminate the possibility that population substructure may contribute to our findings. However, the minor allele frequency for rs16969968 is very low in non-European populations ($<5\%$), suggesting that the effect of this SNP will be far less in other populations.

In summary, we have shown that there are two groups of SNPs with low correlation to each other in the *CHRNA5-CHRNA3-CHRNA4* gene cluster that are associated with heavy smoking with a very high degree of statistical significance. Individuals who have the risk genotype without the protective genotype have over twice the odds of heavy smoking compared with individuals with the opposite genotype combination. These findings strongly argue for further study of the $\alpha 5$ AChR subunit in nicotine dependence-related phenotypes in humans to gain insight into the molecular basis of smoking behavior that confers a high risk of lung cancer and other diseases.

Disclosure of Potential Conflicts of Interest

L.J. Bierut, A. Goate, A.L. Hinrichs, and J.C. Wang: listed as inventors on a patent (U.S. 20070258898) held by Perlegen Sciences, Inc., covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction; L.J. Bierut: Pfizer, Inc. consultant.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Naomi Breslau, Dorothy Hatsukami, and Eric Johnson for providing data to examine the relationship between smoking quantity and nicotine dependence; Robertino Garcia, Maribel Martinez, Oliver Reyes, and Shantia Shears for their technical assistance in setting up DNA samples for genotyping; and Sarah Bertelsen and John Budde for data cleaning.

References

- Centers for Disease Control and Prevention. Annual smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 1997–2001. *MMWR Morb Mortal Wkly Rep* 2005;54:625–8.
- Samet JM. The epidemiology of lung cancer. *Chest* 1993;108:20–9S.
- Heatherton TF, Kozlowski LT, Frecker RC, et al. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. *Br J Addict* 1989; 84:791–800.
- Centers for Disease Control and Prevention. Tobacco use among adults—United States, 2005. *MMWR Morb Mortal Wkly Rep* 2006;55: 1145–8.
- Sullivan PF, Kendler KS. The genetic epidemiology of smoking. *Nic Tob Res* 1999;1:S51–7.
- Berrettini W, Yuan X, Tozzi F, et al. $\alpha 5/\alpha 3$ Nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry* 2008; 13:368–73.

7. Bierut LJ, Madden PA, Breslau N, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007;16:24–35.
8. Saccone SF, Hinrichs AL, Saccone NL, 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.
9. Schlaepfer IR, Hoft, Collins AC, et al. The CHRNA5/A3/B4 gene cluster variability as an imported determinant of early alcohol and tobacco initiation in young adults. *Biol Psychiatry* 2008;63:1039–46.
10. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;452:638–42.
11. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;452:633–7.
12. Amos CI, Wu X, Broderick P, 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.
13. Minna JD. Nicotine exposure and bronchial epithelial cell nicotinic acetylcholine receptor expression in the pathogenesis of lung cancer. *J Clin Invest* 2003;111:31–3.
14. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort. *Cancer* 2002;94:2490–501.
15. Feigelson HS, Rodriguez C, Robertson AS, et al. Determinant of DNA yield and quality from buccal cell samples collected with mouthwash. *Cancer Epidemiol Biomarkers Prev* 2001;10:1005–8.
16. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–9.
17. Hinds DA, Stokowski RP, Patil N, et al. Matching strategies for genetic association studies in structured populations. *Am J Hum Genet* 2004;74:317–25.
18. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–59.
19. Escobedo LG, Peddicord JP. Smoking prevalence in US birth cohorts: the influence of gender and education. *Am J Public Health* 1996;86:231–6.
20. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2007. *CA Cancer J Clin* 2007;57:43–66.
21. Bierut LJ, Stitzel JA, Wang JC, et al. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry* 2008;165:1163–71.
22. Wada E, McKinnon D, Heinemann S, et al. The distribution of mRNA encoded by a new member of the neuronal nicotinic acetylcholine receptor gene family ($\alpha 5$) in the rat central nervous system. *Brain Res* 1996;526:45–53.
23. Salas R, Orr-Urtreger A, Broide RS, et al. The nicotinic acetylcholine receptor subunit $\alpha 5$ mediated short-term effects of nicotine *in vivo*. *Mol Pharmacol* 2003;63:1059–66.
24. Mineur YS, Picciotto MR. Genetics of nicotinic acetylcholine receptors: relevance to nicotine addiction. *Biochem Pharmacol* 2008;75:323–33.
25. Lessov CN, Martin NG, Stratham DJ, et al. Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol Med* 2004;34:865–79.
26. Johnson EO, Morgan-Lopez AA, Breslau N, et al. Test of measurement invariance of the FTND across demographic groups: assessment, effect size, and prediction of cessation. *Drug Alcohol Depend* 2008;93:260–70.