



Published in final edited form as:

Mol Psychiatry. 2016 May ; 21(5): 601–607. doi:10.1038/mp.2015.105.

Rare, low frequency, and common coding variants in *CHRNA5* and their contribution to nicotine dependence in European and African Americans

A full list of authors and affiliations appears at the end of the article.

Abstract

The common nonsynonymous variant rs16969968 in the $\alpha 5$ nicotinic receptor subunit gene (*CHRNA5*) is the strongest genetic risk factor for nicotine dependence in European Americans and contributes to risk in African Americans. To comprehensively examine whether other *CHRNA5* coding variation influences nicotine dependence risk, we performed targeted sequencing on 1582 nicotine dependent cases (Fagerström Test for Nicotine Dependence score ≥ 4) and 1238 non-dependent controls, with independent replication of common and low frequency variants using 12 studies with exome chip data. Nicotine dependence was examined using logistic regression with individual common variants (MAF ≥ 0.05), aggregate low frequency variants ($0.05 > \text{MAF} \geq 0.005$), and aggregate rare variants (MAF < 0.005). Meta-analysis of primary results was performed with replication studies containing 12 174 heavy and 11 290 light smokers. Next-generation sequencing with 180X coverage identified 24 nonsynonymous variants and 2 frameshift deletions in *CHRNA5*, including 9 novel variants in the 2820 subjects. Meta-analysis confirmed the risk effect of the only common variant (rs16969968, European ancestry: OR=1.3, $p=3.5 \times 10^{-11}$; African ancestry: OR=1.3, $p=0.01$) and demonstrated that 3 low frequency variants contributed an independent risk (aggregate term, European ancestry: OR=1.3, $p=0.005$; African ancestry: OR=1.4, $p=0.0006$). The remaining 22 rare coding variants were associated with increased risk of nicotine dependence in the European American primary sample (OR=12.9, $p=0.01$) and in the same risk direction in African Americans (OR=1.5, $p=0.37$). Our results indicate that common, low frequency and rare *CHRNA5* coding variants are independently associated with nicotine dependence risk. These newly identified variants likely influence risk for smoking-related diseases such as lung cancer.

INTRODUCTION

Nicotine is the primary addictive component of tobacco products, and its physiological effects are mediated through neuronal nicotinic acetylcholine receptors.¹ The $\alpha 5/\alpha 3/\beta 4$ nicotinic receptor subunit gene cluster on chromosome 15 harbors the strongest and most replicated genetic risk factor for smoking-related traits. Many independent studies

Correspondence and Requests for Reprints: Laura Jean Bierut, M.D., Washington University School of Medicine, Department of Psychiatry, Box 8134, 660 South Euclid Avenue, St. Louis, MO 63110-1010, Phone: (314) 362-3492, Fax: (314) 362-4247, laura@wustl.edu.

CONFLICT OF INTEREST

LJB, AG, and JCW as well as the spouse of NLS 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.

demonstrated that rs16969968, a common coding variant (D398N) in the $\alpha 5$ nicotinic receptor subunit gene (*CHRNA5*), is associated with nicotine dependence, heaviness of smoking, and smoking cessation, as well as smoking-related illnesses such as lung cancer and chronic obstructive pulmonary disease.²⁻¹⁰ Subsequent large-scale meta-analyses of European ancestry populations identified rs16969968 as unequivocally associated with heaviness of smoking ($p=5.57 \times 10^{-72}$).¹¹⁻¹³ Recently, rs16969968 was shown to have a similar effect in African ancestry populations,^{14,15} where the minor allele is less common. Beyond these robust association studies across ancestry groups, functional studies support the biological role of *CHRNA5* and rs16969968 in the development of nicotine dependence.^{4,16}

We hypothesized that additional low frequency and rare coding variants in *CHRNA5* alter risk for nicotine dependence. To comprehensively assess the relationship between *CHRNA5* coding variation and liability to nicotine dependence, we analyzed targeted sequence data from approximately 3000 nicotine dependent cases and non-dependent controls of European and African descent. Additionally, we used 12 studies with exome chip data to replicate associations of common and low frequency variants with smoking behaviors. Finally, we studied the variance explained in the development of nicotine dependence by the rare, low frequency, and common polymorphisms in *CHRNA5*.

MATERIALS AND METHODS

Primary sample ascertainment and description

Subjects were recruited through the Collaborative Genetic Study of Nicotine Dependence and the Genetic Study of Nicotine Dependence in African Americans.^{9,17} Institutional Review Board approval was obtained at each institution and written informed consent was obtained from all subjects. Community-based recruitment enrolled subjects aged 25-45 years old. All subjects underwent comprehensive phenotypic assessments of smoking behaviors, including the Fagerström Test for Nicotine Dependence (FTND).¹⁸ Nicotine dependent cases were required to be current smokers and have an FTND score of 4 or higher. Non-dependent controls had smoked at least 100 cigarettes (to ensure exposure to nicotine), but had a lifetime maximum FTND score of 1 (Table 1).

Targeted sequencing of *CHRNA5*

The Center for Inherited Disease Research (CIDR) performed next-generation targeted sequencing on *CHRNA5*. Details of the sequencing procedures and quality control measures are provided in the Supplementary Methods. The mean on-target coverage was 180X, and greater than 96% of on-target bases had a depth greater than 20X.

Evaluation of *CHRNA5* coding variants

Genotypic data that passed initial quality control at CIDR were released to the Quality Assurance/Quality Control analysis team at the University of Washington Genetics Coordinating Center. *CHRNA5* coding variants were identified by ANNOVAR¹⁹ and then manually reviewed. This review involved examining summary statistics of the quality control metrics, comparing the quality of novel variants with known variants from dbSNP

and HapMap, as well as inspecting alignments of selected samples with non-reference calls to pass or fail variant sites. Large genetic databases²⁰ and protein prediction programs²¹ were also used to evaluate identified coding variants.

Previously, Haller *et al.*,²² performed pooled sequencing of *CHRNA5* in a sample that contributed 511 participants to the targeted sequencing in this project and identified 4 *CHRNA5* coding variants beyond the well-studied risk variant rs16969968. Targeted sequencing found these 4 coding variants in the same 34 people as pooled sequencing. Furthermore, targeted sequencing identified 6 additional singleton variants among the 511 people included in both analyses. The high quality of the targeted sequencing data was verified using the HumanExome-12v1-1 array. All 2820 individuals included in our primary analysis were genotyped using this array, and the concordance for the common and low frequency coding variants was 99.9%.

Statistical analysis

A total of 1432 European and 1388 African Americans with targeted sequencing of *CHRNA5* and available smoking behaviors were examined. Data were analyzed using the Statistical Analysis System (SAS 9.3, Cary, NC, USA). Logistic regression was used to model case-control status. European and African Americans were analyzed separately. Ancestry groups were verified using EIGENSTRAT²³ and previously collected genome-wide arrays. Ten ancestry-specific principal components (PCs) were developed. Examination of eigenvalues led us to include the first PC in our statistical analyses of both ancestry groups. All models included the standard covariates of sex, age, and first ancestry-specific PC.

Coding variants that passed quality control were divided into three classes based on the derived MAF in the entire sample: rare ($MAF < 0.005$), low frequency ($0.05 > MAF > 0.005$), and common ($MAF > 0.05$). Visual examination of the distribution of the allele frequencies in the sample (Supplementary Figure S1) highlights a natural grouping of these three frequency classes.

In the primary analytic model, low frequency and rare variants were collapsed into an aggregate low frequency variant term and aggregate rare variant term, respectively. Individuals with at least one copy of the minor allele for any of the nonsynonymous or frameshift variants were coded as 1 in each variant class (low frequency or rare), and individuals without any minor allele copies in this class were coded as 0. This collapsing method was based on a burden test²⁴ to increase power to detect the cumulative effect of these variant classes.

Main effects of the one common rs16969968 coding variant, aggregate low frequency variants, and aggregate rare variants were analyzed together in a multivariate model of case-control status (*Multivariate Model Set 1*). This approach was used to examine the effect of low frequency and rare variants conditioned on the effect of the well-established, common coding risk variant rs16969968. The primary logistic regression model was

$$\text{logit}(p) = \beta_0 + \beta_{G_1} \text{rs16969968} + \beta_{G_2} \text{LowFrequencyTerm} + \beta_{G_3} \text{RareTerm} + \beta_C C$$

where C is the vector of standard covariates.

In secondary analyses, we examined the three low frequency variants (rs2229961, rs80087508, rs79109919) as individual terms along with the common rs16969968 variant and aggregate rare variants (*Multivariate Model Set 2*). The secondary logistic regression model was

$$\text{logit}(p) = \beta_0 + \beta_{G_1} \text{rs16969968} + \beta_{G_2} \text{rs2229961} + \beta_{G_3} \text{rs80087508} + \beta_{G_4} \text{rs79109919} + \beta_{G_5} \text{RareTerm} + \beta_C C.$$

Because very few people were homozygous for the minor allele of the low frequency variants (0-5 individuals per variant), the heterozygous and homozygous individuals for each minor allele were collapsed into a single group and compared to the homozygous individuals of the major allele in these analyses.

Explaining phenotypic variation

To examine the variation in nicotine dependence explained by *CHRNA5* coding variants, we used Nagelkerke's adjusted R² from logistic regression of case-control status.²⁵ The variance in phenotype attributed to selected variants was derived as the R² attributable to the full model minus the R² attributable to the base model, including age, sex, and first ancestry-specific PC as predictors of outcome. European and African American samples were analyzed separately.

Replication samples

Replication was sought from cohorts participating in the Gene-Lifestyle Interactions Working Group of CHARGE. The common and three low frequency *CHRNA5* variants were assessed in 12 independent replication datasets with smoking phenotypes and exome chip genotypes. Cigarettes smoked per day (CPD) was used to define the outcome because FTND scores were not available. Our replication analyses compared heavy smokers (CPD > 20) to light smokers (CPD ≤ 10). Previous work has demonstrated that these thresholds of CPD are reasonable proxies for nicotine dependence and non-dependence defined by FTND.^{3,26} European and African Americans were examined in separate logistic regression models that were similar to the primary sample models without the rare variant term. Specifically, the primary replication model was

$$\text{logit}(p) = \beta_0 + \beta_{G_1} \text{rs16969968} + \beta_{G_2} \text{LowFrequencyTerm} + \beta_C C$$

and the secondary replication model was

$$\text{logit}(p) = \beta_0 + \beta_{G_1} \text{rs16969968} + \beta_{G_2} \text{rs2229961} + \beta_{G_3} \text{rs80087508} + \beta_{G_4} \text{rs79109919} + \beta_C C,$$

where C is the vector of standard covariates. For both ancestry groups, each replication study was required to have at least 50 light and 50 heavy smokers to be included in analyses of that group.

Meta-analysis

Meta-analysis involving the 12 replication datasets was performed using PLINK.²⁷ Beta values for the genetic factors obtained from *Multivariable Model Sets 1 and 2* stratified by ancestry were meta-analyzed using weighting by standard errors. Although examination of the Cochran's Q statistic suggested no heterogeneity across studies for the genetic factors in the meta-analyses ($p > 0.1$), except rs16969968 in European Americans ($p = 0.02$), to be consistent, all reported meta-analysis results are from random effects models. An ancestry specific study was included in the meta-analysis of the individual low frequency variants, if the minor allele of that variant occurred in at least 5 subjects in the population. This cut-off was used because random effects meta-analyses may be unstable or undefined for very rare events.

RESULTS

Variants identified in sequencing

Sequencing identified 26 coding variants in *CHRNA5*, including 2 frameshift deletions and 24 nonsynonymous variants (details of these variants are in Figure 1 and Supplementary Table S1). The majority were predicted to be deleterious using a consensus protein prediction method.²¹ Interestingly, the well-studied rs16969968 variant was predicted to be neutral, supporting our approach of including all coding variants in the analyses.

Common *CHRNA5* variant

The only common coding variant identified was the previously well-studied variant rs16969968 located in exon 5 of *CHRNA5*. In the primary sample, the rs16969968 minor allele was associated with increased risk for nicotine dependence in European (OR=1.3, $p = 0.003$) and African Americans (OR=1.5, $p = 0.04$) (*Multivariate Model Set 1*, Table 2). Replication results from 12 independent studies provide strong evidence that the A allele of rs16969968 increases risk for heaviness of smoking (Figure 2 and Supplementary Tables S2-S3). Meta-analyses combining results from the primary and replication datasets demonstrate that rs16969968 has an OR of 1.3 in both European ($p = 3.7 \times 10^{-11}$) and African Americans ($p = 0.01$).

Aggregate low frequency *CHRNA5* variants

Three low frequency, nonsynonymous *CHRNA5* variants were identified. In the primary *Multivariate Model Set 1*, this aggregate low frequency term provided modest evidence for association in both populations (European: OR=1.8, $p = 0.06$; African: OR=1.4, $p = 0.07$) (Table 2). Results from the replication studies demonstrated a significant effect of the aggregated three low frequency variants on heaviness of smoking. The overall meta-analysis combining the primary and replication samples yielded an OR=1.3 in European Americans ($p = 0.005$) and OR=1.4 in African Americans ($p = 0.0006$) for the aggregate low frequency term.

Individual low frequency *CHRNA5* variants

In secondary analyses using *Multivariate Model Set 2*, we examined the independent contributions of the three low frequency variants to nicotine dependence risk controlling for the effect of other *CHRNA5* coding variants. One of these low frequency variants was found primarily in European Americans, and the other two were almost exclusively in African Americans (Table 2).

The low frequency variant rs2229961 causes a valine to isoleucine change at position 134 in exon 4 in the extracellular domain of the receptor. The minor allele principally occurred in European Americans (MAF=0.02) and was rare in African Americans (MAF=0.002). In the primary sample, all 51 individuals with a copy of rs2229961 also possessed at least one copy of the rs16969968 minor (risk) allele, suggesting that these two coding variants are in linkage disequilibrium. Adjusting for the effect of rs16969968, the minor allele of rs2229961 was in the risk direction in European Americans (OR=1.7, p=0.1). Meta-analysis of these primary results and the independent replication samples yielded an OR=1.3 in European Americans (p=0.007) (Supplementary Table S4).

The minor allele of the second low frequency variant rs80087508 causes a lysine to arginine transition at position 167 in the extracellular domain of the receptor. This variant occurred exclusively in African Americans (MAF=0.01). This variant co-occurred with the common rs16969968 minor allele in 5 out of 38 individuals in the primary sample. In *Multivariable Model Set 2* controlling for other coding variants, the minor allele of rs80087508 trended in the risk direction in African Americans (OR=2.1, p=0.06). Meta-analysis of these primary results and the 12 independent replication samples yielded an OR=1.6 in African Americans (p=0.02) (Supplementary Table S5).

The final low frequency variant, rs79109919, causes a leucine to glutamine change at amino acid position 363, which is located in the cytoplasmic domain of the receptor. The minor allele of rs79109919 was common in African Americans (MAF=0.06) and occurred in only one European American individual (MAF=0.0003) in the primary sample. Of the 158 individuals who possessed at least one copy of the rs79109919 minor allele, 7 also possessed a copy of the rs16969968 risk allele and 1 possessed a copy of the low frequency rs80087508 variant, suggesting the independent transmission of these variants. In the primary sample, the minor allele of rs79109919 was in the risk direction in African Americans (OR=1.3, p=0.15). Meta-analysis of the primary and replication results yielded an OR=1.4 in African Americans (p=0.03) (Supplementary Table S5).

Aggregate rare *CHRNA5* variants

Sequencing identified 22 rare coding variants (MAF<0.5%) (20 nonsynonymous variants and 2 frameshift deletions). These variants occurred throughout the protein sequence (Figure 1). Each variant occurred in 1-4 individuals in the primary sample (Supplementary Figure S1). Furthermore, 9 of the 22 rare variants were seen in a single individual and were previously unreported in large reference datasets²⁰ (Exome Variant Server) (Supplementary Table S1).

Because these variants occurred in only a limited number of individuals, we used a collapsing burden test to assess their cumulative effect after adjusting for the effect of rs16969968 and low frequency variants. Overall, 37 individuals possessed at least 1 rare variant (details of these individuals in Supplemental Table S6). In the primary sample, the aggregate rare variant term was associated with a risk effect in the European Americans (OR=12.9, p=0.01) as 12/13 (92%) individuals with at least one rare variant were nicotine dependent cases (Table 2, Table S6). In African Americans, the rare variant term was in the same risk direction but not significant (OR=1.5, p=0.37) as 17/24 (71%) of the individuals with at least one rare variant were nicotine dependent cases.

Phenotypic variation accounted for by testing genetic factors

Nagelkerke's adjusted R^2 was used to assess the proportion of nicotine dependence variation explained by individual SNPs and multivariable models in the primary sample (Table 3). The overall phenotypic variance explained by the genetic variants in this one gene was 2.4% in European Americans and 1.0% in African Americans.

DISCUSSION

This study demonstrates that common, low frequency, and rare *CHRNA5* coding variants independently increase risk for nicotine dependence in both European and African Americans. An important strength of our study was the large sample of African Americans, a population often under-represented in genetic studies. Differences in the genetic architecture of European and African ancestry groups indicate that distinct genetic factors differentially contribute to nicotine dependence in these populations. These differences are highlighted by the fact that the well-established coding variant in *CHRNA5*, rs16969968, is more common in European (MAF=0.35) than African Americans (MAF=0.06). Importantly, we identify and replicate new associations with nicotine dependence for three low frequency, nonsynonymous variants, two of which almost exclusively occur in African Americans (rs80087508 and rs79109919).

Our targeted sequencing of *CHRNA5* in approximately 3000 nicotine dependent cases and non-dependent controls builds upon previous exome chip and sequencing studies of other smoking-related measures.^{22,28-30} Sequencing a large, diverse, unrelated sample (n=2820) enabled us to identify many more coding variants of high quality (n=26), including 9 novel variants, and demonstrate association with nicotine dependence.

Using exome chip data from 12 studies containing over 10 000 heavy and 10 000 light smokers, we replicated our novel associations between low frequency *CHRNA5* coding variants and smoking behaviors observed through our targeted sequencing. Our replication study design of examining phenotypic extremes of smoking quantity (heavy vs light smokers) reduces classification errors and reproduced the robust associations between common and low frequency nonsynonymous *CHRNA5* variants seen in our discovery sample with nicotine dependence.

Because multiple independent *CHRNA5* risk variants were identified, a critical question is what proportion of phenotypic variance is explained by coding variation in this one gene.

Genetic studies of complex traits have identified reproducible associations, but these findings often explain a modest proportion of phenotypic variance.³¹ For nicotine dependence, rs16969968, arguably the single strongest genetic risk factor in European ancestry populations, accounts for 1.0% of variance in European Americans in our study and others.³² *_ENREF_10* The addition of low frequency and rare coding variants increased the estimated phenotypic variance explained by this gene in European Americans to 2.4%. In African Americans, though rs16969968 is less common and therefore explains a smaller proportion of estimated phenotypic variance ($R^2=0.4\%$), adding low frequency and rare coding variants increased this estimate ($R^2=1.0\%$). Since self-reported smoking behaviors are crude measures of exposure to nicotine, the variance explained by these polymorphisms for biomarkers of smoking such as carbon monoxide and cotinine will likely be greater.^{33,34} For example rs16969968 explained four to five times more of the variance in carbon monoxide and cotinine levels compared to self-reported cigarette consumption.^{32,33} Similarly, we hypothesize that these low frequency and rare *CHRNA5* coding variants will explain much more of the variance of these biomarkers.

The findings reported here have limitations. Because the nonsynonymous rs16969968 variant in *CHRNA5* is associated with changes in nicotinic receptor function,^{4,16} we hypothesized that other coding variants will have a similar effect, but functional studies will need to confirm this. Noncoding variants have been previously associated with changes in *CHRNA5* mRNA expression levels in the brain³⁵⁻³⁷ as well as nicotine dependence and lung cancer. Including these non-coding variants in the analyses of our primary sample did not appreciably alter the associations with the coding variants. Because replication analyses were based on exome chip data, the contribution of these non-coding variants could not be further tested. In addition, the majority of rare variants are also not available on the exome chip, and therefore typing of these rare variants is required for replication. Another limitation is that our analysis is restricted to a single gene. Other nicotinic receptors contribute to nicotine dependence.^{30,38}

From a public health perspective, these newly identified low frequency and rare *CHRNA5* coding variants will likely have important prognostic and therapeutic implications. The common *CHRNA5* coding variant (rs16969968) is a strong genetic risk factor for lung cancer and chronic obstructive pulmonary disease^{2,7,8,10} and also influences response to smoking cessation therapies.⁶ Though these low frequency and rare variants will have a smaller impact on a population based level, from an individual viewpoint, the presence of these risk variants has a strong effect on the development of nicotine dependence. An important next step is to test whether these low frequency and rare *CHRNA5* coding variants similarly increase the risk of smoking-related diseases such as lung cancer.

In summary, multiple rare, low frequency and common variants in the *CHRNA5* gene contribute to the development of nicotine dependence. This study is another example of how sequencing a gene associated with a disease can identify additional variants which then explain more of the missing genetic variance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Emily Olfson, BA¹, Nancy L. Saccone, PhD², Eric O. Johnson, PhD³, Li-Shiun Chen, MD MPH ScD¹, Robert Culverhouse, PhD⁴, Kimberly Doheny, PhD⁵, Steven M. Foltz, MS², Louis Fox, BS¹, Stephanie M. Gogarten, PhD⁶, Sarah Hartz, MD, PhD¹, Kurt Hetrick, MS⁵, Cathy C. Laurie, PhD⁶, Beth Marosy, MS⁵, Najaf Amin, PhD⁷, Donna Arnett, PhD⁸, R. Graham Barr, MD DrPH⁹, Traci M. Bartz, MS¹⁰, Sarah Bertelsen, MA JD¹¹, Ingrid B. Borecki, PhD¹², Michael R. Brown, MS¹³, Daniel I. Chasman, PhD^{14,15}, Cornelia M. van Duijn, PhD⁷, Mary F. Feitosa, PhD¹², Ervin R. Fox, MD, MPH¹⁶, Nora Franceschini, MD¹⁷, Oscar H. Franco, MD, PhD⁷, Megan L. Grove, MS¹³, Xiuqing Guo, PhD¹⁸, Albert Hofman, MD, PhD⁷, Sharon L.R. Kardia, PhD¹⁹, Alanna C. Morrison, PhD¹³, Solomon K. Musani, PhD¹⁶, Bruce M. Psaty, MD, PhD^{20,21}, D.C. Rao, PhD²², Alex P. Reiner, MD, MPH^{23,24}, Kenneth Rice, PhD⁶, Paul M. Ridker, MD, MPH^{14,15}, Lynda M. Rose, MS¹⁴, Ursula M. Schick, MS²³, Karen Schwander, MS²², Andre G. Uitterlinden, PhD^{7,25}, Dina Vojinovic, MD DSc⁷, Jen-Chyong Wang, PhD¹¹, Erin B. Ware, PhD¹⁹, Gregory Wilson, MS²⁶, Jie Yao, MD MS¹⁸, Wei Zhao, PhD¹⁹, Naomi Breslau, PhD²⁷, Dorothy Hatsukami, PhD²⁸, Jerry A. Stitzel, PhD²⁹, John Rice, PhD¹, Alison Goate, DPhil¹¹, and Laura J. Bierut, MD¹

Affiliations

¹ Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA ² Department of Genetics, Washington University School of Medicine, St Louis, MO, USA ³ Behavioral Health Epidemiology program, RTI International, Research Triangle Park, NC, USA ⁴ Department of Medicine and Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA ⁵ Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD, USA ⁶ Department of Biostatistics, University of Washington, Seattle, WA, USA ⁷ Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands ⁸ Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA ⁹ Departments of Medicine and Epidemiology, Columbia University Medical Center, New York, NY, USA ¹⁰ Cardiovascular Health Research Unit, Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, USA ¹¹ Department of Neurosciences, Icahn School of Medicine at Mt. Sinai, New York, NY, USA ¹² Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St Louis, MO, USA ¹³ Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA ¹⁴ Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA ¹⁵ Harvard Medical School, Boston, MA, USA ¹⁶ University of Mississippi Medical Center, Jackson, MS, USA ¹⁷ Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA ¹⁸ Institute for Translational Genomics and Population Sciences, Department of

Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA ¹⁹
 Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA ²⁰ Cardiovascular Health Research Unit, Departments of Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA, USA ²¹ Group Health Research Institute, Group Health, Seattle, WA, USA ²² Division of Biostatistics, Washington University School of Medicine in St. Louis, St. Louis, MO, USA ²³ Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA, USA ²⁴ Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA ²⁵ Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands ²⁶ Jackson State University, School of Public Service, Jackson, MS, USA ²⁷ Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA ²⁸ Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA ²⁹ Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health: grant numbers T32GM07200, UL1TR000448, TL1TR000449, and F30AA023685 to E.O.; grant numbers DA030398 and DA038076 to L.C.; and grant number U19CA148172 to L.J.B.

Grant number R01 HL118305 from the National Institutes of Health supported the replication analyses. Grants and contracts from the National Institutes of Health supported the following studies and groups: **COGEND** (P01CA89392), **AAND** (R01DA025888), **CIDR** (HHSN268201100011), **ARIC** (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, HHSN268201100012C, 5RC2HL102419, R01 HL118305), **CHS** (HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, R01HL085251, R01HL068986, R01AG023629, UL1TR000124, DK063491), **COGA** (U10AA008401), **FamHS** (R01HL118305, R01DK089256), **GENOA** (HL054464, HL054457, HL054481, HL071917, NS041558, HL87660, HL119443, HL118305), **HyperGEN** (HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, R01HL55673, R01HL055673, R01HL118305, U01HL54473, R01HL055673, R01HL118305), **JHS** (HSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C, HL103010, HL118305), **MESA** (N01HC95159, N01HC95160, N01HC95161, N01HC95162, N01HC95163, N01HC95164, N01HC95165, N01HC95166, N01HC95167, N01HC95168, N01HC95169, UL1TR000040, UL1RR025005, R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071252, R01HL071258, R01HL071259, UL1RR025005, N02HL64278, UL1TR000124, DK063491), **WGHS** (HL043851, HL080467, CA047988), **WHI** (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, HHSN271201100004C, R21HL123677, R01HL118305).

Erasmus Rucphen Family Study was supported by the following grants: European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947); European Community's Seventh Framework Program (FP7/2007–2013, HEALTH-F4-2007-201413); Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043); ZonMw grant (project 91111025).

Rotterdam Study was supported by Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study was also funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810 and Netherlands Consortium for Healthy Ageing (NCHA).

Please see Supplementary Materials for acknowledgements listed by study.

JAS has received support from Pfizer, Inc. NA is supported by the Hersenstichting Nederland (project number F2013(1)-28). OHF Franco works in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.); Metagenics Inc.; and AXA. Nestlé Nutrition (Nestec Ltd.); Metagenics Inc.; and

AXA had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review or approval of the manuscript. BMP serves on the DSMB of a clinical trial funded by the device manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Supplementary information is available at *Molecular Psychiatry's* website.

References

1. Dani JA, De Biasi M. Cellular mechanisms of nicotine addiction. *Pharmacology, biochemistry, and behavior*. 2001; 70:439–446.
2. Amos CI, Wu X, Broderick P, Gorlov IP, 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; 40:616–622. [PubMed: 18385676]
3. 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; 13:368–373. [PubMed: 18227835]
4. Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Gruzca RA, Xuei X, et al. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry*. 2008; 165:1163–1171. [PubMed: 18519524]
5. Caporaso N, Gu F, Chatterjee N, Sheng-Chih J, Yu K, Yeager M, et al. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One*. 2009; 4:e4653. [PubMed: 19247474]
6. Chen LS, Baker TB, Piper ME, Breslau N, Cannon DS, Doheny KF, et al. Interplay of genetic risk factors (CHRNA5-CHRNA3-CHRNA4) and cessation treatments in smoking cessation success. *Am J Psychiatry*. 2012; 169:735–742. [PubMed: 22648373]
7. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*. 2008; 452:633–637. [PubMed: 18385738]
8. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet*. 2009; 5:e1000421. [PubMed: 19300482]
9. Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*. 2007; 16:36–49. [PubMed: 17135278]
10. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008; 452:638–642. [PubMed: 18385739]
11. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010; 42:436–440. [PubMed: 20418889]
12. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*. 2010; 42:448–453. [PubMed: 20418888]
13. TAG. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010; 42:441–447. [PubMed: 20418890]
14. Chen LS, Saccone NL, Culverhouse RC, Bracci PM, Chen CH, Dueker N, et al. Smoking and genetic risk variation across populations of European, Asian, and African American ancestry--a meta-analysis of chromosome 15q25. *Genet Epidemiol*. 2012; 36:340–351. [PubMed: 22539395]
15. Saccone NL, Wang JC, Breslau N, Johnson EO, Hatsukami D, Saccone SF, et al. The CHRNA5-CHRNA3-CHRNA4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res*. 2009; 69:6848–6856. [PubMed: 19706762]
16. Kuryatov A, Berrettini W, Lindstrom J. Acetylcholine receptor (AChR) alpha5 subunit variant associated with risk for nicotine dependence and lung cancer reduces (alpha4beta2)alpha5 AChR function. *Mol Pharmacol*. 2011; 79:119–125. [PubMed: 20881005]

17. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet.* 2007; 16:24–35. [PubMed: 17158188]
18. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict.* 1991; 86:1119–1127. [PubMed: 1932883]
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38:e164. [PubMed: 20601685]
20. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012; 491:56–65. [PubMed: 23128226]
21. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet.* 2011; 88:440–449. [PubMed: 21457909]
22. Haller G, Druley T, Vallania FL, Mitra RD, Li P, Akk G, et al. Rare missense variants in CHRNA4 are associated with reduced risk of nicotine dependence. *Hum Mol Genet.* 2012; 21:647–655. [PubMed: 22042774]
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–909. [PubMed: 16862161]
24. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet.* 2008; 83:311–321. [PubMed: 18691683]
25. Nagelkerke N. A note on a general definition of the coefficient of determination. *Biometrika.* 1991; 78:691–692.
26. Stevens VL, Bierut LJ, Talbot JT, Wang JC, Sun J, Hinrichs AL, et al. Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:3517–3525. [PubMed: 19029397]
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–575. [PubMed: 17701901]
28. Doyle GA, Chou AD, Saung WT, Lai AT, Lohoff FW, Berrettini WH. Identification of CHRNA5 rare variants in African-American heavy smokers. *Psychiatr Genet.* 2014; 24:102–109. [PubMed: 24682045]
29. Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR, et al. Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biol Psychiatry.* 2014; 75:783–789. [PubMed: 24094508]
30. Wessel J, McDonald SM, Hinds DA, Stokowski RP, Javitz HS, Kennemer M, et al. Resequencing of nicotinic acetylcholine receptor genes and association of common and rare variants with the Fagerstrom test for nicotine dependence. *Neuropsychopharmacology.* 2010; 35:2392–2402. [PubMed: 20736995]
31. Maher B. Personal genomes: The case of the missing heritability. *Nature.* 2008; 456:18–21. [PubMed: 18987709]
32. Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, Perola M, et al. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNA4) on chromosome 15. *Hum Mol Genet.* 2009; 18:4007–4012. [PubMed: 19628476]
33. Bloom AJ, Hartz SM, Baker TB, Chen LS, Piper ME, Fox L, et al. Beyond cigarettes per day. A genome-wide association study of the biomarker carbon monoxide. *Annals of the American Thoracic Society.* 2014; 11:1003–1010. [PubMed: 25072098]
34. Munafo MR, Timofeeva MN, Morris RW, Prieto-Merino D, Sattar N, Brennan P, et al. Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *J Natl Cancer Inst.* 2012; 104:740–748. [PubMed: 22534784]

35. Wang JC, Cruchaga C, Saccone NL, Bertelsen S, Liu P, Budde JP, et al. Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in *CHRNA5*. *Hum Mol Genet.* 2009; 18:3125–3135. [PubMed: 19443489]
36. Wang JC, Gruzca R, Cruchaga C, Hinrichs AL, Bertelsen S, Budde JP, et al. Genetic variation in the *CHRNA5* gene affects mRNA levels and is associated with risk for alcohol dependence. *Mol Psychiatry.* 2009; 14:501–510. [PubMed: 18414406]
37. Wang JC, Spiegel N, Bertelsen S, Le N, McKenna N, Budde JP, et al. Cis-regulatory variants affect *CHRNA5* mRNA expression in populations of African and European ancestry. *PLoS One.* 2013; 8:e80204. [PubMed: 24303001]
38. Saccone NL, Schwantes-An TH, Wang JC, Gruzca RA, Breslau N, Hatsukami D, et al. Multiple cholinergic nicotinic receptor genes affect nicotine dependence risk in African and European Americans. *Genes Brain Behav.* 2010; 9:741–750. [PubMed: 20584212]

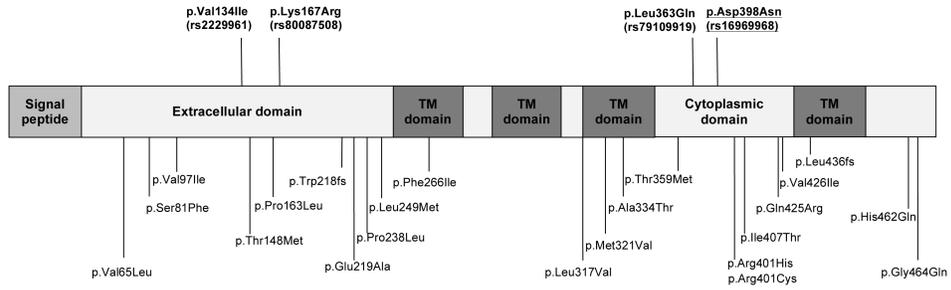


Figure 1. Protein Schematic of *CHRNA5* nonsynonymous and frameshift variants. **Bold underline** indicates the only common variant (MAF>5%); **Bold** indicates low frequency variants (5%>MAF 0.5%); Other variants are rare (MAF<0.5%).

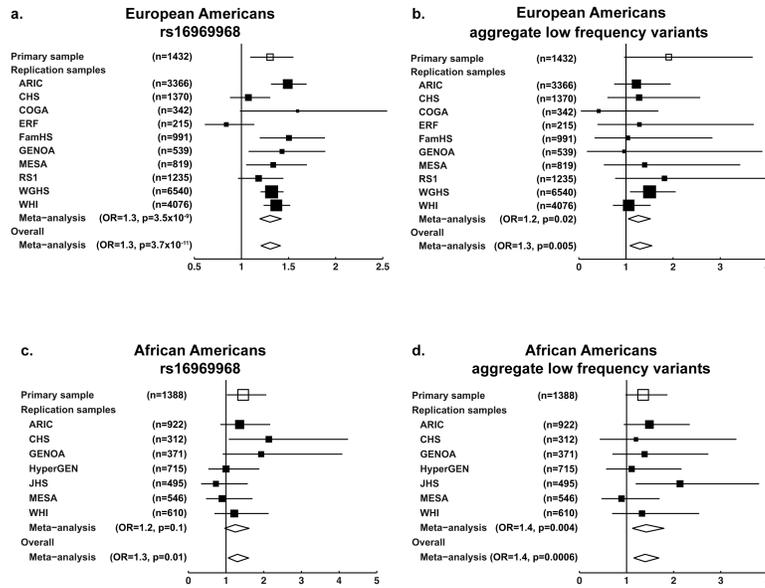


Figure 2. Forest plots showing the primary sample, replication samples, and random effects meta-analyses from *Multivariable Model Set 1*. (a) rs16969968 in European Americans; (b) aggregate low frequency variant term in European Americans; (c) rs16969968 in African Americans; (d) aggregate low frequency variant term in African Americans.

Table 1

Characteristics of primary sample

	European American (n=1432)		African American (n=1388)	
	cases	controls	cases	controls
Sample, n	728	704	854	534
Age, mean (range)	37 (25-45)	36 (25-45)	36 (25-45)	36 (25-45)
Sex				
Female	386 (53%)	482 (68%)	514 (60%)	321 (60%)
Male	342 (47%)	222 (32%)	340 (40%)	213 (40%)
FTND¹ score, mean (range)	6.49 (4-10)	0.02 (0-1)	6.21 (4-10)	0.33 (0-1)
CPD² category, mean (range)	1.94 (0-3)	0.01 (0-1)	1.11 (0-3)	0.03 (0-1)

¹FTND is the Fagerström Test for Nicotine Dependence²CPD is categorical cigarettes per day (1 is <10, 2 is 11-20, 3 is 21-30, 4 is >30);

Table 2

The effect of common, low frequency, and rare *CHRNA5* coding variants on nicotine dependence in primary sample

Variant Class	Variant	European Americans (n=1432)			African Americans (n=1388)		
		MAF ¹	OR (95% CI)	P-value	MAF	OR (95% CI)	P-value
Multivariable Model Set 1							
Common	rs16969968	0.355	1.27 (1.08-1.49)	0.003	0.058	1.46 (1.02-2.07)	0.04
Low Frequency	Aggregate term ²	0.016	1.81 (0.97-3.42)	0.06	0.071	1.35 (0.98-1.87)	0.07
Rare	Aggregate term	0.005	12.90 (1.66-100.54)	0.01	0.009	1.47 (0.60-3.59)	0.40
Multivariable Model Set 2							
Common	rs16969968	0.355	1.28 (1.09-1.50)	0.003	0.058	1.42 (1.00-2.03)	0.05
	rs2229961	0.016	1.71 (0.91-3.23)	0.10	0.002	2.57 (0.28-23.91)	0.41
Low Frequency	rs80087508	0 ³	.	.	0.014	2.00 (0.94-4.27)	0.07
	rs79109919	0.0003 ⁴	.	.	0.057	1.22 (0.86-1.75)	0.26
Rare	Aggregate term	0.005	12.91 (1.66-100.66)	0.01	0.009	1.51 (0.62-3.68)	0.37

This table shows the genetic effect of *CHRNA5* coding variants analyzed jointly in multivariate model sets 1 and 2.

Multivariable Model Set 1 includes rs16969968, the aggregate low frequency variant term, and the aggregate rare variant term;

Multivariable Model Set 2 includes rs16969968, rs2229961, rs80087508, rs79109919, and the aggregate rare variant term;

All models adjusted for sex, age, and first ancestry-specific PC as covariates;

¹MAF stands for minor allele frequency;

²for aggregate terms, the MAF was estimated by the dividing the number of people with at least one low frequency/rare variant by 2 times the total number of people;

³rs80087508 is non-polymorphic in European Americans;

⁴Because the minor allele of rs79109919 occurred less than 5 times in European Americans, the OR and p-value are not presented.

Table 3

Variation in nicotine dependence risk explained by selected variants in primary sample

Variant Class	Variant	European Americans (n=1432)			African Americans (n=1388)		
		MAF	R ²	p-value	MAF	R ²	p-value
<i>Common</i>	rs16969968	0.355	1.0%	0.001	0.058	0.4%	0.04
	rs2229961	0.016	0.4%	0.03	0.002	0.1%	0.24
<i>Low Frequency</i>	rs80087508	0	.	.	0.014	0.3%	0.07
	rs79109919	0.0003	0.2%	0.15	0.057	0.1%	0.34
<i>Rare</i>	aggregate term	0.005	1.0%	0.0009	0.009	0.1%	0.37
<i>All CHRNA5 genetic terms</i>			2.4%	5.5 × 10⁻⁵		1.0%	0.07

This table shows the variance explained by each individual variant and aggregate term by itself, as well as the variance explained by all *CHRNA5* genetic variants examined jointly in the final model. This final model includes rs16969968, rs2229961, rs80087508, rs79109919, and the aggregate rare variant term;

R² is the Nagelkerke's adjusted R² difference from logistic regression, comparing the base model with intercept, sex, age, and ancestry specific PCs to models with genetic variants;

p-values calculated by taking the difference between the -2loglikelihoods in the base model and those with variants as a chi-square statistic.