Rare missense variants in \textit{CHRNB4} are associated with reduced risk of nicotine dependence

Gabe Haller\textsuperscript{1,2}, Todd Druley\textsuperscript{2}, Francesco L. Vallania\textsuperscript{2}, Robi D. Mitra\textsuperscript{2}, Ping Li\textsuperscript{3}, Gustav Akk\textsuperscript{3}, Joe Henry Steinbach\textsuperscript{3}, Naomi Breslau\textsuperscript{4}, Eric Johnson\textsuperscript{5}, Dorothy Hatsuksami\textsuperscript{6}, Jerry Stitzel\textsuperscript{7}, Laura J. Bierut\textsuperscript{1} and Alison M. Goate\textsuperscript{1,2,∗}

\textsuperscript{1}Department of Psychiatry, \textsuperscript{2}Department of Genetics and \textsuperscript{3}Department of Anesthesiology, Washington University, St Louis, MO, USA, \textsuperscript{4}Department of Epidemiology, Michigan State University, East Lansing, MI, USA, \textsuperscript{5}Research Triangle Institute International, Research Triangle Park, NC, USA, \textsuperscript{6}Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA and \textsuperscript{7}Department of Integrative Physiology, University of Colorado, Boulder, CO, USA

Received June 29, 2011; Revised and Accepted October 26, 2011

Genome-wide association studies have identified common variation in the \textit{CHRNA5–CHRNA3–CHRNB4} and \textit{CHRNA6–CHRNB3} gene clusters that contribute to nicotine dependence. However, the role of rare variation in risk for nicotine dependence in these nicotinic receptor genes has not been studied. We undertook pooled sequencing of the coding regions and flanking sequence of the \textit{CHRNA5, CHRNA3, CHRNB4, CHRNA6} and \textit{CHRNB3} genes in African American and European American nicotine-dependent smokers and smokers without symptoms of dependence. Carrier status of individuals harboring rare missense variants at conserved sites in each of these genes was then compared in cases and controls to test for an association with nicotine dependence. Missense variants at conserved residues in \textit{CHRNB4} are associated with lower risk for nicotine dependence in African Americans and European Americans (AA \textit{P} = 0.0025, odds-ratio (OR) = 0.31, 95% confidence-interval (CI) = 0.31–0.72; EA \textit{P} = 0.023, OR = 0.69, 95% CI = 0.50–0.95). Furthermore, these individuals were found to smoke fewer cigarettes per day than non-carriers (AA \textit{P} = 6.6 \times 10^{-5}, EA \textit{P} = 0.021). Given the possibility of stochastic differences in rare allele frequencies between groups replication of this association is necessary to confirm these findings. The functional effects of the two \textit{CHRNB4} variants contributing most to this association (T375I and T91I) and a missense variant in \textit{CHRNA3} (R37H) in strong linkage disequilibrium with T91I were examined \textit{in vitro}. The minor allele of each polymorphism increased cellular response to nicotine (T375I \textit{P} = 0.01, T91I \textit{P} = 0.02, R37H \textit{P} = 0.003), but the largest effect on \textit{in vitro} receptor activity was seen in the presence of both \textit{CHRNB4} T91I and \textit{CHRNA3} R37H (\textit{P} = 2 \times 10^{-6}).

INTRODUCTION

Genetic variation in the cholinergic nicotinic receptor genes (\textit{CHRN}s) has repeatedly been found to be highly associated with nicotine dependence (reviewed in 1). Multiple independent variants have been identified that contribute to smoking-related phenotypes. Among these, the most strongly associated single nucleotide polymorphism (SNP) in several genome-wide association studies (GWAS) of nicotine dependence and correlated traits is a common non-synonymous change (rs16969968/D398N) in \textit{CHRNA5} (2). Despite big differences in the frequency of this SNP across populations, it shows a similar effect on risk, with odds ratios 1.3–2.00 in European, African American and Asian populations (3,4). \textit{In vitro} functional studies have demonstrated that receptors including the amino acid encoded by the minor allele, which is the risk allele, show a reduced response to the nicotine agonist, epibatidine (5). Human carriers of the minor allele show decreased intrinsic resting connectivity strength in the dorsal anterior cingulate–ventral striatum/extended amygdala circuit (6). Additionally, a group of highly correlated variants, including the SNP rs588765 near \textit{CHRNA5}, were shown to...
Neuronal nicotinic acetylcholine receptors are pentameric ion channels produced from various combinations of α and β receptor subunits and are the main target for nicotine in the brain. The vast majority of neuroanally expressed nicotinic receptors contain the α4 and β2 subunits, but many other combinations are formed and have distinct properties and expression patterns. As such, it is thought that the various aspects of nicotine dependence (craving, withdrawal, tolerance, sensitivity, etc.) may be due to these distinct characteristics. For example, recent studies have shown that mice lacking Chrnα5 fail to display the adverse effects of high doses of nicotine and mice lacking Chrnβ4 or Chrnα3 show reduced somatic signs of withdrawal after ceasing long-term nicotine treatment (9,10). Additionally, mice over-expressing Chrnβ4 were recently shown to display a strong aversion to nicotine that can be reversed by virally mediated expression of α5 D398N variant in the medial habenula (11). This work suggests that aversion of nicotine in the mouse is regulated by the balanced activity of B4 and α5 nicotinic receptor subunits in the medial habenula. A recent association study of smoking cessation also suggests that variants in CHRNβ4 may decrease craving and withdrawal symptoms (12). These findings suggest that multiple nicotinic receptor subunit genes likely play a role in the development and maintenance of nicotine dependence and that variants in these genes may have different and in some cases opposing functional effects.

Rare genetic variants have recently been shown to contribute to a number of common human diseases (reviewed in 13). Specifically, multiple rare variants have been reported in genes previously shown to harbor common variants associated with common diseases (14–16). Additionally, studies have demonstrated associations between rare variation in nicotinic receptor genes and nicotine dependence (17,18). As a class, rare SNPs, copy-number variants and small insertion/deletion polymorphisms (indels) constitute the majority of human genetic variation and thus may hold the key to understanding part of the missing heritability of complex traits unaccounted for by recent GWAS. Sequencing is one of the few methods by which these variants can be investigated; however, the cost of sequencing hundreds or thousands of individuals for multiple genes remains high despite advances in sequencing technologies. An efficient alternative approach is to carry out DNA sequencing using DNA pooled from multiple individuals. To further characterize the role of rare genetic variation in the nicotinic acetylcholine receptors on nicotine dependence, we have employed this approach to identify rare variants in the CHRNA5, CHRNA3, CHRNβ4, CHRNA6 and CHRNβ3 genes in individuals with and without nicotine dependence. We report that rare (<5%) missense variants at conserved sites in CHRNβ4 are associated with decreased risk of developing nicotine dependence and decrease the number of cigarettes consumed daily by smokers. These findings suggest that nicotine dependence is mediated in part by rare genetic variation and by variation in CHRNβ4 specifically.

RESULTS

In order to identify novel rare variants associated with nicotine dependence, we sequenced 752 individuals from the extremes of the population distribution of Fagerström Test of Nicotine Dependence (FTND) scores in 6 pools. The protein coding regions of the five human cholinergic nicotinic receptor subunit genes constituting the gene clusters previously reported to harbor common variants associated with nicotine dependence (CHRNA5–CHRNA3–CHRNβ4 and CHRNA6–CHRNβ3) (Fig. 1) were sequenced on an Illumina GAIIx and analyzed using the short indel prediction by large deviation inference and nonlinear true frequency estimation by recursion (SPLINTER) algorithm (19). We obtained >30-fold coverage per allele at all positions within the 28 amplicons designed to cover the protein coding exons of CHRNA3, CHRNA5, CHRNA6, CHRNβ3 and CHRNβ4 and validated 24 (9 novel) missense variants (Fig. 2 and Table 1). We only attempted to validate missense variants as these are
more likely to have a functional impact than non-coding or synonymous coding changes. As the SPLINTER algorithm produces a $P$-value for the probability that a predicted variant is a true positive, we also genotyped a number of variants below our cut-off for significance in order to determine the positive predictive value of the algorithm. A total of 42 variants were genotyped in the sequenced individuals. Of these, 14 were selected for having probabilities of being true positives just below our cut-off value. Only 1 of these 14 variants was validated, while 24 of 28 SNPs predicted to be true positives were validated. This produced a positive predictive value of 86%. Additionally, all variants included as part of the amplified positive control vector were found upon achieving $\sim 30$-fold coverage at mutated sites (sensitivity = 100%) and only $\sim 80$ sites in the 1908 bp negative control vector were predicted to be polymorphic (specificity = $\sim 95\%$). The majority of validated variants, as expected, are rare (22/24 = 92% have minor allele frequency < 5%) in African Americans or Europeans. In fact, many are present in only one sequenced individual (7/24 = 29% are singletons). Additionally, some of the missense variants present in these genes were found at less conserved sites (9/25; phyloP score < 2), suggesting that they are less likely to have a functional impact. Altogether, we identified 15 variants (4 novel) at conserved sites and used these for further analysis. Of these, 12 were also predicted to be damaging by PolyPhen, SIFT or both (Table 1 and Supplementary Material, Table S1) (20,21).

We determined the frequency of each missense variant validated from sequencing in individuals from the Collaborative Study on the Genetics of Nicotine Dependence (COGEND) (710 African Americans and 2055 European Americans) using Sequenom. We then compared the frequency of each missense SNP as well as the number of carriers of conserved missense variants in cases and controls for each of the genes sequenced. To ensure novel findings, we excluded the SNP rs16969968 from analyses in European Americans as this would overshadow the effects of other variants in the CHRNA5 gene and because it is common in European Americans (frequency = 0.35). We find that two SNPs in CHRNA4 (T91I and T375I) and one SNP in CHRNA3 (R37H) are associated with decreased risk of nicotine dependence either in African Americans or in European Americans (frequency = 0.35). We find that two SNPs in CHRNA4 (T91I and T375I) and one SNP in CHRNA3 (R37H) are associated with decreased risk of nicotine dependence either in African Americans or in European Americans (frequency = 0.35). Notably, the T91I and R37H variants are in high linkage disequilibrium (LD) in European Americans ($r^2 = 0.89$, $n = 2035$) and African Americans ($r^2 = 0.59$, $n = 710$), making it difficult to determine the relative contribution of the two variants to the observed association. We do not find an association between rare variants, conserved or otherwise, and nicotine dependence at the other genes studied (CHRNA6, CHRNA5 and CHRNA3) in African Americans or European Americans.

We find that carriers of rare missense variants at conserved sites in CHRNA4 were less frequently nicotine dependent (ND) compared with non-carriers, both in African Americans and European Americans. Therefore, further studies are needed to determine the functional significance of these variants.

**Figure 2.** Schematics of the CHRN genes studied and the locations of missense variants identified in sequenced individuals from the COGEND study. Each protein consists of a signal peptide, two extracellular domains, four transmembrane domains (TM1–4) and three intracellular domains. Conserved sites (phyloP44Vertebrate score $\geq 2$) are shown in grey, while all others are shown in black.
Table 1. Demographic and phenotypic characteristics of COGEND African Americans and European Americans

<table>
<thead>
<tr>
<th>CHRNA4</th>
<th>African Americans</th>
<th>European Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-carriers</td>
<td>Carriers</td>
</tr>
<tr>
<td>n</td>
<td>681</td>
<td>29</td>
</tr>
<tr>
<td>Age</td>
<td>37 ± 6</td>
<td>37 ± 7</td>
</tr>
<tr>
<td>Women-men</td>
<td>250/431</td>
<td>11/18</td>
</tr>
<tr>
<td>Nicotine cases/controls</td>
<td>450/231</td>
<td>11/18</td>
</tr>
<tr>
<td>CPD</td>
<td>17 ± 13</td>
<td>9 ± 6</td>
</tr>
<tr>
<td>FTND Score</td>
<td>4.0 ± 3.2</td>
<td>2.1 ± 2.8</td>
</tr>
<tr>
<td>Withdrawal symptom count</td>
<td>3.7 ± 2.4</td>
<td>2.4 ± 2.9</td>
</tr>
</tbody>
</table>

Carriers are individuals who possess at least one of the following: T91I, T375I, G296S or M456V. ND cases are maximum lifetime FTND ≥ 4, while controls are maximum lifetime FTND < 1. CPD values are mean ± SD and represent the average number of cigarettes smoked per day when the individual was smoking the most. FTND score is maximum lifetime FTND score. P-values were calculated using a two-sided Fisher’s exact test or two-sided Wilcoxon rank-sum test. P-values in bold are those less than 0.05.

Americans and European Americans (African Americans P = 0.0025, odds-ratio (OR) = 0.31, 95% confidence-interval (CI) = 0.31–0.72; European Americans P = 0.023, OR = 0.69, 95% CI = 0.50–0.95) (Table 1). This is significant for African Americans after Bonferroni correction for multiple tests assuming one test for each gene studied (P < 0.01). As we are not able to distinguish between the effects of the T91I and R37H variants, carriers of missense variants at conserved site in either CHRNA3 or CHRNA4 are also less frequently ND compared with non-carriers (African American P = 0.0035; European American P = 0.012). To determine whether the observed association could be explained by the CHRNA3 R37H alone, we performed logistic regression including CHRNA3 R37H genotype as a covariate. CHRNA4 carrier status remained associated with nicotine dependence status in African Americans (P = 0.0025), but not in European Americans (P = 0.37). This is most likely because rs61737499 (CHRNA4 T375I), one of the major contributors to the association SNP in African Americans, is monomorphic in European Americans. Additionally, we investigated the possibility that one of the four comorbid diagnostic and statistical manual of mental disorders IV (DSM IV) defined substance dependence phenotypes measured in our data set (alcohol, cocaine, marijuana and opiates) was contributing to the observed association at CHRNA4. None of these comorbid phenotypes significantly altered the association results observed between nicotine dependence and CHRNA4 carrier status in either African Americans or European Americans when included as a covariate in a logistic regression. These findings suggest that variants in CHRNA4 and possibly CHRNA3 protect against nicotine dependence and that this is not due to a correlation between nicotine dependence and other comorbid phenotypes.

Two groups of common genetic variants within the CHRNA5–CHRNA3–CHRNA4 gene cluster (tagged by rs16969968 and rs588765, respectively) were previously shown to affect nicotine dependence risk and represent the two strongest associations between SNPs in this region and nicotine dependence in European Americans and African Americans (2,7,22). To determine whether our findings could be due to LD with either of these variants, we performed logistic regression including each as a covariate. CHRNA4 carrier status remained associated with nicotine dependence status in African Americans (P = 0.0048), but not in European Americans (P = 0.13) when rs16969968 genotype was included as a covariate. However, we find that CHRNA4 carrier status is associated with nicotine dependence among carriers of the minor allele of rs16969968 (Fisher’s exact test P = 0.03, OR = 0.55) in European Americans. CHRNA4 carrier status was associated with nicotine dependence status in African Americans (P = 0.0031) and in European Americans (P = 0.043) when the rs588765 genotype was included as a covariate. These findings suggest that our observed associations are not due to a correlation between rare variants in CHRNA4 and previously described common variant associations in the region.

To ensure that our findings within the African American portion of the COGEND sample is not due to population stratification, we calculated principal components using EIGENSTRAT (23). We then performed logistic regression to test the association between CHRNA4 carrier status at conserved sites and nicotine dependence, including the first two principal components (PC1 and PC2) as covariates. CHRNA4 carrier status remained associated with nicotine dependence status in African Americans (P = 0.0038). Additionally, we calculated local admixture using local ancestry in admixed populations (24). In the 100 kb encompassing the CHRNA5–CHRNA3–CHRNA4 gene cluster, all African American individuals were found to have local ancestry estimates of 0, 50 or 100% European admixture, while all European Americans had local ancestry estimates of 100% European ancestry. The average European admixture of African Americans in this region was 21.8%, consistent with previous findings (25). CHRNA4 carrier status remained significantly associated with nicotine dependence when we performed logistic regression, including local admixture estimates from the 100 kb encompassing the CHRNA5–CHRNA3–CHRNA4 gene cluster as a covariate in African Americans (P = 0.0029). These findings suggest that the observed association is not due to population stratification, either on a genome scale or on a local scale.

In order to examine which aspects of nicotine dependence are affected by variants in CHRNA4, we tested the association between several nicotine-related phenotypes in addition to case status. When we compared the distribution of cigarettes per day (CPD) in CHRNA4 conserved SNP carriers and non-carriers, we observed that carriers also have a lower mean CPD than non-carriers (African American P = 6.6 × 10⁻³, European American P = 0.021). FTND as a quantitative...
We investigated whether conserved missense variants in \textit{CHRNB4} contribute to withdrawal symptoms among individuals who attempt cessation. We find that carriers of conserved missense variants in \textit{CHRNB4} show fewer signs of withdrawal as measured by the number of withdrawal symptoms endorsed in African Americans but not European Americans (African American \(P = 0.012\), European American \(P = 0.42\)) (Table 1).

To determine whether the three major contributors to this association (\textit{CHRNA4} T375I, \textit{CHRNA4} T91I and \textit{CHRNA3} R37H) affect \(a_3\beta_4\) receptor function, we heterologously expressed the variant forms of \(a_3\) and \(\beta_4\) in HEK293 cells. We examined the effect of the \textit{CHRNB4} T91I variant alone as well as in combination with \textit{CHRNA3} R37H because these two variants are in LD (EA \(r^2 = 0.89\), AA \(r^2 = 0.59\)). We elicited responses with acetylcholine (the endogenous transmitter) or nicotine. The concentration–response relationships did not show significant differences in terms of the concentration producing a half-maximal response, for either acetylcholine or nicotine (Fig. 3A and B; \(t\)-test \(P > 0.2\)). However, each of these variants showed increased maximal response to nicotine relative to the maximal response to acetylcholine (Fig. 3B and C). Both \textit{CHRNB4} variants, the \textit{CHRNA3} R37H variant alone and the combination of \textit{CHRNA3} R37H and \textit{CHRNB4} T91I variants, showed significantly increased relative maximal response to nicotine compared with normal (\(t\)-test \(P < 0.02\)). The maximal response produced from receptors containing \textit{CHRNB4} T91I and \textit{CHRNA3} R37H was also greater (1.4-fold) than the response seen for \textit{CHRNA3} R37H alone, further suggesting an effect of \textit{CHRNB4} T91I independent of \textit{CHRNA3} R37H (\(t\)-test \(P = 0.02\)).

**DISCUSSION**

We find that rare missense variants at evolutionarily conserved sites in \textit{CHRNB4} and \textit{CHRNA3} are associated with reduced risk of nicotine dependence in African Americans and European Americans. Our \textit{in vitro} data suggest that this is likely due to the tendency of these variants to increase cellular response to nicotine, although the exact mechanism is unclear. We do not find an association between rare variants, conserved or otherwise, and nicotine dependence at the other genes \((P = 0.59)\) on stochastic events, namely the sampling of rare allele carriers from a population, than studies of common genetic variants. It is possible that our observed associations may be due to chance fluctuations in rare variant allele.

When considering genetic associations in admixed populations, it is important to ensure that findings are not due to subtle population stratification, either on a genome scale or locally at associated loci. As a result, we calculated both global and local measures of admixture for each of the African American individuals in COGEND. We find that neither global nor local admixture is able to account for the association we observe between rare missense variants at evolutionarily conserved sites in \textit{CHRNB4} and nicotine dependence.
In addition to nicotine dependence, we examined the effect of conserved missense variants in 
*CHRNB4* on several phenotypes designed to tease apart various aspects of dependence. Spe-
cifically, we find that African American and European American rare variant carriers smoke fewer CPD and that African American carriers experience fewer withdrawal symptoms when com-
pared with non-carriers. These data suggest that these 
*CHRNB4* variants are likely contributing to the risk of develop-
ing nicotine dependence by allowing carriers to quit more easily, possibly by allowing them to smoke fewer CPD to combat withdraw-
sal symptoms. It is difficult to determine with confidence which aspect of nicotine dependence variants in 
*CHRNB4* are affecting, however, as each of the phenotypes analyzed are 
highly correlated in our sample (r² from 0.32–0.85) and because our sample was selected to contain the extremes of the 
population distribution of FTND score.

To determine the functional consequence of changes in 
*CHRNB4*, we characterized two variants, as well as a conserved 
*CHRNA3* missense variant correlated with 
*CHRNB4* T911. Both 
*CHRNB4* variants were found to increase maximal nicotine 
response relative to response to acetylcholine (T3751 P = 0.01, 
T911 P = 0.02). Additionally, the 
*CHRNA3* missense variant alone and in combination with 
*CHRNB4* T911 showed significant increases in the relative maximal response to nicotine (P = 0.003 
and 2 × 10⁻⁶). The maximal response produced from receptors 
containing 
*CHRNB4* T911 and 
*CHRNA3* R371H was also greater (>1.4-fold) than the response seen for either 
*CHRNB4* T911 or 
*CHRNA3* R371H alone, further suggesting an effect of 
*CHRNB4* T911 independent of 
*CHRNA3* R371H (P = 0.02). (Table 2)

These results are consistent with findings regarding the in vitro 
function of the previously described 
*CHRNA5* D398N variant (5,27). Increased maximal response to nicotine here confers protection against nicotine dependence, in contrast to a decreased maximal response to epibatidine and an increased risk for nicotine dependence seen for the D398N variant in 
*CHRNA5*. Additionally, assuming increased agonist response is functionally similar to over-expression, our results are consistent with the increased aversion to nicotine seen in mice with targeted overexpression of 
*ChrnB4* (11). We speculate, therefore, that one of the mechan-
isms by which vulnerability to nicotine dependence is conferred is 
through decreased sensitivity and thereby decreased aversion to 
nicotine.

In conclusion, we find an excess of conserved missense var-
iants in 
*CHRNB4* in smokers who are ND compared with 
smokers who are ND. Moreover, the two major genetic contrib-
utors to this association lead to increased functional response to 
nicotine in vitro. We hypothesize that these variants act by 
increasing sensitivity to nicotine and that these changes lead to 
our overall finding of decreased risk for nicotine dependence. 
Further sequencing of individuals for whom detailed 
dependence-related information has been obtained is neces-
sary to confirm and elaborate on these findings.

**MATERIALS AND METHODS**

**Sample selection**

DNA samples were collected as part of the Collaborative 
Genetic Study of Nicotine Dependence (COGEND). All 
members of the COGEND sample underwent a 
semi-structured interview, which assessed smoking behavior, 
other substance use and comorbid psychiatric conditions. 
The COGEND sample includes 710 African Americans [461 
ND cases and 249 smokers with no symptoms of dependence 
(controls)] and 2055 European Americans (1062 ND cases and 
993 controls). As these individuals were ascertained to study 
nicotine dependence, FTND scores (range = 1–10) were 
required to be ≤2 for controls and ≥5 for cases. In all 
cases, lifetime maximum FTND score was used. Additionally, 
to ensure that all individuals in the study had been exposed 
to nicotine, all members were required to have smoked at least 
100 cigarettes in their lifetime. Among individuals with 
FTND scores ≥4, 96% (1458/1523) were also ND using 
DSM-IV nicotine dependence criteria. A total of 352 
African Americans (176 ND cases and 176 smoking controls) 
and 400 European Americans (200 ND cases and 200 smoking 
controls) were sequenced. For African Americans, pools of 88 
individuals were used (two case pools and two control pools). 
For European Americans, pools of 200 were used (one case 
pool and one control pool). The case pools include individuals 
with FTND scores ≥5, whereas the control pools include indi-
viduals with FTND scores ≤1. Individuals for case pools were 
chosen to select the most severely ND subjects in our sample. 
Follow-up genotyping of SNPs identified and validated in the 
sequenced individuals was done in the remaining portion of 
the COGEND sample (310 African Americans and 1662 
European Americans).

**Pooled sequencing**

Pooled DNA sequencing was performed as previously 
described (28). The concentrations of individual DNA 
samples were first measured using Quant-iT™ PicoGreen 
reagent and pooled in equimolar amounts. Each pooled 
DNA sample was then used as the template for the amplifi-
cation of each protein coding exon of the 
*CHRNA3*, 
*CHRNA5*, 
*CHRNA6*, 
*CHRNB3* and 
*CHRNB4* genes. Primers for the 
amplification of each exon were designed using Primer3 and refer-
ence sequences were taken from the human genome reference 
assembly build 36.1 (hg18) (Supplementary Material, 
Table S1). In order to ensure complete coverage of each 
desired exon, a minimum of 50 bp of flanking sequence on 
each side was required for each amplicon. We used Pfu high-
fidelity DNA polymerase in all polymerase chain reactions 
(PCRs) to reduce the identification of SNPs generated as a 
result of the PCR (false positive SNPs). After PCR amplifi-
cation of desired genomic regions, PCR products were cleaned 
using QIAquick PCR purification kits, quantified using 
Quant-iT PicoGreen reagent and ligated in equimolar 
amounts using T4 Ligase and T4 Polynucleotide Kinase. At 
this stage, positive and negative control vectors were amplified 
and added to each pool to serve as internal quality standards 
and to be used in data analysis. After ligation, concatenated 
PCR products were randomly sheared by sonication and pre-
pared for Illumina sequencing on an Illumina Genome Analyz-
er Ix (GAIIx) according to the manufacturer’s specifications. 
As previously shown by Vallania et al. (19), an average cover-
age of 30-fold per allele per pool was shown to correlate to 
optimal positive predictive value for the SNP-calling algo-
thesis and was, therefore, the target level of coverage in this
Table 2. Missense variants in sequenced nicotinic receptor genes present in the COGEND sample

<table>
<thead>
<tr>
<th>Gene</th>
<th>Amino acid position</th>
<th>rs number</th>
<th>CHR</th>
<th>hg18 position</th>
<th>Allele 1</th>
<th>Allele 2</th>
<th>PhyloP score</th>
<th>Sift prediction</th>
<th>Frequency in AA cases</th>
<th>Frequency in AA controls</th>
<th>Frequency in EA cases</th>
<th>Frequency in EA controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3</td>
<td>S326T</td>
<td>–</td>
<td>8</td>
<td>42706583</td>
<td>T</td>
<td>A</td>
<td>0.67</td>
<td>Tolerated</td>
<td>0.002</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B3</td>
<td>H329Y</td>
<td>–</td>
<td>8</td>
<td>42706592</td>
<td>C</td>
<td>T</td>
<td>3.98</td>
<td>Damaging</td>
<td>0</td>
<td>0</td>
<td>0.0009</td>
<td>0.0030</td>
</tr>
<tr>
<td>B3</td>
<td>K451E</td>
<td>rs35327613</td>
<td>8</td>
<td>42710892</td>
<td>G</td>
<td>A</td>
<td>2.35</td>
<td>Tolerated</td>
<td>0.058</td>
<td>0.045</td>
<td>0.0009</td>
<td>0.0005</td>
</tr>
<tr>
<td>B3</td>
<td>H455N</td>
<td>–</td>
<td>8</td>
<td>42710904</td>
<td>C</td>
<td>A</td>
<td>0.97</td>
<td>Tolerated</td>
<td>0</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A6</td>
<td>N447S</td>
<td>rs16891583</td>
<td>8</td>
<td>42730159</td>
<td>C</td>
<td>T</td>
<td>4.96</td>
<td>Damaging</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A6</td>
<td>T231H</td>
<td>–</td>
<td>8</td>
<td>42730807</td>
<td>G</td>
<td>A</td>
<td>6.91</td>
<td>Tolerated</td>
<td>0</td>
<td>0</td>
<td>0.0005</td>
<td>0</td>
</tr>
<tr>
<td>A6</td>
<td>R96H</td>
<td>–</td>
<td>8</td>
<td>42731315</td>
<td>C</td>
<td>T</td>
<td>0.94</td>
<td>Damaging</td>
<td>0</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A6</td>
<td>S43P</td>
<td>–</td>
<td>8</td>
<td>42739457</td>
<td>A</td>
<td>G</td>
<td>1.23</td>
<td>Damaging</td>
<td>0.011</td>
<td>0.018</td>
<td>0</td>
<td>0.0005</td>
</tr>
<tr>
<td>A5</td>
<td>S81F</td>
<td>–</td>
<td>15</td>
<td>76660343</td>
<td>C</td>
<td>T</td>
<td>5.35</td>
<td>Damaging</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A5</td>
<td>V134I</td>
<td>rs2229981</td>
<td>15</td>
<td>76667807</td>
<td>A</td>
<td>G</td>
<td>6.72</td>
<td>Damaging</td>
<td>0.002</td>
<td>0</td>
<td>0.0171</td>
<td>0.0122</td>
</tr>
<tr>
<td>A5</td>
<td>K167R</td>
<td>–</td>
<td>15</td>
<td>76669288</td>
<td>A</td>
<td>G</td>
<td>5.40</td>
<td>Damaging</td>
<td>0.015</td>
<td>0.016</td>
<td>0</td>
<td>0.0005</td>
</tr>
<tr>
<td>A5</td>
<td>L363Q</td>
<td>rs79109919</td>
<td>15</td>
<td>76669876</td>
<td>T</td>
<td>A</td>
<td>4.77</td>
<td>Damaging</td>
<td>0.061</td>
<td>0.054</td>
<td>0.0009</td>
<td>0</td>
</tr>
<tr>
<td>A5</td>
<td>D398N</td>
<td>rs16969098</td>
<td>15</td>
<td>76669980</td>
<td>A</td>
<td>G</td>
<td>3.77</td>
<td>Damaging</td>
<td>0.063</td>
<td>0.032</td>
<td>0.3874</td>
<td>0.3114</td>
</tr>
<tr>
<td>A3</td>
<td>R37H</td>
<td>rs8192745</td>
<td>15</td>
<td>76709629</td>
<td>C</td>
<td>T</td>
<td>2.76</td>
<td>Damaging</td>
<td>0.008</td>
<td>0.010</td>
<td>0.0430</td>
<td>0.0579</td>
</tr>
<tr>
<td>B4</td>
<td>M467V</td>
<td>rs61737502</td>
<td>15</td>
<td>76704628</td>
<td>T</td>
<td>C</td>
<td>1.58</td>
<td>Damaging</td>
<td>0.083</td>
<td>0.068</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B4</td>
<td>M465V</td>
<td>–</td>
<td>15</td>
<td>76704661</td>
<td>T</td>
<td>C</td>
<td>4.71</td>
<td>Damaging</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0005</td>
</tr>
<tr>
<td>B4</td>
<td>T351I</td>
<td>rs61737499</td>
<td>15</td>
<td>76708578</td>
<td>G</td>
<td>A</td>
<td>2.26</td>
<td>Damaged</td>
<td>0.008</td>
<td>0.026</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B4</td>
<td>R349C</td>
<td>–</td>
<td>15</td>
<td>76708657</td>
<td>A</td>
<td>G</td>
<td>1.96</td>
<td>Damaged</td>
<td>0.001</td>
<td>0</td>
<td>0.0007</td>
<td>0.0072</td>
</tr>
<tr>
<td>B4</td>
<td>G296S</td>
<td>–</td>
<td>15</td>
<td>76708816</td>
<td>C</td>
<td>T</td>
<td>5.69</td>
<td>Damaged</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td>0</td>
</tr>
<tr>
<td>B4</td>
<td>S140G</td>
<td>rs56218866</td>
<td>15</td>
<td>76709228</td>
<td>G</td>
<td>A</td>
<td>2.27</td>
<td>Damaged</td>
<td>0.047</td>
<td>0.040</td>
<td>0.0028</td>
<td>0.0051</td>
</tr>
<tr>
<td>B4</td>
<td>R136Q</td>
<td>rs56095004</td>
<td>15</td>
<td>76709295</td>
<td>C</td>
<td>T</td>
<td>0.41</td>
<td>Damaged</td>
<td>0.003</td>
<td>0</td>
<td>0.0122</td>
<td>0.0106</td>
</tr>
<tr>
<td>B4</td>
<td>R136W</td>
<td>–</td>
<td>15</td>
<td>76709296</td>
<td>G</td>
<td>A</td>
<td>0.33</td>
<td>Damaged</td>
<td>0.012</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B4</td>
<td>T91I</td>
<td>rs12914008</td>
<td>15</td>
<td>76710560</td>
<td>A</td>
<td>G</td>
<td>2.16</td>
<td>Damaged</td>
<td>0.005</td>
<td>0.008</td>
<td>0.0378</td>
<td>0.0536</td>
</tr>
<tr>
<td>B4</td>
<td>K57E</td>
<td>–</td>
<td>15</td>
<td>76714871</td>
<td>T</td>
<td>C</td>
<td>1.16</td>
<td>Damaged</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td>0</td>
</tr>
</tbody>
</table>

Cases are individuals with FTND score ≥ 4 and controls are individuals with FTND scores ≤ 1. PhyloP score is the basewise vertebrate conservation score. Positions are from human genome reference assembly build 36.1 (hg18).
study. In order to obtain sufficient coverage per allele per pool, we obtained two lanes of Illumina GAIIx sequencing per pool. Coverage per amplicon was calculated for each pool after the first lane of sequence was obtained and only those amplicons not reaching 30-fold coverage per allele were included for the second lane of sequencing.

**Sequencing analysis**

For analysis, sequencing reads (36 bp reads) were aligned using an alignment algorithm developed by Vallania et al. (19) which aligns sequences allowing for two mismatches or indels of up to 4 bp. In order to quantify the specificity and sensitivity of this method, positive and negative control DNA were introduced as PCR products in the pooled sequencing protocol outlined above. As a positive control to estimate sensitivity for variant calling, a pool of 10 plasmids with a 72 bp insert was generated. One plasmid acts as the ‘wild-type’ insert, while the remaining nine plasmids contain one or two synthetically engineered mutations. All 10 plasmids were combined such that each the allele frequency of each known mutation would mimic either the allele frequency of a single allelic variant or 10 allelic variants in our human DNA pool. Following mixing of the vectors, PCR amplification across the insert sequence was performed and the PCR product was added to the normalized pool of human target PCR reactions during sequencing library preparation. We then obtained more coverage for each amplicon than that which was required to detect all variants in this amplified positive control vector pool in order to ensure a minimal SNP detection false negative rate. As a negative control and to model the sequencing error rate, 1908 bp of the pCMV6-XL5 plasmid was also amplified and included. One half of the pCMV6-XL5 plasmid sequence was used to train the SNP finding algorithm and the other to test the error model. Finally, to identify variants in the pooled sequence data, we used the SPLINTER algorithm (19). The SPLINTER algorithm produces a \( P \)-value for the probability that a predicted variant is a true positive. This probability is produced using large deviation theory to quantify the difference between observed allele frequencies within the sequence data to the background error rate for the same type of sequence changes seen in the amplified pCMV6-XL5 plasmid. A \( P \)-value cut-off value for each lane of Illumina sequencing was defined as the value at which all positive controls were identified. Only those variants falling below this cut-off value were considered ‘predicted’ by SPLINTER. All protein coding or splice site variants predicted by SPLINTER were then validated by individual Sequenom genotyping in each person from the source DNA pools. SNPs validated in the sequenced individuals were then genotyped in all members of the COGEND sample. All individual genotyping was performed using the Sequenom platform as described previously (8).

**Association analysis**

Each missense variant as well as the number of carriers of missense variants at conserved positions for each gene in the genes surveyed was tested for association with nicotine dependence in African Americans and European Americans using Fisher’s exact test. Conservation was determined by basewise vertebrate conservation using PhyloP score (29). A site was called conserved when its phyloP score was \( \geq 2 \), corresponding to a \( P \)-value of 0.01. The SNP rs56218866 (CHRNB4 S140G) was excluded from analyses as it had been shown previously to have no measurable effect on receptor function and did not effect the significance of the association between CHRNA4 and either nicotine dependence or CPD when taken as a covariate (30). For analysis of CPD, FTND score and withdrawal symptom count in carriers and non-carriers, Wilcoxon rank-sum tests were performed.

**In vitro functional assay**

Whole-cell voltage-clamp was performed as previously described to measure intracellular ion concentrations in response to exposure to nicotine (31). For each variant, we co-transfected HEK293 cells with plasmids expressing normal \( \alpha_3 \) and either normal \( \beta_4 \) or variant \( \beta_4 \), or variant \( \alpha_3 \) with normal or variant \( \beta_4 \). Additionally, we co-transfected HEK293T cells with plasmids expressing the minor allele of \( \alpha_3 \) R37H and \( \beta_4 \) T91I, because these SNPs are in very high LD and thus frequently co-occur in human populations. Concentration–response curves were produced for each subunit combination. \( EC_{50} \) values maximal response to agonist were each compared with normal using a two-tailed \( t \)-test.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

In memory of Theodore Reich, founding Principal Investigator of COGEND, we are indebted to his leadership in the establishment and nurturing of COGEND and acknowledge with great admiration his seminal scientific contributions to the field. Lead investigators directing data collection are L.B., N.B., D.H. and E.J. The authors thank Heidi Kromrei and Tracey Richmond for their assistance in data collection. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1RR024992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH.

*Conflict of Interest statement.* L.J.B. and A.M.G. are listed as inventors on the patent ‘Markers for Addiction’ (US 20070258898) covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. L.J.B. acted as a consultant for Pfizer, Inc. in 2008.
FUNDING

The COGEND project is a collaborative research group and part of the National Institute on Drug Abuse (NIDA) Genetics Consortium. Subject collection was supported by NIH grant P01 CA89392 (L.J.B.). Phenytoin and genotypic data are stored in the NIDA Center for Genetic Studies (NCGS) at http://zork.wustl.edu/ under NIDA Contract HHSN271200477451C. Electrophysiological studies supported by NIH grant R21 DA26918 (G.A.).

REFERENCES