Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Hancock, D. B., G. W. Reginsson, N. C. Gaddis, X. Chen, N. L. Saccone, S. M. Lutz, B. Qaiser, et al. 2015. “Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence.” Translational Psychiatry 5 (10): e651. doi:10.1038/tp.2015.149. http://dx.doi.org/10.1038/tp.2015.149.

Published Version
doi:10.1038/tp.2015.149

Citable link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:27822338

Terms of Use

This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
ORIGINAL ARTICLE

Genome-wide meta-analysis reveals common splice site acceptor variant in \textit{CHRNA4} associated with nicotine dependence

DB Hancock1, GW Regnsson2, NC Gaddis3, X Chen4,5, NL Saccone6, SM Lutz7, B Qaiser8, R Sherva9, S Steinberg9, F Zink9, SN Stacey9, C Glasheen2, J Chen8, F Gu10, BN Frederiksen11, A Loukola12, DF Gudbjartsson8, I Brüske13, MT Landi10, H Bickeböller13, P Madden14, L Farrer9,15,16,17,18,19, J Kaprio8,20,21, HR Kranzler22,23, J Gelernter24,25,26,27, TB Baker28, P Kraft29,30, CI Amos31,32,33, NE Caporaso10, JE Hokanson34, LJ Bierut14, TE Thorgeirsson2, EO Johnson35 and K Stefansson2,36

We conducted a 1000 Genomes–imputed genome-wide association study (GWAS) meta-analysis for nicotine dependence, defined by the Fagerström Test for Nicotine Dependence in 17 074 ever smokers from five European-ancestry samples. We followed up novel variants in 7469 ever smokers from five independent European-ancestry samples. We identified genome-wide significant association in the alpha-4 nicotinic receptor subunit (\textit{CHRNA4}) gene on chromosome 20q13: lowest \( P = 8.0 \times 10^{-9} \) across all the samples for rs2273500-C (frequency = 0.15; odds ratio = 1.12 and 95% confidence interval = 1.08–1.17 for severe vs mild dependence), rs2273500-C, a splice site acceptor variant resulting in an alternate \textit{CHRNA4} transcript predicted to be targeted for nonsense-mediated decay, was associated with decreased \textit{CHRNA4} expression in physiologically normal human brains (lowest \( P = 7.3 \times 10^{-4} \)). Importantly, rs2273500-C was associated with increased lung cancer risk (\( N = 28 \, 998 \), odds ratio = 1.06 and 95% confidence interval = 1.00–1.12), likely through its effect on smoking, as rs2273500-C was no longer associated with lung cancer after adjustment for smoking. Using criteria for smoking behavior that encompass more than the single ‘cigarettes per day’ item, we identified a common \textit{CHRNA4} variant with important regulatory properties that contributes to nicotine dependence and smoking-related consequences.

\textit{Translational Psychiatry} (2015) 5, e651; doi:10.1038/tp.2015.149; published online 6 October 2015

INTRODUCTION

Cigarette smoking is a major contributor to cancer, vascular disease and lung disease, and the leading cause of preventable mortality worldwide.\(^1\) Nicotine dependence is heritable,\(^2\) and genome-wide association study (GWAS) analyses of smoking behaviors and nicotine dependence have unequivocally identified single nucleotide polymorphism (SNP) associations with nicotine acetylcholine receptor gene clusters on chromosomes 15q25 (\textit{CHRNA5-CHRNA3-CHRNA4}) and 8p11 (\textit{CHRNA3-CHRNA4}).\(^3,9\) The largest prior GWAS of smoking behavior was conducted using very large sample sizes (\( N \) up to 74 053) and phenotypes such as smoking history (ever vs never), age of onset, smoking cessation (former vs current) and cigarettes per day (CPD).\(^3-7\)
To identify additional genetic loci, we conducted the largest GWAS meta-analysis for nicotine dependence to date. We used the Fagerström Test for Nicotine Dependence (FTND), a six-item questionnaire with scores that range from 0 to 10 and indicate the level of physiological dependence on nicotine. The FTND focuses on the core criteria for dependence, including heavy use/tolerance (for example, CPD) and withdrawal (for example, time to first cigarette in the morning), and although it does not capture some factors such as craving, the FTND remains the strongest predictor of smoking cessation among the primary measures of nicotine dependence. It has also been shown to provide a robust measure of nicotine dependence across different settings and populations.

In our study, we categorized nicotine dependence as mild (FTND score 0–3 or low-level smoking), moderate (FTND score 4–6) or severe (FTND score 7–10) among study participants, all of European-ancestry, who reported smoking more than 100 cigarettes in their lifetime. We conducted a 1000 Genomes–imputed GWAS meta-analysis of nicotine dependence across five study samples (total N = 17,074), identified the alpha-4 nicotinic receptor subunit (CHRNA4) gene as a novel genome-wide significant locus, and conducted follow-up association testing with lung cancer using six study samples (total N = 12,160 cases and 16,838 controls). Our results revealed that rs2273500, a splice site acceptor SNP with important regulatory effects for CHRNA4, was associated with risk of developing both nicotine dependence and lung cancer.

**MATERIALS AND METHODS**

All protocols used in this study received institutional review board approval at their respective sites, and all the study participants or their legal representatives provided informed consent.

**Discovery study samples, the FTND and quality control**

Five samples were used to conduct our GWAS meta-analysis of nicotine dependence: deCODE Genetics, Environment and Genetics in Lung Cancer Etiology Study (EAGLE), Chronic Obstructive Pulmonary Disease Gene (COGEND) Study, Collaborative Genetic Study of Nicotine Dependence (COGEND) Study, and Study of Addiction: Genetics and Environment (SAGE) Study. These study participants, all of European-ancestry, had genome-wide SNP genotypes and FTND scores to define nicotine dependence among participants who reported smoking more than 100 cigarettes in their lifetime. Quality control (QC) was conducted on genotyped participants and SNPs in each sample using PLINK unless otherwise stated.

We used the FTND range of scores to categorize participants’ nicotine dependence as mild (FTND score 0–3), moderate (FTND score 4–6) or severe (FTND score 7–10). In the deCODE Genetics sample only, 4313 low-intensity smokers with no FTND data available but with less than 10 cigarettes per day reported. For EAGLE, age was only available as a categorical variable, so average age could not be calculated. The categorical age distributions were as follows: 23.2% aged 59 or less, 18.2% aged 60–64, 22.4% aged 65–69, 21.4% aged 70–74 and 14.8% aged 75–79.

**Table 1.** Participant characteristics from 10 study samples used for the genome-wide association study (GWAS) meta-analysis or independent replication of follow-up GWAS variants

| Study sample | Total N | No. (%)a | No. (%), male | Mean age (s.d.) |
|--------------|---------|----------|---------------|----------------|
| **GWAS meta-analysis samples** | | | | |
| deCODE | 9090 | 5871 (64.6) | 2074 (22.8) | 1145 (12.6) | 4253 (46.8) | 54.2 (16.7) |
| EAGLE | 3006 | 1416 (47.1) | 1027 (34.2) | 563 (18.7) | 2528 (84.1) | Not availableb |
| COGEND | 2211 | 666 (30.1) | 964 (43.6) | 581 (26.3) | 1214 (54.9) | 57.7 (7.9) |
| COGEND | 1935 | 941 (48.6) | 521 (26.9) | 473 (24.4) | 750 (38.8) | 36.5 (5.5) |
| SAGE* | 832 | 243 (29.2) | 295 (35.5) | 294 (35.3) | 465 (55.9) | 39.4 (11.3) |
| **Independent replication samples** | | | | |
| FTC | 2374 | 1345 (56.7) | 793 (33.4) | 236 (9.9) | 1314 (55.3) | 45.9 (15.6) |
| Yale-Penn | 2116 | 381 (18.0) | 1014 (47.9) | 721 (34.1) | 1247 (58.9) | 37.3 (10.7) |
| UW-TTURC | 1534 | 311 (20.2) | 723 (47.1) | 500 (32.6) | 658 (42.9) | 43.1 (11.5) |
| GAIN | 774 | 327 (42.3) | 280 (36.2) | 167 (21.6) | 389 (50.3) | 53.8 (16.4) |
| nonGAIN | 671 | 298 (44.4) | 234 (34.9) | 139 (20.7) | 349 (52.0) | 52.9 (15.5) |

Abbreviations: COGEND, Collaborative Genetic Study of Nicotine Dependence; COGEND, Chronic Obstructive Pulmonary Disease Gene Study; EAGLE, Environment and Genetics in Lung Cancer Etiology Study; FTC, Finnish Twin Cohort Study; GAIN, Genetic Association Information Network GWAS of schizophrenia; nonGAIN, Molecular Genetics of Schizophrenia—nonGAIN sample; SAGE*, Study of Addiction: Genetics and Environment (* indicates that overlapping COGEND participants were excluded); UW-TTURC, University of Wisconsin-Transdisciplinary Tobacco Use Research Center. aScores on the Fagerström Test for Nicotine Dependence (FTND) were used to categorize nicotine dependence as mild (FTND score 0–3), moderate (FTND score 4–6) or severe (FTND score 7–10). For deCODE only, the mild category included participants with FTND score 0–3 and an additional set of 4313 low-intensity smokers with no FTND data available but with less than 10 cigarettes per day reported. For EAGLE, age was only available as a categorical variable, so average age could not be calculated. The categorical age distributions were as follows: 23.2% aged 59 or less, 18.2% aged 60–64, 22.4% aged 65–69, 21.4% aged 70–74 and 14.8% aged 75–79.
accession number phs000093.v2.p2). Additional analyses were conducted using the specific FTND item scores by the original study investigators. We began by applying all participant-level and SNP-level QC procedures that were recommended as part of the dbGaP release and then applied our own standard set of QC procedures (Supplementary Information). There were 3006 EAGLE participants included in our study. The COGEND observational sample primarily focused on identifying genetic risk factors for COPD, as previously described.16 Recruited participants were non-Hispanic white or African American and aged 45 to 80 years old, who reported a history of smoking (currently or past) and 10 or more cigarette pack-years. FTND was assessed in current smokers only. We used the non-Hispanic white current smoking participants for the analysis. Among the COPD cases, disease severity was staged according to the Global Initiative for Chronic Obstructive Lung Disease criteria, which are based on post-bronchodilator pulmonary function measures. COPD controls had pulmonary function measures in the normal range for their age and height, separately by sex. Exclusion criteria for acute and chronic respiratory disease, cancer and other conditions were used. COPDGene participants were genotyped on the Illumina HumanOmni1-Quad BeadChip array. After applying our standard QC procedures (Supplementary Information), there were 2211 COPDGene participants for analysis.

COGEND, a community-based case–control study of nicotine-dependent smokers vs smokers who never developed nicotine-dependence symptoms, began recruiting participants in 2001 from St. Louis and Detroit through telephone screening to identify current smokers aged 25 to 44 years old.3 The FTND was administered to determine study eligibility. Current smokers with an FTND score of ≥4 were recruited as nicotine-dependent cases, and smokers who reported >100 cigarettes during their lifetime but an FTND score of 0 or 1 were recruited as controls. COGEND participants were genotyped on either the Illumina Human1M-Duo BeadChip array, a part of SAGE,4 or the Illumina HumanOmni2.5 BeadChip array as part of GENEVA.20 In each subset, genotyped SNPs with a call rate >98% and HWE $P > 1 \times 10^{-4}$ were retained. We combined the subsets and removed duplicated participants and first-degree relatives. To circumvent bias that may arise from conducting imputation on subjects genotyped on different arrays, we carried forward only the SNPs genotyped at the intersection of the different arrays.21 After applying our standard QC procedures (Supplementary Information) on the combined COGEND sample, there remained 1935 participants for our study.

The final GWAS sample consisted of the remaining SAGE study participants. The full SAGE sample included participants from COGEND, the Collaborative Study on the Genetics of Alcoholism13 and the Family Study of Cocaine Dependence.7 For our study, we excluded the COGEND participants due to low ascertainment. Before the remaining Collaborative Study on the Genetics of Alcoholism and Family Study of Cocaine Dependence participants were ascertainment as part of case–control studies of addictive disorders and all were ascertainment from sites in the United States, we analyzed them together as done in previous GWAS analyses.5 We henceforth refer to this sample as SAGE5. We obtained their Illumina Human1M-Duo BeadChip genotypes and phenotype data via dbGaP accession number phs000092.v1.p1. After applying our standard QC procedures (Supplementary Information), there remained 832 participants for analysis.

For the four samples of non-isolated populations (EAGLE, COPDGene, COGEND and SAGE5), we used the STRUCTURE program14 to compute the ancestral proportions of all study participants using, based on comparison to the HapMap reference populations of Chinese (denoted CHB) and European Americans (denoted CEU) and African Americans (denoted ASW) using 10,000 SNPs randomly distributed across the genome. We excluded outlying participants with ≥25% Asian and/or African American proportions.

1000 Genomes imputation

Genotype imputation was conducted in each non-isolated sample using IMPUTE2 (ref. 25) with reference to the 1000 Genomes ALL phase I integrated variant set.24,26 Additional details are provided in the Supplementary Information. Following imputation and removal of SNPs and insertions/deletions (indels) with minor allele frequency <0.01 in the 1000 Genomes EUR panel (collection of five European-ancestry populations), we tested 8,548,225 SNPs and 1,395,199 indels for association with nicotine dependence across the samples. We used the info metric to evaluate the SNP/indel imputation quality rather than imposing an imputation quality filter and possibly missing truly associated SNPs/indels.27 The SNP and indel genotype probabilities were converted to dosages and used in the regression model for association testing with nicotine dependence to account for any imputation uncertainty.28 For deCODE, genotype imputation was conducted by long-range phasing of all chip-genotyped individuals with methods described previously.29 Sequence variants were imputed from the deCODE whole-genome sequencing effort into 104,220 chip-genotyped Icelanders, who had been phased with long-range phasing, using the same model as used by IMPUTE2.28

Statistical analyses: testing genome-wide SNP and indel association with nicotine dependence

The genotyped and 1000 Genomes–imputed SNPs and indels were tested for association with categorical nicotine dependence (mild, moderate and severe) using ProbAbel software30 in each sample with linear regression models that included age, sex and sample-specific covariates (if applicable, see Supplementary Information). The four European American and Italian samples also included principal component eigenvectors to minimize bias owing to population stratification; for each sample, we selected the number of eigenvectors needed to account for >75% of the variability in nicotine dependence. The sample-specific GWAS results were combined in METAL31 using inverse variance-weighted meta-analysis. The standard GWAS threshold ($P < 5 \times 10^{-8}$) was used to declare statistically significant results. The I² index was used to assess heterogeneity across samples.25

Follow-up SNP/indel association testing with nicotine dependence in independent samples

For any novel region having genome-wide significant association with nicotine dependence, we selected SNPs and indels associated at meta-analysis $P < 5 \times 10^{-5}$ for follow-up testing across five independent European-ancestry samples: Yale-Penn study,14–15 Finnish Twin Cohort Study (FTC),36,37 University of Wisconsin-Transdisciplinary Tobacco Use Research Center (UW-TTURC),12 Genetic Association Information Network (GAIN) GWAS of Schizophrenia and Molecular Genetics of Schizophrenia—non-Hispanic White Sample. Participants in the Yale-Penn study were from small nuclear families and unrelated indiduates recruited in the eastern United States in the course of studies of the genetics of alcohol, cocaine or opioid dependence. Nicotine dependence had no role in subject selection. Yale-Penn participants were administered the Semi-Structured Assessment for Drug Dependence and Alcoholism and were genotyped on the Illumina HumanOmniQuad v1.0 microarray. Participants with missing rate >2%, HWE $P < 1 \times 10^{-4}$ and with significantly different minor allele frequency across genotyping labs (Yale or CIDR) were set to missing before imputation. SNP genotype imputation was performed with IMPUTE2 (ref. 25) using genotyped SNPs and the March 2012 1000 Genomes EUR reference panel. Genetic relationships were examined by calculating pairwise identity-by-state estimates using PUNK.12 Sample duplicates (identity-by-state >90%) were removed, pairs of individuals whose identity-by-state proportions did not match their reported genetic relationship were assigned to two different families and pairs of individuals who shared >25% of their alleles identity-by-state were assigned to the same family. Participants with gender discordance ($F_2 < 0.2$ for chromosome X SNPs to confirm females and $F_2 > 0.8$ to confirm males) were also removed, unless their true identity could be determined. To verify and correct potential misclassification of self-reported race, we compared the GWAS data from all participants with HapMap phase III reference genotypes. Association tests were performed on N = 2116 using linear regression models adjusted for age, sex and the first three principal component eigenvectors computed using Eigensoft and embedded in generalized estimating equations to correct for correlations among relatives. Altogether, 2374 participants from FTC were included for replication testing. These participants originated from the following cohorts: the Nicotine Addiction Genetics study of adult twins born in 1938–1957 and concordant for ever smoking, and their family members (mainly siblings); a population-lacal family longitudinal study of five consecutive birth cohorts (1983–1987) of Finnish twins (FinnTwin12 sample); and a population-based longitudinal study of five consecutive birth cohorts (1975–1979) of Finnish twins (FinnTwin16 sample).38–41 Genotyping was done with the Illumina Human670-QuadCustom BeadChip (at the Wellcome Trust Sanger Institute) and the Illumina HumanCoreExome BeadChip (at the Wellcome Trust Sanger Institute and at the Broad Institute of MIT and Harvard). FTC
samples were imputed with a large number of population samples, separately by genotyping array, with reference to 1000 Genomes (Phase I integrated variant set release [SHAPEIT2] in National Center for Biotechnology Information (NCBI) build 37 [hg19] coordinates) at the Institute for Molecular Medicine Finland. After imputation, FTG samples were merged together. Standardized residuals of the categorical FTN phenotype, regressed against sex, age, sex, age, birth cohort and the first 10 first principal components (calculated from genome-wide genotype data), were used in QFAM association test in PLINK.12 The resulting regression coefficients are mathematically equivalent to regression coefficients when using raw phenotypes that are linearly regressed on SNP genotypes and covariates, enabling us to combine the FTC results with the others in meta-analysis as done elsewhere.53

UW-TTURC participants were recruited for nicotine dependence and smoking cessation treatment clinical trials in Madison and Milwaukee, Wisconsin beginning in 2001.13 We obtained their Illumina HumanOmni2.5 million genotypes, FTN scores and other phenotypic data via dbGaP accession number phs000404.v1.p1. After applying our standard QC procedures (Supplementary Information), there were 1534 UW-TTURC participants included in our study.

The GAIN and nonGAIN samples originated from the same Molecular Genetics of Schizophrenia study. These companion samples were genotyped separately, using the same platform (Affymetrix 6.0). Data from half of the Molecular Genetics of Schizophrenia study participants genotyped under the auspices of GAIN were obtained via dbGaP accession number phs000021.v3.p2, and data from the other half of the Molecular Genetics of Schizophrenia study participants (nonGAIN) were obtained via dbGaP accession number phs000167.v1.p1. We used only schizophrenia controls from GAIN (N = 774) and nonGAIN (N = 671), for which we applied our standard set of QC procedures (Supplementary Information).

Across the replication samples, the 1000 Genomes-imputed additive genotype dosages for selected SNPs and indels were tested for association with FTN-identified nicotine dependence (mild, moderate and severe) using linear regression models unless otherwise stated. Adjustments were made for age, sex and eigenvectors (first three for Yale-Penn and the number needed to account for > 75% of the phenotypic variability for UW-TTURC and Yale) in all (N = 9187 for FTND score 0–3 or low-level smoking), moderate (N = 4881 with FTND score 4–6) or severe (N = 3056 with FTND score 7–10). We tested genotyped and 1000 Genomes–imputed SNPs and indels for association with categorical nicotine dependence in each sample and then combined the GWAS results using inverse variance-weighted meta-analysis with genomic correction52 applied to each sample. Meta-analysis results for the 9.9 million tested SNPs/indels had no indication of bias (λgc = 0.97, Supplementary Figure 2).

RESULTS

Our GWAS meta-analysis included 17 074 ever smokers from five European-ancestry samples (Table 1). We used the FTND to categorize participants according to nicotine dependence (score 0–3 or low-level smoking), moderate (score 4–6) or severe (score 7–10). We tested genotyped and 1000 Genomes–imputed SNPs and indels for association with categorical nicotine dependence in each sample and then combined the GWAS results using inverse variance-weighted meta-analysis with genomic correction52 applied to each sample. Meta-analysis results for the 9.9 million tested SNPs/indels had no indication of bias (λgc = 0.97, Supplementary Figure 2).

A novel genome-wide significant association was observed on chromosome 20q13 (lowest P = 3.8 × 10−8 for rs4809294, Figure 1a). In total, 23 SNPs/indels on chromosome 20q13 were associated at P < 5 × 10−6 (Supplementary Table 1). We also observed associations within the known gene clusters on chromosomes 15q25 (CHRNA5-CHRNA3-CHRNB4, lowest P = 3.5 × 10−17) and 8p11 (CHRNA3-CHRNA6, lowest P = 1.2 × 10−10) (Figure 1a and Supplementary Tables 2 and 3). None of the 23 top-associated CHRNA4 SNPs interacted with previously established SNPs in the known regions (Supplementary Table 4).

We tested the 23 CHRNA4 SNPs/indels associated with nicotine dependence at P < 5 × 10−6 for independent replication using 7469 ever smokers from five European-ancestry samples (Table 1 and Supplementary Table 5). Two SNPs had genome-wide significant associations with nicotine dependence across all the samples (Table 2): rs2273500 and rs6107779, which are in strong linkage disequilibrium in European-ancestry individuals (D′ = 1.00 and R2 = 0.70, Figure 1b and Supplementary Figure 3). Their minor
alleles (frequency = 0.15 and 0.20) were associated with greater nicotine dependence risk, as demonstrated in Figure 2 for the top SNP rs2273500: meta-analysis odds ratio = 1.06 (95% confidence interval 1.04–1.08) for moderate vs mild nicotine dependence and odds ratio = 1.12 (95% confidence interval 1.08–1.17) for severe vs mild nicotine dependence. None of the tested SNPs/indels showed significant evidence for between-sample heterogeneity (Supplementary Table 6).

The top SNP from the overall meta-analysis, rs2273500, was imputed well across our discovery and replication samples, with info values ranging from 0.8 to 1.0 (Supplementary Tables 1 and 5). See the Supplementary Information for an evaluation of the agreement between rs2273500 imputed dosages and directly observed genotypes. rs2273500 is in some linkage disequilibrium with rs4809294, the top GWAS-identified SNP (D’=0.97 and r²=0.37, Supplementary Figure 3). rs4809294 (replication meta-
Table 2. CHRNA4 SNPs and indels associated with nicotine dependence at genome-wide association study (GWAS) meta-analysis P < 5 × 10⁻⁵ and followed up for independent replication testing

| SNP/indel | Minor allele | Base pair position (NCBI build 37) | SNP type | MAF² | GWAS sample meta-analysis (N = 17 074) | Replication sample meta-analysis (N = 7469) | All sample meta-analysis (N = 24 543) |
|-----------|--------------|------------------------------------|----------|------|---------------------------------------|---------------------------------------------|-----------------------------------------|
| rs2273500 | C            | 61 986 949 Intron                  |          | 0.15 | 0.057 2.3 × 10⁻⁶ 0.061 9.2 × 10⁻⁴ 0.058 | 8.0 × 10⁻⁹                                  |                                          |
| rs6011779 | C            | 61 984 317 Intron                  |          | 0.20 | 0.049 6.0 × 10⁻⁶ 0.059 5.5 × 10⁻⁴ 0.052 | 1.4 × 10⁻⁶                                  |                                          |
| rs6062901 | G            | 61 980 261 Intron                  |          | 0.18 | 0.050 7.0 × 10⁻⁶ 0.055 2.1 × 10⁻³ 0.051 | 5.2 × 10⁻⁶                                  |                                          |
| rs6062899 | G            | 61 979 793 Intron                  |          | 0.19 | 0.049 9.8 × 10⁻⁶ 0.054 2.4 × 10⁻³ 0.050 | 8.6 × 10⁻⁸                                  |                                          |
| rs4809543 | A            | 61 986 950 Intron                  |          |      | 0.074 0.086 3.0 × 10⁻⁷ 0.061 0.070 0.074 | 1.3 × 10⁻⁷                                  |                                          |
| rs4544949 | C            | 61 984 317 Intron                  |          | 0.20 | 0.049 6.0 × 10⁻⁶ 0.059 0.059 0.070 | 1.7 × 10⁻⁷                                  |                                          |
| rs4557732 | G            | 61 983 934 Intron                  |          | 0.078 | 0.082 4.6 × 10⁻⁶ 0.045 0.070 0.071 | 1.9 × 10⁻⁷                                  |                                          |
| rs201806007 | AT         | 61 988 398 Intron                  |          | 0.15 | 0.054 1.2 × 10⁻⁶ 0.054 5.3 × 10⁻³ 0.054 | 2.1 × 10⁻⁷                                  |                                          |
| rs151176846 | C           | 61 989 658 Intron                  |          | 0.080 | 0.078 7.2 × 10⁻⁷ 0.040 0.12 0.071 | 3.5 × 10⁻⁷                                  |                                          |
| rs45623037 | C            | 61 986 950 Intron                  |          | 0.080 | 0.076 8.8 × 10⁻⁵ 0.046 0.070 0.068 | 1.3 × 10⁻⁷                                  |                                          |
| rs45461993 | A            | 61 979 347 Intron                  |          | 0.060 | 0.099 1.5 × 10⁻⁷ 0.028 0.35 0.079 | 7.3 × 10⁻⁷                                  |                                          |
| rs45418932 | G            | 61 978 306 Intron                  |          | 0.080 | 0.076 2.1 × 10⁻⁶ 0.044 0.088 0.067 | 8.4 × 10⁻⁷                                  |                                          |
| rs199666656 | T           | 61 975 634 Intron                  |          | 0.057 | 0.094 3.3 × 10⁻⁷ 0.026 0.39 0.075 | 1.6 × 10⁻⁶                                  |                                          |
| rs45497800 | T            | 61 991 833 Intron                  |          | 0.082 | 0.075 2.2 × 10⁻⁷ 0.035 0.16 0.064 | 2.0 × 10⁻⁶                                  |                                          |
| rs4545294 | G            | 61 974 832 Intron                  |          | 0.062 | 0.092 6.2 × 10⁻⁷ 0.024 0.43 0.073 | 3.2 × 10⁻⁶                                  |                                          |
| rs45508092 | G            | 61 974 731 Intron                  |          | 0.059 | 0.096 5.8 × 10⁻⁷ 0.025 0.40 0.076 | 3.2 × 10⁻⁶                                  |                                          |
| rs45497800 | A            | 61 977 640 Intron                  |          | 0.048 | 0.12 7.2 × 10⁻⁸ 0.010 0.75 0.084 | 3.4 × 10⁻⁶                                  |                                          |
| rs45612034 | A            | 61 974 970 Intron                  |          | 0.062 | 0.090 1.2 × 10⁻⁶ 0.024 0.42 0.072 | 5.0 × 10⁻⁶                                  |                                          |
| rs45470098 | A            | 61 978 306 Intron                  |          | 0.062 | 0.090 1.2 × 10⁻⁶ 0.024 0.42 0.072 | 5.0 × 10⁻⁶                                  |                                          |
| rs45470098 | A            | 61 979 328 Intron                  |          | 0.036 | 0.14 2.4 × 10⁻⁷ 0.0079 0.82 0.090 | 2.1 × 10⁻⁵                                  |                                          |
| rs14498540 | T            | 61 984 931 Intron                  |          | 0.049 | 0.10 4.1 × 10⁻⁶ 0.0045 0.88 0.068 | 1.3 × 10⁻⁴                                  |                                          |

Abbreviations: indel, insertion/deletion; MAF, minor allele frequency; SNP, single nucleotide polymorphism; UTR, untranslated region. ²MAF was weighted by sample size across all 10 samples. SNPs and indels are sorted by meta-analysis P-values across all the samples. P-values surpassing genome-wide significance threshold (P < 5 × 10⁻⁵) are in bold.

Figure 2. Association of the rs2273500 minor allele (C) with severe vs mild nicotine dependence across all the samples. rs2273500 had the lowest P-value for its association with nicotine dependence in meta-analysis across all the samples, which are sorted by their size. Black-filled diamonds indicate the discovery samples, and open diamonds indicate the replication samples. The odds ratio (OR) estimates are shown proportional to the sample size. CI, confidence interval; COGEND, Collaborative Genetic Study of Nicotine Dependence; COPDGene, Chronic Obstructive Pulmonary Disease Gene Study; EAGLE, Environment and Genetics in Lung Cancer Etiology Study; FTC, Finnish Twin Cohort Study; GAIN, Genetic Association Information Network GWAS of schizophrenia; nonGAIN, Molecular Genetics of Schizophrenia—nonGAIN sample; SAGE*, Study of Addiction: Genetics and Environment (* indicates that overlapping COGEND participants were excluded); UW-TTURC, University of Wisconsin-Transdisciplinary Tobacco Use Research Center.

Prior genome-wide studies of CPD did not identify the CHRNA4 region at genome-wide significance, likely owing to differences in phenotype definition and lower SNP coverages. Regarding phenotype definition, we found that our top SNP rs2273500 was most significantly associated with time to first cigarette in the morning ($P = 2.3 \times 10^{-8}$, Supplementary Table 7). In contrast, for the missense CHRNA5 SNP rs16969968 that has reproducible variant associations with nicotine dependence, we found that our top SNP rs2273500 was next evaluated for its association with lung cancer in 28,998 participants from six European-ancestry case-control samples (Supplementary Table 11). However, we were able to evaluate rs2273500 as an eQTL SNP reliably using CHRNA4 mRNA expression levels measured across 10 regions of physiologically normal human brains from 134 European-ancestry participants in the Brain eQTL Almanac. At the transcript level (Supplementary Figure 4), we observed that rs2273500-C resulted in significantly reduced splicing to exon 4.1 compared to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.

Transcripts containing exon 4.1 were also observed in GTEx brain tissue samples, but relative to the liver samples, their frequencies were low (Supplementary Table 10). We observed the same pattern, whereby rs2273500-C carriers had reduced splicing to exon 4.1 and increased splicing to exon 4.2, but these differences were not statistically significant ($P = 0.30$) likely owing to the limited statistical power with the lower split read counts. However, we were able to evaluate rs2273500 as an eQTL SNP targeting for nonsense-mediated decay. To investigate the rs2273500 effect on splicing, we obtained RNA-seq and genotype data from the GTEx project. Liver ($N = 32$) showed the highest CHRNA4 expression among the GTEx tissues, and transcripts containing exon 4.1 were detected. Using split read counts (reads crossing an exon:exon boundary) to test the efficiency of splicing events that utilize the exon 4 splice donor, we found that rs2273500-C resulted in significant reduction of splicing to exon 4.1 in favor of increased splicing to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.

Transcripts containing exon 4.1 were also observed in GTEx brain tissue samples, but relative to the liver samples, their frequencies were low (Supplementary Table 10). We observed the same pattern, whereby rs2273500-C carriers had reduced splicing to exon 4.1 and increased splicing to exon 4.2, but these differences were not statistically significant ($P = 0.30$) likely owing to the limited statistical power with the lower split read counts. However, we were able to evaluate rs2273500 as an eQTL SNP targeting for nonsense-mediated decay. To investigate the rs2273500 effect on splicing, we obtained RNA-seq and genotype data from the GTEx project. Liver ($N = 32$) showed the highest CHRNA4 expression among the GTEx tissues, and transcripts containing exon 4.1 were detected. Using split read counts (reads crossing an exon:exon boundary) to test the efficiency of splicing events that utilize the exon 4 splice donor, we found that rs2273500-C resulted in significant reduction of splicing to exon 4.1 in favor of increased splicing to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.

Transcripts containing exon 4.1 were also observed in GTEx brain tissue samples, but relative to the liver samples, their frequencies were low (Supplementary Table 10). We observed the same pattern, whereby rs2273500-C carriers had reduced splicing to exon 4.1 and increased splicing to exon 4.2, but these differences were not statistically significant ($P = 0.30$) likely owing to the limited statistical power with the lower split read counts. However, we were able to evaluate rs2273500 as an eQTL SNP targeting for nonsense-mediated decay. To investigate the rs2273500 effect on splicing, we obtained RNA-seq and genotype data from the GTEx project. Liver ($N = 32$) showed the highest CHRNA4 expression among the GTEx tissues, and transcripts containing exon 4.1 were detected. Using split read counts (reads crossing an exon:exon boundary) to test the efficiency of splicing events that utilize the exon 4 splice donor, we found that rs2273500-C resulted in significant reduction of splicing to exon 4.1 in favor of increased splicing to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.

Transcripts containing exon 4.1 were also observed in GTEx brain tissue samples, but relative to the liver samples, their frequencies were low (Supplementary Table 10). We observed the same pattern, whereby rs2273500-C carriers had reduced splicing to exon 4.1 and increased splicing to exon 4.2, but these differences were not statistically significant ($P = 0.30$) likely owing to the limited statistical power with the lower split read counts. However, we were able to evaluate rs2273500 as an eQTL SNP targeting for nonsense-mediated decay. To investigate the rs2273500 effect on splicing, we obtained RNA-seq and genotype data from the GTEx project. Liver ($N = 32$) showed the highest CHRNA4 expression among the GTEx tissues, and transcripts containing exon 4.1 were detected. Using split read counts (reads crossing an exon:exon boundary) to test the efficiency of splicing events that utilize the exon 4 splice donor, we found that rs2273500-C resulted in significant reduction of splicing to exon 4.1 in favor of increased splicing to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.

Transcripts containing exon 4.1 were also observed in GTEx brain tissue samples, but relative to the liver samples, their frequencies were low (Supplementary Table 10). We observed the same pattern, whereby rs2273500-C carriers had reduced splicing to exon 4.1 and increased splicing to exon 4.2, but these differences were not statistically significant ($P = 0.30$) likely owing to the limited statistical power with the lower split read counts. However, we were able to evaluate rs2273500 as an eQTL SNP targeting for nonsense-mediated decay. To investigate the rs2273500 effect on splicing, we obtained RNA-seq and genotype data from the GTEx project. Liver ($N = 32$) showed the highest CHRNA4 expression among the GTEx tissues, and transcripts containing exon 4.1 were detected. Using split read counts (reads crossing an exon:exon boundary) to test the efficiency of splicing events that utilize the exon 4 splice donor, we found that rs2273500-C resulted in significant reduction of splicing to exon 4.1 in favor of increased splicing to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.

Transcripts containing exon 4.1 were also observed in GTEx brain tissue samples, but relative to the liver samples, their frequencies were low (Supplementary Table 10). We observed the same pattern, whereby rs2273500-C carriers had reduced splicing to exon 4.1 and increased splicing to exon 4.2, but these differences were not statistically significant ($P = 0.30$) likely owing to the limited statistical power with the lower split read counts. However, we were able to evaluate rs2273500 as an eQTL SNP targeting for nonsense-mediated decay. To investigate the rs2273500 effect on splicing, we obtained RNA-seq and genotype data from the GTEx project. Liver ($N = 32$) showed the highest CHRNA4 expression among the GTEx tissues, and transcripts containing exon 4.1 were detected. Using split read counts (reads crossing an exon:exon boundary) to test the efficiency of splicing events that utilize the exon 4 splice donor, we found that rs2273500-C resulted in significant reduction of splicing to exon 4.1 in favor of increased splicing to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.

Transcripts containing exon 4.1 were also observed in GTEx brain tissue samples, but relative to the liver samples, their frequencies were low (Supplementary Table 10). We observed the same pattern, whereby rs2273500-C carriers had reduced splicing to exon 4.1 and increased splicing to exon 4.2, but these differences were not statistically significant ($P = 0.30$) likely owing to the limited statistical power with the lower split read counts. However, we were able to evaluate rs2273500 as an eQTL SNP targeting for nonsense-mediated decay. To investigate the rs2273500 effect on splicing, we obtained RNA-seq and genotype data from the GTEx project. Liver ($N = 32$) showed the highest CHRNA4 expression among the GTEx tissues, and transcripts containing exon 4.1 were detected. Using split read counts (reads crossing an exon:exon boundary) to test the efficiency of splicing events that utilize the exon 4 splice donor, we found that rs2273500-C resulted in significant reduction of splicing to exon 4.1 in favor of increased splicing to exon 4.2 and to a cryptic splice acceptor in exon 4.1 ($P = 5.4 \times 10^{-38}$, Supplementary Table 10). These results provide strong evidence that rs2273500 functions as a splicing QTL.
**DISCUSSION**

**CHRNA4** has strong biological plausibility for influencing nicotine dependence and consequently its adverse health effects. Nicotinic acetylcholine receptor genes, including **CHRNA4**, encode subunits that assemble together to form ligand-gated ion channels that respond to the neurotransmitter acetylcholine. Nicotine exposure from cigarette smoking also activates the receptors, triggering dopamine release and influencing the reinforcing effect of nicotine.59-61 The subunits encoded by **CHRNA4** and **CHRN2** comprise αβ2 receptors, the most abundantly expressed nicotine acetylcholine receptors in the brain. They have a high affinity for nicotine and serve critical roles in nicotine self-administration and its positive reinforcement.62,63 Knock-out mouse models64,65 and knock-in mice with a hypersensitive receptor66,67 have demonstrated that **CHRNA4** is a necessary and sufficient factor for many characteristic behaviors of nicotine dependence, including nicotine-induced reward, tolerance and anxiety relief. A knock-down rat model suggested that α4-containing receptors have a role in nicotine-mediated analgesia, showing that reduced **CHRNA4** expression in brain significantly attenuated sensitivity to nicotine agonist.68 This is consistent with our finding that rs2273500-C decreases **CHRNA4** expression and, by lowering sensitivity to nicotine’s effects, confers risk for nicotine dependence. Further supporting the relevance of **CHRNA4**, highly effective treatments for smoking cessation, varenicline69,70 and CHRNA4, evidence supporting specific **CHRNA4** variant associations with nicotine dependence, which were not observed in prior studies of CPD.9

Prior studies in humans have supported **CHRNA4** as a susceptibility gene for nicotine dependence and other smoking behaviors. Genome-wide significant linkage signals have been observed in the chromosome 20q13 region containing **CHRNA4** for maximum number of cigarettes smoked in a 24-h period44 and in the nearby chromosome 20q11 region for DSM-IV-defined nicotine dependence.75 Several candidate gene association studies have focused on **CHRNA4**.76-89 Among the common SNPs with reported associations, only rs2236196 was associated in our study (P = 0.027, Table 3): its minor allele (frequency = 0.28) conferring increased risk for nicotine dependence, consistent with prior reports.86,82,83,85-87 rs2236196 shows some linkage disequilibrium with rs2273500 (D’ = 0.86 and r² = 0.50, Supplementary Figure 3). However, in a model including both SNPs, the association remained for rs2273500 (P = 2.4 × 10⁻¹⁴) but not rs2236196 (P = 0.79) in meta-analysis across our GWAS samples, indicating that our signal is distinct from the previously reported SNP association. rs2236196 was presented in one of our prior reports as having a suggestive, but nonsignificant, association with nicotine dependence (N = 1929, P = 0.081).86 By using the multidimensional FTND measure of physiological dependence paired with 1000 Genomes imputation in a large sample, the current study provides the first genome-wide level of significant evidence supporting specific **CHRNA4** variant associations with nicotine dependence, which were not observed in prior studies of CPD with larger sample sizes. Such phenotype differences have similarly been observed for the established **CHRN3** region, whereby genome-wide significant variants were identified when using FTND-defined nicotine dependence but not when using CPD.8

Rare variants in **CHRNA4** have also been implicated as contributing to nicotine dependence risk. A rare missense variant allele in exon 5, R336C (rs56175056), which lowers the sensitivity of the α4 receptor to nicotine exposure,80 has been associated with increased risks of nicotine dependence and smoking-related diseases, including lung cancer, chronic obstructive pulmonary disease, peripheral artery disease and abdominal aortic

---

**Table 3. Results of previously reported **CHRNA4** SNPs in our genome-wide association study meta-analysis of nicotine dependence**

| SNP | Minor allele MAF | Minor allele | rs2236196 | rs2273500 | rs3787137 | rs1044394 | rs1044396 | rs1044397 |
|-----|------------------|--------------|-----------|-----------|-----------|-----------|-----------|-----------|
|     |                  |              | (NCBI build 37) | (NCBI build 37) | (NCBI build 37) | (NCBI build 37) | (NCBI build 37) | (NCBI build 37) |
| rs2236196 | 0.28  | G | 0.95 | 0.14 | 0.04 | 0.94 | 0.25 | 0.92 | 0.0029 | 0.53 | 0.0075 | 0.25 | 0.92 | 0.0075 | 0.53 | 0.0075 |
| rs2273500 | 0.17 | A | 0.98 | 0.0019 | 0.68 | 0.037 | 0.52 | 0.92 | 0.0023 | 0.62 | 0.0027 | 0.49 | 0.0027 | 0.49 | 0.0027 | 0.49 | 0.0027 |
| rs3787137 | 0.45 | G | 0.0054 | 0.98 | 0.0043 | 0.98 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 |
| rs1044394 | 0.063 | A | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 |
| rs1044396 | 0.46 | G | 0.95 | 0.0054 | 0.98 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 |
| rs1044397 | 0.45 | C | 0.99 | 0.0054 | 0.98 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 |
| rs3787137 | 0.45 | G | 0.0054 | 0.98 | 0.0043 | 0.98 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 |
| rs1044394 | 0.063 | A | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 | 0.99 | 0.0020 |
| rs1044396 | 0.46 | G | 0.95 | 0.0054 | 0.98 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 |
| rs1044397 | 0.45 | C | 0.99 | 0.0054 | 0.98 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 | 0.42 | 0.0057 |

**Abbreviations:** COGEND, Collaborative Genetic Study of Addictive Disorders; EAGLE, Environment and Genetics in Lung Cancer Etiology Study; dbGenome, database of genetic variants; EAGLE, European Genome-phenome Archive; FTND, Fagerstrom Test for Nicotine Dependence; GWAS, genome-wide association study; MA, multiple and alternate; N, number; NOS, not otherwise specified; SAGE, Sandler Adolescent Genome Education; SNPs, single nucleotide polymorphisms; T, trend.
aneurysms.91 Another rare variant in exon 5, P451L (rs55915440), was nominated for its association with decreased risk of nicotine dependence92 but not independently corroborated91 rs2273500 is located 5.2 kb from R336C and 5.5 kb from P451L; neither rare variant was captured in our study owing to their frequencies. Our study of common variants identified a splice site acceptor variant allele (rs2273500-C) as being associated with (1) increased risk of nicotine dependence at genome-wide significance, (2) decreased CHRNA4 expression in human brain and (3) increased lung cancer risk likely through its effect on smoking. Future studies with a large sample of brain-specific CHRNA4 exon-specific sequences and FTND measurements are needed to validate and elucidate the splicing mechanism involving rs2273500 and its effect on nicotine dependence risk. Nonetheless, our new evidence revealing a common SNP with important regulatory features, along with a newly discovered functional rare variant,91 firmly establish CHRNA4 as an important susceptibility gene for nicotine dependence and its adverse health consequences.

CONFLICT OF INTEREST
LJB and the spouse of NLS are listed as inventors on U.S. Patent 8080371, ‘Markers for Addiction’ covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. GWR, SS, FZ, SN5, DFG, TET, and KS are members of deCODE Genetics/Amgen. Although unrelated to this research, HRK has been a consultant or advisory board member for Alkermes, Lilly, Lundbeck, Otsuka and Pfizer and is a member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, supported by AbbVie, Ethypharm, Lilly, Lundbeck and Pfizer. JKR consulted for Pfizer from 2012 to 2014 on nicotine dependence. The remaining authors declare no conflict of interest.

ACKNOWLEDGMENTS
This work was supported by the National Institutes of Health (NIH), National Institute on Drug Abuse (NIDA) grant number RO1 DA035825. Acknowledgments for the study samples used to conduct the GWAS meta-analysis, independent replication testing for nicotine dependence and follow-up association testing with lung cancer are included in the Supplementary Information.

REFERENCES

1 World Health Organization. WHO Report on the Global Tobacco Epidemic, 2009. World Health Organization: Geneva, Switzerland; 2008.

2 Sullivan PF, Kendler KS. The genetic epidemiology of smoking. Nicotine Tob Res 1999; 1: 551–557.

3 Bierut LJ, Madden PA, Breslau N, Johnson EO, Hattisumari D, Poromeier OF et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet 2007; 16: 24–35.

4 Thorgerisson 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.

5 Thorgerisson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Gudjonsson SA et al. Identification of low-frequency variants associated with gout and serum uric acid levels. Nat Genet 2011; 43: 1127–1130.

6 Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol 2010; 34: 591–602.

7 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.

8 Celik C, Gudeboom M, Sestak J, Knott J, Luan J et al. A meta-analysis of tobacco consumption variants in smoking initiation and dependence. Hum Mol Genet 2013; 22: 3964–3972.

9 Loukola A, Widensoja J, Keskitalo-Vuokko K, Brosms U, Korhonen T, Ripatti S et al. Genome-wide association study on detailed profiles of smoking behavior and nicotine dependence in a twin sample. Mol Psychiatry 2014; 19: 615–624.
49 Kutal A, Varghese V, Smith C, Walker R et al. Genetic associations between nicotinic acetylcholine receptor subunit genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15.

50 Li MD, Lou XY, Chen G, Ma JZ, Elston RC. Gene-gene interactions among CHRNA4, CHRNA5, and CHRNB4 on chromosome 15.

51 Timofeeva MN, Hung RJ, Rafnar T, Christiani DC, Field JK, Bickeboller H et al. 6p21.33 variants in BRCA2 and CHEK2 affect risk of lung cancer.

52 Devlin B, Roeder K. Genomic control for association studies.

53 Zollner S, Pritchard JK. Overcoming the winner’s curse: estimating penetrance parameters from case-control data. Am J Hum Genet 2007; 80: 605–615.

54 Yoon D, Kim YJ, Cui WY, Van der Vaart A, Cho YS, Lee JY et al. Large-scale genome-wide association study of Asian population reveals genetic factors in FRMD4A and CHRNA4 for nicotine dependence in Finns.

55 Koob GF. Drugs of abuse: anatomy, pharmacology and function of reward systems. Annu Rev Pharmacol Toxicol 2009; 49: 57–71.

56 Keskitalo-Kuokko V, Hallfors J, Broms U, Pergadia ML, Saccone SF, Loukola A et al. Chromosome 20 shows increased OSAM-N nicotine dependence in Finnish adult smokers. Tobacco Tob Res 2012; 14: 153–160.

57 Gu F, Wacholder S, Kovalchik S, Panagiotou OA, Reyes-Guzman C, Freedman ND et al. Time to smoke first morning cigarette and lung cancer in a case-control study. J Natl Cancer Inst 2014; 106: dju118.

58 Tapper AR, McKinney SL, Marks MJ, Lester HA. Nicotine responses in hypersensitive and knockout alpha4 mice account for tolerance to both hypothermia and locomotor suppression in wild-type mice. Physiol Genomics 2007; 31: 422–428.

59 Bitner RS, Nikkel AL, Curzon P, Donnelly-Roberts DL, Puttfarcken PS, Namovic M et al. Reduced nicotine receptor-mediated antinociception following in vivo antiseizure knock-down in rat. Brain Res 2000; 871: 66–74.

60 Jemalby DE, Bays HT, Rigotti NA, Azoulay S, Watsky EJ, Williams KE et al. Efficacy of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation: a randomized controlled trial. JAMA 2006; 296: 56–63.

61 Benowitz NL. Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. Annu Rev Pharmacol Toxicol 2009; 49: 57–71.

62 Barrett JC. Haploview: visualization and analysis of SNP genotype data. Cold Spring Harbor Protoc 2009; pdb ip71.

63 Flieck P, Amode MR, Barrett D, Beal K, Bills K, Brent S et al. Ensembl 2014. Nucleic Acids Res 2014; 42: D749–D755.

64 Barrett JC. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013; 45: 580–585.

65 GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multigenic tissue regulation in humans. Science 2015; 348: 648–660.

66 Ramasamy A, Trabzuni D, Gueflri S, Varghese V, Smith C, Walker R et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 2014; 17: 1418–1428.

67 Tapper AR, McKinney SL, Marks MJ, Lester HA. Nicotine responses in hypersensitive and knockout alpha4 mice account for tolerance to both hypothermia and locomotor suppression in wild-type mice. Physiol Genomics 2007; 31: 422–428.

68 Bitner RS, Nikkel AL, Curzon P, Donnelly-Roberts DL, Puttfarcken PS, Namovic M et al. Reduced nicotine receptor-mediated antinociception following in vivo antiseizure knock-down in rat. Brain Res 2000; 871: 66–74.

69 Jemalby DE, Bays HT, Rigotti NA, Azoulay S, Watsky EJ, Williams KE et al. Efficacy of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, vs placebo or sustained-release bupropion for smoking cessation: a randomized controlled trial. JAMA 2006; 296: 56–63.

70 Benowitz NL. Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. Annu Rev Pharmacol Toxicol 2009; 49: 57–71.

71 West R, Zatorowski W, Cedronia M, Lewandowska D, Pazik J, Aveyard P et al. Placebo-controlled trial of cytisine for smoking cessation. N Engl J Med 2011; 365: 1393–1400.

72 Gu F, Wacholder S, Kovalchik S, Panagiotou OA, Reyes-Guzman C, Freedman ND et al. Time to smoke first morning cigarette and lung cancer in a case-control study. J Natl Cancer Inst 2014; 106: dju118.

73 Guertin KA, Gu F, Wacholder S, Freedman ND, Panagiotou OA, Reyes-Guzman C et al. Time to first morning cigarette and risk of chronic obstructive pulmonary disease: smokers in the PLCO cancer screening trial. PLoS One 2015; 10: e0125973.

74 Han S, Gelernter J, Lu X, Yang BZ. Meta-analysis of 15 genome-wide linkage scans of smoking behavior. Biol Psychiatry 2010; 67: 12–19.

75 Keskitalo-Vuokko K, Hallofors J, Broms U, Pergadia ML, Saccone SF, Loukola A et al. Chromosome 20 shows increased OSAM-N nicotine dependence in Finnish adult smokers. Tobacco Tob Res 2012; 14: 153–160.

76 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.

77 Feng Y, Niou T, Xing H, Xu X, Chen C, Peng S et al. A common haplotype of the nicotine acetylcholine receptor alpha 4 subunit gene is associated with vulnerability to nicotine addiction in men. Am J Hum Genet 2004; 75: 112–121.

78 Li MD, Beuten J, Ma JZ, Payne TJ, Lou XY, Garcia V et al. Ethnic- and gender-specific association of the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) with nicotine dependence. Hum Mol Genet 2005; 14: 1211–1219.

79 Lou XY, Chen GB, Yan, L, Ma JZ, Zhu L, Elston RC et al. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. Am J Hum Genet 2007; 80: 1125–1137.

80 Hutchison KE, Allen DL, Filbey FM, Jepson C, Lerman C, Benowitz NL et al. CHRNA4 and tobacco dependence: from gene regulation to treatment outcome. Arch Gen Psychiatry 2007; 64: 1078–1086.

81 Li MD, Lou XY, Chen G, Ma JZ, Elston RC. Gene-gene interactions among CHRNA4, CHRNA2, BDNF, and NTR2 in nicotine dependence. Biol Psychiatry 2008; 64: 951–957.

82 Saccone NL, Saccone SF, Hinrichs AL, Stitzel JA, Duan W, Penglia MI et al. Multiple distinct risk loci for nicotine dependence identified by dense coverage of the complete family of nicotine receptor subunit (CHRN) genes. Am J Med Genet B Neuropsychiatr Genet 2009; 150B: 432–466.

83 Breitling LP, Dahmen N, Mittelstaedt D, Rujescu D, Gallinat J, Fehr C et al. Association of nicotinic acetylcholine receptor subunit 4 polymorphisms with nicotine dependence in 5000 Germans. Pharmacogenomics J 2009; 9: 219–224.

84 Han S, Yang BZ, Kranzler HR, Osin D, Anton R, Gelernter J. Association of CHRNA4 polymorphisms with smoking behavior in two populations. Am J Med Genet B Neuropsychiatr Genet 2011; 156B: 421–429.

85 Schwanke NL, Schwantes-Am TH, Wang JC, Gruca 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.

86 Wei J, Chu C, Wang Y, Yang Y, Wang Q, Li T et al. Association study of 45 candidate genes in nicotine dependence in Han Chinese. Addict Behav 2012; 37: 622–626.

87 Kamens HM, Corley RP, McQueen MB, Stallings MC, Hopfer CJ, Crowley TJ et al. Possible association of nicotinic acetylcholine receptor gene expression with smoking behavior. J Neurosci 2011; 1329–1337.

88 Chen H, Shinkai T, Utsunomiya K, Yamada K, Sakata S, Fukunaka Y et al. Possible association of nicotinic acetylcholine receptor gene (CHRNA4 and CHRN2B) polymorphisms with nicotine dependence in Japanese males: an exploratory study. Pharmacopsychiatry 2013; 46: 77–82.

89 Keskitalo-Kuokko V, Pitkaranta J, Broms U, Helloavaa M, Aromaa A, Perola M et al. Associations of nicotine intake measures with CHRN genes in Finnish smokers. Nicotine Tob Res 2011; 13: 686–690.

90 McClure-Begley TD, Papke RL, Stone KL, Stokes C, Levy AD, Gelernter J et al. Rare nicotine nicotinic acetylcholine receptor alpha4 subunit (CHRNA4) variants affect expression and function of high-affinity nicotinic acetylcholine receptors. J Pharmacol Exp Ther 2014; 348: 410–420.

Translational Psychiatry (2015), 1 – 11
91 Thorgeirsson TE, Steinberg S, Reginsson GW, Bjomsdottir G, Rafnar T, Jonsdottir I et al. A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. Submitted.

92 Xie P, Kranzler HR, Krauthammer M, Cosgrove KP, Oslin D, Anton RF et al. Rare nonsynonymous variants in alpha-4 nicotinic acetylcholine receptor gene protect against nicotine dependence. *Biol Psychiatry* 2011; 70: 528–536.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)