Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence

Hancock, D. B.; Guo, Y.; Reginsson, G. W.; Gaddis, N. C.; Lutz, S. M.; Sherva, R.; Loukola, A.; Minica, C. C.; Markunas, C. A.; Han, Y.; Young, K. A.; Gudbjartsson, D. F.; Gu, F.; McNeil, D. W.; Qaiser, B.; Glasheen, C.; Olson, S.; Landi, M. T.; Madden, P. A.F.; Farrer, L. A.

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

Link to publication in VU Research Portal

citation for published version (APA)
Hancock, D. B., Guo, Y., Reginsson, G. W., Gaddis, N. C., Lutz, S. M., Sherva, R., Loukola, A., Minica, C. C., Markunas, C. A., Han, Y., Young, K. A., Gudbjartsson, D. F., Gu, F., McNeil, D. W., Qaiser, B., Glasheen, C., Olson, S., Landi, M. T., Madden, P. A. F., ... Johnson, E. O. (2018). Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. Molecular Psychiatry, 23(9), 1911–1919. https://doi.org/10.1038/mp.2017.193

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 02. Oct. 2023
ORIGINAL ARTICLE

Genome-wide association study across European and African American ancestries identifies a SNP in *DNMT3B* contributing to nicotine dependence

DB Hancock1, Y Guo2, GW Reginsson3, NC Gaddis4, SM Lutz5, R Sherva6, A Loukola7, CC Minica8, CA Markunas9, Y Han9, KA Young10, DF Gudbjartsson3,11, FG u 12, DW McNeill13,14, B Qaiser7, C Glasheen1, S Olson15, MT Landi12, PAF Madden16, LA Farrer6,17,18,19,20, J Vink8,21, NL Saccone22, MC Neale23,24, HR Kranzler25,26, J McKay27, RJ Hung28, CI Amos9, ML Marazita29, DI Boomsma8, TB Baker30, J Gelernter31,32,33,34, J Kaprio7,35, NE Caporaso12, TE Thorgeirsson3, JE Hokanson10, LJ Bierut16, K Stefansson3 and EO Johnson36

Cigarette smoking is a leading cause of preventable mortality worldwide. Nicotine dependence, which reduces the likelihood of quitting smoking, is a heritable trait with firmly established associations with sequence variants in nicotine acetylcholine receptor genes and at other loci. To search for additional loci, we conducted a genome-wide association study (GWAS) meta-analysis of nicotine dependence, totaling 38,602 smokers (28,677 Europeans/European Americans and 9925 African Americans) across 15 studies. In this largest-ever GWAS meta-analysis for nicotine dependence and the largest-ever cross-ancestry GWAS meta-analysis for any smoking phenotype, we reconfirmed the well-known *CHRNAS-CHRNA3-CHRNB4* genes and further yielded a novel association in the DNA methyltransferase gene *DNMT3B*. The intronic *DNMT3B* rs910083-C allele (frequency = 44–77%) was associated with increased risk of nicotine dependence at \( P = 3.7 \times 10^{-8} \) (odds ratio (OR) = 1.06 and 95% confidence interval (CI) = 1.04–1.07 for severe vs mild dependence). The association was independently confirmed in the UK Biobank (\( N = 48,931 \)) using heavy vs never smoking as a proxy phenotype (\( P = 3.6 \times 10^{-4} \), OR = 1.05, and 95% CI = 1.02–1.08). Rs910083-C is also associated with increased risk of squamous cell lung carcinoma in the International Lung Cancer Consortium (\( N = 60,586 \), meta-analysis \( P = 0.0095 \), OR = 1.05, and 95% CI = 1.01–1.09). Moreover, rs910083-C was implicated as a cis-methylation quantitative trait locus (QTL) variant associated with higher *DNMT3B* methylation in fetal brain (\( N = 166, P = 2.3 \times 10^{-26} \)) and a cis-expression QTL variant associated with higher *DNMT3B* expression in adult cerebellum from the Genotype-Tissue Expression project (\( N = 103, P = 3.0 \times 10^{-4} \)) and the independent Brain eQTL Almanac (\( N = 134, P = 0.028 \)). This novel *DNMT3B* cis-acting QTL variant highlights the importance of genetically influenced regulation in brain on the risks of nicotine dependence, heavy smoking and consequent lung cancer.

Molecular Psychiatry advance online publication, 3 October 2017; doi:10.1038/mp.2017.193

INTRODUCTION

Cigarette smoking is a leading cause of preventable death, resulting annually in nearly 6 million premature deaths worldwide.1 Smoking-related deaths are most often attributed to increased rates of cancer, cardiovascular disease and chronic obstructive pulmonary disease.2 Despite the well-known adverse health effects, an estimated 45.3 million US adults smoke cigarettes, of whom over 68% report wanting to quit.3,4

1Behavioral and Urban Health Program, Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC, USA; 2Center for Genomics in Public Health and Medicine, RTI International, Research Triangle Park, NC, USA; 3deCODE Genetics/Amgen, Reykjavik, Iceland; 4Research Computing Division, RTI International, Research Triangle Park, NC, USA; 5Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; 6Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA; 7Institute for Medicinal Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 8Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands; 9Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, NH, USA; 10Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; 11Department of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland; 12Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, United States Department of Health and Human Services, Bethesda, MD, USA; 13Department of Psychology, West Virginia University, Morgantown, WV, USA; 14Department of Dental Practice and Rural Health, West Virginia University, Morgantown, WV, USA; 15Public Health Informatics Program, eHealth, Quality and Analytics Division, RTI International, Research Triangle Park, NC, USA; 16Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA; 17Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 18Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; 19Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 20Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands; 21Department of Genetics, Washington University, St. Louis, MO, USA; 22Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; 23Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; 24Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 25Crescenc VA Medical Center, Philadelphia, PA, USA; 26International Agency for Research on Cancer, World Health Organization, Lyon, France; 27Lunenfeld-Tanenbaum Research Institute, Sinai Health System, University of Toronto, Toronto, ON, Canada; 28Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, USA; 29Center for Tobacco Research and Intervention, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA; 30Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; 31Department of Genetics, Yale University School of Medicine, New Haven, CT, USA; 32Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA; 33VA CT Healthcare Center; Department of Psychiatry, West Haven, CT, USA; 34Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland and 35Fellow Program and Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC, USA. Correspondence: Dr DB Hancock, RTI International, 3040 East Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709, USA.

E-mail: dhancock@rti.org

Received 28 March 2017; revised 14 July 2017; accepted 17 July 2017
Cigarette smoking is a complex multi-step behavior involving initiation, regular smoking, nicotine dependence, cessation and relapse. Some regular smokers maintain low-level smoking without developing symptoms of dependence,\textsuperscript{5} while others become heavily dependent smokers and experience the most difficulty with cessation and the highest risk of relapse.\textsuperscript{6,7} Nicotine dependence has high heritability (estimates up to 75%),\textsuperscript{8,9} and besides reducing the likelihood of quitting smoking, it is predictive of withdrawal severity,\textsuperscript{10} response to treatment\textsuperscript{11} and smoking-related health outcomes.\textsuperscript{12,13}

Genome-wide association study (GWAS) analyses of nicotine dependence phenotypes\textsuperscript{14–24} have firmly established associations with several loci, including nicotinic acetylcholine receptor genes on chromosomes 15q25 (CHRNA5-CHRNA3-CHRNB4)\textsuperscript{18} and 20q13 (CHRNA4). The largest GWAS meta-analyses relied on widely ascertained phenotypes such as cigarettes per day (CPD),\textsuperscript{16–18} which represents only one of several components of nicotine dependence.\textsuperscript{25} Focusing GWAS on nicotine dependence rather than CPD may improve statistical power for identifying variants that influence the broader construct of dependence.\textsuperscript{19} This idea is supported by our prior nicotine dependence GWAS meta-analysis (total \(N = 17,074\) ever-smokers of European/European American ancestry (EUR)) that discovered associations with CHRNA4 single-nucleotide polymorphisms (SNPs) that were driven by time to first cigarette in the morning (TTFC) and had not been detected in GWAS meta-analyses of CPD with much larger sample sizes.\textsuperscript{23} To improve statistical power further and to search for additional loci, we more than doubled our sample size to perform the largest GWAS meta-analysis of nicotine dependence to date, including 38,602 ever-smokers (28,677 of EUR and 9925 of African American (AA) ancestries) across 15 studies. We extended our study to include correlates with DNA methylation (DNAm) and RNA expression (RNAexp) of nearby genes across human brain tissues and evaluated associations with a critical smoking-related outcome: lung cancer (total \(N = 81,821\) cases and controls).

**MATERIALS AND METHODS**

Study protocols received institutional review board approval at their respective sites. All study participants provided written informed consent. We included the five studies from our prior GWAS meta-analysis\textsuperscript{23} and 10 additional studies. Details of their study design, genotyping, quality control, 1000 Genomes (1000G) imputation and analysis are provided in Supplementary Methods and Supplementary Table 1.

Nicotine dependence phenotype

We included studies with SNP genotypes and Fagerström Test for Nicotine Dependence (FTND) data\textsuperscript{26} collected among smokers. FTND scores range from 0 (no dependence) to 10 (highest dependence level). As before,\textsuperscript{23} we used FTND to categorize nicotine dependence as mild (scores 0–3), moderate (scores 4–6) or severe (scores 7–10). Two of the 15 studies additionally included low-intensity smokers who reported CPD as \(\leq 10\) but had no data available on the other FTND items and were defined as mildly dependent. Concordance rates between these FTND and CPD categories showed minimal phenotype misclassification (Supplementary Methods).

Nicotine dependence GWAS meta-analysis and independent follow-up

We used linear regression to test SNP/indel associations with categorical nicotine dependence (mild = 0/moderate = 1/severe = 2) in each separate study and ancestry group. Covariates included age, sex, principal components and study-specific covariates (as needed); additional adjustment for family structure was made in studies with relatives included (Supplementary Methods).

We combined GWAS results, using METAL\textsuperscript{27} with fixed-effects inverse variance-weighting meta-analysis, across all studies with FTND data to maximize statistical power. Genomic control was applied to the deCODE results to adjust for inflation due to relatedness among participants; all other studies had low inflation values (\(\lambda < 1.02\)). We excluded SNPs/indels with minor allele frequency \(< 1\%\) in 1000G EUR or African (AFR) panels, depending on the ancestry group analyzed. The standard threshold \((P < 5 \times 10^{-8}\)) originally based on 1 million independent tests genome-wide as computed using HapMap-based imputation for EUR studies,\textsuperscript{28} has been validated for 1000G-imputed GWAS of common variants.\textsuperscript{29} Rather than imposing a more stringent, yet to be consistently determined, threshold when analyzing common variants across EUR and AA studies,\textsuperscript{29–32} we carried forward novel variation implicated at \(P < 5 \times 10^{-8}\) and relied on confirmation in an independent study to declare genome-wide significance. For this confirmation step, we utilized UK Biobank (\(N = 48,931\) EUR participants) results with heavy, defined as pack-years (ICPD/20 cigarettes per pack × years smoked) \(\geq 10\), vs never smoking as a proxy phenotype.\textsuperscript{30} This prior GWAS was designed as a nested case-control analysis that sampled the extremes of smoking habit, and thus did not encompass light smoking.

Regional association plots were created using LocusZoom\textsuperscript{33} with linkage disequilibrium (LD) estimates of \(r^2\) and \(D^\prime\) based on 1000G EUR and AFR panels. Allele frequencies were weighted by sample size. Odds ratio (OR) estimates were computed using the \(\beta\) estimate from the SNP term in the linear regression model \((\hat{g} \times \text{rank})\) for severe vs mild dependence) and then compared across studies using the Forest Plot Viewer.\textsuperscript{34} Heterogeneity across studies was assessed using the \(I^2\) index.\textsuperscript{35}

**RESULTS**

We performed GWAS analyses across 15 studies, totaling 38,602 (28,677 EUR and 9925 AA) ever-smokers (Supplementary Table 2), with nicotine dependence defined as mild (\(N = 17,796\); 46.1%), moderate (\(N = 13,527\); 35.0%) or severe (\(N = 7279\); 18.9%). More than 99% of the participants were \(\geq 18\) years old. Males constituted 53.2% of the total sample size.

Our GWAS meta-analysis tested nearly 18 million genotyped and 1000G-imputed SNPs/indels for association with mild/
DNMT3B SNP associations with nicotine dependence
Rs910083, an intronic DNMT3B SNP, was identified across both ancestries: meta-analysis $P = 3.7 \times 10^{-8}$ and $\beta$ (standard error) = 0.032 (0.0057) for the C allele (Table 1), corresponding to OR (95% confidence interval (CI)) of 1.06 (1.04–1.07) for severe vs mild dependence (Supplementary Figure 3). Rs910083 was imputed well (quality scores = 0.98–1), and it showed no evidence of heterogeneity across studies ($P = 0.71$).

Rs910083-C, the minor allele for EUR (frequency = 44%) but the major allele for AA (frequency = 77%), is associated with increased nicotine dependence risk (Table 1): EUR-specific meta-analysis $P = 4.1 \times 10^{-8}$ and OR (95% CI) = 1.06 (1.03–1.08); and AA-specific meta-analysis $P = 7.3 \times 10^{-5}$ and OR (95% CI) = 1.10 (1.05–1.15). Many SNPs/indels were in moderate to high LD ($r^2 > 0.4$) with rs910083 in 1000G EUR, spanning 220 kb (chr20:31,268,924–31,488,466) and including DNMT3B and its neighboring genes, microtubule-associated protein, RP/EB family, member 1 (MAPRE1) and COMMD domain containing 7 (COMMD7). However, in 1000G AFR, SNPs in LD ($r^2 > 0.4$) with rs910083 were localized to a 47 kb region (chr20:31,356,560–31,403,394) including only DNMT3B (Figures 2A–B, NCBI build 37 positions).

No DNMT3B variants have been implicated previously for any substance use disorder (SUD) phenotype. Upstream of DNMT3B, chromosome 20q11 also harbors the nucleolar protein 4-like (NOL4L) gene, which was reported at genome-wide significance for heavy vs never smoking in the UK Biobank for the indel rs57342388.\textsuperscript{24} This indel was associated at meta-analysis $P = 0.0017$ in our study (Table 1): OR (95% CI) = 1.04 (1.02–1.07) for severe vs mild dependence for the insertion allele, consistent with the prior result. Rs57342388 is located 216 kb upstream of our top DNMT3B SNP rs910083. The two variants are weakly correlated ($r^2 = 0.11$ in 1000G EUR where minor allele frequency = 2% for rs57342388 vs 18% for rs910083), $r^2 = 0.0022$ in AFR where minor allele frequency = 19% vs 42%) but are in moderate to high LD ($D^\prime = 0.57$ in EUR, $D^\prime = 1$ in AFR). In follow-up testing with both SNPs included in the same model, both were associated with nicotine dependence (meta-analysis $P = 1.7 \times 10^{-6}$ for rs910083 and $8.3 \times 10^{-3}$ for rs57342388), showing that our observed DNMT3B association signal is not explained by the previously reported NOL4L signal.

We tested rs910083-C for association with each of the specific FTND items, as presented in Supplementary Table 3. The rs910083-C association was driven most strongly by TTFC (meta-analysis $P = 1.2 \times 10^{-8}$). Its next most significantly associated FTND item was CPD (meta-analysis $P = 0.0011$). TTFC is an indicator of withdrawal severity upon awakening\textsuperscript{40} and behavioral automaticity (habitual smoking without awareness or cognitive control)\textsuperscript{7}. Although TTFC has its distinct features (for example, strongest
Nicotine dependence GWAS identifies DNMT3B  

DB Hancock et al

DNMT3B SNP association with heavy smoking in an independent study

Using heavy vs never smoking GWAS results from the UK Biobank (N = 24,457 heavy and 24,474 never smokers), we found that rs910083-C is associated at P = 3.6 × 10^{-4} and OR (95% CI) = 1.05 (1.02–1.08) for heavier smoking. Although effect sizes were small, its associations with risks of nicotine dependence and heavy smoking were consistent (Table 1).

DNMT3B SNP associations with DNAm and RNAexp

Using a genome-wide meQTL study of 166 fetal brain samples, we found that rs910083-C associates with higher DNAm levels of the probe cg136366640, located 252 base pairs upstream of the DNMT3B gene (β = 0.082, P = 2.3 × 10^{-24}). This probe was the only one observed in the region with significant mQTL variants (Supplementary Figure 4). There were a total of 140 SNPs associated with DNAm of this probe (smallest P = 1.8 × 10^{-44}), and we observed associations of these SNPs with nicotine dependence with meta-analysis P-values ranging from 3.66 × 10^{-8} for rs910083 to 0.0051 (Supplementary Table 4). To our knowledge, there are no similar data to assess cis-meQTL effects in brain tissues from adults with no psychiatric disease. However, we assessed the cis-acting effect of rs910083 on RNAexp across several adult brain tissues using two independent data sets, GTEx37 followed by Brain eQTL Almanac.38 Across the 13 brain tissues in GTEx, we observed the highest DNMT3B gene expression levels in cerebellum (median log_{10} RPKM = 0.304) and cerebellar hemisphere (median log_{10} RPKM = 0.298), compared to median log_{10} RPKM ≤ 0.325 for all other brain tissues (Supplementary Figure 5). Moreover, across the brain tissues, rs910083 was most significantly associated with DNMT3B gene expression levels in cerebellum (P = 3.0 × 10^{-5}) and cerebellar hemisphere (P = 7.0 × 10^{-7}), with the C allele associated with higher DNMT3B expression (Figure 3 and Supplementary Table 5). We replicated this pattern in the Brain eQTL Almanac, where DNMT3B mRNA transcript expression levels were highest in cerebellar cortex (Supplementary Figure 6)—the outer layer of the cerebellum that comprises most of its volume. Consistent with GTEx, rs910083-C was associated with increased DNMT3B mRNA transcript expression levels specifically in cerebellar cortex (P = 0.028).

Beyond the brain tissues in GTEx, rs910083 is associated with RNAexp of other genes within 1 MB (Supplementary Table 6): MAPRE1 (smallest P = 7.3 × 10^{-17} in sun exposed skin), COMMD7 (smallest P = 1.0 × 10^{-6} in colon) and BPI fold containing family B, member 2 (BPIFB4, smallest P = 5.4 × 10^{-5} in artery).

DNMT3B SNP associations with lung cancer

We assessed rs910083 for association with lung cancer using a GWAS meta-analysis of EUR studies from TRICL-LCCO39 (N = 27,349 cases and 54,472 controls, Supplementary Table 7). Rs910083-C is significantly associated with increased risk of squamous cell carcinoma (N = 6937 cases and 53,649 controls, meta-analysis P = 0.0095 and OR [95% CI] = 1.05 [1.01–1.09]), consistent with the increased nicotine dependence risk. Rs910083 is not associated with adenocarcinoma.

We evaluated the effect of adjusting for smoking on the rs910083 association with squamous cell carcinoma in the studies with smoking data readily available. We found that the level of statistical significance and magnitude of association were both weakened with adjustment for ever/never smoking and pack-years (Supplementary Table 8), suggesting that the rs910083 predictor of cessation success among the FTND items), it is correlated with heaviness of smoking as captured by CPD.7

Table 1. SNP associations in the novel DNMT3B gene and previously reported genes that were identified at P < 5 × 10^{-7} in the nicotine dependence GWAS meta-analysis

| Chr | Gene/variant | Prior GWAS reported smoking phenotype | UK Biobank heavy vs never smokers (N = 48,931) | EUR and NA studies (N = 38,662) |
|-----|--------------|--------------------------------------|---------------------------------|---------------------------------|
|     |              |                                      |                                |                                 |
| 20q11 | DNMT3B       |                                      |                                |                                 |
| 20q11.3 | NOL4L     |                                      |                                |                                 |
| 20q13 | CHRNA4      |                                      |                                |                                 |
| 17 | DBH        |                                      |                                |                                 |
| 2q14.1 | RP4        |                                      |                                |                                 |

Abbreviations: AA, African American; EUR, European/European American; GWAS, genome-wide association study; LD, linkage disequilibrium; SE, standard error; SNP, single-nucleotide polymorphism. The SNP with the smallest nicotine dependence GWAS meta-analysis P is shown, along with nearby SNPs previously reported in other GWAS of smoking phenotypes. Results from the UK Biobank heavy vs never smoking GWAS are also presented. NA, not available due to minor allele frequency < 0.02 where they exist in moderate LD (r^2 > 0.64 and D' > 0.95 in 1000G EUR). Rs6062901 had r^2 = 0.047 (0.0065) 7.0 × 10^{-6} and D' = 0.73–0.84 with rs2273500 and rs11697662 in 1000G EUR and r^2 = 0.47–0.51 and D' = 0.92–0.96 in 1000G AFR. Rs56116178 and rs111280114 had r^2 = 0.98 and D' = 0.72–0.73 and rs3025343 in 1000G EUR. Using a genome-wide meQTL study of 166 fetal brain samples,36 DNMT3B smoking were consistent (Table 1).
Figure 2. Novel DNMT3B SNP associations with nicotine dependence from GWAS meta-analysis of EUR and AA studies. SNP and indel associations are shown across DNMT3B and its 100 kb flanking region (NCBI build 37 positions presented). \( r^2 \) values between the top SNP rs910083 and all other SNPs are shown in reference to 1000 Genomes panels: (a) EUR and (b) AFR. Indels with missing \( r^2 \) values are indicated in gray. The \( P \)-value threshold of \( 5 \times 10^{-8} \) is marked by the solid black line. AA, African American; EUR, European/European American; GWAS, genome-wide association study; SNP, single-nucleotide polymorphism.
DISCUSSION

This largest-ever GWAS meta-analysis for nicotine dependence, and the largest-ever cross-ancestry GWAS meta-analysis for any smoking phenotype, identified rs910083 as a novel SNP that regulates DNMT3B in human brain and contributes to risks of nicotine dependence and heavy smoking. Rs910083 was discovered via meta-analysis of two ancestry groups; the association signal includes SNPs in high LD with rs910083 across the COMMD7, DNMT3B and MAPRÉ1 genes in EUR ancestry, but LD is localized to the DNMT3B gene in AA ancestry. Moreover, rs910083 was implicated as a cis-acting QTL SNP that influences DNMT3B DNAm in fetal brain and DNMT3B RNAexp in adult cerebellum, with the C allele being associated with higher DNAm and RNAexp levels. While this pattern might contrast the traditional view of higher DNAm being correlated with lower RNAexp, the observed effects reflect temporal and spatial differences, thus limiting our ability to draw direct correlations. Nonetheless, genome-wide QTL comparisons in human brain have shown that almost half of SNPs that act as both an eQTL and meQTL show the same direction of association for DNAm and RNAexp,41 as we observed here for DNMT3B and before for CHRNA5.42 The previously established nicotinic acetylcholine receptor genes also harbor nicotine dependence-associated SNPs with important consequences for gene regulation, including noncoding SNPs that correlate with DNAm,43 splicing,42 and/or RNAexp43,44 in brain tissues that are frequently studied for nicotine and other SUDs because of their role in primary reward pathways and executive function, such as prefrontal cortex.45 This newly identified DNMT3B SNP association highlights changes in DNAm in fetal brain and with RNAexp specifically in cerebellum, a part of the brain that has often been overlooked despite some indications for its involvement in the neurobiology of addiction.46–48

DNMT3B encodes a DNA methyltransferase prominently involved in de novo DNAm that establishes patterns early in development; it may also contribute to maintenance DNAm.49 Although initially reported for maintaining DNAm at imprinted loci,50 later evidence showed that other members of the DNMT family (DNMT3A and DNMT3L) are required for imprinting, while DNMT3B may not play an essential role.51 Mouse models have shown that complete loss of DNMT3B function is embryonically lethal.50 However, recessive inheritance of rare mutations that render DNMT3B partially functional are known to cause Immunodeficiency, Centromeric instability and Facial dysmorphism syndrome, which manifests with growth and neurodevelopmental abnormalities.49 Because of the critical role that DNMT3B plays in establishing methylation, altered expression has been associated with Immunodeficiency, Centromeric instability and Facial dysmorphism syndrome at >700 genes involved in brain development and other processes.52 These and other genes that are regulated by DNMT3B methylation represent candidate genes that may directly contribute to nicotine dependence susceptibility. DNMT3B has not previously been connected with the biology underlying the risk of nicotine dependence or any other SUD. However, there is evidence that in vitro cigarette smoke exposure leads to increased DNMT3B expression in human respiratory epithelial cells, and DNMT3B overexpression results in downstream hypermethylation that has been widely implicated in lung cancer.53 Given this indication, DNMT inhibition has been an active area of research for cancer treatment; two inhibitor agents are currently approved by the US Food and Drug Administration (decitabine, which shows high affinity for DNMT3A/3B over DNMT1,54 and azacytidine), and at least one other promising agent (zebularine) awaits clinical trial testing.55 With the DNMT3B variant discovery for nicotine dependence, DNMT3B inhibition...
may merit future study for smoking cessation treatment. Rs910083 resides in an active promoter marked by H3K9ac in several brain regions examined in the Roadmap Epigenomics Project.\textsuperscript{36} and it associates with DNMT3B RNAexp specifically in the cerebellum. This finding does not negate the importance of other brain regions known to be involved in SUDs but brings up the possibility of altered gene regulation in the cerebellum contributing to the complex neurobiological pathway leading to dependence. A main function of the cerebellum is motor coordination, but it is also involved in non-motor functions relevant to SUD, including reward.\textsuperscript{46–48} The cerebellum responds to acute and long-term exposures to nicotine\textsuperscript{57–62} and other substances,\textsuperscript{47} and it makes functional connections with the prefrontal cortex and other brain tissues that are widely recognized for their involvement in SUDs.\textsuperscript{47}

Our discovery of the rs910083 association with nicotine dependence was made possible by assembling the largest possible sample size of FTND studies comprised of ever-smokers to maximize statistical power. No large FTND studies were left for replication, but because CPD is a central component of both the FTND and the heavy vs never smoking definition used in the UK Biobank, we tested for independent confirmation of the novel association with heavy vs never smoking and found that rs910083-C also conferred risk in the UK Biobank.

Larger GWAS meta-analyses have been reported for other smoking phenotypes such as ever vs never smoking, but these studies were comprised only of EUR participants and based on HapMap imputation.\textsuperscript{16–18} Rs910083, a 1000G-imputed SNP, was not captured in these studies, but 9 of the 18 DNMT3B SNPs associated with nicotine dependence at meta-analysis \( P < 5 \times 10^{-7} \) in the present study were HapMap-imputed (Supplementary Table 9); these SNPs were in strong LD with rs910083 among EURs \( (r^2 = 0.78–0.99 \text{ in } 1000G \text{ EUR}) \) but weaker LD among AAs \( (r^2 = 0.29–0.76 \text{ in } 1000G \text{ AFR}) \). Using results from the largest GWAS meta-analysis of CPD (Tobacco and Genetics (TAG) consortium, \( N = 38,181 \text{ EUR ever smokers independent of the ones included here} \)), we found that the nine HapMap-imputed DNMT3B SNPs were associated with CPD at \( P \)-values ranging from \( 0.027 \) to \( 0.059 \) and a consistent direction of association with nicotine dependence; in comparison, \( P \)-values for ever vs never smoking \( (N = 74,035 \text{ in the TAG consortium}) \) ranged from \( 0.049 \) to \( 0.34 \) (Supplementary Table 9). We caution that the best DNMT3B signal in the TAG consortium was observed for CPD at only nominal significance \( (\text{smallest } P = 0.027) \), despite having a nearly equivalent sample size as our study. However, our study yielding more statistically significant DNMT3B SNP associations with nicotine dependence \( (\text{smallest } P = 3.7 \times 10^{-8}) \) is likely due to a combination of factors, including (1) reliance on FTND, a multi-dimensional phenotype that encompasses CPD and other important features of smoking behavior including TTF that drove the rs910083 association, and (2) 1000G imputation which has been shown to strengthen association signals for some loci due to the finer mapping available.\textsuperscript{53,64} We have similarly observed more statistically significant associations with nicotine dependence and stronger effect sizes, compared to CPD, in prior studies of CHRNA4 and CHRNA5,\textsuperscript{43,44}

Until now, the only common DNMT3B variant implicated by GWAS was identified in a study of inflammatory bowel disease (intrinsic rs4911259).\textsuperscript{65} This SNP was associated with nicotine dependence in our meta-analysis \( (P = 3.5 \times 10^{-7}) \) and is in LD with our top SNP rs910083 \( (r^2 = 0.76 \text{ and } D' = 1 \text{ in } 1000G \text{ EUR}, r^2 = 0.16 \text{ and } D' = 1 \text{ in } 1000G \text{ AFR}) \). Cigarette smoking is the environmental factor most consistently associated with inflammatory bowel disease,\textsuperscript{66} and these shared SNP association signals suggest that DNMT3B SNPs may exert pleiotropic effects. Alternatively, it is possible that smoking mediates the DNMT3B SNP association with inflammatory bowel disease, but the inflammatory bowel disease GWAS did not include adjustment for smoking.\textsuperscript{65}

Beyond finding DNMT3B and (as expected) CHRNA5-CHRNA3-CHRNA4, our GWAS meta-analysis resulted in \( P < 5 \times 10^{-7} \) for two other loci previously implicated in smoking, CHRNA4 on chromosome 20q13\textsuperscript{23,24} and dopamine \( \beta \)-hydroxylase (DBH) on chromosome 9q34.\textsuperscript{47,74} Our prior nicotine dependence GWAS meta-analysis of EUR studies identified the CHRNA4 splice site SNP rs2773500.\textsuperscript{23} Rare CHRNA4 variants have also been found to associate with nicotine dependence.\textsuperscript{67,68} Our study supported common CHRNA4 SNP associations among EUR samples, but no association was detected for these SNPs among AA studies (Table 1).

DBH is a strong functional candidate for influencing nicotine dependence. The dopaminergic system lies at the core of the brain's reward pathway, and the DBH enzyme converts dopamine into norepinephrine. An upstream DBH SNP (rs32053543) was identified in a GWAS of smoking cessation (current vs former smokers)\textsuperscript{16–18} and later independently replicated.\textsuperscript{69,70} Consistent with rs3205343-A being associated with reduced success of quitting smoking, its phenotypic profile has been expanded to include associations with: (1) heavier smoking \( (N = 48,931 \text{ in the UK Biobank}, P = 1.2 \times 10^{-5}) \); (2) higher FTND scores \( (N = 1430 \text{ EUR participants, } P = 0.023) \); and (3) higher nicotine dependence risk in our EUR studies \( (N = 28,677, \text{meta-analysis } P = 1.7 \times 10^{-7}) \). Smaller \( P \)-values were found for other 1000G-imputed upstream DBH SNPs in the UK Biobank and our study \((rs111280114 \text{ and rs56116178, respectively; Table 1)} \); the minor alleles of these SNPs were similarly associated with increased risks among EUR studies (Supplementary Figure 7 for rs56116178). Because these DBH SNP all occur at <1% frequency among AAs, studying DBH variation on nicotine dependence risk in this ancestry group will require larger sample sizes or an alternative study design.

Nicotine dependence-associated variants in CHRNA5-CHRNA3-CHRNA4\textsuperscript{45,72} and CHRNA4\textsuperscript{46,51} have been previously shown to associate with lung cancer and other smoking-related diseases. Our study shows that the nicotine dependence-associated SNPs in DNMT3B and DBH are also associated with lung cancer (Supplementary Table 7). These findings may reflect the SNPs acting indirectly on lung through their influence on smoking (Supplementary Table 8). Alternatively, because DBH is expressed in the lung\textsuperscript{73} and DNMT3B overexpression has been shown in lung cancer, we cannot exclude the possibility that either of these SNPs act directly to promote lung cancer through an unknown mechanism.\textsuperscript{53} The DNMT3B and DBH SNPs were both associated with squamous cell lung carcinoma. This histological subtype has a strong association with smoking and occurs infrequently in never-smokers. In contrast, neither SNP is associated with adenocarcinoma, a subtype that has a weaker association with smoking\textsuperscript{74} and an increasing prevalence over time among never-smokers.\textsuperscript{75} Histology-specific associations are not uncommon for lung cancer genetic loci.\textsuperscript{39}

Our findings expand the known genetic architecture of nicotine dependence by showing that the DNMT3B SNP rs910083 increases the likelihood of developing nicotine dependence as observed across two different ancestries, smoking heavily, and consequently incurring a heightened risk of lung cancer.\textsuperscript{53,55} The convergence of prior and current findings indicate that the complex neurobiology underlying nicotine dependence involves several sequence variants with functional and regulatory effects across distinct brain tissues.
ACKNOWLEDGMENTS

We thank the many study participants. We also thank Michael E. Hall for reviewing the manuscript. This work was supported by the National Institute on Drug Abuse grant numbers R01 DA035825, R01 DA036583 and R01 DA042090. Acknowledgements for the nicotine dependence studies are included in the Supplementary Information. Funding for lung cancer studies was provided by the National Cancer Institute grant number U19 CA148127.

REFERENCES

1 World Health Organization. WHO report on the global tobacco epidemic, 2011: Warning about the dangers of tobacco. World Health Organization: Geneva, Switzerland, 2011.
2 U.S. Department of Health and Human Services. The health consequences of smoking—50 years of progress: a Report of the Surgeon General. U.S. Department of Health and Human Services: Atlanta, GA, 2014.
3 Centers for Disease Control and Prevention (CDC). Vital signs: current cigarette smoking among adults aged >18 years—United States, 2005–2010. MMWR Morb Mortal Wkly Rep 2011; 60: 1207–1212.
4 Centers for Disease Control and Prevention (CDC). Quitting smoking among adults—United States, 2001–2010. MMWR Morb Mortal Wkly Rep 2011; 60: 1513–1519.
5 Shiffman S, Tindle H, Li X, Scholl S, Dunbar M, Mitchell-Milard C. Characteristics and smoking patterns of intermittent smokers. Exp Clin Psychopharmacol 2012; 20: 264–277.
6 Breslau N, Johnson ED. Predicting smoking cessation and major depression in nicotine-dependent smokers. Am J Public Health 2000; 90: 1122–1127.
7 Baker TB, Piper ME, Schlam TR, Cook JW, Smith SS, Loh WT et al. Are tobacco dependence and withdrawal related amongst heavy smokers? Relevance to conceptualizations of dependence. J Abnorm Psychol 2012; 121: 909–921.
8 Zelman DC, Brandon TH, Jorenby DE, Baker BA. Measures of affect and nicotine dependence predict differential response to smoking cessation treatments. J Consult Clin Psychol 1992; 60: 943–952.
9 Gu F, Wacholder S, Kovalich S, Panagiotou OA, Reyes-Guzman C, Freedman ND et al. Time to smoke first cigarette in the morning as an index of ability to quit smoking: implications for nicotine dependence. Nicotine Tob Res 2007; 9(Suppl 4): S555–S570.
10 Sullivan PF, Kendler KS. The genetic epidemiology of smoking. Nicotine Tob Res 1999; 1: 551–557.
11 Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behav Genet 2005; 35: 397–406.
12 Baker TB, Piper ME, Schlam TR, Cook JW, Smith SS, Loh WT et al. Are tobacco dependence and withdrawal related amongst heavy smokers? Relevance to conceptualizations of dependence. J Abnorm Psychol 2012; 121: 909–921.
13 Zelman DC, Brandon TH, Jorenby DE, Baker BA. Measures of affect and nicotine dependence predict differential response to smoking cessation treatments. J Consult Clin Psychol 1992; 60: 943–952.
14 Gu F, Wacholder S, Kovalich S, Panagiotou OA, Reyes-Guzman C, Freedman ND et al. Time to smoke first cigarette in the morning and lung cancer in a case-control study. J Natl Cancer Inst 2014; 106: dju118.
15 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 2013; 8: e612597.
16 Bierut LJ, Madden PA, Breslau N, Johnson EO, Hattuck S, Pommerleau OF et al. Novel genes identified in a high-density genome-wide association study for nicotine dependence. Hum Mol Genet 2007; 16: 24–35.
17 Thorgerisson TE, Geller F et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008; 452: 638–642.
18 Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton LE et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 2010; 42: 436–440.
19 Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010; 42: 441–447.
20 Thorgerisson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 2010; 42: 448–453.
21 Rice JP, Hartz SM, Agrawal A, Malysa L, Bennett S, Breslau N et al. CHRNB3 is more strongly associated with Fagerstrom Test for Cigarette Dependence-based nicotine dependence than cigarettes per day: phenotype definition changes genome-wide association studies results. Addiction 2012; 107: 2019–2028.
Nicotine dependence GWAS identifies DNMT3B
DB Hancock et al

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