Nicotine & Tobacco Research, 2019, 707–713
doi:10.1093/ntr/ntx262
Original investigation
Received June 23, 2017; Editorial Decision November 21, 2017; Accepted November 28, 2017

Original investigation

An Exome-Wide Association Study Identifies New Susceptibility Loci for Age of Smoking Initiation in African- and European-American Populations

Keran Jiang, PhD,[1] Zhongli Yang, PhD,[1] Wenyan Cui, PhD,[1] Kunkai Su, PhD,[1] Jennie Z. Ma, PhD,[2] Thomas J. Payne, PhD,[3] Ming D. Li, PhD[1,4,5]

[1]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University School of Medicine, Hangzhou, China; [2]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA; [3]ACT Center for Tobacco Treatment, Education and Research, Department of Otolaryngology and Communicative Sciences, University of Mississippi Medical Center, Jackson, MS; [4]Research Center for Air Pollution and Health, Zhejiang University, Hangzhou, China; [5]Institute of Neuroimmune Pharmacology, Seton Hall University, South Orange, NJ, USA

Corresponding Author: Ming D. Li, PhD, State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China and Institute of Neuroimmune Pharmacology, Seton Hall University, South Orange, NJ, USA. E-mail: ml2km@zju.edu and limd586@outlook.com

Abstract

Introduction: Cigarette smoking is one of the largest causes of preventable death worldwide. This study aimed to identify susceptibility loci for age at smoking initiation (ASI) by performing an exome-wide association analysis.

Methods: A total of 2510 smokers of either African-American (AA) or European-American (EA) origin were genotyped and analyzed at both the single nucleotide polymorphism (SNP) and gene levels. After removal of those SNPs with a minor allele frequency (<0.01), 48 091 and 34 933 SNPs for AAs and EAs, respectively, were used to conduct a SNP-based association analysis. Gene-based analyses were then performed for all SNPs examined within each gene. Further, we estimated the proportion of variance explained by all common SNPs included in the analysis.

Results: The strongest signals were detected for SNPs rs17849904 in the pitrilysin metallopeptidase 1 gene (*PITRM1*) in the AA sample ($p = 9.02 \times 10^{-7}$) and rs34722354 in the discoidin domain of the receptor tyrosine kinase 2 gene (*DDR2*) in the EA sample ($p = 9.74 \times 10^{-7}$). Both SNPs remained significant after Bonferroni correction for the number of SNPs tested. Subsequently, the gene-based association analysis revealed a significantly associated gene, *DHRS7*, in the AA sample ($p = 5.00 \times 10^{-6}$), a gene previously implicated in nicotine metabolism.

Conclusions: Our study revealed two susceptibility loci for age of smoking initiation in the two ethnic samples, with the first being *PITRM*1 for AA smokers and the second *DDR2* for EA smokers. In addition, we found *DHRS7* to be a plausible candidate for ASI in the AA sample from our gene-based association analysis.

Implications: *PITRM1* and *DHRS7* for African-American smokers and *DDR2* for European-American smokers are new candidate genes for smoking initiation. These genes represent new additions to smoking initiation, an important but less studied phenotype in nicotine dependence research.

Introduction

Tobacco is the most widely used addictive substance and the leading preventable cause of diseases, disability, and deaths throughout the world. Tobacco smoking has been associated with a higher risk of several cancers and cardiovascular and respiratory diseases.[1] According to the recent World Health Organization survey, more than 1.3 billion persons in the world smoke, and smoking-related diseases are responsible for approximately 6 million deaths each year worldwide, a number predicted to increase to 8.3 million by 2030.[2]

Smoking behaviors, including age at smoking initiation (ASI), smoking dependence (SD), and smoking cessation (SC), are all complex phenotypes determined by both genetic and environmental factors as well as their interactions.[3] Twin, family, and adoption studies have indicated that genetic factors play a significant role in smoking initiation (SI), dependence, and cessation.[3–5] Previous studies have estimated the heritability of different smoking phenotypes with a range of 21% to 84%.[6–8]

To identify susceptibility loci for each smoking phenotype, numerous studies have been conducted, with approaches including genome-wide linkage scans,[9] candidate gene-based association analyses, and the genome-wide association study (GWAS). Of those loci identified by various approaches, especially with GWAS, the most robust findings are for the variants in the *CHRNA5*/*CHRNA3*/*CHRNB4* cluster on chromosome 15q25.1,[10–12] such as rs16969968 and rs1051730, which are significantly associated with the number of cigarettes smoked per day (CPD) and the Fagerström Test for Nicotine Dependence (FTND) score. The evidence supporting the involvement of nAChRs in the etiology of ND is indisputable, in part, because of their essential role in mediating the rewarding effects of nicotine.[13] Further, *CHRNA6/CHRNB3* on chromosome 8p11[12] and *CYP2A6* on chromosome 19q13[12] are consistently associated with CPD at the genome-wide significance level.[12] Previous studies have implicated *CHRNA6/CHRNB3* subunits in nicotine-induced dopamine release,[14] and variants in *CYP2A6* reduce the enzymatic activity of CYP2A6.[15] For other smoking behaviors, GWAS revealed that *BDNF* on chromosome 11p13 and *RGS17* on chromosome 6q25 are significantly associated with SI[16,17] and *DBH* on chromosome 9q34 and *PARD3* on chromosome 10p11 with SC in multiracial populations.[16,18] Of these important SNPs, less than 5% of the variance can be explained by those variants for each phenotype of interest. To explore the missing heritability, one common approach is to detect multiple alleles in the same gene region that affect the same trait. Gene-based association testing can improve statistical power in the presence of allelic heterogeneity by combining single variants from GWAS into a gene-based score.[19] In this study, we employed the "Versatile Gene-Based Association Study 2" (VEGAS2) approach[20] to assign variants to genes and calculate gene-based *p* values based on computer simulation.

The majority of reported GWAS for smoking phenotypes were conducted in the people of European ancestry.[21] However,[22] African-Americans generally initiate smoking later and smoke fewer cigarettes per day, but they are less likely to be able to quit smoking and have a higher risk of smoking-related lung cancer than European-Americans. Therefore, to conduct GWAS for smoking behavior, especially for ASI in African populations, is greatly needed.

Nearly 90% of adult smokers begin their smoking before or at the age of 18, and, at present, one-fourth of young adults are smokers.[23] Although the majority of smokers try to quit, smoking cessation cannot be achieved easily, primarily because of the addictive physiological and psychological properties of nicotine. According to a cross-sectional study, early SI is associated with a higher probability of becoming a heavier smoker and a lower rate of smoking cessation success.[24] More importantly, in current smokers, early SI is independently associated with a higher lung cancer risk after adjusting for CPD.[39]

Because of the high cost of obtaining sufficient statistical power, whole-genome sequencing analysis was not conducted in this study. To reveal the molecular mechanism underling each smoking phenotype, we used high-throughput approaches such as exome-based association study to identify genetic variants that contribute to ASI and other smoking-related phenotypes.

Material and Methods

Subjects and Demographic Characteristics

A total of 2510 smokers selected from the Mid-South Tobacco Case–Control (MSTCC) study were included in this study. A detailed description of the inclusion and exclusion criteria used for the recruitment of the MSTCC sample has been published previously.[26] Among them, individuals who exhibited other substance dependence or abuse (such as marijuana, cocaine, morphine) except for alcohol use or abuse (<10%) were excluded. All research protocols were approved by each participating Institutional Review Board, and written informed consent was obtained from all participants.

This sample set consisted of 1654 unrelated AA smokers (857 males and 797 females) and 856 unrelated EA smokers (422 males and 434 females). All participants had smoked at least 100 cigarettes in their lifetimes.[26] Detailed characteristics of the MSTCC AA and EA samples are presented in Table 1.

Phenotype of Interest

Almost all answers were self-reported, with the structured questionnaire being administered by a specially trained clinical study researcher. Questionnaires used to assess smoking-related behaviors included the FTND,[27] the CPD, SI (in ever vs. never smokers), ASI, and smoking cessation (former vs. current smokers). For the objectives of this study, only ASI was analyzed. The question used to measure ASI was "How old were you when you started smoking regularly?" This phenotype was treated as a continuous variable among ever smokers.

Genotyping and Quality Control

Genomic DNA was extracted from EDTA-treated peripheral venous blood of each subject using the Qiagen DNA purification kit. All DNA samples were treated with RNase A to remove potential

Table 1. Characteristics of Study Samples by Ethnicities

Characteristic	African-Americans (*N* = 1654)	European-Americans (*N* = 856)
No. Females (%)	797 (48.2)	434 (50.7)
CPD (Mean ± SD)	26.4 ± 6.38	27.8 ± 7.9
FTND (Mean ± SD)	8.4 ± 1.5	7.9 ± 2.0
Age of Smoking Initiation (Mean ± SD)	18.4 ± 2.7	17.7 ± 3.6
≤18 (years)	1115 (67.4)	636 (74.3)
19–30 (years)	526 (31.8)	214 (25.0)
≥31 (years)	13 (0.8)	6 (0.7)

contaminating RNA, and DNA quality and the concentration of each sample was determined by the A260/A280 absorbance ratio. All samples were genotyped using the Illumina Infinium Human Exome BeadChip (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's protocol. This chip aims to detect the association of rare variants with larger effect size and was developed from functional exonic variants (> 90%) and disease-associated tag markers found at least three times in more than two datasets from the whole-exome sequencing of more than 12 000 individuals (www.illumina.com).

After genotyping, we performed the following quality control analyses on all genotyped SNPs. First, among 242 901 genotyped variants, we excluded all insertions and deletions to ensure that all base pair positions were unique and referred to the same variant. Second, any SNP with a call rate of <95% was excluded. Third, we removed those SNPs that were not in Hardy–Weinberg equilibrium at a p value of $< 1.0 \times 10^{-6}$. Finally, we used a 1% minor allele frequency (MAF) as a threshold to define rare and common variants, and any SNP with an MAF of <1% was excluded from the analysis.

Strict and rigorous quality control was implemented in both sample selection and population substructure assessments. The samples with incomplete phenotypic information were removed. Meanwhile, to evaluate the population structure and identify potential outliers, we performed principal components analysis (PCA) using EIGENSTRAT.[28]

All individual SNP- and gene-based association analyses were performed for the AA and EA samples separately. After the quality control steps, a final total of 48 095 SNPs in the 1654 AA sample and 34 992 SNPs in the 856 EA sample were retained for the association analysis.

Association Analysis

For individual SNP-based association analysis, we used the PLINK (v. 1.07)[29] to perform multiple linear regression under the additive genetic model with sex, age, and the first three principle components as covariates in the AA and EA samples separately. Considering the type of chip used in the study and the number of SNPs included in each chip, Bonferroni correction rather than the commonly adapted genome-wide significant p value of 10^{-8} was adopted for this study. To predict the function of the intronic SNPs, the online bioinformatic tools SNPnexus,[30] SIFT,[31] and Polyphen[32] were used.

Gene-based Analysis Using VEGAS2

The VEGAS2 approach was used for the gene-based association test.[20] This tool summarizes association signals from all the SNPs within each gene, considering linkage disequilibrium (LD) between markers based on HapMap data and calculated as the sum of all χ^2-converted SNP p values within the gene that were generating by PLINK. The significance of each gene was determined with the Bonferroni-corrected p value for the number of genes examined.

SNP Heritability

The heritability of the joint effect of all SNPs (ie, SNP heritability or h^2_{SNP}) was estimated using the restricted maximum likelihood analysis implemented in the genome-wide complex trait analysis package (GCTA).[33] After calculation of the genetic relation matrix (GRM), h^2_{SNP} was estimated using a linear mixed model in which the measure of genetic similarity was included as a random effect to predict the phenotype of interest.

Results

As shown in Table 1, this study included 1654 AA and 856 EA smokers. For these smokers, the CPD (mean ± SD) was 26.4 ± 6.4 and 27.8 ± 7.0, and the average ASI (mean ± SD) was 18.4 ± 2.7 and 17.7 ± 3.6 years, for the AA and EA smokers, respectively. To define the relation between CPD and ASI, a correlation test was conducted, which revealed a negative correlation in both AAs ($p = -0.047$) and EAs ($p = -0.051$).

SNP-based Association Analysis

As stated earlier, after various QC steps, 48 091 and 34 933 autosomal SNPs remained for genome-wide association analysis in the AA and EA samples, respectively. We observed no evidence of systematic genomic inflation for the test statistic (ie, $\lambda = 1.000$) for all genotyped SNPs secondary to population stratification.

Of those SNPs identified by EWAS, one remained statistically significant after Bonferroni correction in the AA sample and another in the EA sample. For the AA sample, the strongest associated signal was achieved for rs17849904 in the pitrilysin metallopeptidase 1 gene on chromosome 10p15 (*PITRM1*; $p = 9.02 \times 10^{-7}$; MAF = 0.012), a gene involved in modulating both metalloendopeptidase and enzyme activator activity that has been associated with Alzheimer's disease.[34] In the EA sample, the most significant association was observed for rs34722354 ($p = 9.74 \times 10^{-7}$; MAF = 0.02), located in the 5′-untranslated region (UTR) of *DDR2* on chromosome 1q23. Tables 2 and 3 provide a list of the other SNPs associated with ASI at a p value of $<10^{-5}$, and Figure 1 shows the Manhattan plots for the AA and EA samples. The QQ plots of these two populations can be found in Supplementary Figures S1 and S2. To determine population difference, the identified top SNPs in AA and EA samples were compared, which revealed a P value of 0.054 for rs17849904 in the EA population. Because of low MAF (<0.01), rs34722354 was removed from the association analysis for the AA sample.

Gene-based Analysis

To evaluate the association between ASI and all common SNPs within a gene, we performed a gene-based analysis using the VEGAS2 method.[20] The p value for 48 091 and 34 933 autosomal SNPs in the AA and EA samples were employed for the gene-based analysis. A total of 8569 and 8967 genes were tested in the AA and EA samples, respectively. Supplementary Tables S1 and S2 show the top genes with p values < 0.005 ranked by their p values from the VEGAS2 analysis for the AA and EA samples, respectively. In the AA, the only gene that remained statistically significant after Bonferroni correction was dehydrogenase/reductase 7 (*DHRS7*; $p = 5.00 \times 10^{-6}$), which encodes a member of the short-chain dehydrogenases/reductase (SDR) protein family and functions as an enzyme from the SDR superfamily to contribute to the metabolism of xenobiotics. In the EA sample, the top hit was forkhead box N1 (*FOXN1*; $p = 1.20 \times 10^{-5}$), which has been associated with rudimentary thymus in a previous report.[35] Furthermore, *C14ORF135* was identified in both the SNP-based and the gene-based analyses. In the individual SNP-based association analysis, the most significant SNP located in PITRM1 gene, the p value for this gene was 0.02 in the gene-based association analysis.

SNP Heritability

The univariate GCTA-GREML analysis was used to estimate the proportion of variance explained by all common SNPs for the ASI

Table 2. SNPs Associated With Age of Smoking Initiation a *p* Value of <1 × 10^{-5} in the AA Sample

SNP	Chromosome (bp position)	Gene	Effect Allele	Ref Allele	MAF	β	SE	*p* value	Prediction (SIFT)	Prediction (Polyphen)
rs17849904	10 (3181126)	*PITRM1*	*A*	*G*	0.012	1.98	0.42	9.02 × 10^{-7}	Tolerated	Possibly damaging
rs2003417	7 (21778429)	*DNAH11*	*G*	*A*	0.012	1.99	0.44	5.93 × 10^{-6}	Damaging	Probably damaging
rs35675573	6 (43572453)	*POLH*	A	G	0.018	1.51	0.36	2.51 × 10^{-5}	Tolerated	Benign
rs2997211	10 (28378758)	*MPP7*	A	G	0.040	0.97	0.24	3.49 × 10^{-5}	Tolerated	Benign
rs138051249	2 (71361161)	*MPHOSPH10*	C	A	0.011	1.88	0.46	4.03 × 10^{-5}	Tolerated	Benign
rs167437	14 (60591887)	*C14orf135*	G	A	0.46	-0.37	0.093	5.80 × 10^{-5}	Tolerated	Benign
rs150688	14 (60582053)	*C14orf135*	A	G	0.46	-0.37	0.093	5.84 × 10^{-5}	Tolerated	Benign
rs3735099	7 (2472429)	*CHST12*	A	C	0.015	1.53	0.39	8.54 × 10^{-5}	Tolerated	Benign
rs308998	14 (60585131)	*C14orf135*	A	G	0.49	-0.37	0.093	8.59 × 10^{-5}	Tolerated	Benign
rs2074506	6 (30890483)	*VARS2*	A	C	0.10	-0.62	0.16	8.86 × 10^{-5}	Tolerated	Possibly damaging

MAF: minor allele frequency; NA: not applicable.

Table 3. SNPs Associated With Age of Smoking Initiation at a *p* value of <1 × 10^{-5} in the EA Sample

SNP	Chromosome (bp position)	Gene	Effect allele	Ref allele	MAF	β	SE	*p* value	Prediction (SIFT)	Prediction (Polyphen)
rs34722354	1 (162740121)	*DDR2*	A	G	0.020	2.77	0.56	9.74 × 10^{-7}	Damaging	Benign
rs140717526	1 (109779073)	*SARS*	A	G	0.011	3.48	0.82	2.43 × 10^{-5}	Tolerated	Benign
rs41277210	1 (216144049)	*USH2A*	A	G	0.028	2.18	0.52	2.83 × 10^{-5}	Tolerated	Benign
rs17229382	1 (117568217)	*CD101*	A	G	0.023	2.30	0.57	5.18 × 10^{-5}	Tolerated	Benign
rs55874520	18 (64172434)	*CDH19*	C	G	0.021	2.27	0.56	5.61 × 10^{-5}	Damaging	Probably damaging
rs11568466	17 (26817537)	*SLC13A2*	A	G	0.21	0.85	0.20	5.72 × 10^{-5}	NA	NA
rs139345781	13 (33017869)	*N4BP2L2*	G	A	0.020	2.52	0.63	6.43 × 10^{-5}	Damaging	Benign
rs36051007	2 (179545859)	*TTN*	A	G	0.28	0.78	0.20	7.55 × 10^{-5}	NA	NA
rs41303285	1 (215914751)	*USH2A*	A	C	0.019	2.40	0.60	7.72 × 10^{-5}	Tolerated	Probably Damaging
rs12463674	2 (179432185)	*TTN*	G	A	0.28	0.78	0.20	7.83 × 10^{-5}	NA	NA

MAF: minor allele frequency; NA: not applicable.

phenotype in both the AA and EA samples. The estimated heritability of ASI was 0.129 (SE = 0.16) for AAs and 0.07 (SE = 0.29) for the EAs, which represents the upper limit of the amount of phenotypic variance explained by all the SNPs included in our EWAS.

Discussion

In this study, we performed a genome-wide association analysis to link about 40 000 genotyped common SNPs to the ASI for both AA and EA smokers. We obtained evidence of significant ethnicity-specific associations with ASI for non-synonymous variants rs17849904 in *PITRM1* and rs34722354 in *DDR2* in the AA and EA smokers, respectively. Furthermore, our gene-based association analysis indicated a putative role of *DHRS7* in the AA sample, a gene that has been implicated in nicotine metabolism.[36,37]

Consistent with other reports,[38] the majority of smokers included in our samples started their regular smoking at around 18 years of age (18.3 ± 2.9 years). Compared with EA smokers (17.7 ± 3.6), AA smokers started their regular smoking at a slightly later age (18.4 ± 2.7 years). In addition, AA smokers smoke fewer cigarettes per day but are less likely to quit smoking.[39] Together, these phenotypic differences likely imply genetic differences between the AA and EA smokers. Besides, there was a negative correlation between ASI and CPD in AA (p = −0.047) and EA (p = −0.051). These results indicate that the distributions of our samples are consistent with previous epidemiological surveys.[40]

So far, a number of GWAS have been conducted for various smoking behaviors,[41,42] with several chromosomal regions repeatedly observed to be associated with ASI. Several SNPs near the HLA region on chromosome 6p21[42] and others in the nicotinic receptor candidate genes *CHRNA3* on chromosome 15q25 and *CHRNA1* on chromosome 2q31[41] and in an intergenic region on chromosome 2q21[42] show associations with ASI. Moreover, *CYP2A6* exhibited nominally significant associations with ASI.[41] These results could be explained by the relation between *CYP2A6* variation and both nicotine metabolism[43] and smoking behavior.[44] However, most of these findings were not replicated in our study, which may be a consequence of the smaller sample.

In this study, allele A of rs17849904 in *PITRM1* was negatively associated with ASI, indicating a potential protective role for this locus. This implies that carriers of the minor allele would have a late smoking initiation. The *PITRM1* gene encodes a 117-kDa mitochondrial matrix enzyme (also known as presequence peptidase; *PreP*), which contributes to digestion of mitochondrial amyloid beta (Aβ)[45] and interaction with mitochondrial targeting sequence (MTS) of proteins imported from the inner mitochondrial membrane,[46] such that mutations lead to mitochondrial dysfunction. This dysfunction is a hallmark of neurodegeneration[47] and of many psychiatric disorders such as Alzheimer's dementia (AD),[48] which is characterized by the accumulation of the Aβ peptide as plaques in the neuropil.[49] Carriers of *PITRM1*-R138Q missense mutations have a slowly progressive neurodegenerative phenotype.[34] Although smoking is a risk factor for Alzheimer's disease, and several important markers such as *APOE* have been identified,[50] the mechanism

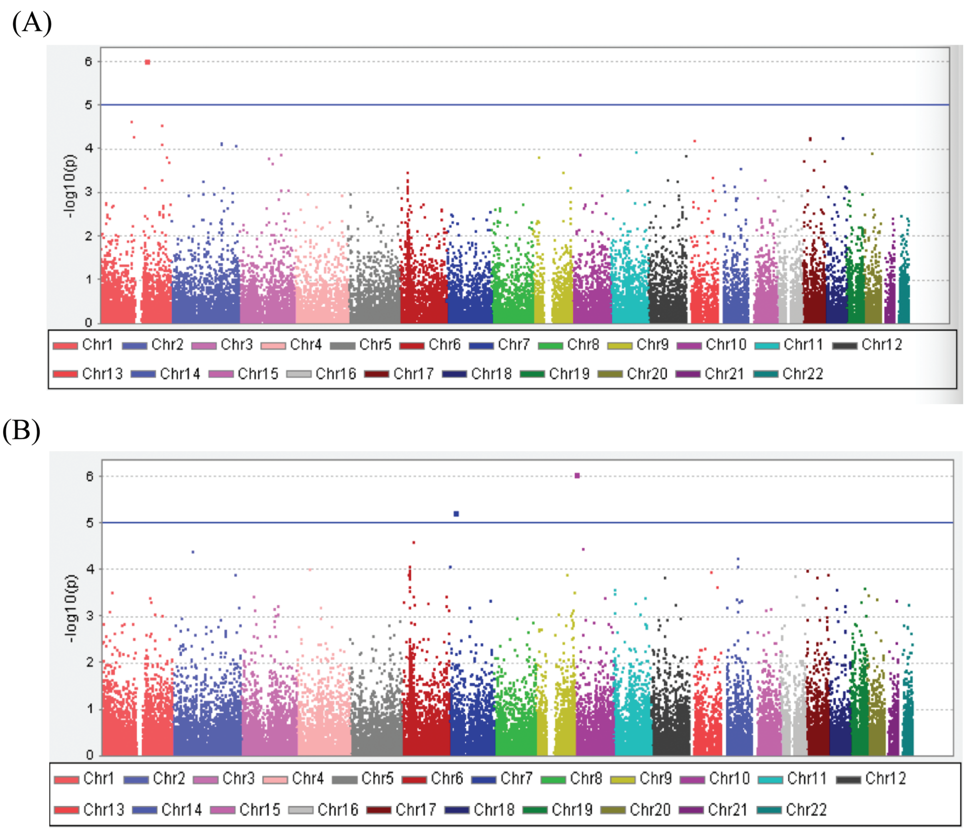

Figure 1. Manhattan plots for EWAS of smoking initiation for the EA (A) and AA (B) subjects.

underlying the correlation between Alzheimer's disease and smoking is not fully understood. In our study, this SNP was predicted to be possibly damaging by PolyPhen, implying that rs17849904 may contribute to the functional modification of the protein.

Another association finding in the EA sample revealed a protective allele, rs34722354, in *DDR2*. This gene encodes a membrane-bound receptor tyrosine kinase that binds collagen and is involved in regulation of cell proliferation and survival.[51] Lung cancer-associated mutations have been identified in *DDR2*,[52] and overexpression of the gene correlates with clinicopathologic features of a poor prognosis.[53] Lung cancer is one of the most commonly observed diseases associated with tobacco smoking. Meanwhile, for rs34722354, *in silico* programs predict a functional modification by SIFT, which is based on the conservation of amino acid residues in sequence alignments from closely related sequences across species. The secondary structure of the protein may be destroyed by this non-synonymous mutation. Although the functional relevance of this genetic variant in smoking and smoking-related diseases such as lung cancer is still unclear, our finding suggests that rs34722354 is a plausible locus for ASI and other smoking-related phenotypes, a finding that deserves further investigation. Besides, the two loci located in the *TTN* gene are in high linkage disequilibrium (r^2 > 0.89) with each other. This gene encodes a large, abundant protein of striated muscle, which has been reported to be associated with a higher risk for cardiovascular death.[54]

To evaluate the association between ASI and all SNPs in the gene of interest, a gene-based association test was performed using VEGAS2, which uses a list of SNP p values as inputs and then considers the underlying SNP–SNP correlation pattern employing the genotype data from the HapMap Project or other databases.[55] As reported elsewhere,[56] VEGAS2 is stable across different boundaries and remains powerful even with the inclusion of non-significant SNPs. We used the default gene boundary in VEGAS2 (± 50 kb) and focused on the top individual SNPs and previously reported genes that are related to ASI. Through this analysis, we identified a potential candidate gene, *DHRS7*, in the AA sample, which encodes a member of the short-chain dehydrogenases/reductase (SDR) protein family.[57] Human SDR members play important roles in various biochemical pathways, including those of intermediary metabolism and biotransformation of xenobiotics.[36] The *DHRS7* gene is located on chromosome 14 and has two isoforms. Isoform 1 consists of 339 amino acids (38 kDa), and isoform 2 consists of 289 amino acids (32 kDa). Although cytochrome P450 2A6 (*CYP2A6*) encodes the enzyme responsible for the majority of nicotine metabolism reactions,[58] the *DHRS7* product metabolizes NNK 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone to 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), a secondary metabolite of nicotine *in vitro*.[36,37] A previous study has demonstrated that the poor-metabolizer genotypes of *CYP2A6* are associated with a later ASI.[59] Together, our gene-based analysis indicated that *DHRS7* is associated with ASI, likely by changing nicotine metabolism.

Further, our gene-based analysis showed a significant association of *PITRM1* with ASI, suggesting that *PITRM1* represents a candidate gene for ASI, which warrants replication. Another novel candidate gene for ASI identified in this study with both individual SNP-based and gene-based analyses is *C14ORF135*, although there is no report on the role of this gene in smoking.

Further, to investigate the contribution of common SNPs to ASI, we estimated SNP heritability using GCTA software. We found that about 12.9% of the phenotypic variance in ASI was tagged by common SNPs in the AA sample and 6.8% in the EA sample. Such findings imply that a significant number of SNPs remain to be identified for their association with ASI. Although the exact reasons are unknown, the small sample of this study may have contributed to the low detection power. Future study with larger samples is greatly needed to replicate the findings and unravel the biological mechanisms underlying these associations.

There are a few limitations of this study that should be considered. First, although our samples are rather large, they are still too small with limited power for a EWAS. More studies with significantly a large sample or meta-analysis of multiple independent samples are greatly needed to final more susceptibility variants for ASI. Second, because a limited number of SNPs within a gene were included in the exome chip used for many genes, our gene-based association analysis was not as powerful as we wanted, which might limit the number of candidate genes identified. Nevertheless, the susceptibility SNPs and genes for ASI revealed by this study are not only statistically significant but also biological meaningful, and the findings deserve to be replicated in a future study.

In sum, our individual SNP-based association analysis revealed two novel non-synonymous SNPs with one located in *PITRM1* and another in *DDR2* that are significantly associated with ASI in AA and EA smokers, respectively. Furthermore, our gene-based analysis revealed that *DHRS7* is a novel candidate gene for ASI. Given the documented biological roles of these genes, more replications with a large sample and molecular studies are needed in the future.

Supplementary Materials

Supplementary Tables S1–S3 and Figures S1 and S2 can be found online at https://academic.oup.com/ntr/.

Funding

This study was supported in part by the China Precision Medicine Initiative (2016YFC0906300), the Research Center for Air Pollution and Health of Zhejiang University, and National Institutes of Health grant DA012844 to MDL.

Declaration of Interests

None declared.

Acknowledgments

We thank all the subjects who participated this study. We also thank Dr. David L Bronson for his excellent editing of this paper.

References

1. Gandini S, Botteri E, Iodice S, et al. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer*. 2008;122(1):155–164.
2. World Health Organization. WHO (2008) Report on the Global Tobacco Epidemic. 2008. http://www.who.int/gho/tobacco/en/. Accessed November 19, 2017.
3. Li MD. The genetics of smoking related behavior: a brief review. *Am J Med Sci*. 2003;326(4):168–173.
4. Lessov CN, Martin NG, Statham DJ, et al. Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol Med*. 2004;34(5):865–879.
5. Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction*. 2003;98(1):23–31.
6. Hall W, Madden P, Lynskey M. The genetics of tobacco use: methods, findings and policy implications. *Tob Control*. 2002;11(2):119–124.
7. Goode EL, Badzioch MD, Kim H, et al.; Framingham Heart Study. Multiple genome-wide analyses of smoking behavior in the Framingham Heart Study. *BMC Genet*. 2003;4(suppl 1):S102.
8. Horimoto AR, Oliveira CM, Giolo SR, et al. Genetic analyses of smoking initiation, persistence, quantity, and age-at-onset of regular cigarette use in Brazilian families: the Baependi Heart Study. *BMC Med Genet*. 2012;13:9.
9. Han S, Gelernter J, Luo X, Yang BZ. Meta-analysis of 15 genome-wide linkage scans of smoking behavior. *Biol Psychiatry*. 2010;67(1):12–19.
10. Saccone NL, Wang JC, Breslau N, et al. The CHRNA5-CHRNA3-CHRNB4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res*. 2009;69(17):6848–6856.
11. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008;452(7187):638–642.
12. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al.; ENGAGE Consortium. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*. 2010;42(5):448–453.
13. Benowitz NL. Nicotine addiction. *N Engl J Med*. 2010;362(24):2295–2303.
14. Mineur YS, Picciotto MR. Genetics of nicotinic acetylcholine receptors: Relevance to nicotine addiction. *Biochem Pharmacol*. 2008;75(1):323–333.
15. Ray R, Tyndale RF, Lerman C. Nicotine dependence pharmacogenetics: role of genetic variation in nicotine-metabolizing enzymes. *J Neurogenet*. 2009;23(3):252–261.
16. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42(5):441–447.
17. Yoon D, Kim YJ, Cui WY, et al. Large-scale genome-wide association study of Asian population reveals genetic factors in FRMD4A and other loci influencing smoking initiation and nicotine dependence. *Hum Genet*. 2012;131(6):1009–1021.
18. Loukola A, Wedenoja J, Keskitalo-Vuokko K, et al. Genome-wide association study on detailed profiles of smoking behavior and nicotine dependence in a twin sample. *Mol Psychiatry*. 2014;19(5):615–624.
19. Huang H, Chanda P, Alonso A, Bader JS, Arking DE. Gene-based tests of association. *PLoS Genet*. 2011;7(7):e1002177.
20. Mishra A, Macgregor S. VEGAS2: software for more flexible gene-based testing. *Twin Res Hum Genet*. 2015;18(1):86–91.
21. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet*. 2010;11(5):356–366.
22. CDC. Racial/Ethnic disparities and geographic differences in lung cancer incidence --- 38 States and the District of Columbia, 1998–2006. *MMWR Morb Mortal Wkly Rep*. 2010;59(44):1434–1438.
23. Jamal A, King BA, Neff LJ, Whitmill J, Babb SD, Graffunder CM. Current cigarette smoking among adults - United States, 2005-2015. *MMWR Morb Mortal Wkly Rep*. 2016;65(44):1205–1211.
24. Andreeva TI, Krasovsky KS, Semenova DS. Correlates of smoking initiation among young adults in Ukraine: a cross-sectional study. *BMC Public Health*. 2007;7:106.
25. Prizment AE, Yatsuya H, Lutsey PL, et al. Smoking behavior and lung cancer in a biracial cohort: the Atherosclerosis Risk in Communities study. *Am J Prev Med*. 2014;46(6):624–632.
26. Yang J, Wang S, Yang Z, et al. The contribution of rare and common variants in 30 genes to risk nicotine dependence. *Mol Psychiatry*. 2015;20(11):1467–1478.
27. Fagerström KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addict Behav*. 1978;3(3–4):235–241.
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–909.

29. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–575.
30. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*. 2009;25(5):655–661.
31. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812–3814.
32. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30(17):3894–3900.
33. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
34. Brunetti D, Torsvik J, Dallabona C, et al. Defective PITRM1 mitochondrial peptidase is associated with Aβ amyloidotic neurodegeneration. *EMBO Mol Med*. 2016;8(3):176–190.
35. Chou J, Massaad MJ, Wakim RH, Bainter W, Dbaibo G, Geha RS. A novel mutation in FOXN1 resulting in SCID: a case report and literature review. *Clin Immunol*. 2014;155(1):30–32.
36. Stambergova H, Skarydova L, Dunford JE, Wsol V. Biochemical properties of human dehydrogenase/reductase (SDR family) member 7. *Chem Biol Interact*. 2014;207:52–57.
37. Skarydova L, Zverinova M, Stambergova H, Wsol V. A simple identification of novel carbonyl reducing enzymes in the metabolism of the tobacco specific carcinogen NNK. *Drug Metab Lett*. 2012;6(3):174–181.
38. Sartor CE, Jackson KM, McCutcheon VV, et al. Progression from first drink, first intoxication, and regular drinking to alcohol use disorder: a comparison of African American and European American Youth. *Alcohol Clin Exp Res*. 2016;40(7):1515–1523.
39. Giovino GA. Epidemiology of tobacco use in the United States. *Oncogene*. 2002;21(48):7326–7340.
40. Kendler KS, Myers J, Damaj MI, Chen X. Early smoking onset and risk for subsequent nicotine dependence: a monozygotic co-twin control study. *Am J Psychiatry*. 2013;170(4):408–413.
41. Caporaso N, Gu F, Chatterjee N, et al. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One*. 2009;4(2):e4653.
42. Siedlinski M, Cho MH, Bakke P, et al.; COPDGene Investigators; ECLIPSE Investigators. Genome-wide association study of smoking behaviours in patients with COPD. *Thorax*. 2011;66(10):894–902.
43. Mwenifumbo JC, Sellers EM, Tyndale RF. Nicotine metabolism and CYP2A6 activity in a population of black African descent: impact of gender and light smoking. *Drug Alcohol Depend*. 2007;89(1):24–33.
44. Strasser AA, Malaiyandi V, Hoffmann E, Tyndale RF, Lerman C. An association of CYP2A6 genotype and smoking topography. *Nicotine Tob Res*. 2007;9(4):511–518.
45. Pinho CM, Teixeira PF, Glaser E. Mitochondrial import and degradation of amyloid-β peptide. *Biochim Biophys Acta*. 2014;1837(7):1069–1074.
46. Teixeira PF, Pinho CM, Branca RM, Lehtiö J, Levine RL, Glaser E. *In vitro* oxidative inactivation of human presequence protease (hPreP). *Free Radic Biol Med*. 2012;53(11):2188–2195.
47. Johri A, Beal MF. Mitochondrial dysfunction in neurodegenerative diseases. *J Pharmacol Exp Ther*. 2012;342(3):619–630.
48. Manczak M, Anekonda TS, Henson E, Park BS, Quinn J, Reddy PH. Mitochondria are a direct site of A beta accumulation in Alzheimer's disease neurons: implications for free radical generation and oxidative damage in disease progression. *Hum Mol Genet*. 2006;15(9):1437–1449.
49. Falkevall A, Alikhani N, Bhushan S, et al. Degradation of the amyloid beta-protein by the novel mitochondrial peptidasome, PreP. *J Biol Chem*. 2006;281(39):29096–29104.
50. van Duijn CM, Havekes LM, Van Broeckhoven C, de Knijff P, Hofman A. Apolipoprotein E genotype and association between smoking and early onset Alzheimer's disease. *BMJ*. 1995;310(6980):627–631.
51. Ikeda K, Wang LH, Torres R, et al. Discoidin domain receptor 2 interacts with Src and Shc following its activation by type I collagen. *J Biol Chem*. 2002;277(21):19206–19212.
52. Hammerman PS, Sos ML, Ramos AH, et al. Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov*. 2011;1(1):78–89.
53. Xie BH, Lin WH, Ye JM, et al. DDR2 facilitates hepatocellular carcinoma invasion and metastasis via activating ERK signaling and stabilizing SNAIL1. *J Exp Clin Canc Res*. 2015;34.
54. Zhang C, Zhang H, Wu G, et al. Titin-truncating variants increase the risk of cardiovascular death in patients with hypertrophic cardiomyopathy. *Can J Cardiol*. 2017;33(10):1292–1297.
55. Hong MG, Pawitan Y, Magnusson PK, Prince JA. Strategies and issues in the detection of pathway enrichment in genome-wide association studies. *Hum Genet*. 2009;126(2):289–301.
56. Petersen A, Alvarez C, DeClaire S, Tintle NL. Assessing methods for assigning SNPs to genes in gene-based tests of association using common variants. *PLoS One*. 2013;8(5):e62161.
57. Štambergová H, Zemanová L, Lundová T, et al. Human DHRS7, promising enzyme in metabolism of steroids and retinoids? *J Steroid Biochem Mol Biol*. 2016;155(Pt A):112–119.
58. Benowitz NL, St Helen G, Dempsey DA, Jacob P 3rd, Tyndale RF. Disposition kinetics and metabolism of nicotine and cotinine in African American smokers: impact of CYP2A6 genetic variation and enzymatic activity. *Pharmacogenet Genomics*. 2016;26(7):340–350.
59. Liu T, David SP, Tyndale RF, et al. Associations of CYP2A6 genotype with smoking behaviors in southern China. *Addiction*. 2011;106(5):985–994.
60. Krishna Kumar S, Feldman MW, Rehkopf DH, Tuljapurkar S. Limitations of GCTA as a solution to the missing heritability problem. *Proc Natl Acad Sci U S A*. 2016;113(1):E61–E70.