Published in final edited form as:
Hum Genet. 2012 June ; 131(6): 1009–1021. doi:10.1007/s00439-011-1102-x.

Large-scale genome-wide association study of Asian population reveals genetic factors in *FRMD4A* and other loci influencing smoking initiation and nicotine dependence

Dankyu Yoon,
Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Korea

Young-Jin Kim,
Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Korea; Lee Center for Genome Science, Korea National Institute of Health, Osong 363-951, Korea

Wen-Yan Cui,
State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; Department of Psychiatry and Neurobehavioral Sciences, University of Virginia, 1670 Discovery Drive, Suite 110, Charlottesville, VA 22911, USA

Andrew Van der Vaart,
Department of Psychiatry and Neurobehavioral Sciences, University of Virginia, 1670 Discovery Drive, Suite 110, Charlottesville, VA 22911, USA

Yoon Shin Cho,
Center for Genome Science, Korea National Institute of Health, Osong 363-951, Korea

Jong-Young Lee,
Center for Genome Science, Korea National Institute of Health, Osong 363-951, Korea

Jennie Z. Ma,
Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22903, USA

Thomas J. Payne,
Department of Otolaryngology and Communicative Sciences, University of Mississippi Medical Center, Jackson, MS 39216, USA

Ming D. Li, and
State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; Department of Psychiatry and Neurobehavioral Sciences, University of Virginia, 1670 Discovery Drive, Suite 110, Charlottesville, VA 22911, USA

Taesung Park

Ming_Li@virginia.edu, tspark@stats.snu.ac.kr.

Conflict of interest The authors declare that they have no conflict of interest.

Department of Statistics, College of Natural Science, Seoul National University, Seoul 151-742, Korea

Abstract

Diseases related to smoking are the second leading cause of death in the world. Cigarette smoking is a risk factor for several diseases such as cancer and cardiovascular and respiratory disorders. Despite increasing evidence of genetic determination, the susceptibility genes and loci underlying various aspects of smoking behavior are largely unknown. Moreover, almost all reported genome-wide association studies (GWASs) have been performed on samples of European origin, limiting the applicability of the results to other ethnic populations. In this first GWAS on smoking behavior in an Asian population, after analyzing 8,842 DNA samples from the Korea Association Resource project with 352,228 single nucleotide polymorphisms (SNPs) genotyped for each sample, we identified 8 SNPs significantly associated with smoking initiation (SI) and 4 with nicotine dependence (ND). Because of the current unavailability of an independent Asian smoking sample, we replicated the discoveries in independent samples of European-American and African-American origin. Of the 12 SNPs examined in the replicated samples, we identified two SNPs, in the regulator of G-protein signaling 17 gene (rs7747583, p $\text{value}_{\text{meta}} = 6.40 \times 10^{-6}$; rs2349433, p $\text{value}_{\text{meta}} = 5.57 \times 10^{-6}$), associated with SI. Also, we found two SNPs significantly associated with ND; one in the FERM domain containing 4A (rs4424567, p $\text{value}_{\text{meta}} = 2.30 \times 10^{-6}$) and the other at 7q31.1 (rs848353, p $\text{value}_{\text{meta}} = 9.16 \times 10^{-8}$). These SNPs represent novel targets for examination of smoking behavior and warrant further investigation using independent samples.

Introduction

Cigarette smoking is one of the most significant risk factors for premature death, costing nearly six million human lives per year throughout the world (Warren et al. 2008). According to a recent report from the World Health Organization (WHO), smoking-related diseases rank as the second leading cause of death in the world, and it is expected that the prolonged pattern of current tobacco consumption will double the number of deaths annually by 2020 (WHO; http://www.who.int). Further, cigarette smoking is a well-known risk factor for cardiovascular and respiratory diseases (Giovino 2007; Muller and Wehb 2008; Vianna et al. 2008).

Many twin studies in the US and other countries hav consistently implied that genetic factors contribute to th risk of becoming a regular smoker. Initial evidence for genetic influences on nicotine dependence (ND) cam from cross-sectional studies of twins that showed a mean heritability of 0.53 (range 0.28–0.84) for cigarette smoking (for reviews, see Carmelli et al. 1992; Hughes 1986). Our meta-analysis of the genetic parameter estimates for smoking initiation (SI) and ND based on 17 twin studies determined that the weighted mean heritability is 0.50 and 0.59, respectively, for all smokers (Li et al. 2003). Complex segregation analyses of smoking behavior in 49 three-generation families supported a dominant major gene effect with residual familial correlation (Cheng et al 2000).

Although twin studies provide strong evidence for involvement of genetic factors in various stages of smoking behavior, identifying these susceptibility genes has been a challenge.

During the past few years, a significant numbe of genome-wide association studies (GWAS) have been conducted for various smoking phenotypes, including SI and current smoking (Caporaso et al. 2009; TGC 2010 Thorgeirsson et al. 2010; Vink et al. 2009), age at smoking initiation (Siedlinski et al. 2011), ND (Bierut et al. 2007 Liu et al. 2009; Uhl et al. 2007), smoking quantity (Siedlinski et al. 2011; Berrettini et al. 2008; Caporaso et al 2009; Liu et al. 2009, 2010; TGC 2010; Thorgeirsson et al 2010, 2008), and smoking cessation (Siedlinski et al. 2011 TGC 2010; Uhl et al. 2008, 2010). However, most of these studies focused primarily on samples of European origin. Given the distinct differences in linkage disequilibrium (LD) patterns across ethnic populations (Li and Burmeiste 2009), a GWAS in other ethnic samples is warranted.

In this study, we performed a GWAS on two smoking-related behaviors, SI and ND, in a cohort from the Korea Association Resource (KARE) project (Cho et al. 2009). A total of 8,842 individuals with 352,228 SNPs genotyped per individual were examined to find any association between smoking behavior and genetic factors. Our initia analysis at the discovery stage identified 12 SNPs ($p < 1 \times 10^{-5}$) significantly associated with smokin behavior. Although they did not reach the genome-wid significance level ($p < 1 \times 10^{-7}$), they represented severa genomic regions possibly associated with smoking behavior. Among 12 SNPs, 8 were significantly associated with SI and 4 with ND.

Because of the unavailability of an independent Asian smoking sample, the replication study was performed in a sample of European-American or African-American origin. Of the 12 SNPs examined in independent samples, 4 were replicated. At the 5% significance level, two SNPs, rs7747583 and rs2349433, in *RGS17* were replicated in association with SI, and rs4424567 in *FRMD4A* and rs848353 at 7q31.1 were replicated for association with ND. Finally, we performed meta-analysis on the association results from both the discovery and replication samples.

Materials and methods

Subjects

Discovery sample—All participants provided informed consent on a document that had been approved by all participating Institutional Review Boards. The relevant information has been described in a previous paper (Cho et al. 2009). Briefly, samples from the 10,038 participants in the Korea Association Resource (KARE) project were recruited from two areas, Ansung and Ansan, in South Korea. Participant ages ranged from 40 to 69 years. Two population based cohorts in this study were established as part of Korean Genome and Epidemiology Study (KoGES), which started in 2001, that provides extensive phenotype data over 260 traits through epidemiological surveys, physical examination and clinical examination.

Replication sample—The subjects used in our replication sample were from the Mid-South Tobacco Family (MSTF) study with either African-American (AA) or European-American (EA) origin. They were recruited primarily from Tennessee, Mississippi, and Arkansas during 1999-2004. Proband smokers were required to be at least 21 years of age, to have smoked for at least the last 5 years, and to have consumed an average of 20

cigarettes per day for the last 12 months. Once a proband was identified, siblings and biological parents were recruited full sibs per family. Detailed information of the recruitment and characteristics of study participants have been reported (Li 2008; Li et al. 2005,2006; Ma et al. 2005). Together, a total of 200 EA and 402 AA families were recruited. Informed consent was obtained from all participants and approval from the appropriate Institutional Review Boards.

DNA extraction and genotyping

Discovery sample—Genomic DNA was isolated from peripheral blood. All DNA samples from the 10,004 participants were genotyped with the Affymetrix Genome-Wide Human SNP Array 5.0. Genotypes were called by Bayesian Robust Linear Modeling using the Mahalanobis Distance (BRLMM) algorithm. After removing samples with low call rates (n = 401), contamination (n = 11), sex inconsistencies (n = 41), cryptic relatedness (n = 608), or serious concomitant illness (n = 101), 8,842 samples remained for use. Similarly, we removed those SNPs with Hardy–Weinberg equilibrium (HWE) p values $<10^{-6}$, minor allele frequencies >0.01, and genotype call rates >95%, leaving 352,228 SNPs. More detailed statistical analysis of the GWAS data can be found in a previous report (Cho et al. 2009). Considering the samples used for GWAS were recruited from different geographical locations of Korea, we used the magnitude of the genomic inflation factor and multidimensional scaling (MDS)/principal component analysis (PCA) plots to examining the presence of the population stratification, which revealed no evidence of this effect in the KARE samples (Cho et al. 2009).

Replication samples—The DNA was extracted from peripheral blood using a kit from Qiagen Inc. (Valencia, CA). On the basis of the GWAS results from the discovery sample, 12 SNPs were selected for testing in the replication samples. All SNPs were genotyped using the *Taq*Man SNP Genotyping Assay in a 384-well microplate format (Applied Biosystems, Foster, CA). Briefly, 15 ng of DNA was amplified in a total volume of 7 μl containing an MGB probe and 2.5 μl of *Taq*Man universal PCR master mix. Allelic discrimination analysis was performed on the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster CA). To ensure the quality of the genotyping, SNP-specific control samples were added to each 384-well plate.

Definition of smoking-related phenotypes

Discovery sample—Definition of smoking-related phenotypes was essentially the same as in our previous study on the investigation of association of *CHRNA5/A3/B4* with smoking in the same sample (Li et al. 2010). Briefly, on the basis of the survey questionnaire, from which information on smoking status (i.e., never smoker, former smoker who quit one year before being enrolled in the study, occasional smoker who smoke irregularly or only on specific occasions such as when drinking, or habitual smoker who smoke regularly on a daily basis) and number of cigarettes smoked per day (CPD) for the habitual smoker was drawn, three smoking-related phenotypes were defined for SI and ND. Regarding SI, the first measure (called SI-1) was defined as a binary trait comparing “never smoked” and “having regular smoking experiences (former, occasional, or habitual smoker),” and the second measure (called SI-2) was defined as an ordinal trait with four categories: never,

former, light, and habitual. The ND phenotype was defined as an ordinal trait with five categories according to CPD: non-smoking, <10 CPD, 11–20 CPD, 21–30 CPD, and >31 CPD. Such assessment of ND not only has been commonly used in the literature but also is the most replicative in terms of positive findings across populations (Li 2008; Thorgeirsson et al. 2008). The detailed demographic information as well as smoking-related phenotype information is given in Table 1.

Replication sample—Smoking initiation was defined as the age at which the person began smoking regularly. The presence of ND was ascertained by the three measures most commonly used in the literature: CPD; the Heaviness of Smoking Index (HSI 0-6 scale), which includes CPD and smoking urgency (i.e., how soon after waking up does the subject smoke the first cigarette?); and the FTND (0-10 scale) (Heatherton et al. 1991). CPD in the MSTF sample is categorized CPD, as used in the KARE samples. A detailed description of the demographic and clinical characteristics of the MSTF sample is presented in Table 2.

Statistical analysis

Discovery sample—Statistical analyses were performed using PLINK (Purcell et al. 2007) and R software. A logistic regression model was fit for the binary phenotype, SI-1, adjusting for age, sex, and geographic area under additive, dominant, and recessive models. The cumulative *logit* model was fit to the ordinal phenotype SI-2 and ND (Bender and Grouven 1997).

To find supporting evidence for the 12 associated SNPs, we performed imputation analysis for untyped SNPs within ±500 kb of these SNPs. Imputation was carried out using the IMPUTE program on the basis of the NCBI (build 36) and dbSNP (build 126) databases. We initially used 90 individuals from JPT and CHB founders in HapMap as a reference panel comprising 3.99 million SNPs (release 22) (Marchini et al. 2007). We removed the imputed SNPs with <0.5 genotype information content, posterior probability scores <0.90, and low call rates (<0.90), MAF < 0.01, and HWE ($p < 1 \times 10^{-7}$).

Replication samples—The PedCheck program was used to determine genotyping consistency for Mendelian inheritance of all 11 SNPs selected for replication. To verify the quality of our genotyping, we also checked the SNP data for any significant departure from HWE. The HWE at each locus was assessed by the χ^2 test. The allele frequencies for each genetic marker were calculated using the FREQ program of S.A.G.E. (v. 5.0). Associations between individual SNPs and smoking behavior phenotypes, including the ND measures and age at onset, were determined by the PBAT program using generalized estimating equations (Lange et al. 2003). Three genetic models (additive, dominant, and recessive) were tested, with sex and age as covariates in the AA and EA samples; sex, age, and ethnicity served as covariates for the combined sample. Relations with the three ND measures and age at onset were examined individually.

Meta-analysis—We conducted the meta-analysis using Fisher's combining *p* value method (Fisher 1932) in order to combine the results from the discovery and replication samples. Considering that the PBAT approach used in the replication study provides only

score statistics with effect directions and *p* values, we used an equal weight for both studied samples. Prior to conducting meta-analysis, we compared the risk allele of each SNP individually in the two samples and found that all SNPs selected for replication have the same risk allele for the discovery and replication samples. Any variant with discordant direction of effect was not considered as having been replicated.

Results

Description of discovery and replication samples in relation to smoking behaviors

Discovery sample—In total, 8,842 subjects with 352,228 SNPs each were analyzed. The average ages of the participants were 52.22 ± 8.91 (SD) years: 51.78 ± 8.78 for men and 52.61 ± 9.01 for women. Female participants were 52.7% (*N* = 4,659) of all subjects, however, only 4.93% were smokers, being categorized as former (1.34%), occasional (1.29%), or habitual (2.30%) smokers. In contrast, 80.62% of the male participants smoked. Among them, 31.05% were former smokers, 4.78% occasional smokers, and 44.79% habitual smokers. For habitual smokers, the average number of cigarettes smoked per day (CPD) was 19.51 ± 8.74 for male and 11.93 ± 7.28 for female smokers.

Replication sample—The average age of the subjects was 40.8 ± 14.8 years for EAs and 39.9 ± 14.6 years for AAs. The average family size was 3.16 ± 0.69 for EAs and 3.10 ± 0.75 for AAs. Of these families, 60.5% had at least one biological parent recruited, 50.4% had two full sibs, and 44.6% had three or more full sibs. If a family had no biological parents recruited, we had at least three full sibs. The average Fagerstrom test for ND (FTND) score was 6.4 ± 2.2 for EA and 6.3 ± 2.1 for AA smokers. The average CPD was 19.8 ± 13.6 for EA and 19.5 ± 13.0 for AA smokers.

GWAS results from the discovery sample

We conducted GWAS for both SI and ND using either total or male samples. We did not perform this analysis on the female sample alone, because there were so few female smokers in the KARE sample.

The analysis of the total sample identified 4 SNPs: 1 for SI-1, 1 for SI-2, and 2 for ND, with *p* values $< 10^{-5}$ (Table 3). The analysis of the male sample identified 10 SNPs: 1 for SI-1, 5 for SI-2, and 4 for ND with *p* values $< 10^{-5}$ (Table 3). Among the total 14 SNPs identified, two overlapped. For these 12 SNPs brought to replication, the quality of genotyping was investigated through cluster plots, which revealed no evidence of genotyping error. Figure 1 shows the chromosomal plots and quantile-quantile (QQ) plots of the association signals for three smoking-related phenotypes. The values of genomic inflation factor λ (ranging from 0.97 to 1.01) and QQ plot of signals showed no evidence of the presence of population stratification in our discovery sample.

Smoking initiation—In the total sample, two SNPs, rs1442815 and rs11134474, located in the gene desert region on chromosomes 15q26 and 5q33, respectively, were associated with SI-1 and SI-2, respectively. Further analysis of the male sample revealed six more SNPs with SI-1 or SI-2 with *p* values $<10^{-5}$: rs7761503, rs7747583, and rs2349433, located

in *RGS17* on chromosome 6; rs10511914 in *DCAF12*, located on chromosome 9; and rs1385706 and rs12534697, located in an intergenic region at 7q11.

Nicotine dependence—In the total sample, two SNPs, rs1404697 and rs848353, located on chromosome 7q31.1, were associated with ND, with p values $<10^{-5}$. Further analysis of the male sample revealed four additional SNPs associated with ND with p values $<10^{-5}$: rs1404697 and rs848353 on chromosome 7, rs12536963 in *EXOC4* on chromosome 7, and rs4424567 in *FRMD4A* on chromosome 10. Two SNPs, rs1404697 and rs848353, were identified in both the total and male samples.

Association analysis results from replication samples

The association analysis of the total sample identified two SNPs for SI and two for ND, and that of the male sample identified six SNPs for SI and four for ND. Among the 14 SNPs, 12 unique ones were selected for the replication study: 8 for SI and 4 for ND.

The replication study was performed for the EA and AA samples separately. Among the eight SNPs for SI, two, rs7747583 and rs2349433 at *RGS17*, passed the replication significance threshold ($p < 0.05$) and were replicated in the EA male sample (Table 4). Among the four SNPs for ND, two passed the replication significance threshold: rs4424567 at *FRMD4A* was replicated in the EA male sample, and rs848353 at 7q31.1, close to the hypothetical protein-coding genes *C7orf66* and *LOC154907*, was replicated in the AA total sample. One SNP, at the intronic region of *EXOC4*, could not be genotyped in the replication study.

Meta-analysis results of discovery and replication samples

Meta-analysis was performed for the selected 11 SNPs by combining the results from the discovery and replication samples. Three meta-analyses were performed: (1) KAR-E + EA, (2) KARE + AA, and (3) KARE + EA + AA. Among the 11 replicated SNPs, 10 provided concordant direction of genetic effects, and one (rs10511914) showed discordant direction of effect for SI. All four replicated SNPs provided concordant direction of effects. The results are summarized in Table 4. Through meta-analysis, only 1 SNP rs848353 reached genome-wide significance ($p = 9.16 \times 10^{-8}$). In general, the meta-analysis of KARE + AA decreased the p value for rs848353. For the remaining three SNPs tested, however, meta-analysis did not reduce the p values enough to reach a genome-wide significance level.

Figure 2 shows the regional association plots for three independent signals near rs4424567 at *FRMD4A*, rs848353 at 7q31.1, and rs2349433 (rs7747583) at *RGS17*. We found that two SNPs, rs7747583 and rs2349433 at *RGS17*, are in complete LD, with $r^2 = 0.99$. The regional association plots using both genotypes and imputed SNPs support our findings. As shown in Fig. 2a, SNP rs4424567 hovers in the intronic region of *FRMD4A*. Figure 2b shows that SNP rs848353 is part of a large cluster of associated SNPs near *C7orf66*. Figure 2c shows the dense cluster of associated SNPs, including the lead SNP rs2349433, in *RGS17*.

Discussion

This study represents the first large-scale GWAS on smoking-related behaviors, SI and ND, in an Asian population. Among 352,228 genotyped SNPs, our initial GWA analysis identified 12 ($p < 1 \times 10^{-5}$), among which 8 were significantly associated with SI and 4 with ND. In the replication study with American samples of both EA and AA origin, rs7747583 and rs2349433 in *RGS17* were replicated in association with SI, and rs4424567 in *FRMD4A* and rs848353 at 7q31.1 were replicated in association with ND at the 5% significance level. In an additional meta-analysis for these variants, rs848353 reached genome-wide significance (p value$_{meta} = 9.16 \times 10^{-8}$), whereas the three did not.

Three replicated loci have some biologically plausible interpretation supported by previous studies. First, consider FERM domain containing 4A (*FRMD4A*), which is homologous in mouse EpH4 epithelial cells. This gene encodes a scaffolding protein controlling the activation of Arf6 by connecting Par-3, the Arf6 guanine-nucleotide exchange factor (GEF), and cytohesin-1 (Ikenouchi and Umeda 2010). Previously, a hypermethylation event was detected in *FRMD4A* in an analysis of DNA methylation of benzopyrene (BaP)-treated MCF-7 cells (Sadikovic et al. 2007). BaP is a common polycyclic aromatic hydrocarbon, an important class of environmental contaminants, that is also present in cigarette smoke. Moreover, Johnson et al. (2008) reported that *FRMD4A* was associated with substance dependence phenotypes defined by Diagnostic and Statistical Manual (DSM) diagnoses of substance abuse and FTND diagnoses of ND. *FRMD4A* is responsible for variation of the electrocardiographic RR interval (Marroni et al. 2009).

Second, we should consider the following genes near rs848353 at 7q31.1: *C7orf66*, *LOC154907*, and *PNPLA8-THAP5-DNAJB9.* Although the function of the hypothetical protein-coding genes *C7orf66* and *LOC154907* remains to be determined, some biological interpretation of *PNPLA8-THAP5-DNAJB9* is plausible. Patatin-like phospholipase domain containing 8 (*PNPLA8*), also known as phospholipase A2, catalyzes cleavage of fatty acids from phospholipids. The encoded product of *PNPLA8* regulates membrane physical properties and the release of lipid second messengers and growth factors. *THAP5* is a THAP domain containing 5 isoform 2, and *DNAJB9* is a DnaJ (Hsp40) homolog, subfamily B, member 9. Numerous previous association findings at 7q31.1 for various phenotypes, such as death among heart failure patients (Morrison et al. 2010), ulcerative colitis (Asano et al. 2009; Barrett et al. 2009; McGovern et al. 2010; Silverberg et al. 2009), cognitive performance (Need et al. 2009), attention deficit hyperactivity disorder (Anney et al. 2008; Lasky-Su et al. 2008), and serum metabolites (Gieger et al. 2008), may imply an important role for this locus in the etiology of these complex disorder/phenotypes.

Third, the encoded protein of *RGS17* is a member of the regulator of G-protein signaling family. The gene product of *DCAF12*, *DDB1*, and *CUL4* associated factor 12 interacts with the COP9 signalosome, a macromolecular complex that interacts with cullin-RING E3 ligases and regulates their activity by hydrolyzing cullin-Nedd8 conjugates. The locus 6q25.2 containing *RGS17* was previously reported to be associated with major depressive and bipolar disorders (Ferreira et al. 2008; Liu et al. 2011), general cognitive ability (Davis et al. 2010), and tonometry (Levy et al. 2007).

Finally, a recent GWAS identified *FRMD4A* associated with substance dependence (Johnson et al. 2008). It is interesting that our result is consistent with that of Johnson et al., as we focused on ND, whereas Johnson et al. focused on substance dependence phenotypes assessed by DSM and FTND. However, Johnson et al. did not report any replication result. As far as we know, the present study is the first GWAS to show that *FRMD4A* is directly associated with ND with an independent replication.

Interestingly, our replication results appeared to be sex or population specific. For example, the SNPs at *FRMD4A* and *RGS17* were replicated only in the male group of the EA sample. An SNP at 7q31.1 was replicated in the total AA sample but not the EA sample. The analysis result suggests that *FRMD4A* and *RGS17* are associated with ND and SI in a sex-specific manner in conjunction with environmental and epigenetic factors. The complex interaction of genetic, environmental, and epigenetic factors may contribute to the disparity of the results for the total and male samples in the replication study. The discrepancy between the results of the EA and AA samples in the replication samples may be secondary to differences in genetic architecture across populations. For example, two ND-associated SNPs, rs1404697 and rs848353 at 7q31.1, are 4.0 kb apart. The two variants are highly correlated in the Korean population ($r^2 = 0.98$) and the European population ($r^2 = 0.85$, HapMap CEU). However, a low correlation was observed in an African population ($r^2 = 0.013$, HapMap YRI). As shown in Table 2, rs848353 and rs1404697 demonstrate a great difference in allele frequency in the AA sample (0.05 for rs1404697 and 0.32 for rs848353). Also note that rs848353 was replicated in the AA sample, whereas rs1404697 was not.

Of a note, we found that the mode of inheritance that provided the most significant association results also differed from population to population. This is probably attributable to the fact that genetic diversity among populations accounts for the replicated association under different genetic inheritance models in samples with different ancestry. For 7q31.1, the additive model provided the strongest signal in the discovery stage, whereas the dominant model provided the strongest signal in the AA replication sample. For *RGS17*, the additive model provided the strongest signal in the discovery sample, whereas the recessive model provided the strongest signal in the EA replication sample.

Because of the recruitment criteria used in the MSTF study, which differed from the KARE sample, there were only limited former smokers in the MSTF samples. Thus, we could not find an exactly same measure for both samples on smoking initiation phenotype. The age of onset in the MSTF sample is the age at which regular smoking was started, providing information on smoking initiation. Therefore, we used the age of onset phenotype in the replication study based on the available clinical information in MSTF. Also, the CPD in the MSTF sample is categorized CPD, as used in the KARE samples. In addition, we analyzed HSI and FTND for MSTF sample.

In this study, we used both ordinal phenotype (SI-2) and binary phenotype (SI-1) for smoking initiation. SI-2 is the modified SI-1 (never vs. others) considering the subcategories of “others.” Because the individuals in “former” group were those who smoked in the past but had stopped at the time of recruiting, we placed the “former smoker” between “never” and “occasional” smoker in the context of smoking initiation and current smoking interest.

However, there would be an argument for the order of the ordinal phenotypes. Our previous candidate gene-based association study using the same ordinal phenotype definition has replicated the previously well-known smoking-related gene *CHRAN5/A3/B5* (Li et al. 2010), which demonstrates the appropriateness of ordinal phenotype definition. In addition, the statistical test for checking the proportional odds assumption with respect to SNP variables showed that the proportionality assumption is reasonable.

For phenotype definition, of course, it would be great if we had biomarker information on those samples, which is impossible for the time being, given the assay cost associated with biomarkers and the unavailability of materials for many of those samples in the Korean population. As far as we know, however, smoking behaviors in almost all reported studies were based on questionnaires, with the number of cigarettes smoked per day being the most commonly used (Li 2008), and only limited studies used the biomarkers. Thus, further GWAS study on ND in an Asian sample with better assessment of smoking behavior, including biomarker information, is warranted.

Acknowledgments

This work was supported by a grant from the National Institute of Health of Korea 2009-E73007-00 and National Research Foundation NRF-2008-313-C00146 to TP; and National Institutes of Health grant DA-12844 to MDL. The KARE data were obtained from the Korean Genome Analysis Project 4845-301, which was funded by a grant from the Korea National Institute of Health (Korea Center for Disease Control, Ministry for Health, Welfare and Family Affairs), Republic of Korea. We are grateful for the invaluable contributions of clinical information and tissue samples by the participants in this study, as well as for the dedicated work of the research staff at the clinical sites and molecular research laboratory. We thank Nam H. Cho (Department of Preventive Medicine, Ajou University, South Korea) and Chol Shin (Department of Internal Medicine, Korea University Ansan Hospital, South Korea) for their great efforts in generating and providing these valuable data. We also thank Dr. David L. Bronson for his excellent editing of this manuscript.

References

Anney RJ, Lasky-Su J, O'Dushlaine C, Kenny E, Neale BM, Mulligan A, Franke B, Zhou K, Chen W, Christiansen H, Arias-Vasquez A, Banaschewski T, Buitelaar J, Ebstein R, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E, Steinhausen H, Asherson P, Faraone SV, Gill M. Conduct disorder and ADHD: evaluation of conduct problems as a categorical and quantitative trait in the international multicentre ADHD genetics study. Am J Med Genet B Neuropsychiatr Genet. 2008; 147B(8):1369–1378. [PubMed: 18951430]

Asano K, Matsushita T, Umeno J, Hosono N, Takahashi A, Kawaguchi T, Matsumoto T, Matsui T, Kakuta Y, Kinouchi Y, Shimosegawa T, Hosokawa M, Arimura Y, Shinomura Y, Kiyohara Y, Tsunoda T, Kamatani N, Iida M, Nakamura Y, Kubo M. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. Nat Genet. 2009; 41(12):1325–1329. [PubMed: 19915573]

Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A, Wesley E, Parnell K, Zhang H, Drummond H, Nimmo ER, Massey D, Blaszczyk K, Elliott T, Cotterill L, Dallal H, Lobo AJ, Mowat C, Sanderson JD, Jewell DP, Newman WG, Edwards C, Ahmad T, Mansfield JC, Satsangi J, Parkes M, Mathew CG, Donnelly P, Peltonen L, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin A, Craddock N, Deloukas P, Duncanson A, Jankowski J, Markus HS, McCarthy MI, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Samani N, Trembath RC, Viswanathan AC, Wood N, Spencer CC, Bellenguez C, Davison D, Freeman C, Strange A, Langford C, Hunt SE, Edkins S, Gwilliam R, Blackburn H, Bumpstead SJ, Dronov S, Gillman M, Gray E, Hammond N, Jayakumar A, McCann OT, Liddle J, Perez ML, Potter SC, Ravindrarajah R, Ricketts M, Waller M, Weston P, Widaa S, Whittaker P, Attwood AP, Stephens J, Sambrook J, Ouwehand WH, McArdle WL, Ring SM, Strachan DP. Genome-wide association study of ulcerative colitis identifies three new

susceptibility loci, including the HNF4A region. Nat Genet. 2009; 41(12):1330–1334. [PubMed: 19915572]

Bender R, Grouven U. Ordinal logistic regression in medical research. J R Coll Physicians Lond. 1997; 31(5):546–551. [PubMed: 9429194]

Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H, Waterworth D, Muglia P, Mooser V. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. Mol Psychiatry. 2008; 13(4):368–373. [PubMed: 18227835]

Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, Fugman D, Goate AM, Hinrichs AL, Konvicka K, Martin NG, Montgomery GW, Saccone NL, Saccone SF, Wang JC, Chase GA, Rice JP, Ballinger DG. Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet. 2007; 16(1):24–35. [PubMed: 17158188]

Caporaso N, Gu F, Chatterjee N, Sheng-Chih J, Yu K, Yeager M, Chen C, Jacobs K, Wheeler W, Landi MT, Ziegler RG, Hunter DJ, Chanock S, Hankinson S, Kraft P, Bergen AW. Genome-wide and candidate gene association study of cigarette smoking behaviors. PLoS ONE. 2009; 4(2):e4653. [PubMed: 19247474]

Carmelli D, Swan GE, Robinette D, Fabsitz R. Genetic influence on smoking—a study of male twins. N Engl J Med. 1992; 327(12):829–833. [PubMed: 1508241]

Cheng LS, Swan GE, Carmelli D. A genetic analysis of smoking behavior in family members of older adult males. Addiction. 2000; 95(3):427–435. [PubMed: 10795363]

Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, Yoon D, Lee MH, Kim DJ, Park M, Cha SH, Kim JW, Han BG, Min H, Ahn Y, Park MS, Han HR, Jang HY, Cho EY, Lee JE, Cho NH, Shin C, Park T, Park JW, Lee JK, Cardon L, Clarke G, McCarthy MI, Lee JY, Oh B, Kim HL. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet. 2009; 41(5):527–534. [PubMed: 19396169]

Davis OS, Butcher LM, Docherty SJ, Meaburn EL, Curtis CJ, Simpson MA, Schalkwyk LC, Plomin R. A three-stage genome-wide association study of general cognitive ability: hunting the small effects. Behav Genet. 2010; 40(6):759–767. [PubMed: 20306291]

Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, MacIntyre DJ, MacLean AW, St CD, Robinson M, Van Beck M, Pereira AC, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Ferrier IN, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans PA, Blackwood DH, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat Genet. 2008; 40(9):1056–1058. [PubMed: 18711365]

Fisher, RA. Statistical methods for research workers. 4th edn. Oliver and Boyd; Edinburgh: 1932.

Gieger C, Geistlinger L, Altmaier E, Hrabe de Angelis M, Kronenberg F, Meitinger T, Mewes HW, Wichmann HE, Weinberger KM, Adamski J, Illig T, Suhre K. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet. 2008; 4(11):e1000282. [PubMed: 19043545]

Giovino GA. The tobacco epidemic in the United States. Am J Prev Med. 2007; 33(6 Suppl):S318–S326. [PubMed: 18021906]

Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict. 1991; 86(9): 1119–1127. [PubMed: 1932883]

Hughes JR. Genetics of smoking: a brief review. Behav Ther. 1986; 17:335–345.

Ikenouchi J, Umeda M. FRMD4A regulates epithelial polarity by connecting Arf6 activation with the PAR complex. Proc Natl Acad Sci USA. 2010; 107(2):748–753. [PubMed: 20080746]

Johnson C, Drgon T, Liu QR, Zhang PW, Walther D, Li CY, Anthony JC, Ding Y, Eaton WW, Uhl GR. Genome wide association for substance dependence: convergent results from epidemiologic and research volunteer samples. BMC Med Genet. 2008; 9:113. [PubMed: 19094236]

Lange C, Silverman EK, Xu X, Weiss ST, Laird NM. A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics. 2003; 4(2):195–206. [PubMed: 12925516]

Lasky-Su J, Anney RJ, Neale BM, Franke B, Zhou K, Maller JB, Vasquez AA, Chen W, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E, Steinhausen HC, Taylor E, Daly M, Laird N, Lange C, Faraone SV. Genome-wide association scan of the time to onset of attention deficit hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet. 2008; 147B(8):1355–1358. [PubMed: 18937294]

Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang SJ, Vasan RS, Mitchell GF. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. BMC Med Genet. 2007; 8(Suppl 1):S3. [PubMed: 17903302]

Li MD. Identifying susceptibility loci for nicotine dependence: 2008 update based on recent genome-wide linkage analyses. Hum Genet. 2008; 123(2):119–131. [PubMed: 18205015]

Li MD, Burmeister M. New insights into the genetics of addiction. Nat Rev Genet. 2009; 10(4):225–231. [PubMed: 19238175]

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. [PubMed: 12492752]

Li MD, Beuten J, Ma JZ, Payne TJ, Lou XY, Garcia V, Duenes AS, Crews KM, Elston RC. Ethnic- and gender-specific association of the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) with nicotine dependence. Hum Mol Genet. 2005; 14(9):1211–1219. [PubMed: 15790597]

Li MD, Payne TJ, Ma JZ, Lou XY, Zhang D, Dupont RT, Crews KM, Somes G, Williams NJ, Elston RC. A genomewide search finds major susceptibility loci for nicotine dependence on chromosome 10 in African Americans. Am J Hum Genet. 2006; 79(4):745–751. [PubMed: 16960812]

Li MD, Yoon D, Lee JY, Han BG, Niu T, Payne TJ, Ma JZ, Park T. Associations of variants in CHRAN5/A3/B4 gene cluster with smoking behavior in a Korean population. PloS One. 2010; 5(8):e12183. 16. [PubMed: 20808433]

Liu YZ, Pei YF, Guo YF, Wang L, Liu XG, Yan H, Xiong DH, Zhang YP, Levy S, Li J, Haddock CK, Papasian CJ, Xu Q, Ma JZ, Payne TJ, Recker RR, Li MD, Deng HW. Genome-wide association analyses suggested a novel mechanism for smoking behavior regulated by IL15. Mol Psychiatry. 2009; 14(7):668–680. [PubMed: 19188921]

Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G, Vollenweider P, Preisig M, Wareham NJ, Zhao JH, Loos RJ, Barroso I, Khaw KT, Grundy S, Barter P, Mahley R, Kesaniemi A, McPherson R, Vincent JB, Strauss J, Kennedy JL, Farmer A, McGuffin P, Day R, Matthews K, Bakke P, Gulsvik A, Lucae S, Ising M, Brueckl T, Horstmann S, Wichmann HE, Rawal R, Dahmen N, Lamina C, Polasek O, Zgaga L, Huffman J, Campbell S, Kooner J, Chambers JC, Burnett MS, Devaney JM, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein S, Wilson JF, Wild SH, Campbell H, Vitart V, Reilly MP, Li M, Qu L, Wilensky R, Matthai W, Hakonarson HH, Rader DJ, Franke A, Wittig M, Schafer A, Uda M, Terracciano A, Xiao X, Busonero F, Scheet P, Schlessinger D, St Clair D, Rujescu D, Abecasis GR, Grabe HJ, Teumer A, Volzke H, Petersmann A, John U, Rudan I, Hayward C, Wright AF, Kolcic I, Wright BJ, Thompson JR, Balmforth AJ, Hall AS, Samani NJ, Anderson CA, Ahmad T, Mathew CG, Parkes M, Satsangi J, Caulfield M, Munroe PB, Farrall M, Dominiczak A, Worthington J, Thomson W, Eyre S, Barton A, Mooser V, Francks C, Marchini J. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet. 2010; 42(5):436–440. [PubMed: 20418889]

Liu Y, Blackwood DH, Caesar S, de Geus EJ, Farmer A, Ferreira MA, Ferrier IN, Fraser C, Gordon-Smith K, Green EK, Grozeva D, Gurling HM, Hamshere ML, Heutink P, Holmans PA, Hoogendijk WJ, Hottenga JJ, Jones L, Jones IR, Kirov G, Lin D, McGuffin P, Moskvina V, Nolen WA, Perlis RH, Posthuma D, Scolnick EM, Smit AB, Smit JH, Smoller JW, St Clair D, van Dyck R, Verhage M, Willemsen G, Young AH, Zandbelt T, Boomsma DI, Craddock N, O'Donovan MC, Owen MJ, Penninx BW, Purcell S, Sklar P, Sullivan PF. Meta-analysis of genome-wide

association data of bipolar disorder and major depressive disorder. Mol Psychiatry. 2011; 16(1):2–4. [PubMed: 20351715]

Ma JZ, Beuten J, Payne TJ, Dupont RT, Elston RC, Li MD. Haplotype analysis indicates an association between the DOPA decarboxylase (DDC) gene and nicotine dependence. Hum Mol Genet. 2005; 14(12):1691–1698. [PubMed: 15879433]

Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007; 39(7):906–913. [PubMed: 17572673]

Marroni F, Pfeufer A, Aulchenko YS, Franklin CS, Isaacs A, Pichler I, Wild SH, Oostra BA, Wright AF, Campbell H, Witteman JC, Kaab S, Hicks AA, Gyllensten U, Rudan I, Meitinger T, Pattaro C, van Duijn CM, Wilson JF, Pramstaller PP. A genome-wide association scan of RR and QT interval duration in 3 European genetically isolated populations: the EUROSPAN project. Circ Cardiovasc Genet. 2009; 2(4):322–328. [PubMed: 20031603]

McGovern DP, Gardet A, Torkvist L, Goyette P, Essers J, Taylor KD, Neale BM, Ong RT, Lagace C, Li C, Green T, Stevens CR, Beauchamp C, Fleshner PR, Carlson M, D'Amato M, Halfvarson J, Hibberd ML, Lordal M, Padyukov L, Andriulli A, Colombo E, Latiano A, Palmieri O, Bernard EJ, Deslandres C, Hommes DW, de Jong DJ, Stokkers PC, Weersma RK, Sharma Y, Silverberg MS, Cho JH, Wu J, Roeder K, Brant SR, Schumm LP, Duerr RH, Dubinsky MC, Glazer NL, Haritunians T, Ippoliti A, Melmed GY, Siscovick DS, Vasiliauskas EA, Targan SR, Annese V, Wijmenga C, Pettersson S, Rotter JI, Xavier RJ, Daly MJ, Rioux JD, Seielstad M. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat Genet. 2010; 42(4):332–337. [PubMed: 20228799]

Morrison AC, Felix JF, Cupples LA, Glazer NL, Loehr LR, Dehghan A, Demissie S, Bis JC, Rosamond WD, Aulchenko YS, Wang YA, Haritunians T, Folsom AR, Rivadeneira F, Benjamin EJ, Lumley T, Couper D, Stricker BH, O'Donnell CJ, Rice KM, Chang PP, Hofman A, Levy D, Rotter JI, Fox ER, Uitterlinden AG, Wang TJ, Psaty BM, Willerson JT, van Duijn CM, Boerwinkle E, Witteman JC, Vasan RS, Smith NL. Genomic variation associated with mortality among adults of European and African ancestry with heart failure: the cohorts for heart and aging research in genomic epidemiology consortium. Circ Cardiovasc Genet. 2010; 3(3):248–255. [PubMed: 20400778]

Muller F, Wehbe L. Smoking and smoking cessation in Latin America: a review of the current situation and available treatments. Int J Chron Obstruct Pulmon Dis. 2008; 3(2):285–293. [PubMed: 18686737]

Need AC, Attix DK, McEvoy JM, Cirulli ET, Linney KL, Hunt P, Ge D, Heinzen EL, Maia JM, Shianna KV, Weale ME, Cherkas LF, Clement G, Spector TD, Gibson G, Goldstein DB. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. Hum Mol Genet. 2009; 18(23):4650–4661. [PubMed: 19734545]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–575. [PubMed: 17701901]

Sadikovic B, Andrews J, Rodenhiser DI. DNA methylation analysis using CpG microarrays is impaired in benzopyrene exposed cells. Toxicol Appl Pharmacol. 2007; 225(3):300–309. [PubMed: 17904174]

Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas DA, Anderson W, Kong X, Rennard SI, Beaty TH, Hokanson JE, Crapo JD, Silverman EK. Genome-wide association study of smoking behaviours in patients with COPD. Thorax. Jun 16.2011 Epub ahead of print.

Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, Achkar JP, Goyette P, Scott R, Xu W, Barmada MM, Klei L, Daly MJ, Abraham C, Bayless TM, Bossa F, Griffiths AM, Ippoliti AF, Lahaie RG, Latiano A, Pare P, Proctor DD, Regueiro MD, Steinhart AH, Targan SR, Schumm LP, Kistner EO, Lee AT, Gregersen PK, Rotter JI, Brant SR, Taylor KD, Roeder K, Duerr RH. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet. 2009; 41(2):216–220. [PubMed: 19122664]

TGC. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet. 2010; 42(5):441–447. [PubMed: 20418890]

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsater A, Flex A, Aben KK, de Vegt F, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452(7187):638–642. [PubMed: 18385739]

Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Magi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tonjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Doring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Jarvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, McCarthy MI, Peltonen L, Thorsteinsdottir U, Stefansson K. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet. 2010; 42(5):448–453. [PubMed: 20418888]

Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE. Molecular genetics of nicotine dependence and abstinence: whole genome association using 520, 000 SNPs. BMC Genet. 2007; 8:10. [PubMed: 17407593]

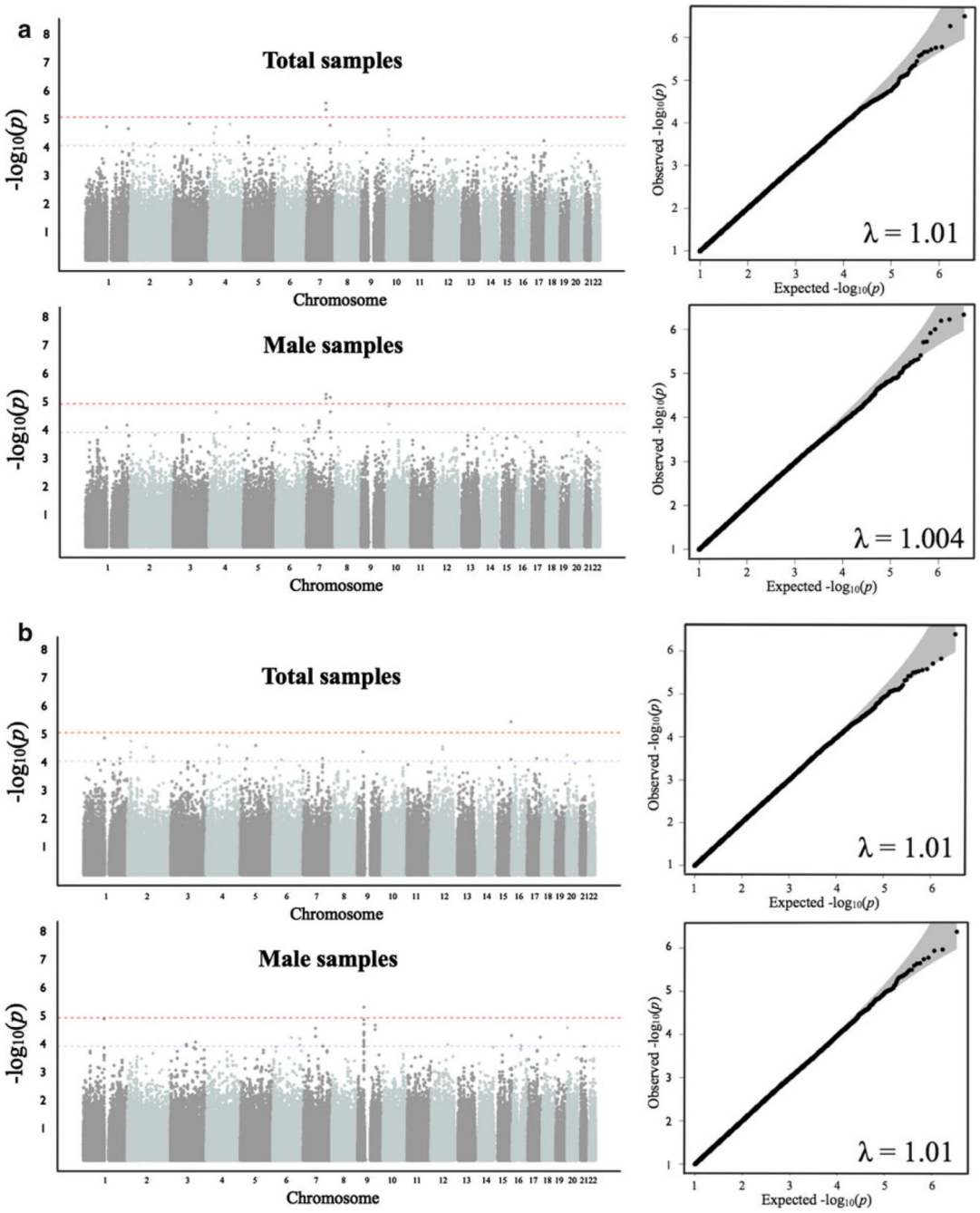
Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, David SP, Niaura R, Lerman C. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. Arch Gen Psychiatry. 2008; 65(6):683–693. [PubMed: 18519826]

Uhl GR, Drgon T, Johnson C, Ramoni MF, Behm FM, Rose JE. Genome-wide association for smoking cessation success in a trial of precessation nicotine replacement. Mol Med. 2010; 16(11-12):513–526. [PubMed: 20811658]

Vianna EO, Gutierrez MR, Barbieri MA, Caldeira RD, Bettiol H, Da Silva AA. Respiratory effects of tobacco smoking among young adults. Am J Med Sci. 2008; 336(1):44–49. [PubMed: 18626235]

Vink JM, Smit AB, de Geus EJ, Sullivan P, Willemsen G, Hottenga JJ, Smit JH, Hoogendijk WJ, Zitman FG, Peltonen L, Kaprio J, Pedersen NL, Magnusson PK, Spector TD, Kyvik KO, Morley KI, Heath AC, Martin NG, Westendorp RG, Slagboom PE, Tiemeier H, Hofman A, Uitterlinden AG, Aulchenko YS, Amin N, van Duijn C, Penninx BW, Boomsma DI. Genome-wide association study of smoking initiation and current smoking. Am J Hum Genet. 2009; 84(3):367–379. [PubMed: 19268276]

Warren CW, Jones NR, Peruga A, Chauvin J, Baptiste JP, Costa de Silva V, el Awa F, Tsouros A, Rahman K, Fishburn B, Bettcher DW, Asma S. Global youth tobacco surveillance, 2000-2007. MMWR Surveill Summ. 2008; 57(1):1–28. [PubMed: 18219269]



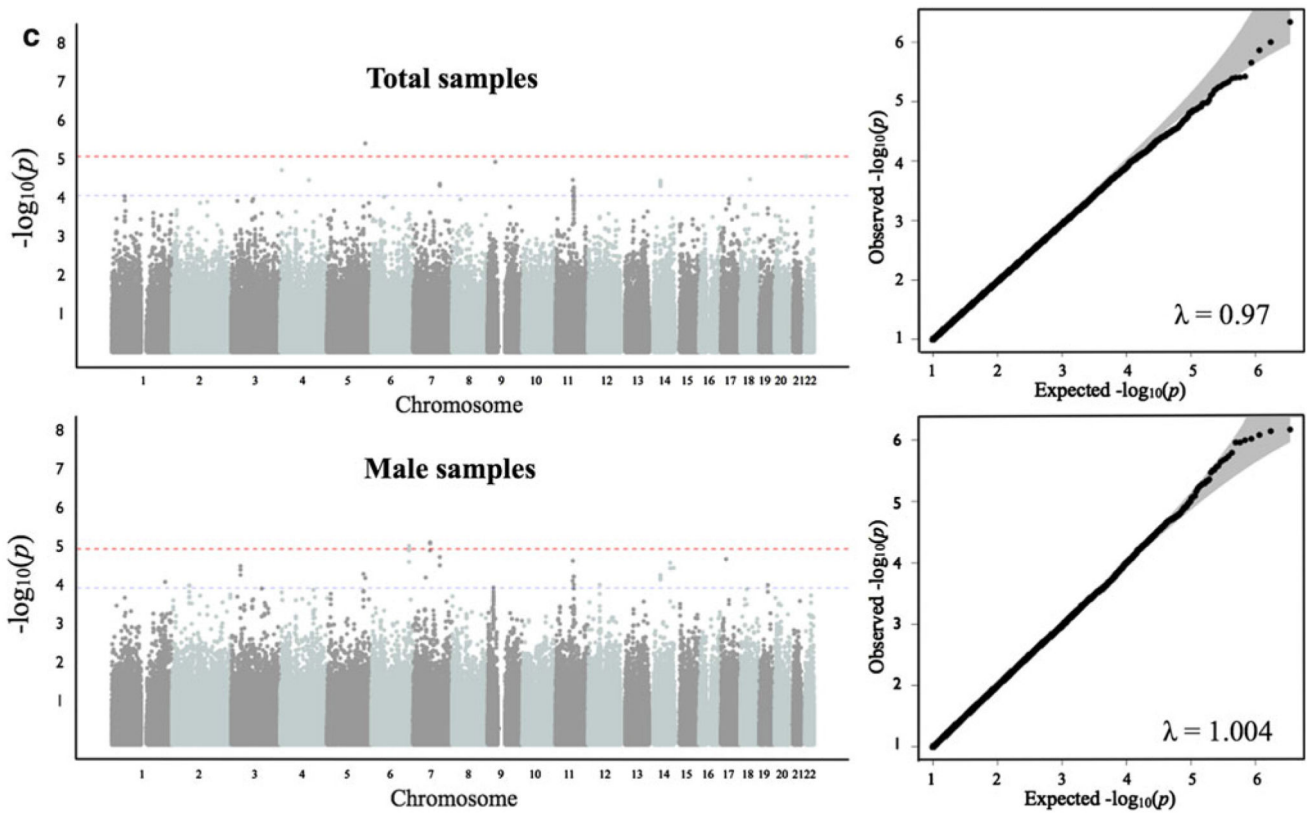


Fig. 1.

Chromosomal and QQ plots for smoking behaviors. Statistical significance of each SNP is scaled as $-\log_{10}$ (*p* value). *Red dotted line* indicates threshold in Discovery sample ($p < 10^{-5}$), and *blue dotted line* indicates *p* value 10^{-4}. We plotted the observed $-\log_{10}$-scaled association results versus $-\log_{10}$-scaled expected association results in QQ plots. Chromosomal and QQ plots for ND phenotype **(a)**, SI-1 phenotype **(b)**, and SI-2 phenotype **(c)**

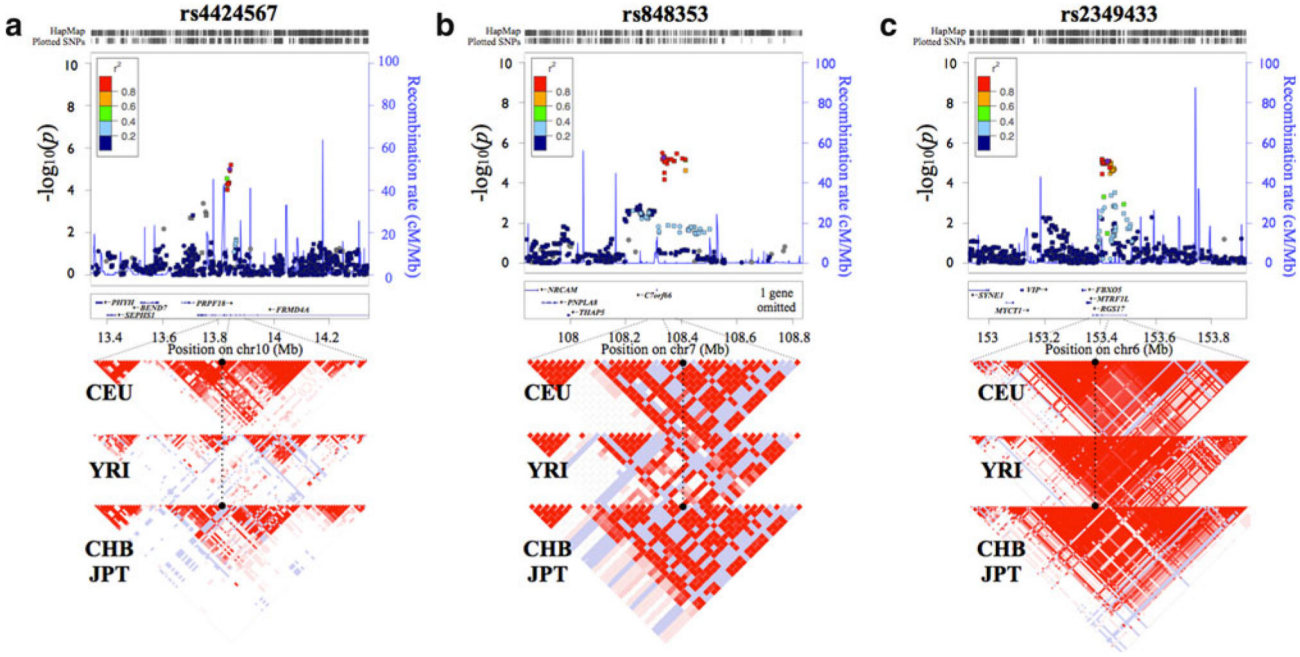


Fig. 2.

Regional plots of significantly associated SNPs. We show the locus ±500 kb around the lead associated SNP. Statistical significance of each SNP is scaled as $-\log_{10}$ (*p* value) along the chromosomal position (NCBI build 36). The lead SNP at each locus is shown in a *purple diamond.* Direct genotyping in SNP genotyping chip is in a *circle* and imputed SNP in a *square.* Pairwise correlation between the lead and other SNPs at locus is shown on a scale from minimum *(blue*) to maximum *(red).* Estimated recombination rates from HapMap are plotted in aqua. The *box* below the association signals contains gene information within each locus. The direction of each gene is shown as a left or right *arrow*, and exon is shown as a *filled square*, whereas intron is indicated as a *line.* The LD pattern is at the bottom of each regional plot. A *dotted line* indicates the ±50-kb boundary from the lead SNP. The *black circle* stands for the position of the lead associated SNP. We show the LD pattern drawn from HapMap data for the European (CEU), African (YRI), and Asian (CHB/JPT) populations. R^2 of SNPs is drawn on a scale from minimum *(white*) to maximum *(red).* Regional association plot for ND-associated SNPs, rs4424567 **(a)** and rs848353 **(b)** and SI-associated SNP, rs2349433 **(c)**

Table 1

Characteristics of KARE discovery sample

Characteristic	Total	Male	Female
Sample size (%)	8,842	4,183 (47.31)	4,659 (52.69)
Mean age (years) ± SE	52.22 ± 8.91	51.78 ± 8.78	52.61 ± 9.01
SI-1 (%)			
Controls	5,143	807 (15.69)	4,336 (84.31)
Cases	3,582	3,357 (93.72)	225 (6.28)
SI-2 (%)			
Non-smokers	5,143	807 (15.69)	4,336 (84.31)
Former smokers	1,354	1,293 (95.49)	61 (4.51)
Occasional smokers	258	199 (77.13)	59 (22.87)
Habitual smokers	1,970	1,865 (94.67)	105 (5.33)
CPD (%)			
1–10	441	375 (85.03)	66 (14.97)
11–20	1,179	1,148 (97.37)	31 (2.63)
21–30	209	206 (98.56)	3 (1.44)
>30	129	127 (98.45)	2 (1.55)

Table 2

Characteristics of African-American and European-American samples

Characteristic	African-American	European-American
No. of nuclear families	402	200
Avg. members/family	3.14 ± 0.75	3.17 ± 0.69
No. of subjects	1,366	671
% female	66.1	69.5
Age (years)	39.4 ± 14.4	40.5 ± 15.5
No. of smokers	1,053	515
Age of smoking onset (years)	17.3 ± 4.7	15.5 ± 4.4
Years smoked	20.4 ± 12.5	23.2 ± 13.5
CPD	19.4 ± 13.3	19.5 ± 13.4
HSI	3.7 ± 1.4	3.9 ± 1.4
FTND score	6.26 ± 2.15	6.33 ± 2.22

CPD number of cigarettes smoked per day, *HSI* Heaviness of Smoking Index, *FTND* Fagerström test for ND

Table 3

Result of Genome-Wide Association Study for smoking initiation and nicotine dependence for total and male samples in discovery stage

Sample	Phenotype	Chr.	SNP ID	Related gene	Alleles[a]	AF	KARE discovery	
						KARE	OR	*p* value*
Total	SI-1	15	rs1442815	-	G/A	0.44	1.254	4.10×10^{-6}
	SI-2	5	rs11134474	-	A/G	0.04	0.645	4.60×10^{-6}
	ND	7	rs1404697	-	C/G	0.13	1.384	3.19×10^{-6}
		7	rs848353	-	T/C	0.13	1.37	5.43×10^{-6}
Male	SI-1	9	rs10511914	DCAF12	T/C	0.28	0.757	4.23×10^{-6}
	SI-2	6	rs7761503	RGS17	T/C	0.43	0.833	9.46×10^{-6}
		6	rs7747583	RGS17	C/A	0.43	0.833	9.96×10^{-6}
		6	rs2349433	RGS17	G/A	0.43	0.831	8.19×10^{-6}
		7	rs1385706	-	G/A	0.32	1.221	6.69×10^{-6}
		7	rs12534697	-	G/C	0.32	1.219	7.12×10^{-6}
	ND	7	rs1404697	-	C/G	0.13	1.411	4.63×10^{-6}
		7	rs848353	-	T/C	0.13	1.401	6.40×10^{-6}
		7	rs12536963	EXOC4	T/G	0.29	1.29	5.92×10^{-6}
		10	rs4424567	FRMD4A	G/A	0.46	1.249	9.90×10^{-6}

AF allele frequency; the frequency is for the minor allele in KARE sample, *OR* odds ratio

*Additive model

[a]Major/minor alleles in KARE

Table 4

Result of replication and meta-analysis for smoking initiation and nicotine dependence for total and male samples

Sample	Chr	rs ID	Related gene	Allele*	AF		Pheno-type	Replication		Meta-analysis *p* value		
					EA	AA		EA	AA	KARE + EA	KARE + AA	KARE + EA + AA
Total	15	rsl442815	-	G/A	0.3	0.18	Age onset	4.37×10^{-1} [r]	9.22×10^{-1} [r]	2.55×10^{-5}	5.10×10^{-5}	1.70×10^{-4}
	5	rs11134474	-	A/G	0.01	0.12		-1.92×10^{-1} [a]	-5.69×10^{-2} [r]	1.32×10^{-5}	4.23×10^{-6}	7.99×10^{-6}
	7	rs1404697	-	C/G	0.05	0.05	CPD	1.98×10^{-1} [a]	-1.45×10^{-1} [r]	9.65×10^{-6}	7.21×10^{-6}	1.36×10^{-5}
							HSI	2.66×10^{-1} [a]	-6.06×10^{-1} [r]	1.27×10^{-5}	3.18×10^{-6}	8.16×10^{-6}
							FTND	1.66×10^{-1} [a]	-5.37×10^{-2} [r]	8.18×10^{-6}	2.84×10^{-6}	4.82×10^{-6}
	7	rs848353	-	T/C	0.06	0.32	CPD	-3.17×10^{-1} [r]	4.57×10^{-3} [d]	2.46×10^{-5}	4.59×10^{-7}	1.52×10^{-6}
							HSI	-3.17×10^{-1} [r]	2.63×10^{-3} [d]	2.46×10^{-5}	2.72×10^{-7}	9.27×10^{-7}
							FTND	-3.17×10^{-1} [r]	8.35×10^{-4} [d]	2.46×10^{-5}	9.16×10^{-8}	3.29×10^{-7}
Male	9	rs10511914	DCAF12	T/C	0.74	0.64	Age onset	3.97×10^{-1} [r]	6.67×10^{-1} [d]	2.40×10^{-5}	3.89×10^{-5}	1.22×10^{-4}
	6	rs7761503	RGS17	T/C	0.72	0.48		7.84×10^{-2} [r]	-3.43×10^{-1} [r]	1.12×10^{-5}	4.43×10^{-5}	3.34×10^{-5}
	6	rs7747583	RGS17	C/A	0.72	0.48		-4.09×10^{-2} [r]	-3.54×10^{-1} [r]	6.40×10^{-6}	4.78×10^{-5}	2.03×10^{-5}
	6	rs2349433	RGS17	G/A	0.72	0.48		-4.29×10^{-2} [r]	-6.61×10^{-1} [r]	5.57×10^{-6}	7.11×10^{-5}	3.09×10^{-5}
	7	rs1385706	-	G/A	0.16	0.19		3.75×10^{-1} [a]	-2.27×10^{-1} [d]	3.49×10^{-5}	2.19×10^{-5}	6.76×10^{-5}
	7	rs12534697	-	G/C	0.15	0.13		3.96×10^{-1} [a]	3.13×10^{-1} [r]	3.89×10^{-5}	3.12×10^{-5}	9.89×10^{-5}
	7	rs1404697	-	C/G	0.05	0.05	CPD	4.67×10^{-1} [a]	-4.91×10^{-1} [a]	3.04×10^{-5}	3.18×10^{-5}	1.16×10^{-4}
							HSI	5.57×10^{-1} [a]	-6.90×10^{-1} [a]	3.58×10^{-5}	4.36×10^{-5}	1.81×10^{-4}
							FTND	6.99×10^{-1} [a]	-6.62×10^{-1} [a]	4.41×10^{-5}	4.20×10^{-5}	2.13×10^{-4}
	7	rs848353	-	T/C	0.06	0.32	CPD	7.23×10^{-1} [a]	-1.14×10^{-1} [r]	6.15×10^{-5}	1.10×10^{-5}	6.33×10^{-5}
							HSI	5.65×10^{-1} [a]	-2.40×10^{-1} [r]	4.89×10^{-5}	2.21×10^{-5}	9.75×10^{-5}
							FTND	4.84×10^{-1} [a]	-4.32×10^{-1} [r]	4.24×10^{-5}	3.82×10^{-5}	1.42×10^{-4}
	7	rs12536963	EXOC4	T/G	-	-	-	-	-	-	-	-

Sample	Chr	rs ID	Related gene	Allele*	AF		Pheno-type	Replication		Meta-analysis *p* value		
					EA	AA		EA	AA	KARE + EA	KARE + AA	KARE + EA + AA
	10	rs4424567	FRMD4A	G/A	0.73	0.83	CPD	1.38×10^{-2} [a]	-7.45×10^{-1} [r]	2.30×10^{-6}	9.45×10^{-5}	1.49×10^{-5}
							HSI	1.51×10^{-2} [a]	5.62×10^{-1} [d]	2.50×10^{-6}	7.29×10^{-5}	1.26×10^{-5}
							FTND	4.75×10^{-2} [a]	4.09×10^{-1} [d]	7.32×10^{-6}	5.43×10^{-5}	2.62×10^{-5}

AF allele frequency; the frequency is for the minor allele in KARE

-: These SNPs showed a protective effect in replication study

*

[a] Additive

[d] dominant, and

[r] recessive models were used in the association