

# Smoking and Genetic Risk Variation Across Populations of European, Asian, and African American Ancestry—A Meta-Analysis of Chromosome 15q25

Li-Shiun Chen,<sup>1\*</sup> Nancy L. Saccone,<sup>2</sup> Robert C. Culverhouse,<sup>3</sup> Paige M. Bracci,<sup>4</sup> Chien-Hsiun Chen,<sup>5,6</sup> Nicole Dueker,<sup>7</sup> Younghun Han,<sup>8</sup> Hongyan Huang,<sup>9</sup> Guangfu Jin,<sup>10</sup> Takashi Kohno,<sup>11</sup> Jennie Z. Ma,<sup>12</sup> Thomas R. Przybeck,<sup>1</sup> Alan R. Sanders,<sup>13</sup> Jennifer A. Smith,<sup>14</sup> Yun Ju Sung,<sup>9</sup> Angie S. Wenzlaff,<sup>15</sup> Chen Wu,<sup>16</sup> Dankyu Yoon,<sup>17,18</sup> Ying-Ting Chen,<sup>5</sup> Yu-Ching Cheng,<sup>19</sup> Yoon Shin Cho,<sup>20,21</sup> Sean P. David,<sup>22,23,24</sup> Jubao Duan,<sup>13</sup> Charles B. Eaton,<sup>24</sup> Helena Furberg,<sup>25</sup> Alison M. Goate,<sup>1</sup> Dongfeng Gu,<sup>26,27</sup> Helen M. Hansen,<sup>28</sup> Sarah Hartz,<sup>1</sup> Zhibin Hu,<sup>10</sup> Young Jin Kim,<sup>17,22</sup> Steven J. Kittner,<sup>29</sup> Douglas F. Levinson,<sup>30</sup> Thomas H. Mosley,<sup>14</sup> Thomas J. Payne,<sup>31</sup> D. C. Rao,<sup>9</sup> John P. Rice,<sup>1</sup> Treva K. Rice,<sup>9</sup> Tae-Hwi Schwantes-An,<sup>2</sup> Sanjay S. Shete,<sup>8</sup> Jianxin Shi,<sup>32</sup> Margaret R. Spitz,<sup>8</sup> Yan V. Sun,<sup>14</sup> Fuu-Jen Tsai,<sup>33</sup> Jen C. Wang,<sup>1</sup> Margaret R. Wrensch,<sup>28</sup> Hong Xian,<sup>3</sup> Pablo V. Gejman,<sup>13</sup> Jiang He,<sup>34</sup> Steven C. Hunt,<sup>35</sup> Sharon L. Kardia,<sup>14</sup> Ming D. Li,<sup>36</sup> Dongxin Lin,<sup>16</sup> Braxton D. Mitchell,<sup>20</sup> Taesung Park,<sup>37</sup> Ann G. Schwartz,<sup>15</sup> Hongbing Shen,<sup>10</sup> John K. Wiencke,<sup>28</sup> Jer-Yuarn Wu,<sup>5,6</sup> Jun Yokota,<sup>38</sup> Christopher I. Amos,<sup>8</sup> and Laura J. Bierut<sup>1</sup>

<sup>1</sup>Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri

<sup>2</sup>Department of Genetics, Washington University School of Medicine, St. Louis, Missouri

<sup>3</sup>Department of Internal Medicine, Washington University School of Medicine, St. Louis, Missouri

<sup>4</sup>Department of Epidemiology and Biostatistics, UCSF, San Francisco, California

<sup>5</sup>National Genotyping Center Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

<sup>6</sup>Graduate Institute of Chinese Medical Science, China Medical University, Taichung, Taiwan

<sup>7</sup>Department of Epidemiology and Public Health, University of Maryland, Baltimore, Maryland

<sup>8</sup>Department of Epidemiology Anderson Cancer Center, University of Texas M.D., Houston, Texas

<sup>9</sup>Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri

<sup>10</sup>Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing, China

<sup>11</sup>Division of Genome Biology, National Cancer Center Research Institute, Tokyo, Japan

<sup>12</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia

<sup>13</sup>Department of Psychiatry and Behavioral Sciences North Shore University Health System Research Institute, University of Chicago, Chicago, Illinois

<sup>14</sup>Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan

<sup>15</sup>Karmanos Cancer Institute, Wayne State University, Detroit, Michigan

<sup>16</sup>Department of Etiology & Carcinogenesis Cancer Institute, Chinese Academy of Medical Sciences, Beijing, China

<sup>17</sup>Interdisciplinary Program in Bioinformatics College of Natural Science, Seoul National University, Seoul, Korea

<sup>18</sup>Center for Immunology and Pathology, National Institute of Health, Seoul, Korea

<sup>19</sup>Department of Medicine, University of Maryland Medical Center, Baltimore, Maryland

<sup>20</sup>Center for Genome Science, National Institute of Health, Seoul, Korea

<sup>21</sup>Department of Biomedical Science, Hallym University, Chuncheon, Korea

<sup>22</sup>Center for Health Sciences, SRI International, Menlo Park, California

<sup>23</sup>Department of Medicine, Stanford University School of Medicine, California

<sup>24</sup>Department of Family Medicine, Brown University, Providence, Rhode Island

<sup>25</sup>Department of Epidemiology, Memorial Sloan Kettering Cancer Center, New York, New York

<sup>26</sup>Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

<sup>27</sup>Chinese National Center for Cardiovascular Disease Control and Research, Beijing, China

<sup>28</sup>Neurological Surgery Division of Epidemiology, Helen Diller Family Cancer Center, San Francisco, California

<sup>29</sup>Department of Neurology University of Maryland Baltimore, Maryland

<sup>30</sup>Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, California

<sup>31</sup>University of Mississippi Medical Center, Jackson, Mississippi

<sup>32</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

<sup>33</sup>School of Post-Baccalaureate Chinese Medicine, China Medical University, Taiwan

<sup>34</sup>Department of Epidemiology, Tulane School of Public Health and Tropical Medicine, New Orleans, Louisiana

<sup>35</sup>Department of Internal Medicine, University of Utah, Salt Lake City, Utah

<sup>36</sup>Department of Psychiatry and Neurobehavioral Sciences, University of Virginia, Charlottesville, Virginia

<sup>37</sup>Department of Statistics College of Natural Science, Seoul National University, Seoul, Korea

<sup>38</sup>Division of Multistep Carcinogenesis, National Cancer Center Research Institute, Tokyo, Japan

Recent meta-analyses of European ancestry subjects show strong evidence for association between smoking quantity and multiple genetic variants on chromosome 15q25. This meta-analysis extends the examination of association between distinct genes in the *CHRNA5-CHRNA3-CHRNA4* region and smoking quantity to Asian and African American populations to confirm and refine specific reported associations. Association results for a dichotomized cigarettes smoked per day phenotype in 27 datasets (European ancestry (N = 14,786), Asian (N = 6,889), and African American (N = 10,912) for a total of 32,587 smokers) were meta-analyzed by population and results were compared across all three populations. We demonstrate association between smoking quantity and markers in the chromosome 15q25 region across all three populations, and narrow the region of association. Of the variants tested, only rs16969968 is associated with smoking ( $P < 0.01$ ) in each of these three populations (odds ratio [OR] = 1.33, 95% CI = 1.25–1.42,  $P = 1.1 \times 10^{-17}$  in meta-analysis across all population samples). Additional variants displayed a consistent signal in both European ancestry and Asian datasets, but not in African Americans. The observed consistent association of rs16969968 with heavy smoking across multiple populations, combined with its known biological significance, suggests rs16969968 is most likely a functional variant that alters risk for heavy smoking. We interpret additional association results that differ across populations as providing evidence for additional functional variants, but we are unable to further localize the source of this association. Using the cross-population study paradigm provides valuable insights to narrow regions of interest and inform future biological experiments. *Genet. Epidemiol.* 36:340–351, 2012. © 2012

Wiley Periodicals, Inc.

**Key words:** smoking; genetics; meta-analysis; cross-population

Supporting Information is available in the online issue at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).

\*Correspondence to: Li-Shiun Chen, Department of Psychiatry (Box 8134), Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110. E-mail: [chenli@psychiatry.wustl.edu](mailto:chenli@psychiatry.wustl.edu)

Received 24 August 2011; Revised 24 January 2012; Accepted 30 January 2012

Published online 16 April 2012 in Wiley Online Library ([wileyonlinelibrary.com/journal/gepi](http://wileyonlinelibrary.com/journal/gepi)).

DOI: 10.1002/gepi.21627

## INTRODUCTION

Recent genetic meta-analyses, including tens of thousands of subjects of European ancestry, show strong evidence of association between smoking quantity (cigarettes smoked per day; CPD) and multiple genetic markers on chromosome 15q25 [Liu et al., 2010; Saccone et al., 2010; TAG, 2010; Thorgeirsson et al., 2010]. Those studies synthesized evidence across many independent datasets to highlight specific variants in the region of the *CHRNA5-CHRNA3-CHRNA4* gene cluster associated with smoking behavior in European ancestry subjects. It is important to determine the biological mechanisms underlying these associations; however, the high linkage disequilibrium (LD) in this region among individuals of European ancestry makes it difficult to differentiate potentially causal variants from the many correlated variants. Because the genetic architecture of chromosome 15q25 varies across populations, comparing associations across diverse populations with differing genetic architecture can help refine the region of association and point to variants more likely to have functional relevance [Rotimi and Jorde 2010; Saccone et al., 2008; Zaitlen et al., 2010].

The most robust genetic finding on chromosome 15q25 in subjects of European ancestry is the region tagged by rs16969968, rs1051730, and other correlated variants. This finding has been replicated for smoking-related traits in multiple distinct datasets [Baker et al., 2009; Berrettini et al., 2008; Keskitalo et al., 2009; Saccone et al., 2007, 2009; Sherva et al., 2008; Stevens et al., 2008; Thorgeirsson et al., 2008; Weiss et al., 2008] and has now been reported as the most significant genome-wide association in recent meta-analyses of European ancestry subjects (e.g. rs16969968,  $P = 5.57 \times 10^{-72}$ , or rs1051730,  $P = 2.75 \times 10^{-73}$ ) [Liu et al., 2010; Saccone et al., 2010; TAG, 2010; Thorgeirsson et al., 2010]. We will use the term “bin” to de-

note a group of correlated single nucleotide polymorphisms (SNPs) ( $r^2 \geq 0.7$ ) that may constitute the same association signal in European ancestry samples [Carlson et al., 2004]. Under this definition and using the 1000 Genomes Pilot 1 CEU as the European ancestry reference sample [Durbin et al., 2010], the single bin tagged by rs16969968 and rs1051730 includes 52 known variants. This bin, which we will call bin A, groups together and unifies the most significant meta-analysis findings as well as individual dataset reports of SNPs associated with nicotine dependence, heavy smoking, lung cancer, and other smoking-related diseases in European ancestry datasets.

There are additional markers of interest in this region that are not strongly correlated with bin A. Because of the clear association between smoking behavior and bin A, each of the large-scale meta-analyses of European ancestry samples carried out association tests conditional on bin A variants for other SNPs to determine whether additional genetic markers in 15q25 are associated after adjusting for effects of bin A [Liu et al., 2010; Saccone et al., 2010; TAG, 2010; Thorgeirsson et al., 2010]. After conditioning on bin A, the meta-analyses identified additional SNPs in this region associated with smoking behavior. These SNPs can be grouped into three distinct bins (B, C, D) (Table I). Bin B, tagged by rs588765 and rs880395, is associated with genome-wide significance among heavy vs. light smokers but only in analyses conditioning on bin A ( $P = 1.2 \times 10^{-9}$ ) [Saccone et al., 2010]. Notably, bin B is also associated with mRNA levels of *CHRNA5* in brain and lung [Falvella et al., 2010; Smith et al., 2010; Wang et al., 2009]. Bin C, tagged by rs6495308 [Liu et al., 2010], rs2036534 [Thorgeirsson et al., 2010], rs7163730, rs9788682, rs684513 [TAG, 2010], and rs578776 [Saccone et al., 2010], is associated with heavy smoking after conditioning on bin A ( $P$ -values from  $9.1 \times 10^{-5}$  to  $6.3 \times 10^{-9}$ ). In contrast to bin B, bin C is less significant in conditional analysis compared to single

**TABLE I. Genetic variants associated with smoking quantity reported in meta-analyses in subjects of European ancestry**

Bin	SNP	References	Coded allele	Coded allele frequency		
				CEU	JPT/CHB	ASW/YRI
A	rs16969968	Saccone et al. (2010)	A	0.42	0.03	0.07
A	rs1051730	TAG (2010), Liu et al. (2010), Thorgeirsson et al. (2010)	T	0.42	0.03	0.12
B	rs588765/rs880395	Saccone et al. (2010)	T	0.39	0.05	0.23
C	rs6495308	Liu et al. (2010)	C	0.20	0.80	0.31
C	rs2036534	Thorgeirsson et al. (2010)	C	0.19	0.50	0.24
C	rs7163730	TAG (2010)	G	0.19	0.49	0.25
C	rs9788682	TAG (2010)	A	0.19	0.28	0.20
C	rs684513	TAG (2010)	G	0.20	0.25	0.21
C	rs578776	Saccone et al. (2010)	T	0.24	0.87	0.55
D	rs2869046	Thorgeirsson et al. (2010)	C	0.45	0.39	0.17

Allele frequencies based on 1000 Genomes Pilot 1 and HapMap 3 Release 2.

SNP analysis. Bin D is represented by rs2869046, which also displayed residual association after conditioning on bin A ( $P = 4.8 \times 10^{-5}$ ) [Thorgeirsson et al., 2010]. Markers from these different bins (A, B, C, and D) are only modestly correlated with one another, with  $r^2 \leq 0.52$  in the 1000 Genomes Pilot 1 CEU ( $N = 180$ ; Table II).

Differences in the correlational structure of markers spanning the region 15q25 between populations result in distinct sub-bins of correlated markers among Asian and African American populations that provide an opportunity to refine the source of the previously reported signals. For example, bin A, consisting of 52 variants including rs16969968, separates into 20 sub-bins in Asians (based on 1000 Genomes Pilot 1 JPT/CHB) and 38 sub-bins in African Americans (based on combined information from the 1000 Genomes Pilot 1 YRI and HapMap 3 Release 2 ASW) [Altshuler et al., 2010]. In particular, rs16969968 and rs1051730 are highly correlated in European ancestry ( $r^2 = 1$ ) and Asian populations ( $r^2 = 1$ ), but display only moderate correlation ( $r^2 = 0.40$ ) in the African American population. These differences in genetic architecture can be used to dissect the association signals.

The purpose of this meta-analysis is to determine if bins A, B, C, and D shows consistent association with smoking behavior across populations and, if so, to leverage these differences in genetic correlation across populations to refine the genetic associations in this region previously reported in subjects of European ancestry. We expect a sub-bin showing consistent evidence across all three populations to be more likely to contain a variant altering a biological mechanism. We performed meta-analyses of results from a total of 27 datasets: nine European ancestry samples (used to evaluate consistency with previous results), seven Asian samples, and 11 African American samples. We tested for association between smoking phenotypes and the four distinct bins (A through D) across all three populations. This cross-population study therefore improves our understanding of genetic risk for smoking by highlighting potentially functional variants.

## METHODS

### SAMPLES

Results from 27 datasets, containing a total of 32,587 smokers with measures of CPD, contributed to the meta-

analyses. Of these datasets, nine consisted of European ancestry subjects ( $N = 14,786$ ), seven consisted of Asians ( $N = 6,889$ ), and 11 consisted of African Americans ( $N = 10,912$ ). Twenty datasets were samples of unrelated individuals. The remaining seven datasets were family-based studies, for which the primary analyses involved an extraction of unrelated individuals. To be included in the analyses, each subject was required to have reported smoking cigarettes in his/her lifetime. Genotyping varied among studies from extensive coverage based on genome-wide association genotyping to only a limited number of candidate SNPs genotyped in this 15q25 region. Text S1 provides additional details for each dataset, including recruitment, primary phenotypes, definitions for smokers and CPD, DNA source, genotyping platforms, and genotyping quality control. Table S1 shows the sample size and demographics for each participating dataset. Four of nine datasets of European ancestry were included in the previous report [Saccone et al., 2010] (see Table S1 for the overlap, which involves only European-ancestry samples). The informed consent from participants and approval from the appropriate institutional review boards were obtained.

### PHENOTYPES

Smoking quantity was assessed with cigarettes smoked per day (CPD). The primary phenotype for analysis was a dichotomous trait contrasting light smoking controls ( $CPD \leq 10$ ) to heavy smoking cases ( $CPD > 20$ ). In addition, a four-level ordered trait ( $CPD \leq 10$ ;  $11 \leq CPD \leq 20$ ;  $21 \leq CPD \leq 30$ ;  $CPD \geq 31$  coded as 0, 1, 2, 3, respectively) was developed for confirmatory analysis. The only exception was one study (Women's Health Initiative) that measured smoking amount with different threshold levels ( $CPD \leq 14$ ,  $15 \leq CPD \leq 24$ ,  $25 \leq CPD \leq 34$ ,  $CPD \geq 35$ ), and  $CPD \leq 14$  defined the light smoking controls that was contrasted with  $CPD \geq 25$  as heavy smoking cases.

### VARIANTS FOR ANALYSES

Multiple SNPs in 15q25 have been identified as associated with smoking behavior in studies of European ancestry subjects. We focused on the results highlighted in the most powerful studies, namely the large meta-analyses [Liu et al., 2010; Saccone et al., 2010; TAG, 2010; Thorgeirsson

**TABLE II. Correlation between the examined variants ( $r^2$ ) in European, Asian, and African American ancestry populations: Bins A, B, C, and D denote four groups of correlated SNPs ( $r^2 \geq 0.7$ ) in the European ancestry reference sample (1000 Genomes Pilot 1 CEU)**

	European ancestry										
	rs16969968	rs1051730	rs588765	rs880395	rs6495308	rs2036534	rs7163730	rs9788682	rs684513	rs578776	rs2869046
rs16969968	1.0										
rs1051730	1.0	1.0									
rs588765	0.44	0.44	1.0								
rs880395	0.46	0.46	0.76	1.0							
rs6495308	0.18	0.18	0.11	0.05	1.0						
rs2036534	0.17	0.17	0.1	0.15	0.75	1.0					
rs7163730	0.17	0.17	0.1	0.15	0.75	1.0	1.0				
rs9788682	0.17	0.17	0.1	0.15	0.75	1.0	1.0	1.0			
rs684513	0.1	0.1	0.11	0.16	0.7	0.85	0.85	0.85	1.0		
rs578776	0.23	0.23	0.04	0.01	0.78	0.66	0.66	0.66	0.61	1.0	
rs2869046	0.18	0.18	0.54	0.52	0.07	0.14	0.14	0.14	0.15	0.04	1.0
	Asian ancestry										
	rs16969968	rs1051730	rs588765	rs880395	rs6495308	rs2036534	rs7163730	rs9788682	rs684513	rs578776	rs2869046
rs16969968	1.0										
rs1051730	1.0	1.0									
rs588765	0	0	1.0								
rs880395	0	0	0.1	1.0							
rs6495308	0.1	0.1	0.21	0.08	1.0						
rs2036534	0.03	0.03	0.05	0.04	0.07	1.0					
rs7163730	0.03	0.03	0.05	0.04	0.06	0.97	1.0				
rs9788682	0.01	0.01	0.02	0	0.1	0.38	0.39	1.0			
rs684513	0.01	0.01	0.02	0.02	0.08	0.28	0.29	0.64	1.0		
rs578776	0.17	0.17	0.11	0.06	0.62	0.02	0.02	0.02	0.02	1.0	
rs2869046	0.04	0.04	0	0.01	0.01	0.11	0.13	0.13	0.09	0.02	1.0
	African American ancestry										
	rs16969968	rs1051730	rs588765	rs880395	rs6495308	rs2036534	rs7163730	rs9788682	rs684513	rs578776	rs2869046
rs16969968	1.0										
rs1051730	0.4	1.0									
rs588765	–	0.04	1.0								
rs880395	–	0.02	0.68	1.0							
rs6495308	0.04	0.06	0.14	0.11	1.0						
rs2036534	0.03	0.05	0.06	0.04	0.09	1.0					
rs7163730	0.03	0.05	0.06	0.04	0.09	1.0	1.0				
rs9788682	0.02	0.03	0.06	0.05	0.03	0.6	0.6	1.0			
rs684513	–	0.04	0.08	0.06	0.16	0.62	0.62	0.48	1.0		
rs578776	0.1	0.01	0.55	0.34	0.22	0.08	0.08	0.05	0.13	1.0	
rs2869046	–	0.03	0.03	0.01	0.13	0	0	0.04	0.02	0.02	1.0

We used SNAP to obtain LD values from HapMap 3 ASW and 1000 Genomes Pilot 1 reference populations. –, data unavailable.

et al., 2010] (Table I). Table II lists the 11 targeted SNPs and illustrates how LD (measured by  $r^2$ ) structure varies across different populations. We used SNAP [Johnson et al., 2008] with 1000 Genomes Pilot 1 reference samples [Altshuler et al., 2010; Durbin et al., 2010] and HapMap3 ASW to obtain LD estimates for our three populations: CEU for European ancestry, JPT/CHB for Asians, and ASW/YRI for African Americans.

It is important to examine not just the 11 previously identified SNPs listed in Table I, but all SNPs correlated with these 11 SNPs in Europeans. We used a two-step process to define distinct groups of correlated SNPs, which we call bins. First, we grouped previously identified SNPs by their correlation in the 1000 Genomes Pilot 1 CEU (i.e. European ancestry) reference sample, using  $r^2 \geq 0.7$  as our threshold [Durbin et al., 2010]. Under this strategy, the 11

previously identified SNPs listed in Table I are partitioned into four groups: Group A (rs16969968, rs1051730), Group B (rs588765, rs880395), Group C (rs6495308, rs2036534, rs7163730, rs9788682, rs684513, rs578776), and Group D (rs2869046) (Table II). From these four groups, we established the bins by including all SNPs correlated ( $r^2 \geq 0.7$ ) in the European ancestry reference sample with at least one of the SNPs defining the bin. The threshold of 0.7 was chosen to provide an inclusive collection of tested SNPs. Using SNAP to obtain correlated variants in a bin based on 1000 Genomes Pilot 1 CEU, we identified 52 SNPs in bin A, 111 SNPs in bin B, 82 SNPs in bin C, and 15 SNPs in bin D.

Next, we partitioned these SNPs within a bin into “sub-bins” based on  $r^2 \geq 0.8$  in the Asian and African American populations. The higher threshold of 0.8 was used for sub-bins in the other populations to refine the focus of the



analyses. In Asians, we identified 20 sub-bins for bin A, 39 sub-bins for bin B, 24 sub-bins for bin C, and seven sub-bins for bin D. In African Americans, we identified 38 sub-bins for bin A, 37 sub-bins for bin B, 26 sub-bins for bin C, and seven sub-bins for bin D.

## STATISTICAL ANALYSES AND META-ANALYSES

We evaluated the genetic associations between heavy smoking and each genotyped SNP in three populations. Standardized scripts were developed centrally by the coordinating site (Washington University) for analyses of all participating datasets at each individual research center. Results were returned to the coordinating site for quality checks and meta-analyses. Individual SNP analyses were performed using SAS (SAS Institute, Cary, NC).

In each dataset, association between heavy vs. light smoking based on CPD and all SNPs was evaluated with logistic regression models as the primary analysis. Genotypes were coded additively as the number of nonreference alleles, where the reference allele was defined as the major allele in the European ancestry population in dbSNP [Sherry et al., 2001]; consistency of allelic coding was confirmed by comparing allele labels and allele frequencies across all datasets within each population. Age as a continuous variable and gender were included as covariates. Secondary analyses of the four-level CPD trait used linear regression models with the same covariates, assuming that the trait has a simple linear relationship with the predictors.

Analyses were stratified by ancestry: European, Asian, and African American. We evaluated the effect of each bin A SNP using single SNP association analyses. For bins B, C, and D, both single SNP association and conditional analyses controlling for bin A were performed. Analyses conditional on bin A (rs16969968) served as our primary analysis model for bins B, C, and D because they were targeted due to previously reported results of analyses conditional on bin A in European ancestry meta-analyses.

For each ancestry group, every dataset with at least one genotyped SNP in a given sub-bin contributed to the meta-analysis of that sub-bin. For each sub-bin, a SNP was selected as the target. In samples where the target SNP was missing, we used the results from the SNP with highest correlation ( $r^2$ ) with the target SNP in the sub-bin defined by the 1000 Genome Pilot 1 JPT/CHB for Asians, and the 1000 Genome Pilot 1 YRI or HapMap3 ASW project for African Americans.

We used PLINK to perform meta-analyses and generate overall summary odds ratios (ORs), standard errors, and  $P$ -values [Purcell et al., 2007]. The R package, rmeta, was used to confirm results and generate meta-analysis plots [Lumley, 2009]. Meta-analyses results were based on fixed effect models to determine the evidence for association within our collected samples, so we are not making a general inference about what might be observed in other samples.

## MULTIPLE TEST CORRECTION

Our primary analysis was to determine if any intersecting sub-bins across Asians and African Americans would display evidence of consistent association when comparing heavy vs. light smokers, where we defined a consistent association as having the same direction and  $P$ -value

$< 0.01$  in both populations. Our binning strategy resulted in 100 single sub-bin tests and 67 conditional association tests across the four bins: a total of 167 tests. Because the probability of any particular test resulting in a  $P$ -value  $< 0.01$  in both non-European populations by chance would be 0.0001 ( $= 0.01 \times 0.01$ ), results consistently associated in both populations would remain significant after Bonferroni correction ( $167 \times 0.0001 < 0.05$ ).

## RESULTS

### GENETIC ASSOCIATIONS IN BIN A

Bin A (tagged by rs16969968 and rs1051730 in Europeans) includes 52 SNPs correlated ( $r^2 \geq 0.7$ ) in the European ancestry reference sample. This bin separates into 20 sub-bins in Asian populations and 38 sub-bins in African American populations. We had adequate coverage to test nine of these 20 sub-bins in Asian data and 27 of these 38 sub-bins in African American data.

We detected a strong association between the dichotomous phenotype of heavy smoking vs. light smoking and bin A in European ancestry data (OR = 1.31, 95% CI = 1.22–1.40,  $P = 1.3 \times 10^{-14}$ ). The only sub-bin showing consistent association with heavy smoking across the other two populations is tagged by rs16969968 (Asian population: OR = 1.64, 95% CI = 1.15–2.32,  $P = 5.8 \times 10^{-3}$ ; African American population: OR = 1.62, 95% CI = 1.21–2.17,  $P = 1.1 \times 10^{-3}$ ). As noted in the “Methods,” because the probability of any particular test resulting in a  $P < 0.01$  in both populations by chance alone would be 0.0001, this result of consistent association in both populations remained significant after Bonferroni correction.

Figure 1 shows all SNPs in bin A, and the only consistently associated sub-bins ( $P < 0.01$  in both Asians and African Americans). Bin A variants span six genes in the European ancestry population, the sub-bin tagged by rs16969968 in the Asian population spans three genes, and the sub-bin tagged by rs16969968 in the African American population spans only one gene (CHRNA5). Figure 2 provides a forest plot summary of the stratified meta-analyses for the bin/sub-bin tagged by rs16969968, the only consistent association for bin A, in all three populations. Each plot lists ORs for each contributing sample. The overall cross-population meta-analysis across all datasets gave an OR of 1.33 (95% CI = 1.25–1.42,  $P = 1.1 \times 10^{-17}$ ).

In European and Asian populations, rs16969968 and rs1051730 are highly correlated. However, due to the different LD structure in African Americans, rs16969968 and rs1051730 represent two different sub-bins ( $r^2 = 0.40$  in HapMap 3 Release 2 ASW). In our analysis of African Americans, there is stronger evidence of association between the dichotomous phenotype heavy smoking vs. light smoking and the sub-bin tagged by rs16969968 (OR = 1.62, 95% CI = 1.21–2.17,  $P = 1.1 \times 10^{-3}$ ), compared to the sub-bin tagged by rs1051730 (OR = 1.15, 95% CI = 1.03–1.28,  $P = 1.1 \times 10^{-2}$ ). This stronger finding is seen despite the lower minor allele frequency (MAF) and much smaller available sample and for rs16969968 (MAF = 0.06, 667 cases/1,140 controls) compared to that for rs1051730 (MAF = 0.12, 1,712 cases/5,640 controls).

For bin A, no tested sub-bin other than the one tagged by rs16969968 shows consistent association across populations. The meta-analyzed genetic associations between all

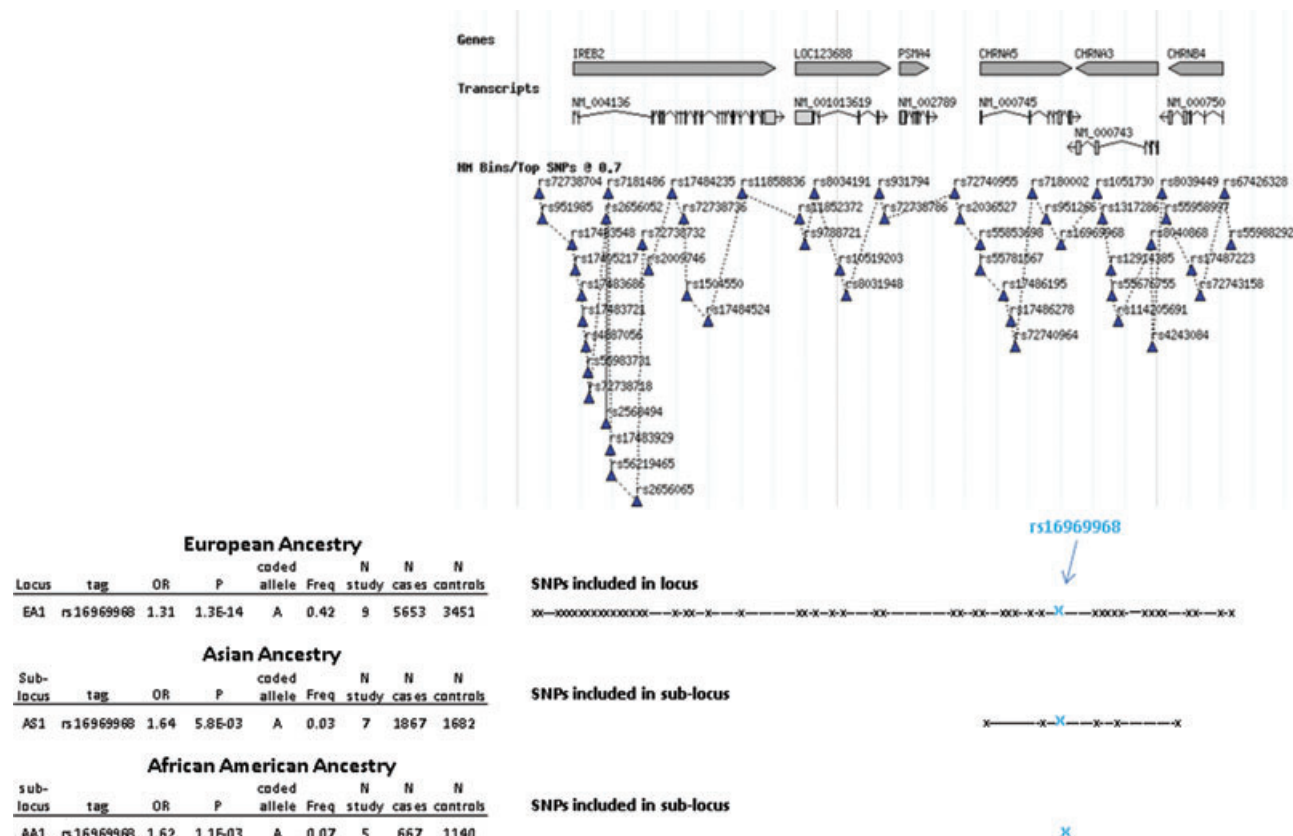


Fig. 1. Top associations with heavy vs. light smoking; Bin A across three populations. This figure shows all 52 SNPs in bin A, and also the only consistently associated sub-bins ( $P < 0.01$  in both Asians and African Americans). This figure also lists corresponding odds ratios for the association between the sub-bins and heavy smoking. Bin A variants span across six genes in the European ancestry population, the sub-bin tagged by rs16969968 in the Asian population spans across two genes, and the sub-bin tagged by rs16969968 in the African American population spans across only one gene (CHRNA5).

available constituent sub-bins and heavy smoking are shown in Table S2.

GENETIC ASSOCIATIONS IN BIN B

Bin B (tagged by rs588765 and rs880395 in Europeans) includes 111 SNPs correlated ( $r^2 \geq 0.7$ ) in the European ancestry reference sample, which was partitioned into 39 sub-bins in Asian and 37 sub-bins in African American ancestry reference samples. We had adequate coverage to test 10 of these 39 sub-bins in Asian samples and 22 of these 37 sub-bins in African American samples. Consistent with the previous report [Saccone et al., 2010] that used some of these same data (see Table S1 for the overlap, which involves only European-ancestry samples), we find that in European ancestry samples, bin B is associated (OR = 1.27, 95% CI = 1.16–1.38,  $P = 8.7 \times 10^{-8}$ ) with heavy smoking in conditional analyses with rs16969968; bin B is not associated in single SNP analyses (OR = 1.0, 95% CI = 0.94–1.07,  $P = 0.99$ ). In Asian samples, testing for SNP association conditioning on rs16969968 show an association between heavy smoking and bin B, with the strongest result for the sub-bin tagged by rs514743 (OR = 1.30, 95% CI = 1.07–1.58,  $P = 9.7 \times 10^{-3}$ ), which is similar to the single SNP test (OR = 1.28, 95% CI = 1.05–1.56,  $P = 0.014$ ). In African American subjects, there is a trend of association for the same sub-bin in conditional

association (OR = 1.16, 95% CI = 0.99–1.36,  $P = 0.064$ ; Table S3), compared to the single SNP association (OR = 1.05, 95% CI = 0.96–1.15,  $P = 0.24$ ). Thus, we found evidence of association in the Asian samples consistent with the association observed in the samples of European ancestry, but only a trend toward association in the African American subjects. The meta-analyzed conditional and single SNP associations between these constituent sub-bins and heavy smoking are shown in Tables S3 and S6.

GENETIC ASSOCIATIONS IN BIN C

Bin C (tagged by rs6495308, rs2036534, rs7163730, rs9788682, rs684513, and rs578776 in Europeans) includes 82 SNPs correlated ( $r^2 \geq 0.7$ ) in the European ancestry reference sample, which was partitioned into 24 sub-bins in Asian and 26 sub-bins in African American reference samples. We had adequate coverage to test 12 of these 24 sub-bins in Asian samples and 19 of these 26 sub-bins in African American samples. Consistent with the previous studies [Liu et al., 2010; Saccone et al., 2010; TAG, 2010; Thorgerisson et al., 2010], in European ancestry samples, there is an association between heavy smoking and bin C (OR = 0.79, 95% CI = 0.72–0.86,  $P = 2.5 \times 10^{-7}$ ) in association tests conditioning on rs16969968 as well as an association in a

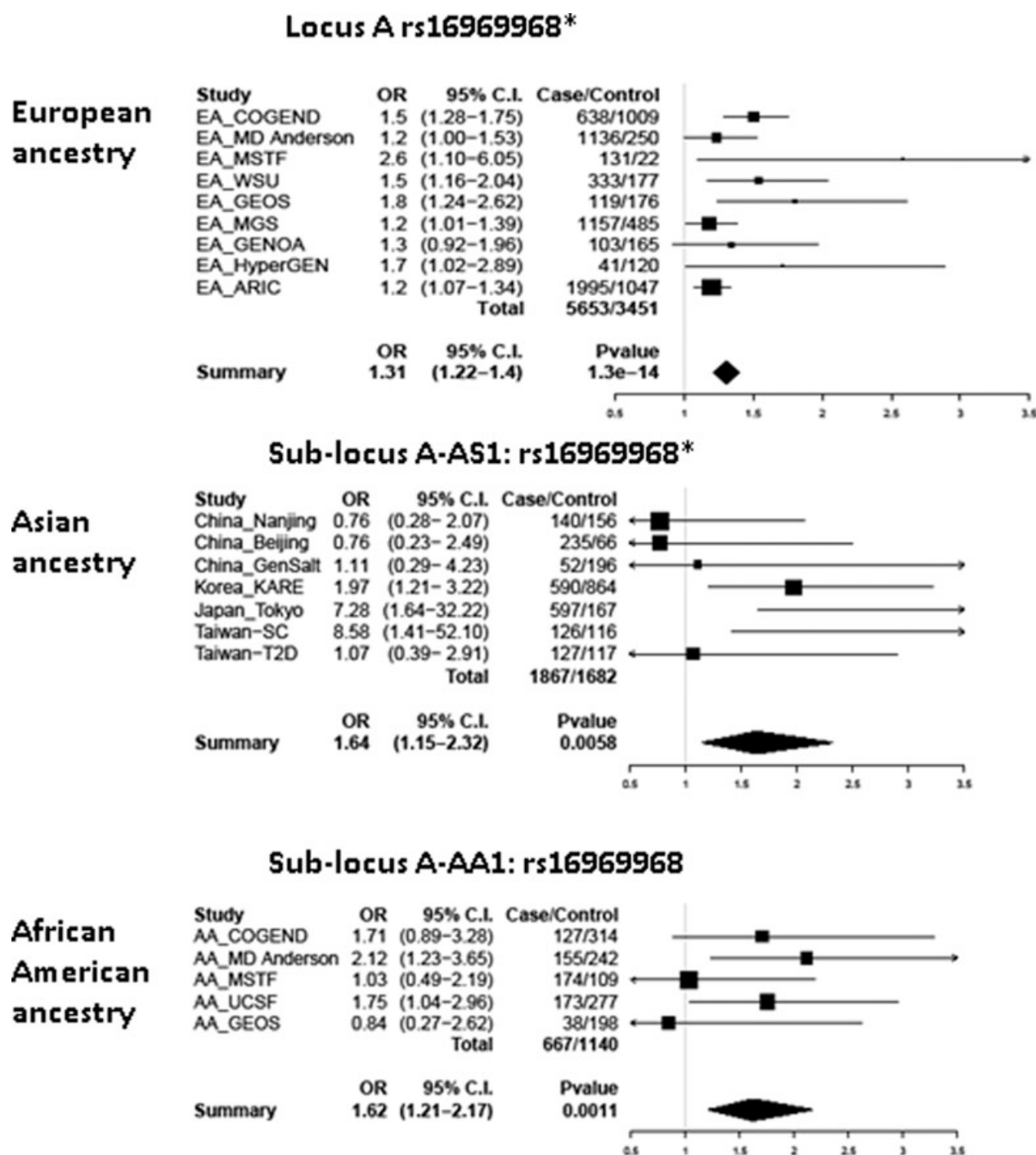


Fig. 2. Rs16969968 and heavy smoking in samples of European, Asian, and African American ancestry. The ORs and 95% CIs are for the effect per allele using additive coding in the logistic regression with age and sex as covariates. \*When rs16969968 is not available, rs1051730 and rs951266 are used as proxy SNPs in European and Asian ancestry samples.

single SNP analysis (OR = 0.77, 95% CI = 0.71–0.83,  $P = 4.0 \times 10^{-11}$ ).

Neither the Asian nor African American populations provide strong evidence of association with heavy smoking in any tested sub-bin in bin C under conditional association tests (all  $P > 0.01$ ). In the Asian data, the strongest single SNP signal was the sub-bin tagged by rs6495308 (OR = 0.83,

95% CI = 0.72–0.96,  $P = 9.8 \times 10^{-3}$ ). In the African American data, there was no evidence of consistent association in either single SNP or conditional analyses ( $P > 0.01$ ) for the sub-bin tagged by rs6495308 or any other sub-bin. The meta-analyzed conditional and single SNP associations between tested sub-bins and heavy smoking are shown in Tables S4 and S7.



## GENETIC ASSOCIATIONS IN BIN D

Bin D (tagged by rs2869046 in Europeans) includes 15 SNPs correlated ( $r^2 \geq 0.7$ ) in the European ancestry reference sample, which was partitioned into seven sub-bins in Asians and seven sub-bins in African Americans. We had adequate coverage to test two of these seven sub-bins in Asian samples and three of the seven sub-bins in African American samples. We found no evidence of association between bin D and heavy smoking in European ancestry data, or across populations in single SNP or conditional association analyses ( $P > 0.1$ ). The meta-analyzed genetic associations between available sub-bins and heavy smoking conditional and single SNP associations are shown in Tables S5 and S8.

All bins were tested in secondary analyses using the four level phenotype measured by CPD and results were similar.

## DISCUSSION

This collaborative genetic meta-analysis of smoking behavior is the first to show consistent association in the chromosome 15q25 region with heavy smoking, across samples representing three genetically distinct populations—European ancestry, Asian, and African American. Previous meta-analyses examined only European ancestry data to definitively identify associations between chromosome 15q25 and smoking behavior. Smaller individual studies of Asians and African Americans have previously examined this region for association with smoking and related phenotypes. Smoking quantity has been reported as associated with variants correlated with rs16969968 in subjects of Asian and African American descent [Amos et al., 2010; Li et al., 2005; Li et al., 2010; Saccone et al., 2009; Schwartz et al., 2010; Shiraishi et al., 2009; Wu et al., 2009]. Our meta-analysis synthesizes reported findings of individual SNP associations and compares genetic associations across multipopulation samples to take the correlations between genetic variants within each population into account. Our meta-analysis strengthens the evidence of association between the specific SNP rs16969968 in bin A and heavy smoking across these diverse populations.

The strongest association signal seen in this gene cluster in European ancestry populations is represented by a group of 52 correlated variants, including rs16969968, which we call bin A. Due to these high correlations, the ability to statistically refine the association between smoking and these SNPs is very limited when using only European ancestry subjects. However, the LD structure between these 52 variants breaks down into 20 sub-bins in Asians and 38 sub-bins in African Americans.

By requiring consistent genetic effects across the three populations, we can refine a genetic association to variants that are more likely to reflect potential functional variants. Two SNPs in bin A are the most frequently reported from previous meta-analyses of smoking behavior in European ancestry subjects: rs16969968 and rs1051730. They are highly correlated ( $r^2 = 1$ ) in European ancestry and Asian populations, but display only modest correlation in African Americans ( $r^2 = 0.40$ ; HapMap 3 Release 2 ASW). We can leverage this difference in LD architecture to differentiate the association of heavy smoking with these two variants.

In our meta-analysis of African Americans, rs16969968 is more strongly associated with heavy smoking (OR = 1.62,

95% CI = 1.21–2.17,  $P = 0.0011$ ,  $N = 1,807$ ) than rs1051730 (OR = 1.15, 95% CI = 1.03–1.28,  $P = 0.011$ ,  $N = 7,352$ ). SNP rs16969968 is the most strongly associated polymorphism across all three populations and the only variant meeting the consistent association threshold in our study. In addition, SNP rs16969968 causes an amino acid change in the nicotinic receptor  $\alpha 5$  subunit and alters function of its receptor [Bierut et al., 2008]. The observed consistent associations across diverse populations, combined with the results of biological experiments on rs16969968, provide converging evidence that rs16969968, rather than rs1051730, is most likely one causative variant in this region driving the strongest association signal.

Prior meta-analyses in European ancestry populations have reported additional association signals distinct from bin A, and they cluster into three groups. Bin B, a group of 111 variants highly correlated in Europeans, includes the previously reported associated SNPs rs588765 and rs880395. The association with bin B previously reported in Europeans was seen only in association analyses conditioning on rs16969968. Bin B consists of 39 sub-bins in Asian subjects and 37 sub-bins in African American subjects. In conditional analyses, we found evidence of association between bin B and heavy smoking in the Asian data (OR = 1.30, 95% CI = 1.07–1.58,  $P = 9.7 \times 10^{-3}$ ) as well as reproducing the European ancestry finding (OR = 1.27, 95% CI = 1.16–1.38,  $P = 8.7 \times 10^{-8}$ ). In the African American data, there was a trend toward association in the same direction (OR = 1.16, 95% CI = 0.99–1.36,  $P = 0.064$ ).

Bin B variants, located upstream of the coding region of *CHRNA5*, are associated with variability in *CHRNA5* mRNA levels in European ancestry samples [Falvella et al., 2010; Smith et al., 2010; Wang et al., 2009]. Low levels of *CHRNA5* mRNA expression are associated with lower risk for nicotine dependence. No data exist on *CHRNA5* mRNA expression in other populations, and further work to examine expression data and smoking behavior in other populations is needed. Because the risk allele of rs16969968 occurs primarily on the low mRNA expression alleles represented by bin B, conditional SNP analysis controlling for bin A (rs16969968) is important to distinguish between these two distinct mechanisms [Saccone et al., 2010; Wang et al., 2009]. This is an important example to demonstrate how a genetic effect could be better detected and characterized when additional related variants are taken into account.

Bin C, a group of 82 variants correlated in Europeans, consisted of 24 sub-bins in subjects of Asians and 26 sub-bins in African Americans. Variants reported in previous meta-analyses of European ancestry (rs6495308 [Liu et al., 2010], rs2036534 [Thorgeirsson et al., 2010], rs7163730, rs9788682, rs684513 [TAG, 2010], rs578776 [Saccone et al., 2010]) were all examined. However, no tested SNP in bin C was consistently associated with heavy smoking with  $P < 0.01$  in both Asians and African Americans. Similarly, we have no consistent associations with bin D, which contains 15 SNPs correlated in Europeans and consists of seven sub-bins in Asians and seven sub-bins in African Americans.

In undertaking this project, we faced numerous challenges. First, smoking behavior differs substantially across populations. Smoking quantity distributions differ across populations; smokers of European ancestry smoke more heavily than do Asians or African Americans. As a result, we decided to compare heavy ( $>20$  CPD) vs. light smoking ( $\leq 10$  CPD) in our primary association analysis to more closely capture the contrast between nicotine



dependent smokers and nondependent smokers. We then confirmed the consistency of results using the full distribution of smoking quantity in subsequent analyses.

Second, genotyping coverage varied between studies, and several studies in our meta-analysis had only a few variants genotyped. As a result, not all sub-bins were tested and the sample size varied across the tested sub-bins. For example, SNP rs55853698 that was imputed and reported as highly associated with smoking quantity in a previous meta-analysis of European ancestry subjects [Liu et al., 2010], lies in bin A, but no genotyping data were available for testing this SNP or the sub-bin it represents in Asian and African American populations. Use of imputed data has the potential to mitigate these problems. However, imputation was not possible for our low-coverage studies. Therefore, the concerns about having untested SNPs and unequal subjects in the region would remain even with imputed data. We believe it is important to report our findings based on directly genotyped variants, and the interpretation of the consistent associations is not expected to change with imputation.

In addition, we were not able to perform thorough admixture tests in all datasets due to variable genotyping. Population stratification unaccounted for by our stratified analyses of self-identified ancestry—European, Asian, and African American—could be a confounder in our results. Although we are leveraging the admixture to separate the effects of different genetic variants, there may be differential admixture in the cases and controls among African Americans. The impact of varied genetic architecture within given, broadly defined populations as well as within and across populations represented by individual sites (e.g., Japanese, Chinese, and Korean) needs to be elucidated in future larger scale studies with sufficient representation of individuals from different population backgrounds and more comprehensive genotyping.

Lastly, an association seen in one population that is not consistent across all three may nonetheless represent a true biological signal. Lack of consistency for the association may simply reflect differences in power for our population samples. Issues that can affect power across our three diverse populations include sample size, MAF, and even population-specific effect size. The last factor could arise in a variety of ways, including differences in LD structure, background variation, and marker information content. Thus, we suggest caution when interpreting the negative or non-consistent association results from this study. Though these results strengthen the evidence for rs16969968 as a likely causal variant, this region remains in need of further interrogation with additional genotyping and standardized imputation across all populations.

Despite the limitations of this study, this meta-analysis refines the association signals with heavy smoking across samples representing European ancestry, Asian, and African American populations. In particular, for bin A, we present evidence showing rs16969968 is a likely causal variant for heavy smoking among the common SNPs in the bin. Our evidence also suggests there are additional distinct genetic variants in the chromosome 15q25 region associated with smoking, but we are unable to clearly identify these other associations across all three populations. For example, we extend the finding of association with bin B in European ancestry samples to an association in Asians, and a trend toward association in African Americans.

This consistent pattern of cross-population association despite many unmeasured genetic and environmental differences has provided important evidence to support true causal variants. It also provides critical information by narrowing a region of interest so laboratory experiments that must follow association studies can focus on a smaller number of variants. Thus, this study represents an important step on the pathway from association to function.

## ACKNOWLEDGMENTS

*Collaborative Genetic Study of Nicotine Dependence (COGEND)*: We thank the subjects who participated in this study. We wish to thank Hilary Davidson, Sherri Fisher, Tracey Richmond, and Heidi Kromrei for administrative support; and Louis Fox for data analysis. The Collaborative Genetic Study of Nicotine Dependence (COGEND) study investigators are Laura Bierut (PI), Michael Brent, Naomi Breslau, Robert Culverhouse, Alison Goate, Richard Grucza, Dorothy Hatsukami, Anthony Hinrichs, Eric Johnson, Sharon Murphy, John Rice, Nancy Saccone, Scott Saccone, Joe Henry Steinbach, Jerry Stitzel, and Jen-Chyong Wang. *MD Anderson*: 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 different clinical sites. Also, we wish to thank Dr. Qing Xu for performing SNP-typing on our samples. *Mid-south Tobacco Family Study (MSTF)*: 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 different clinical sites. *Molecular Genetics of Schizophrenia*: We thank the study participants and the research staff at the study sites. *Genetic Epidemiology Network of America (GENOA)*: Mayo Clinic (Rochester Field Center and Genotyping Center): Stephen T. Turner, Mariza de Andrade, Julie Cunningham. University of Texas Health Sciences Center (DNA lab): Eric Boerwinkle, Megan L. Grove-Gaona. University of Michigan (Analysis Center): Patricia Peyser, Lawrence Bielak, Wei Zhao. *Hypertension Genetic Epidemiology Network (HyperGEN)*: University of Utah (Network Coordinating Center, Field Center, and Molecular Genetics Lab): Steven C. Hunt, Ph.D. (Network Director and Field Center P.I.); Mark F. Leppert, Ph.D. (Molecular Genetics P.I.); Jean-Marc Lalouel, M.D., D.Sc.; Robert B. Weiss, Ph.D.; Roger R. Williams, M.D. (late); Janet Hood. University of Alabama at Birmingham (Field Center): Cora E. Lewis, M.D., M.S.P.H. (P.I.); Albert Oberman, M.D., M.P.H.; Donna Arnett, Ph.D.; Phillip Johnson; Christie Oden. Boston University (Field Center): Richard H. Myers, Ph.D. (P.I.); R. Curtis Ellison, M.D.; Yuqing Zhang, M.D.; Jemma B. Wilk, D.Sc.; Luc Djouss, M.D., D.Sc.; Jason M. Laramie; Greta Lee Splansky, M.S. University of Minnesota (Field Center and Biochemistry Lab): James S. Pankow, Ph.D. (Field Center P.I.); Michael B. Miller, Ph.D.; Michael Li, Ph.D.; John H. Eckfeldt, M.D., Ph.D.; Anthony a. Killeen, M.D., Ph.D.; Catherine Leiendecker-Foster, M.S.; Jean Bucks; Greg Rynders. University of North Carolina (Field Center): Kari E. North, Ph.D. (P.I.); Barry I. Freedman, M.D.; Gerardo Heiss, M.D. Washington University (Data Coordinating Center): D.C. Rao, Ph.D. (P.I.); Charles Gu, Ph.D.; Treva Rice, Ph.D.; Aldi T. Kraja, D.Sc., Ph.D.; Gang Shi, Ph.D.; Yun Ju Sung, Ph.D.; Karen L. Schwander, M.S.;

Matthew Brown; Michael A. Province, Ph.D.; Ingrid Borecki, Ph.D. Weil Cornell Medical College (Echo Reading Center): R.B. Devereux, M.D.; Giovanni de Simone, M.D.; Jonathan N. Bella, M.D. National Heart, Lung, & Blood Institute: Cashell Jaquish, Ph.D.; Dina Paltoo, Ph.D. ARIC: The authors thank the staff and participants of the ARIC study for their important contributions. *Japan*: We thank Dr. Kouya Shiraishi for his help on genotype data processing. *GenSalt*: The GenSalt Study Steering Committee: Dongfeng Gu, Jiang He (Chair), James E. Hixson, Cashell E. Jaquish, Depei Liu, DC Rao, Paul K. Whelton, and Zhijian Yao. *GenSalt Collaborative Research Group*: Tulane University Health Sciences Center, New Orleans, USA: Jiang He (PI), Lydia A. Bazzano, Chung-Shiuan Chen, Jing Chen, Lee Hamm, Paul Muntner, Kristi Reynolds, Jaqueline R. Reuben, Paul K. Whelton, and Wenjie Yang. Washington University School of Medicine, St. Louis, USA: DC Rao (PI), Matthew Brown, Charles Gu, Hongyan Huang, Treva Rice, Karen Schwander, Gang Shi, and Yun Ju Sung. Chinese Academy of Medical Sciences, Beijing, China: Dongfeng Gu (PI), Jie Cao, Jichun Chen, Xiufang Duan, Jianfeng Huang, Jinghan Huang, Jianxin Li, Depei Liu, Donghua Liu, Enchun Pan, Yang Wei, and Xiqui Wu. Shandong Academy of Medical Sciences, Shandong, China: Fanghong Lu (PI), Shikuan Jin, Qingjie Meng, Fan Wu, and Yingxin Zhao; Shandong Center for Diseases Control and Prevention, Shandong, China: Jixiang Ma (PI), Weika Li, and Jiyu Zhang; Zhengzhou University: Dongsheng Hu (PI), Yaxin Ding, Hongwei Wen, Meixi Zhang, and Weidong Zhang; Xinle Traditional Chinese Medicine Hospital, Hebei, China: Xu Ji (PI), Rongyan Li, Haijun Zu; Nanjing University of Medical Sciences, Jiangsu, China: Cailliang Yao (PI), Yongchao Li, Chong Shen, and Jiayi Zhou; Xi'an Jiaotong University, Shanxi, China: Jianjun Mu (PI), Enrang Chen, Qinzhou Huang, and Man Wang. Chinese National Human Genome Center at Beijing: Zhi-Jian Yao (PI), Shufeng Chen, Dongfeng Gu, Hongfan Li, Laiyuan Wang, Penghua Zhang, Qi Zhao. University of Texas Health Sciences Center at Houston: James E. Hixson (PI) and Lawrence C. Shimmin. National Heart, Lung, and Blood Institute: Cashell E. Jaquish. *Women's Health Initiative*: This manuscript was prepared in collaboration with investigators of the WHI and has been reviewed and approved (MS1453) by the Women's Health Initiative (WHI) Publications & Presentations Committee. The authors thank Charles Kooperberg and the WHI investigators and staff and study participants for making the program possible. WHI investigators are listed at [http://www.whiscience.org/publications/WHI\\_investigators\\_shortlist.pdf](http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf). *Collaborative Genetic Study of Nicotine Dependence (COGEND)*: The COGEND contribution was supported by the National Cancer Institute (NCI; P01 CA089392), The National Human Genome Research Institute (NHGRI; U01 HG04422-01), and the National Institute on Drug Abuse (NIDA; K02 DA021237). COGEND genotyping was in part performed under NIDA Contract HHSN271200477471C; phenotypic and genotypic data are stored in the NIDA Center for Genetic Studies (NCGS) at <http://zork.wustl.edu/> under NIDA Contract HHSN271200477451C (PIs J Tischfield and J Rice); genotyping services were also provided by the Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health (NIH) to The Johns Hopkins University, contract number HHSN268200782096. Dr. LiShiun Chen was supported by KL2RR024994 and K08DA030398. Support was also provided by R01DA026911, R03DA023166,

and R21DA033827. *MD Anderson*: Research was supported by NIH grants U19CA148127, R01CA121197S2, R01CA141716, R01CA127219, CA121197, R01CA133996, P30CA16672, P50CA70907, R01CA55769 and Cancer Prevention Research Institute of Texas grant RP10043. *Mid-south Tobacco Family Study (MSTF)*: This project is supported in part by NIH Grant R01 DA012844 (PI: Ming Li). *Wayne State University and the Karmanos Cancer Institute: Family Health Study; Women's Epidemiology of Lung Disease; EXHALE (Exploring Health Ancestry and Lung Epidemiology)*: This research was supported through NIH grants R01CA060691 and R01CA87895, and NIH contract PC35145. *University of Maryland: GEOS*: The GEOS Study was supported by the National Institutes of Health Genes, Environment and Health Initiative (GEI) Grant U01 HG004436, as part of the GENEVA consortium under GEI, with additional support provided by the Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488); and the Office of Research and Development, Medical Research Service, and the Baltimore Geriatrics Research, Education, and Clinical Center of the Department of Veterans Affairs. Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to the Johns Hopkins University (contract number HHSN268200782096C). Assistance with data cleaning was provided by the GENEVA Coordinating Center (U01 HG 004446; PI Bruce S Weir). Study recruitment and collection of datasets were supported by a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and by grants from the National Institute of Neurological Disorders and Stroke (NINDS) and the NIH Office of Research on Women's Health (R01 NS45012, U01 NS069208-01). *Molecular Genetics of Schizophrenia*: This study was supported by NIH R01 grants (MH67257 to Nancy G. Buccola, MH59588 to Bryan J. Mowry, MH59571 to Pablo V. Gejman, MH59565 to Robert Freedman, MH59587 to Farooq Amin, MH60870 to William F. Byerley, MH59566 to Donald W. Black, MH59586 to Jeremy M. Silverman, MH61675 to Douglas F. Levinson, MH60879 to C. Robert Cloninger, and MH81800 to Pablo V. Gejman), NIH U01 grants (MH46276 to C. Robert Cloninger, MH46289 to Charles Kaufmann, MH46318 to Ming T. Tsuang, MH79469 to Pablo V. Gejman, and MH79470 to Douglas F. Levinson), the Genetic Association Information Network (GAIN), and by The Paul Michael Donovan Charitable Foundation. Genotyping was carried out by the Center for Genotyping and Analysis at the Broad Institute of Harvard and MIT (Stacy Gabriel and Daniel B. Mirel), which is supported by grant U54 RR020278 from the National Center for Research Resources. Genotyping of half of the EA sample and almost all the AA sample was carried out with support from GAIN. The GAIN quality control team (Gonçalo R. Abecasis and Justin Paschall) made important contributions to the project. We thank Shaun Purcell for assistance with PLINK. *Genetic Epidemiology Network of America (GENOA)*: The Genetic Epidemiology Network of Arteriopathy phenotyping and genome-wide genotyping is supported by the National Heart Lung and Blood Institute (NHLBI) of the National Institutes of Health (HL54457, HL68737, and HL087660). *Hypertension Genetic Epidemiology Network (HyperGEN)*: The Hypertension Genetic Epidemiology Network is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515. *ARIC*: The Atherosclerosis

Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. *Nanjing and Beijing*: This work was supported by the Chinese National Natural Science Foundation grant 30230080 (Hongbing Shen) and the State Key Basic Research Program grants 2002CB512902 (Hongbing Shen). Dr. Dongxin Lin was supported by State Key Basic Research Program grant 2004CB518701. *KARE (Korea Association Resource)*: The KARE data analyzed in this study 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. The work of TP was supported by the Consortium for Large Scale Genome Wide Association Study, the National Research Foundation (KRF-2008-313-C00086) and the Brain Korea 21 Project of the Ministry of Education. *Japan*: Grants-in-Aid from the Ministry of Health, Labor and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control and for Cancer Research (19-9 and 19S-1). *Taiwan*: This study was supported by Academia Sinica Genomic Medicine Multicenter Study and National Research Program for Genomic Medicine, National Science Council, Taiwan (National Clinical Core, NSC97-3112-B-001-014 and National Genotyping Center, NSC97-3112-B-001-015). *GenSalt*: The Genetic Epidemiology Network of Salt Sensitivity is supported by a cooperative agreement project Grant (U01HL072507) from the National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland. *University of California San Francisco*: This work was supported by the National Institute of Environmental Health Sciences [R01 ES06717]; and the National Cancer Institute [R01 CA 52689 to MW]. *Women's Health Initiative*: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. SPD is supported by DA 017441 and 02733. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **CONFLICT OF INTEREST DISCLOSURE**: LB Bierut, AM Goate, JP Rice, and JC Wang are listed as inventors on Issued U.S. Patent. 8,080,371, "Markers for Addiction" covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Ming Li serves as a scientific advisor to ADial Pharmaceuticals.

## REFERENCES

- Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang Q, Ghorri MJ, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VA, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311):52-58.
- Amos CI, Gorlov IP, Dong Q, Wu X, Zhang H, Lu EY, Scheet P, Greisinger AJ, Mills GB, Spitz MR. 2010. Nicotinic acetylcholine receptor region on chromosome 15q25 and lung cancer risk among African Americans: a case-control study. *J Natl Cancer Inst* 102(15):1199-1205.
- Baker TB, Weiss RB, Bolt D, von Niederhausen A, Fiore MC, Dunn DM, Piper ME, Matsunami N, Smith SS, Coon H, McMahon WM, Scholand MB, Singh N, Hoidal JR, Kim SY, Leppert MF, Cannon DS. 2009. Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine Tob Res* 11(7):785-796.
- Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H, Waterworth D, Muglia P, Mooser V. 2008. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry* 13(4):368-373.
- Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Gruzca RA, Xuei X, Saccone NL, Saccone SF, Bertelsen S, Fox L, Horton WJ, Morgan SD, Breslau N, Budde J, Cloninger CR, Dick DM, Foroud T, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Kuperman S, Madden PAF, Mayo K, Nurnberger JJ, Pomerleau O, Porjesz B, Reyes O, Schuckit M, Swan GT, J. A, Edenberg HJ, Rice JP, Goate AM. 2008. Nicotine dependence and the a5-a3-b4 nicotinic receptor gene cluster: variants in the nicotinic receptors alter the risk for nicotine dependence. *Am J Psychiatry* 165(11):1163-1171.
- Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 74(1):106-120.
- Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061-1073.
- Falvella FS, Galvan A, Colombo F, Frullanti E, Pastorino U, Dragani TA. 2010. Promoter polymorphisms and transcript levels of nicotinic receptor CHRNA5. *J Natl Cancer Inst* 102(17):1366-1370.
- Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24(24):2938-2939.
- Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, Perola M, Pitkaniemi J, Peltonen L, Aromaa A, Kaprio J. 2009. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNA4) on chromosome 15. *Hum Mol Genet* 18(20):4007-4012.
- Li MD, Beuten J, Ma JZ, Payne TJ, Lou XY, Garcia V, Duenes AS, Crews KM, Elston RC. 2005. Ethnic- and gender-specific association of the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) with nicotine dependence. *Hum Mol Genet* 14(9):1211-1219.
- Li MD, Yoon D, Lee JY, Han BG, Niu T, Payne TJ, Ma JZ, Park T. 2010. Associations of variants in CHRNA5/A3/B4 gene cluster with smoking behaviors in a Korean population. *PLoS One* 5(8):e12183.
- 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, Laos 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. 2010. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 42(5):436–440.
- Lumley T. 2009. *rmeta: Meta-analysis*, R package version 2.15.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
- Rotimi CN, Jorde LB. 2010. Ancestry and disease in the age of genomic medicine. *N Engl J Med* 363(16):1551–1558.
- Saccone NL, Saccone SF, Goate AM, Gruzza RA, Hinrichs AL, Rice JP, Bierut LJ. 2008. In search of causal variants: refining disease association signals using cross-population contrasts. *BMC Genet* 9:58–68.
- Saccone NL, Wang JC, Breslau N, Johnson EO, Hatsukami D, Saccone SF, Gruzza RA, Sun L, Duan W, Budde J, Culverhouse RC, Fox L, Hinrichs AL, Steinbach JH, Wu M, Rice JP, Goate AM, Bierut LJ. 2009. The CHRNA5-CHRNA3-CHRNA4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res* 69(17):6848–6856.
- Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, Giegling I, Han S, Han Y, Keskitalo-Vuokko K, Kong X, Landi MT, Ma JZ, Short SE, Stephens SH, Stevens VL, Sun L, Wang Y, Wenzlaff AS, Aggen SH, Breslau N, Broderick P, Chatterjee N, Chen J, Heath AC, Heliövaara M, Hoft NR, Hunter DJ, Jensen MK, Martin NG, Montgomery GW, Niu T, Payne TJ, Peltonen L, Pergadia ML, Rice JP, Sherva R, Spitz MR, Sun J, Wang JC, Weiss RB, Wheeler W, Witt SH, Yang BZ, Caporaso NE, Ehringer MA, Eisen T, Gapstur SM, Gelemtier J, Houlston R, Kaprio J, Kendler KS, Kraft P, Leppert MF, Li MD, Madden PA, Nothen MM, Pillai S, Rietschel M, Rujescu D, Schwartz A, Amos CI, Bierut LJ. 2010. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet* 6(8).
- Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau O, Swan GE, Goate AM, Rutter J, Bertelsen S, Fox L, Fugman D, Martin NG, Montgomery GW, Wang JC, Ballinger DG, Rice JP, Bierut LJ. 2007. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 16(1):36–49.
- Schwartz AG, Wenzlaff AS, Bock CH, Ruterbusch JJ, Chen W, Cote ML, Artis AS, Van Dyke AL, Land SJ, Harris CC, Pine SR, Spitz MR, Amos CI, Levin AM, McKeigue PM. 2010. Admixture mapping of lung cancer in 1812 African-Americans. *Carcinogenesis* 32(3):312–317.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29(1):308–311.
- Sherva R, Wilhelmsen K, Pomerleau CS, Chasse SA, Rice JP, Snedecor SM, Bierut LJ, Neuman RJ, Pomerleau OF. 2008. Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit alpha 5 (CHRNA5) with smoking status and with 'pleasurable buzz' during early experimentation with smoking. *Addiction* 103(9):1544–1552.
- Shiraishi K, Kohno T, Kunitoh H, Watanabe S, Goto K, Nishiwaki Y, Shimada Y, Hirose H, Saito I, Kuchiba A, Yamamoto S, Yokota J. 2009. Contribution of nicotine acetylcholine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. *Carcinogenesis* 30(1):65–70.
- Smith RM, Alachkar H, Papp AC, Wang D, Mash DC, Wang JC, Bierut LJ, Sadee W. 2010. Nicotinic alpha5 receptor subunit mRNA expression is associated with distant 5' upstream polymorphisms. *Eur J Hum Genet* 19:76–83.
- Stevens VL, Bierut LJ, Talbot JT, Wang JC, Sun J, Hinrichs AL, Thun MJ, Goate A, Calle EE. 2008. Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev* 17(12):3517–3525.
- TAG. 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 42(5):441–447.
- 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 PFA, 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, Kiemeny LA, Matthiasson SE, Oskarsson H, Tyrfinngsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. 2008. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452(3):638–642.
- 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, Gudjonsson 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, Willemssen G, Vogelzangs N, Doring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, GarciaPrats 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, Tyrfinngsson T, Kiemeny 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. 2010. Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* 42(5):448–453.
- Wang JC, Cruchaga C, Saccone NL, Bertelsen S, Liu P, Budde JP, Duan W, Fox L, Gruzza RA, Kern J, Mayo K, Reyes O, Rice J, Saccone SF, Spiegel N, Steinbach JH, Stitzel JA, Anderson MW, You M, Stevens VL, Bierut LJ, Goate AM. 2009. Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in CHRNA5. *Hum Mol Genet* 19(6):3125–3135.
- Weiss RB, Baker TB, Cannon DS, von Niederhausen A, Dunn DM, Matsunami N, Singh NA, Baird L, Coon H, McMahon WM, Piper ME, Fiore MC, Scholand MB, Connett JE, Kanner RE, Gahring LC, Rogers SW, Hoidal JR, Leppert MF. 2008. A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet* 4(7):e1000125.
- Wu C, Hu Z, Yu D, Huang L, Jin G, Liang J, Guo H, Tan W, Zhang M, Qian J, Lu D, Wu T, Lin D, Shen H. 2009. Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. *Cancer Res* 69(12):5065–5072.
- Zaitlen N, Pasaniuc B, Gur T, Ziv E, Halperin E. 2010. Leveraging genetic variability across populations for the identification of causal variants. *Am J Hum Genet* 86(1):23–33.