HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2019 July 15.

Published in final edited form as:

Nat Genet. ; 44(8): 895–899. doi:10.1038/ng.2351.

Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population

A full list of authors and affiliations appears at the end of the article.

Abstract

To find additional susceptibility loci for lung cancer, we tested promising associations from our previous genome-wide association study (GWAS)[1] of lung cancer in the Chinese population in an extended validation sample size of 7,436 individuals with lung cancer (cases) and 7,483 controls. We found genome-wide significant ($P < 5.0 \times 10^{-8}$) evidence for three additional lung cancer susceptibility loci at 10p14 (rs1663689, close to *GATA3*, $P = 2.84 \times 10^{-10}$), 5q32 (rs2895680 in *PPP2R2B-STK32A-DPYSL3*, $P = 6.60 \times 10^{-9}$) and 20q1 3.2 (rs4809957 in *CYP24A1*, $P = 1.20 \times 10^{-8}$). We also found consistent associations for rs247008 at 5q31.1 *(IL3-CSF2-P4HA2*, $P = 7.68 \times 10^{-8}$) and rs9439519 at 1p36.32 (*AJAP1-NPHP4*, $P = 3.65 \times 10^{-6}$). Four of these loci showed evidence for interactions with smoking dose ($P = 1.72 \times 10^{-10}$, $P = 5.07 \times 10^{-3}$, $P = 6.77 \times 10^{-3}$ and $P = 4.49 \times 10^{-2}$ for rs2895680, rs4809957, rs247008 and rs9439519, respectively). These results advance our understanding of lung cancer susceptibility and highlight potential pathways that integrate genetic variants and smoking in the development of lung cancer.

Thus far, GWAS has achieved considerable success in deciphering the genetic basis of lung cancer in both European-ancestry and Asian populations[1–10]. We recently reported a multistage GWAS of lung cancer in the Chinese population, with two newly identified (13q12.12 and 22q12.2) and two replicated loci (3q28 and 5p15.33) for lung cancer susceptibility[1]. Although previous GWAS efforts have provided valuable clues regarding the etiology of lung cancer, additional genetic factors remain to be discovered.

Correspondence should be addressed to H.S. (hbshen@njmu.edu.cn). .

AUTHOR CONTRIBUTIONS

H.S. directed the study, obtained financial support and was responsible for study design, interpretation of results and manuscript writing. J. Dong and Z.H. performed overall project management, along with C. Wu, and drafted the initial manuscript. J. Dong, Z.H., Z.L., J. Dai and R.Z. performed statistical analyses. D. Lin, T.W., Y. Shi, D. Lu, L.J., B.Z., J.L. and K.C. directed each participating study and jointly organized this study. M.C., C. Wang, Y.J., S.C., Z.Q., J.G. and C.S. were responsible for sample processing and managed the genotyping data. H.M., G.J., Z.P., Y.C., Y. Shu and L.X. were responsible for subject recruitment and sample preparation of the Nanjing samples. C. Wu, D.Y., X.L. and W.T. were responsible for subject recruitment and sample preparation of the Beijing samples. H.G., Q.D., L.L. and P.X. were responsible for subject recruitment and sample preparation of the Wuhan samples. X.Z., J.W., G.Z., H.C., B.H. and C.B. were responsible for subject recruitment and sample preparation of the Shanghai samples. Z.Y., W.W., P.G., Y.Z., H. Zhang and Y.Y. were responsible for subject recruitment and sample preparation of the Shenyang samples. L.Y. was responsible for subject recruitment and sample preparation of the Guangzhou samples. H. Zheng was responsible for subject recruitment and sample preparation of the Tianjin samples. C.I.A. was responsible for the scientific editing. F.C. oversaw the statistical analyses process. All authors approved the final manuscript.

Note: Supplementary information is available in the online version of the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Therefore, we conducted an evaluation of promising associations in an extended two-stage replication consisting of 7,436 cases and 7,483 controls, focusing on SNPs with *P* values ranging from 10^{-4} to 10^{-6} in our initial GWAS (Supplementary Fig. 1 and Supplementary Table 1). A total of 95 SNPs met these criteria (Online Methods). We used linkage disequilibrium (LD) analyses to reduce the number of SNPs for genotyping and selected for further analysis those with the lowest *P* values among multiple SNPs in high LD ($r^2 \geq 0.8$), yielding 67 SNPs for genotyping in the first replication stage (Supplementary Table 2). We excluded from further analysis the remaining 28 SNPs (Supplementary Table 3), which were in high LD with these 67 SNPs. The results from the 67 SNPs in the GWAS and replication studies are shown in Supplementary Tables 2 and 4. Among these 67 SNPs, 6 were consistently replicated in the first-stage replication of 2,283 cases and 2,243 controls with the same direction of significant associations as in the GWAS (Supplementary Table 4 and Supplementary Fig. 1). The second-stage replication of 5,153 cases and 5,240 controls further verified the significant associations with five loci at 10p14 (rs1663689), 5q32 (rs2895680), 20q13.2 (rs4809957), 5q31.1 (rs247008) and 1p36.32 (rs9439519) (Table 1 and Supplementary Table 4). In the combined analysis, we observed associations for rs1663689 ($P_{combined} = 2.84 \times 10^{-10}$, odds ratio (OR) = 0.88) at 10p14, rs2895680 ($P_{combined} = 6.60 \times 10^{-9}$, OR = 1.14) at 5q32, rs4809957 ($P_{combined} = 1.20 \times 10^{-8}$, OR = 1.13) at 20q13.2, rs247008 ($P_{combined} = 7.68 \times 10^{-8}$, OR = 0.89) at 5q31.1 and rs9439519 ($P_{combined} = 3.65 \times 10^{-6}$, OR = 1.11) at 1p36.32 (Table 1 and Supplementary Table 5), three of which (rs1663689, rs2895680 and rs4809957) achieved genome-wide significance ($P = 5.0 \times 10^{-8}$).

After imputation, we tested 13,849 imputed SNPs (imputed $r^2 > 0.5$, minor allele frequency (MAF) > 0.05, located in the 1,000 kb around each side of the five most significant marker SNPs) (Fig. 1). For 10p14, a series of SNPs were in high LD with rs1663689 ($r^2 = 0.826$ to $r^2 = 1$, $P = 8.76 \times 10^{-5}$ to $P = 6.00 \times 10^{-5}$) (Fig. 1a and Supplementary Table 6). Similarly, we found several SNPs in high LD with rs4809957 at 20q13.2 ($r^2 = 0.967$ to $r^2 = 1$, $P = 7.08 \times 10^{-5}$ to $P = 5.54 \times 10^{-5}$), including one SNP that was genotyped in the GWAS (rs1570669) and one synonymous SNP in exon 6 of *CYP24A1* (rs2296239) (Fig. 1c and Supplementary Table 6). rs9439526, rs9439527 and rs9439528 at 1p36.32 were in high LD with rs9439519 ($r^2 = 0.855$, $P = 7.91 \times 10^{-4}$ to $P = 4.22 \times 10^{-4}$), and rs9439527 had also been directly genotyped in the GWAS (Fig. 1e). Although we detected residual associations at many SNPs in the regions for 5q31.1 and 5q32, the most significant signal occurred at the two index SNPs, rs247008 at 5q31.1 and rs2895680 at 5q32 (Fig. 1b,d and Supplementary Table 6).

Stratification analyses showed heterogeneities in the ORs for rs2895680 among gender ($P = 0.002$), smoking status ($P = 0.002$), smoking dose ($P < 0.001$) and histology type ($P = 0.015$), for rs4809957 among histology type ($P = 0.025$), for rs247008 among age subgroups ($P = 0.006$) and for rs9439519 among gender ($P = 0.011$), smoking status ($P = 0.023$), smoking dose ($P = 0.044$) and histology type ($P < 0.001$) (Supplementary Table 7). We also performed stratification analyses by three study centers (south China: Guangzhou; central China: Nanjing and Shanghai; and north China: Beijing, Shenyang and Tianjin) (Supplementary Fig. 2). We used meta-analyses to combine the results from the three centers and test for heterogeneity. We found similar *P* values for all five SNPs in the metaanalysis

(Supplementary Fig. 2). Interaction analyses showed that rs2895680, rs4809957, rs247008 and rs9439519 interacted multiplicatively with smoking dose to contribute to lung cancer risk (interaction $P = 1.72 \times 10^{-10}$, $P = 5.07 \times 10^{-3}$, $P = 6.77 \times 10^{-3}$ and $P = 4.49 \times 10^{-2}$ for rs2895680, rs4809957, rs247008 and rs9439519, respectively) (Supplementary Table 8). Further analyses in heavy smokers (pack years > 24) and light smokers (0 < pack years ≤ 24) showed that rs2895680 was significantly associated with smoking dose (OR = 1.09, $P = 9.48 \times 10^{-3}$), but this association was not present for any other SNPs (data not shown).

We also performed expression quantitative trait loci (eQTL) analyses (Online Methods) for these five SNPs. We found that rs247008 was a *cis*-eQTL of *SLC22A5* (encoding solute carrier family 22, member 5; $P < 3.96 \times 10^{-7}$, the defined threshold) in a study[11] that measured global gene expression in monocytes in 1,490 unrelated individuals using the Illumina HT-12 v3 BeadChip and that genotyped the subjects using the Affymetrix Genome-Wide Human SNP Array 6.0. For *SLC22A5,* we found 76 *cis*-eQTLs in the data from that study[11], and the most significant *cis*-eQTL was rs2631360 ($P = 6.66 \times 10^{-85}$), which is located 260 kb downstream of rs247008. However, we found no associations with known eQTLs for the four other SNPs.

SNP rs1663689 is located 908 kb downstream of the transcription factor *GATA3* (encoding GATA binding protein 3) at 10p14. *GATA3* is a target for c-Myc and contributes to the development and prognosis of various cancers, including lung cancer[12–14]. Moreover, GATA3 interacts with BRCA1 and functions as a regulator of tumor initiation by targeting caspase-14, a putative tumor suppressor[15,16]. In addition, the *GATA3* region was reported to be associated with Hodgkin's lymphoma in a previous GWAS[13].

SNP rs2895680 is within intron 2 of *STK32A* (encoding serine/threonine kinase 32A) and is between *PPP2R2B* (encoding protein phosphatase 2, regulatory subunit B, β) and *DPYSL3* (encoding dihydropyrimidinase-like 3) at 5q32. *STK32A* and *PPP2R2B* encode members of the serine/threonine kinase family that has a paramount role in cellular homeostasis, transcription factor phosphorylation and cell-cycle regulation[17]. PPP2R2B was also previously reported to interact with PDK1 (pyruvate dehydrogenase kinase, isozyme 1) and antagonize PDK1-Myc signaling to regulate rapamycin sensitivity in cancer[18]. *DPYSL3* was previously reported to be a predictor of response to interferon-γ (IFN-γ) in patients with lung cancer after α-galactosylceramide–pulsed treatment of dendritic cells[19]. Notably, our stratification analysis showed that the effect of rs2895680 on lung cancer risk was more pronounced in females, never smokers and lung adenocarcinoma, a pattern similar to genetic variants in the *TERT-CLPTM1L* locus[20], although the mechanism behind this is unknown.

SNP rs4809957, located in the 3′ untranslated region of *CYP24A1* (encoding cytochrome P450, family 24, subfamily a, polypeptide 1/25-hydroxyvitamin D-24-hydroxylase) at 20q13.2, interacted with smoking dose to contribute to lung cancer risk. *CYP24A1* is a member of the cytochrome P450 superfamily and is an enzyme that is involved in the metabolism of 1,25-dihydroxyvitamin D3 ($1,25(OH)_2D_3$). $1,25(OH)_2D_3$ exerts antiproliferative effects on human cancer cells that are mediated by the vitamin D receptor[21]. *CYP24A1* was previously reported as a highly probable diagnostic biomarker for non–small-cell lung cancer and is overexpressed in lung tumors[22,23]. High expression of CYP24A1 also

abrogates the antiproliferative effects of $1,25(OH)_2D_3$ in lung adenocarcinoma[24]. Furthermore, benzo[*a*]pyrene (BaP), a causative agent oflung cancer, effectively enhances the $1,25(OH)_2D_3$-dependent induction of CYP24A1 (ref. 25).

Although the main effect of the rs247008 variant does not quite reach genome-wide significance, the interaction between this SNP and cigarette smoking suggests that the association might be biologically plausible. The 5q31.1 region is where the T helper type 2 (T_H2) cytokine cluster is located. Extensive evidence has shown a crucial function for T_H2 cytokines in tumor growth, tumor immunogenicity and the host immune response in various cancer types, including lung cancer[26,27]. Furthermore, the immune response of T_H2 cytokines can be induced by cigarette smoking, which promotes widespread inflammatory and mutagenic effects in the lungs[28,29]. Among the genes encoding T_H2 cytokines, *P4HA2* (encoding prolyl-4 hydoxylase-2), *CSF2* (encoding colony stimulating factor 2, also known as *GM-CSF*) and *IL3* (encoding interleukin-3) are near rs247008. The increased production of P4HA2 that is regulated by the tumor suppressor p53 can inhibit angiogenesis and tumor growth, whereas mice with combined loss of IFN-γ, IL-3 and CSF2 develop chronic pulmonary inflammation and lung tumors at high frequencies[30,31]. In addition, rs247008 is a *cis*-eQTL for *SLC22A5*. *SLC22A5* encodes a carnitine transporter, and genetic variants in this gene have been reported to be associated with asthma[32]. *SLC22A5* is also a direct target of PPAR-α (peroxisomal proliferator-activated nuclear receptor-α)[33]. The expression of PPAR-α changes before the formation of lung tumors, and the activation of PPAR-α favors lung tumor development *in vivo*[34]. rs247008 is about 99 kb upstream of *ACSL6* (encoding acyl-CoA synthetase long-chain family member 6). The expression of *ACSL6* mRNA is stimulated by chronic nicotine exposure *in vitro* and *in vivo*, and genetic variants in *ACSL6* show significant associations with the number of cigarette smoked per day[35]. These findings indicate the potential role of the 5q31.1 region in both lung cancer development and the effects of cigarette smoking.

We also observed an interaction between rs9439519 and cigarette smoking at a borderline significant level. SNP rs9439519 is located at 1p36.32 between *AJAP1*, encoding adherens junctions associated protein 1, and *NPHP4*, encoding nephrocystin-4. *AJAP1* was recently described as a putative tumor suppressor gene and is involved in the inhibition of cell adhesion and migration[36]. *NPHP4* encodes a cilia-associated protein and acts as a potent negative regulator of Hippo signaling, which has an essential role in tumor suppression and the control of cell proliferation[37]. However, rs9439519 did not reach genome-wide significance in our analyses, and these results should be treated with caution.

In our previous lung cancer GWAS in the Chinese population[1], we replicated two loci at 3q28 and 5p15.33, which were identified by other GWAS, but not loci at 15q25.1 and 20p12.1. Three additional loci at 3q29 (rs2131877) and 18p11.22 (rs11080466 and rs11663246) were recently identified in a Korean population[8,9]. However, we did not observe consistent associations in our GWAS data with these three loci. We here identify three new susceptibility loci (5q32, 10p14 and 20q13.2) for lung cancer risk and two additional suggestive loci at 1p36.32 and 5q31.1 in the Chinese population. More crucially, we highlight four genetic modulators (5q32, 20q13.2, 5q31.1 and 1p36.32) of lung cancer risk that interact with cigarette smoking. However, the SNPs analyzed in our study were

limited: we did not take SNPs with $1.0 \times 10^{-4} < P \leq 1.0 \times 10^{-2}$ forward for replication. Further studies are required to provide a more comprehensive understanding of the genetics of lung cancer by combining GWAS with large samples drawn from diverse ethnic populations.

METHODS

Methods and any associated references are available in the online version of the paper.

ONLINE METHODS

Study design.

All participating subjects provided informed consent, and this study was approved by the institutional review boards of each participating institution. The study populations from the GWAS and the first-stage replication have been described previously[1]. For the second-stage replication, we enlarged the sample size by testing additional 1,123 cases and 1,074 controls from Guangzhou (248 cases and 64 controls), Shenyang (107 cases and 224 controls) and Tianjin (768 cases and 786 controls). As a result, the second-stage replication samples consisted of 5,153 cases and 5,240 controls from Nanjing and Shanghai (941 cases and 1,069 controls), Beijing (932 cases and 936 controls), Tianjin (768 cases and 786 controls), Shenyang (1,133 cases and 1,251 controls) and Guangzhou (1,379 cases and 1,198 controls).

The GWAS was conducted using an Affymetrix Genome-Wide Human SNP Array 6.0, followed by a systematic quality control step before the association analysis[1]. In brief, SNPs were excluded if they did not map on autosomal chromosomes, had a call rate <95%, MAF < 0.05, $P < 1 \times 10^{-5}$ for Hardy-Weinberg equilibrium in all GWAS samples or $P < 1 \times 10^{-4}$ for Hardy-Weinberg equilibrium in either the Nanjing or Beijing study samples. We removed samples with low call rates (13 subjects with call rates <95%), ambiguous gender (7 cases), familial relationships (89 subjects), extreme heterozygosity rate (22 samples) and outliers (4 subjects). As a result, a total of 2,331 lung cancer cases and 3,077 controls with 591,370 SNPs were included in the final analysis of the GWAS.

The statistical analysis of the GWAS and two-stage replication data has also been described elsewhere[1]. Briefly, the genome-wide association analysis was performed using an additive model in a logistic regression analysis in PLINK 1.07 (see URLs). We used the Minimac software (see URLs) to impute untyped SNPs using the LD information from the hg18/1000 Genomes database (with CHB+JPT as the reference set, released June 2010). Chromosome region was plotted using LocusZoom 1.1 (see URLs). The potential gene-environment interaction was evaluated by using the 'SNPassoc' package in R (see URLs). General analyses were performed with R software (version 2.11.1; The R Foundation for Statistical Computing).

URLs.
PLINK 1.07, http://pngu.mgh.harvard.edu/~purcell/plink/; R 2.11.1 statistical environment, http://www.cran.r-project.org/; Minimac, http://genome.sph.umich.edu/wiki/Minimac; LocusZoom1.1, http://csg.sph.umich.edu/locuszoom/; GTEx (Genotype-Tissue Expression) eQTL Browser, http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi; eQTL.chicago.edu, http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/; Gene Expression Analysis Based on Imputed Genotypes, http://www.sph.umich.edu/csg/liang/imputation/.

SNP selection and genotyping for replication.

We selected SNPs meeting the following criteria for the first stage replication: (i) $1.0 \times 10^{-6} < P \leq 1.0 \times 10^{-4}$ for all GWAS samples and a consistent association at $P \leq 1.0 \times 10^{-2}$ in both the Nanjing and Beijing studies; (ii) not located in the same chromosome region of SNPs reported in our previous GWAS study; (iii) having clear genotyping clusters; (iv) and only the SNP with the lowest *P* value was selected when multiple SNPs were observed in strong LD ($r^2 \geq 0.8$). As a result, a total of 95 SNPs survived in the conditions of criteria (i), (ii) and (iii), and among these, 67 SNPs met criteria (iv) (Supplementary Table 2); the remaining 28 SNPs were in high LD with the 67 selected SNPs (Supplementary Table 3). Therefore, 67 SNPs were included the first-stage replication. Significantly associated SNPs ($P < 0.05$) in the first-stage replication were genotyped in the second-stage validation samples.

Genotyping analyses in the first-stage replication were performed using the middle-throughput TaqMan OpenArray Genotyping Platform (Applied Biosystems, Inc.) and the iPLEX Sequenom MassARRAY platform (Sequenom, Inc). A TaqMan allelic discrimination assay (Applied Biosystems, Inc.) was used for the second-stage validation samples. The primers and probes used are available on request. A series of methods was used to control the quality of genotyping: (i) case and control samples were mixed on each plate; (ii) genotyping was performed without knowing case or control status; (iii) two water controls were used in each plate as blank controls; (iv) 5% of the samples were randomly selected to repeat the genotyping, as blind duplicates, and the reproducibility was 100%; and (v) 2,421 samples were randomly selected and examined using both the TaqMan OpenArray platform and the TaqMan assay for rs1663689, rs2895680, rs247008, rs4809957 and rs9439519, yielding concordance rates of between 99.95% and 99.98%.

Cis-eQTL analysis.

For the *cis*-eQTL analysis, we applied the publicly available data from GTEx (Genotype-Tissue Expression) eQTL Browser, eQTL.chicago. edu and Gene Expression Analysis Based on Imputed Genotypes (see URLs). The *cis* associations between SNPs at the five identified loci and the expressions of nearby genes were evaluated in a variety of cells and tissues, including lymphoblastoid cell lines[38–43], monocytes[11], fibroblasts[43] and liver[44] and brain tissues[45].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Jing Dong[1,21], Zhibin Hu[1,2,3,21], Chen Wu[4,21], Huan Guo[5], Baosen Zhou[6], Jiachun Lv[7], Daru Lu[8], Kexin Chen[9], Yongyong Shi[10], Minjie Chu[1], Cheng Wang[1], Ruyang Zhang[1], Juncheng Dai[1], Yue Jiang[1], Songyu Cao[1], Zhenzhen Qin[1], Dianke Yu[4], Hongxia Ma[1], Guangfu Jin[1,2], Jianhang Gong[1], Chongqi Sun[1], Xueying Zhao[8], Zhihua Yin[6], Lei Yang[7], Zhiqiang Li[10], Qifei Deng[5], Jiucun Wang[8], Wei Wu[6], Hong Zheng[9], Guoquan Zhou[10], Hongyan Chen[8], Peng Guan[6], Zhihang Peng[1], Yijiang Chen[11], Yongqian Shu[11], Lin Xu[12], Xiangyang Liu[13], Li Liu[14], Pin Xu[15], Baohui

Han[16], Chunxue Bai[17], Yuxia Zhao[18], Haibo Zhang[19], Ying Yan[19], Christopher I Amos[20], Feng Chen[1], Wen Tan[4], Li Jin[8], Tangchun Wu[5], Dongxin Lin[4], and Hongbing Shen[1,2,3]

Affiliations

[1]Department of Epidemiology and Biostatistics and Ministry of Education (MOE) Key Lab for Modern Toxicology, School of Public Health, Nanjing Medical University, Nanjing, China.

[2]Section of Clinical Epidemiology, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Cancer Center, Nanjing Medical University, Nanjing, China.

[3]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China.

[4]State Key Laboratory of Molecular Oncology and Department of Etiology and Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

[5]Institute of Occupational Medicine and Ministry of Education, Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

[6]Department of Epidemiology, School of Public Health, China Medical University, Shenyang, China.

[7]The Institute for Chemical Carcinogenesis, State Key Laboratory of Respiratory Disease, Guangzhou Medical College, Guangzhou, China.

[8]State Key Laboratory of Genetic Engineering, Center for Fudan-VARI Genetic Epidemiology and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China.

[9]Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Hospital and Institute, Tianjin, China.

[10]Bio-X Center and Affiliated Changning Mental Health Center, Ministry of Education Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Shanghai Jiao Tong University, Shanghai, China.

[11]Department of Thoracic Surgery and Oncology, First Affiliated Hospital of Nanjing Medical University, Nanjing, China.

[12]Department of Thoracic Surgery, Affiliated Cancer Hospital of Nanjing Medical University, Jiangsu Cancer Hospital, Nanjing, China.

[13]Department of Thoracic Surgery, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

[14]Cancer Center of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

[15]Department of Oncology, Wuhan Iron and Steel Group/Corporation Staff-Worker Hospital, Wuhan, China.

[16]Department of Respiratory Disease, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China.

[17]Department of Respiratory Disease, Zhongshan Hospital, Fudan University, Shanghai, China.

[18]Department of Radiation Oncology, First Affiliated Hospital of China Medical University, Shenyang, China.

[19]Department of Radiotherapy, Shenyang Northern Hospital, Shenyang, China.

[20]Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

[21]These authors contributed equally to this work.

ACKNOWLEDGMENTS

This work is funded by the China National High-Tech Research and Development Program Grant (2009AA022705) and partly funded by the National Key Basic Research Program Grant (2011CB503805) and the National Natural Science Foundation of China (30730080, 30972541 and 30901233), Jiangsu Natural Science Foundation (BK2011028), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (11KJA330001), the US National Institutes of Health Grant (U19 CA148127) and the Priority Academic Program Development of Jiangsu Higher Education Institutions. The authors thank all the study subjects, research staff and students who participated in this work.

References

1. Hu Z et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat. Genet. 43, 792–796 (2011). [PubMed: 21725308]
2. McKay JD et al. Lung cancer susceptibility locus at 5p15.33. Nat. Genet. 40, 1404–1406 (2008). [PubMed: 18978790]
3. Wang Y et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat. Genet. 40, 1407–1409 (2008). [PubMed: 18978787]
4. Hung RJ et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452, 633–637 (2008). [PubMed: 18385738]
5. Miki D et al. Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. Nat. Genet. 42, 893–896 (2010). [PubMed: 20871597]
6. Amos CI et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat. Genet. 40, 616–622 (2008). [PubMed: 18385676]
7. Li Y et al. Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. Lancet Oncol. 11, 321–330 (2010). [PubMed: 20304703]
8. Ahn MJ et al. The 18p11.22 locus is associated with never smoker non-small cell lung cancer susceptibility in Korean populations. Hum. Genet. 131, 365–372 (2012). [PubMed: 21866343]
9. Yoon KA et al. A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population. Hum. Mol. Genet. 19, 4948–4954 (2010). [PubMed: 20876614]
10. Thorgeirsson TE et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452, 638–642 (2008). [PubMed: 18385739]
11. Zeller T et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. PLoS ONE 5, e10693 (2010).

12. Maurice D, Hooper J, Lang G & Weston K c-Myb regulates lineage choice in developing thymocytes via its target gene Gata3. EMBO J. 26, 3629–3640 (2007). [PubMed: 17641686]
13. E nciso-Mora V et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat. Genet. 42, 1126–1130 (2010). [PubMed: 21037568]
14. Yang Y et al. The Notch ligand Jagged2 promotes lung adenocarcinoma metastasis through a miR-200-dependent pathway in mice. J. Clin. Invest. 121, 1373–1385 (2011). [PubMed: 21403400]
15. Tkocz D et al. BRCA1 and GATA3 corepress FOXC1 to inhibit the pathogenesis of basal-like breast cancers. Oncogene published online, doi:10.1038/onc.2011.531 (28 11 2011).
16. Asselin-Labat ML et al. Gata-3 negatively regulates the tumor-initiating capacity of mammary luminal progenitor cells and targets the putative tumor suppressor caspase-14. Mol. Cell. Biol. 31, 4609–4622 (2011). [PubMed: 21930782]
17. Edelman AM, Blumenthal DK & Krebs EG Protein serine/threonine kinases. Annu. Rev. Biochem. 56, 567–613 (1987). [PubMed: 2956925]
18. Tan J et al. B55β-associated PP2A complex controls PDK1-directed myc signaling and modulates rapamycin sensitivity in colorectal cancer. Cancer Cell 18, 459–471 (2010). [PubMed: 21075311]
19. Okita K et al. A set of genes associated with the interferon-γ response of lung cancer patients undergoing a-galactosylceramide-pulsed dendritic cell therapy. Cancer Sci. 101, 2333–2340 (2010). [PubMed: 20804502]
20. Hsiung CA et al. The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. PLoS Genet. 6, e1001051 (2010).
21. Peehl DM et al. Antiproliferative effects of 1,25-dihydroxyvitamin D3 on primary cultures of human prostatic cells. Cancer Res. 54, 805–810 (1994). [PubMed: 7508338]
22. Kim B et al. Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data. Cancer Res. 67, 7431–7438 (2007). [PubMed: 17671213]
23. Anderson MG, Nakane M, Ruan X, Kroeger PE & Wu-Wong JR Expression of VDR and CYP24A1 mRNA in human tumors. Cancer Chemother. Pharmacol. 57, 234–240 (2006). [PubMed: 16180015]
24. Chen G et al. CYP24A1 is an independent prognostic marker of survival in patients with lung adenocarcinoma. Clin. Cancer Res. 17, 817–826 (2011). [PubMed: 21169243]
25. Matsunawa M et al. The aryl hydrocarbon receptor activator benzo[a]pyrene enhances vitamin D3 catabolism in macrophages. Toxicol. Sci. 109, 50–58 (2009). [PubMed: 19244278]
26. Yamamura M, Modlin RL, Ohmen JD & Moy RL Local expression of antiinflammatory cytokines in cancer. J. Clin. Invest. 91, 1005–1010 (1993). [PubMed: 8450029]
27. Huang M et al. Human non-small cell lung cancer cells express a type 2 cytokine pattern. Cancer Res. 55, 3847–3853 (1995). [PubMed: 7641203]
28. Sorrentino R et al. Plasmacytoid dendritic cells prevent cigarette smoke and Chlamydophila pneumoniae–induced Th2 inflammatory responses. Am. J. Respir. Cell Mol. Biol. 43, 422–431 (2010). [PubMed: 19901344]
29. Vassallo R, Tamada K, Lau JS, Kroening PR & Chen L Cigarette smoke extract suppresses human dendritic cell function leading to preferential induction of Th-2 priming. J. Immunol. 175, 2684–2691 (2005). [PubMed: 16081845]
30. Teodoro JG, Parker AE, Zhu X & Green MR p53-mediated inhibition of angiogenesis through up-regulation of a collagen prolyl hydroxylase. Science 313, 968–971 (2006). [PubMed: 16917063]
31. Dougan M et al. A dual role for the immune response in a mouse model of inflammation-associated lung cancer. J. Clin. Invest. 121, 2436–2446 (2011). [PubMed: 21537082]
32. Moffatt MF et al. A large-scale, consortium-based genomewide association study of asthma. N. Engl. J. Med. 363, 1211–1221 (2010). [PubMed: 20860503]
33. Wen G, Ringseis R & Eder K Mouse OCTN2 is directly regulated by peroxisome proliferator–activated receptor α (PPARα) via a PPRE located in the first intron. Biochem. Pharmacol. 79, 768–776 (2010). [PubMed: 19819229]

34. Li MY et al. Roles of peroxisome proliferator-activated receptor-α and -γ in the development of non-small cell lung cancer. Am. J. Respir. Cell Mol. Biol. 43, 674–683 (2010). [PubMed: 20081051]
35. Chen J et al. ACSL6 is associated with the number of cigarettes smoked and its expression is altered by chronic nicotine exposure. PLoS ONE 6, e28790 (2011).
36. McDonald JM et al. The SHREW1 gene, frequently deleted in oligodendrogliomas, functions to inhibit cell adhesion and migration. Cancer Biol. Ther. 5, 300–304 (2006). [PubMed: 16410724]
37. Habbig S et al. NPHP4, a cilia-associated protein, negatively regulates the Hippo pathway. J. Cell Biol. 193, 633–642 (2011). [PubMed: 21555462]
38. Dixon AL et al. A genome-wide association study of global gene expression. Nat. Genet. 39, 1202–1207 (2007). [PubMed: 17873877]
39. Stranger BE et al. Population genomics of human gene expression. Nat. Genet. 39, 1217–1224 (2007). [PubMed: 17873874]
40. Veyrieras JB et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 4, e1000214 (2008).
41. Pickrell JK et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768–772 (2010). [PubMed: 20220758]
42. Montgomery SB et al. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464, 773–777 (2010). [PubMed: 20220756]
43. Dimas AS et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325, 1246–1250 (2009). [PubMed: 19644074]
44. Schadt EE et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 6, e107 (2008).
45. Myers AJ et al. A survey of genetic human cortical gene expression. Nat. Genet. 39, 1494–1499 (2007). [PubMed: 17982457]

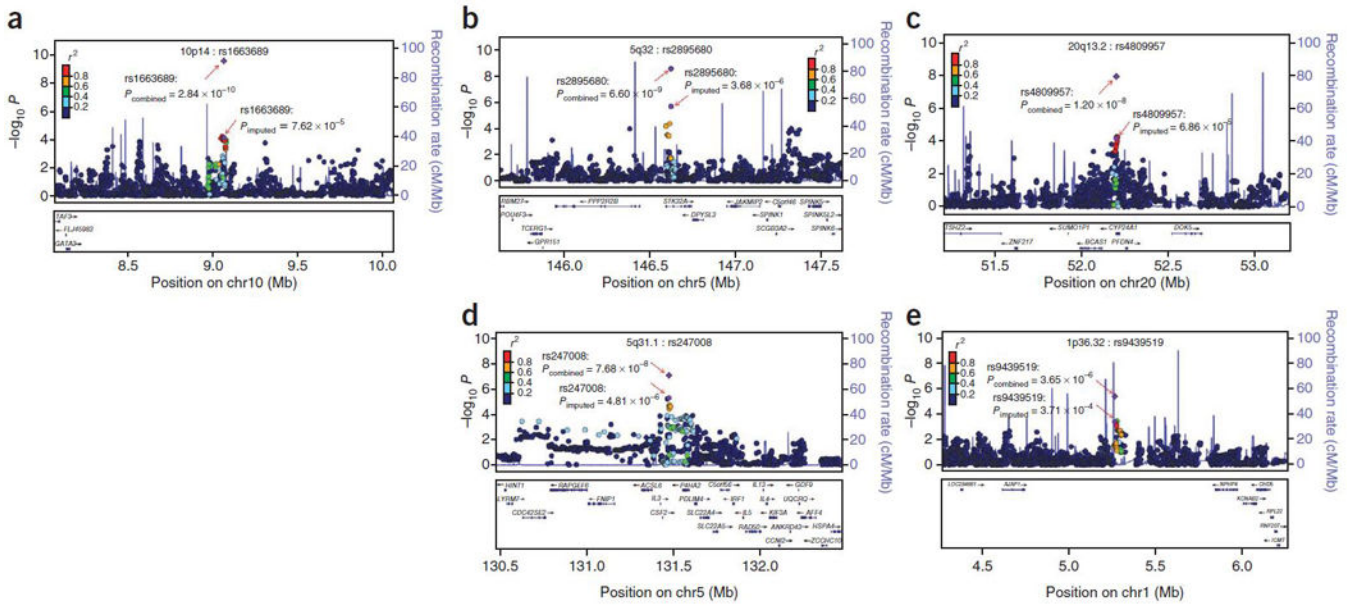

Figure 1.

Regional plots of the five identified marker SNPs (rs1663689 at 10p14, rs2895680 at 5q32, rs4809957 at 20q13.2, rs247008 at 5q31.1 and rs9439519 at 1p36.32). Results ($-\log_{10} P$) are shown for SNPs in the region flanking 1,000 kb on either side of the marker SNPs. The marker SNPs are shown in purple, and the r^2 values for the other SNPs are indicated in different colors. The genes within the region of interest are annotated, with arrows indicating the direction of transcription.

Table 1

summary of GWAS and replication studies for the five associated SNPs

SNP	Study	MAF		OR_{add}	
		Cases	Controls	(95% CI)[c]	P_{add}[c]
rs1663689	GWAS	0.382	0.421	0.84 (0.77–0.91)	3.03×10^{-5}
10p14 A/G[a]	Replication 1	0.384	0.421	0.83 (0.76–0.91)	7.73×10^{-5}
906,520[b]	Replication 2	0.394	0.416	0.91 (0.86–0.96)	1.14×10^{-3}
	Combined all	0.389	0.419	0.88 (0.84–0.91)	2.84×10^{-10}
	SC	0.390	0.419	0.87 (0.82–0.93)	1.90×10^{-5}
	AC	0.387	0.419	0.87 (0.83–0.92)	1.12×10^{-7}
	SCC	0.384	0.419	0.85 (0.77–0.94)	2.11×10^{-3}
	Other	0.402	0.419	0.92 (0.80–1.05)	1.93×10^{-1}
rs2895680	GWAS	0.322	0.281	1.22 (1.12–1.33)	4.59×10^{-6}
5q32 T/C[a]	Replication 1	0.335	0.297	1.16 (1.05–1.29)	3.95×10^{-3}
146,624,308[b]	Replication 2	0.314	0.299	1.07 (1.00–1.13)	3.53×10^{-2}
	Combined all	0.321	0.293	1.14 (1.09–1.19)	6.60×10^{-9}
	SC	0.299	0.293	1.05 (0.98–1.12)	1.53×10^{-1}
	AC	0.341	0.293	1.19 (1.13–1.25)	3.22×10^{-11}
	SCC	0.309	0.293	1.06 (0.95–1.18)	3.18×10^{-1}
	Other	0.300	0.293	1.01 (0.88–1.16)	8.69×10^{-1}
rs4809957	GWAS	0.387	0.347	1.19 (1.09–1.30)	7.62×10^{-5}
20q13.2 C/T[a]	Replication 1	0.379	0.357	1.10 (1.01–1.21)	3.85×10^{-2}
52,204,578[b]	Replication 2	0.398	0.374	1.11 (1.05–1.17)	4.83×10^{-4}
	Combined all	0.391	0.362	1.13 (1.08–1.18)	1.20×10^{-8}
	SC	0.387	0.362	1.11 (1.05–1.19)	7.41×10^{-4}
	AC	0.392	0.362	1.13 (1.07–1.18)	3.12×10^{-6}
	SCC	0.400	0.362	1.17 (1.05–1.29)	3.80×10^{-3}
	Other	0.389	0.362	1.11 (0.97–1.27)	1.18×10^{-1}
rs247008	GWAS	0.430	0.475	0.83 (0.76–0.90)	1.14×10^{-5}
5q31.1 G/A[a]	Replication 1	0.447	0.466	0.88 (0.81–0.97)	6.84×10^{-3}
131,475,003[b]	Replication 2	0.451	0.466	0.93 (0.87–0.98)	8.36×10^{-3}
	Combined all	0.445	0.469	0.89 (0.86–0.93)	7.68×10^{-8}
	SC	0.443	0.469	0.87 (0.82–0.92)	5.29×10^{-6}
	AC	0.441	0.469	0.89 (0.85–0.93)	1.98×10^{-6}
	SCC	0.459	0.469	0.93 (0.84–1.03)	1.88×10^{-1}
	Other	0.471	0.469	0.98 (0.86–1.11)	7.07×10^{-1}
rs9439519	GWAS	0.307	0.275	1.21 (1.10–1.32)	7.69×10^{-5}

SNP	Study	MAF		OR_{add}	
		Cases	Controls	(95% CI)[c]	P_{add}[c]
1p36.32 A/G[a]	Replication 1	0.301	0.270	1.13 (1.03–1.25)	1.38×10^{-2}
5,264,494[b]	Replication 2	0.292	0.276	1.08 (1.01–1.15)	1.95×10^{-2}
	Combined all	0.297	0.274	1.11 (1.06–1.16)	3.65×10^{-6}
	SC	0.297	0.274	1.12 (1.05–1.20)	7.70×10^{-4}
	AC	0.275	0.274	1.00 (0.95–1.06)	9.84×10^{-1}
	SCC	0.319	0.274	1.25 (1.12–1.40)	1.19×10^{-4}
	Other	0.484	0.274	2.41 (2.12–2.75)	4.86×10^{-40}

[a] Major/minor alleles.

[b] Base pair position of the SNP.

[c] Adjusted by age, gender, pack years of smoking and principal components for the GWAS and adjusted by age, gender and pack years of smoking for other analyses. OR_{add} and P_{add} were calculated using the additive model.

MAF, minor allele frequency; SC, squamous-cell carcinoma; AC, adenocarcinoma; SCC, small-cell carcinoma;

'other' includes large-cell lung cancer and mixed cell carcinoma.