Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia

Qing Lan, Chao A Hsiung, Keitaro Matsuo, Yun-Chul Hong, Adeline Seow, Zhaoming Wang, H Dean Hosgood III, Kexin Chen, Jiu-Cun Wang, Nilanjan Chatterjee, Wei Hu, Maria Pik Wong, Wei Zheng, Neil Caporaso, Jae Yong Park, Chien-Jen Chen, Yeul Hong Kim, Young Tae Kim, Maria Teresa Landi, Hongbing Shen, Charles Lawrence, Laurie Burdette, Meredith Yeager, Jeffrey Yuenger, Kevin B Jacobs, I-Shou Chang, Tetsuya Mitsudomi, Hee Nam Kim, Gee-Chen Chang, Bryan A Bassig, Margaret Tucker, Fusheng Wei, Zhihua Yin, Chen Wu, She-Juan An, Biyun Qian, Victor Ho Fun Lee, Daru Lu, Jianjun Liu, Hyo-Sung Jeon, Chin-Fu Hsiao, Jae Sook Sung, Jin Hee Kim, Yu-Tang Gao, Ying-Huang Tsai, Yoo Jin Jung, Huan Guo, Zhibin Hu, Amy Hutchinson, Wen-Chang Wang, Robert Klein, Charles C Chung, In-Jae Oh, Kuan-Yu Chen, Sonja I Berndt, Xingzhou He, Wei Wu, Jiang Chang, Xu-Chao Zhang, Ming-Shyan Huang, Hong Zheng, Junwen Wang, Xueying Zhao, Yuqing Li, Jin Eun Choi, Wu Chou Su, Kyong Hwa Park, Sook Whan Sung, Xiao-Ou Shu, Yuh-Min Chen, Li Liu, Chang Hyun Kang, Lingmin Hu, Chung-Hsing Chen, William Pao, Young-Chul Kim, Tsung-Ying Yang, Jun Xu, Peng Guan, Wen Tan, Jian Su, Chih-Liang Wang, Haixin Li, Alan Dart Loon Sihoe, Zhenhong Zhao, Ying Chen, Yi Young Choi, Jen-Yu Hung, Jun Suk Kim, Ho-Il Yoon, Qiuyin Cai, Chien-Chung Lin, Inkyu Park, Ping Xu, Jing Dong, Christopher Kim, Qincheng He, Reury-Perng Perng, Takashi Kohno, Sun-Seog Kweon, Chih-Yi Chen, Roel Vermeulen, Junjie Wu, Wei-Yen Lim, Kun-Chieh Chen, Wong-Ho Chow, Bu-Tian Ji, John K C Chan, Minjie Chu, Yao-Jen Li, Jun Yokota, Jihua Li, Hongyan Chen, Yong-Bing Xiang, Chong-Jen Yu, Hideo Kunitoh, Guoping Wu, Li Jin, Yen-Li Lo, Kouya Shiraishi, Ying-Hsiang Chen, Hsien-Chih Lin, Tangchun Wu, Yi-Long Wu, Pan-Chyr Yang, Baosen Zhou, Min-Ho Shin, Joseph F Fraumeni Jr., Dongxin Lin, Stephen J Chanock, Nathaniel Rothman

Supplemental Note

Study Information:

Supplemental Table 1 describes the 14 studies involved in this GWAS. Nine of the 14 lung cancer studies in East Asia pooled for this effort were described previously in our first lung cancer GWAS[1] and in an effort to replicate additional lung cancer GWAS findings[2]. These studies include the Chinese Academy of Medical Sciences Cancer Hospital Study (CAMSCH)[3], the Fudan Lung Cancer Study (FLCS), the Genetic Epidemiological Study of Lung Adenocarcinoma (GELAC)[4], the Genes and Environment in Lung Cancer, Singapore Study (GEL-S)[5], the South Korea Multi-Center Lung Cancer Study (SKLCS)[6-8], the Shenyang Lung Cancer Study (SLCS)[9], the Shanghai Women's Health Cohort Study (SWHS)[10, 11], the Tianjin Lung Cancer Study (TLCS)[12], and the Wuhan lung cancer study (WLCS)[13]. Additional studies included in this GWAS are the Guangdong Study (GDS), the Hong Kong Study (HKS), the Japanese Female Lung Cancer Collaborative Study (JLCS)[14], the Chonnam National University Lung Cancer Study (CNULCS) Study[15], and the Yunnan Lung Cancer Study (YLCS). Besides one prospective cohort (SWHS), all studies were case–control by design. Questionnaire data on environmental tobacco smoke (ETS) was available for five studies (GEL-S, GELAC, SLCS, SWHS, YLCS). Supplemental Table 1 includes the total number of participants included in the study, by study center. Ages were similar in cases and controls across all studies, except for the GDS (Supplemental Table 1). Each study was approved by their local institutional review board and all study participants provided informed consent.

The CAMSCH study has been previously described[3]. Briefly, cases were recruited from Beijing city and surrounding provinces at the Chinese Academy of Medical Sciences Cancer Hospital, which is located in Beijing, China. Controls were from a community nutritional survey. Subjects were cancer-free individuals in the same region during the same period as the cases were collected. All subjects were ethnic Chinese.

The FLCS enrolled cases and controls from four hospitals in the Shanghai area of China. Cases and controls were from four hospitals, including the Shanghai Changhai Hospital, the Jiangsu Taizhou People's Hospital, the Zhongshan Hospital Fudan University, and the Shanghai Chest Hospital. Cases were patients who presented to these hospitals with primary lung cancer. Controls were patients of these hospitals that were diagnosed with diseases unrelated to any cancer. All subjects were ethnic Chinese.

The GDS enrolled cases and controls from the Guangdong area of China. Cases were incident in-patients diagnosed with lung cancer at the Guangdong Lung Cancer Institute and Guangdong General Hospital. All of the cases were histologically confirmed primary lung cancer (all histological types). Controls were healthy, cancer-free patients selected from health examination clinics of the same hospitals. All subjects were ethnic Chinese.

The GEL-S study is a case-control study carried out in the 5 major public-sector hospitals in Singapore[5] (Changi General Hospital, National Cancer Centre, National University Hospital, Singapore General Hospital and Tan Tock Seng Hospital). Eligible cases were Chinese women with a diagnosis of primary lung carcinoma (all histological types). Controls were selected from the same hospital, frequency-matched by 10-year age groups and admitted or seen in the hospital within 30 days of the date of diagnosis of the corresponding case. Exclusion criteria were admissions for the diagnosis or management of malignancy or chronic respiratory disease (excluding tuberculosis). All subjects were ethnic Chinese.

The GELAC study recruited cases from six hospitals, including the National Taiwan University Hospital, Taipei Veterans General Hospital, Chang-Gung Memorial Hospital, Taichung Veterans General Hospital, National Cheng-Kung University Hospital, and Kaohsiung Medical University Hospital in Taiwan. Cases were 18 years or more of age with incident primary lung cancer. Control subjects were cancer-free, randomly selected from the health examination clinics of the same hospitals during the same time period of case recruitment and frequency matched by age, gender. The study has been previously described[4, 16]. All subjects were ethnic Chinese.

The HKS enrolled cases and controls from the Queen Mary Hospital and the Grantham Hospital in Hong Kong. Cases were lung cancer patients who underwent tumor resection in the Grantham Hospital and the Queen Mary Hospital, ranged from at least 18 years to 80 years old. Controls were patients with non-cancer diagnoses at Queen Mary Hospital, comparable in age to cases. All subjects were ethnic Chinese.

The JLCS enrolled cases and cancer-free controls at the Aichi Cancer Center Research Institute and the National Cancer Center Hospital. Case subjects were patients with no history of cancer who were histologically diagnosed with lung cancer. Control subjects were age- and sex-matched first-visit outpatients during the same period who were confirmed to have no cancer or history of neoplasia. The initial phase on enrollment for this study has been previously described[14]. All subjects were ethnic Japanese.

The SKLCS study of lung cancer in Korea consists of several studies conducted at multiple study centers, which have been previously described[6-8], as follows: The SNU study of lung cancer in Korea was conducted between 2001 and 2008 and is described in. The lung cancer patients were recruited from patients at Seoul National University Hospitals in Seoul and Bundang, while the age- and sex-matched controls were participants in the health checkup program of Inha University Hospital. Cases were recruited from Seoul National University Hospitals in Seoul and Bundang between 2005 and 2007, and between 2001 and 2008, respectively with no age, histological or stage restrictions. The controls were without lung cancer admitted to the health checkup program of Inha University Hospital between 2001 and 2004 were recruited. Eligibility criteria only included subjects without any history of lung cancer. All study subjects were ethnic Koreans. The KUMC study was carried out between August 2001 and February 2008 and is described in. Lung cancer patients were recruited from the patient pool at the Genomic Research Center for Lung and Breast/Ovarian Cancer and the Inha University Medical Center. Age- and sex-matched controls were participants in the health checkup program of Inha University Hospital. All study subjects were ethnic Koreans. The KNUH study was conducted between January 2006 and December 2007 and is described in. Cases were recruited from the patient pool at the Kyungpook National University Hospital. The control subjects were randomly selected from a pool of healthy volunteers who visited the general health check-up center at the Hospital. All study subjects were ethnic Koreans.

The SLCS was carried out in the Liaoning Province in Northeast China and has been previously described[9]. Cases were never smoking females with histologically confirmed lung cancers. Controls were selected from cancer-free patients with other lung diseases but free of cancer history and symptom. Controls were all non-smoking females and frequency matched to cases on age (± 5 years). All patients were ethnic Chinese.

The CNULCS was conducted in the Chonnam National University Hwasun Hospital in Jeollanam-do, South Korea and has been previously described[15]. Consecutive, newly diagnosed incident lung cancer cases were enrolled into this study. Cases were histologically confirmed. Cases with secondary or recurrent tumors were excluded. All subjects were ethnic Koreans.

The SWHS has been previously described in detail[10, 11]. We used a nested case-control study design in this cohort. Briefly, a total of 75,221 women from urban Shanghai, China, ages 40 to 70 years, participated in the study and completed the baseline surveys. Of those, 279 women who were found to be younger than age 40 years or older than age 70 years, 1,490 women who had a prevalent case of cancer, and 10 women who did not accrue any follow-up time, were excluded. The remaining women were followed through December 2008. Three biennial in-person follow-ups for all living cohort members were conducted by in-home visits with response rates of greater than ~95%. Incident cases diagnosed with malignant neoplasm of the bronchus or lung were included in this study. Controls were selected among the study participants in the cohort who were cancer free at the time of cancer diagnosis of the matched cases. One control was randomly selected and matched with each case by age at baseline (±2 years). All subjects were ethnic Chinese.

The TLCS enrolled patients with newly diagnosed and histologically confirmed primary non-small cell lung cancer from Tianjin Medical University Cancer Hospital in Tianjin, China. This study has been previously described[12]. Patients with a previous medical history of cancer, or previous radiotherapy or chemotherapy were excluded. Controls were recruited from patients who underwent regular health check-up during the same time when cases were recruited and lived in the same neighborhoods or nearby communities. The controls were frequency-matched to cases on gender and age. All subjects were ethnic Chinese.

The WLCS study was carried out in Wuhan, China and has been described elsewhere[13]. Cases were consecutively recruited from the Wuhan Zhongnan Hospital. There were no restrictions based on age, sex, and histology, but patients with a previous cancer history or with unknown conditions of radiotherapy or chemotherapy were excluded. Controls were cancer-free outpatients from other clinics in the same hospital during the same period when the cases were recruited. All controls were frequency matched to the cases by age, sex, and residential area (urban or countryside). All subjects were ethnic Chinese.

The YLCS enrolled never smoking female cases and never smoking female controls in the Yunnan Province of China. Incident lung cancer cases and controls without history of cancer were enrolled. Cases and controls had to be at least 18 years old to be eligible. All subjects were ethnic Chinese.

Support for individual studies:

SKLCS (Y.T.K.) - National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (2011-0016106). (J.C.) - This work was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health &Welfare, Republic of Korea (grant no. 0720550-2). (J.S.S) – grant number is A010250.
WLCS (T.W.) - National Key Basic Research and Development Program (2011CB503800)
SLCS (B.Z.) - National Nature Science Foundation of China (81102194). Liaoning Provincial Department of Education (LS2010168). China Medical Board (00726).
GDS (Y.L.W.) - Foundation of Guangdong Science and Technology Department (2006B60101010, 2007A032000002, 2011A030400010). Guangzhou Science and Information Technology Bureau (2011Y2-00014). Chinese Lung Cancer Research Foundation, National

Natural Science Foundation of China (81101549). Natural Science Foundation of Guangdong Province (S2011010000792).
TLCS (K.C., B.Q) - Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT), China (IRT1076). Tianjin Cancer Institute and Hospital. National Foundation for Cancer Research US.
FLCS (J.C.W., D.R., L.J.) - Ministry of Health (201002007). Ministry of Science and Technology (2011BAI09B00). National S&T Major Special Project (2011ZX09102-010-01). China National High-Tech Research and Development Program (2012AA02A517, 2012AA02A518). National Science Foundation of China (30890034). National Basic Research Program (2012CB944600). Scientific and Technological Support Plans from Jiangsu Province (BE2010715).
NLCS (H.S.) - China National High-Tech Research and Development Program Grant (2009AA022705). Priority Academic Program Development of Jiangsu Higher Education Institution. National Key Basic Research Program Grant (2011CB503805).
GEL-S (A.S.) - National Medical Research Council Singapore grant (NMRC/0897/2004, NMRC/1075/2006). (J.Liu) - Agency for Science, Technology and Research (A*STAR) of Singapore.
GELAC (C.A.H.) - National Research Program on Genomic Medicine in Taiwan (DOH98-TD-G-111-015). National Research Program for Biopharmaceuticals in Taiwan (DOH 100-TD-PB-111-TM013). National Science Council, Taiwan (NSC 100-2319-B-400-001).
YLCS (Q.L.) - Supported by the intramural program of U.S. National Institutes of Health, National Cancer Institute.
SWHS (W.Z., W-H.C., N.R.) - The work was supported by a grant from the National Institutes of Health (R37 CA70867) and the National Cancer Institute intramural research program, including NCI Intramural Research Program contract (N02 CP1101066).
JLCS (K.M, T.K.) - Grants-in-Aid from the Ministry of Health, Labor, and Welfare for Research on Applying Health Technology and for the 3rd-term Comprehensive 10-year Strategy for Cancer Control; by the National Cancer Center Research and Development Fund; by Grant-in-Aid for Scientific Research on Priority Areas and on Innovative Area from the Ministry of Education, Science, Sports, Culture and Technology of Japan. (W.P.) - NCI R01-CA121210.
HKS (J.W.) - General Research Fund of Research Grant Council, Hong Kong (781511M)

Reference List

1. Hsiung,C.A. *et al.* The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS. Genet* **6**, (2010).

2. Hosgood,H.D., III *et al.* Genetic variant in TP63 on locus 3q28 is associated with risk of lung adenocarcinoma among never-smoking females in Asia. *Hum. Genet* [Epub ahead of print] (2012).

3. Wu,C. *et al.* Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. *Cancer Res.* **69**, 5065-5072 (2009).

4. Jou,Y.S. *et al.* Association of an EGFR intron 1 SNP with never-smoking female lung adenocarcinoma patients. *Lung Cancer* **64**, 251-256 (2009).

5. Tang,L. *et al.* Lung cancer in Chinese women: evidence for an interaction between tobacco smoking and exposure to inhalants in the indoor environment. *Environ. Health Perspect.* **118**, 1257-1260 (2010).

6. Kim,J.H. *et al.* Genetic polymorphisms of ataxia telangiectasia mutated affect lung cancer risk. *Hum. Mol. Genet.* **15**, 1181-1186 (2006).

7. Jung,H.Y. *et al.* Association study of TP53 polymorphisms with lung cancer in a Korean population. *J. Hum. Genet.* **53**, 508-514 (2008).

8. Park,J.Y. *et al.* Polymorphisms of the DNA repair gene xeroderma pigmentosum group A and risk of primary lung cancer. *Cancer Epidemiol. Biomarkers Prev.* **11**, 993-997 (2002).

9. Yin,Z. *et al.* ERCC2, ERCC1 polymorphisms and haplotypes, cooking oil fume and lung adenocarcinoma risk in Chinese non-smoking females. *J Exp. Clin. Cancer Res.* **28**, 153 (2009).

10. Zhang,Y. *et al.* Family history of cancer and risk of lung cancer among nonsmoking Chinese women. *Cancer Epidemiol. Biomarkers Prev.* **16**, 2432-2435 (2007).

11. Zheng,W. *et al.* The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am. J. Epidemiol.* **162**, 1123-1131 (2005).

12. Qian,B. *et al.* Association of genetic polymorphisms in DNA repair pathway genes with non-small cell lung cancer risk. *Lung Cancer* **73**, 138-146 (2011).

13. Guo,H. *et al.* Functional promoter -1271G>C variant of HSPB1 predicts lung cancer risk and survival. *J Clin. Oncol.* **28**, 1928-1935 (2010).

14. Ito,H. *et al.* Association between a Genome-Wide Association Study-Identified Locus and the Risk of Lung Cancer in Japanese Population. *J Thorac. Oncol.* **7**, 790-798 (2012).

15. Piao,J.M. *et al.* p53 codon 72 polymorphism and the risk of lung cancer in a Korean population. *Lung Cancer* **73**, 264-267 (2011).

16. Lo Y.L *et al.* Risk factors for primary lung cancer among never smokers by gender in a matched case-control study. *Cancer Causes Control* In press (2012).

Supplementary Table 1. Demographic characteristics of cases and controls from participating study centers in a GWAS of lung cancer among never-smoking Asian females

										Environmental Tobacco Smoke (ETS)			
		Subject				Age [mean(std)]				Ever[‡]		Never	
Study Group[†]	Region	All Cases	Adeno-carcinomas	Squamous Cell	Controls	All Cases	Adeno-carcinomas	Squamous Cell	Controls	Cases	Controls	Cases	Controls
CAMSCH	Mainland China	642	555	32	334	57.0(10.1)	56.8(10.2)	58.6(10.2)	45.9(12.8)	NA[*]			
FLCS	Mainland China	295	212	49	386	59.6(11.1)	59.3(11.1)	61.3(11.7)	60.1(8.4)	NA[*]			
GDS	Mainland China	634	535	7	123	57.7(11.8)	58.3(11.3)	65.3(10.4)	40.3(10.7)	NA[*]			
GEL-S	Singapore	187	120	8	296	63.4(11.9)	63.3(10.9)	53.5(11.7)	64.6(11.7)	96	163	84	128
GELAC	Taiwan	1302	1059	75	1095	59.4(11.5)	59.5(11.5)	58.5(10.6)	58.0(11.9)	962	722	295	351
HKS	Hong Kong	298	226	0	666	60.7(10.9)	61.5(11.1)	NA[*]	51.8(19.4)	NA[*]			
JLCS	Japan	425	407	10	549	59.1(9.3)	59.3(8.9)	58.3(10.0)	59.1(9.4)	NA[*]			
SKLCS	South Korea	546	419	28	1082	61.6(11.0)	61.6(10.7)	63.6(11.2)	51.1(14.7)	NA[*]			
SLCS	Mainland China	574	378	98	1024	56.7(11.7)	56.2(11.5)	58.4(11.1)	52.3(14.8)	487	226	60	37
CNULCS	South Korea	612	498	51	480	62.7(11.0)	62.1(11.1)	64.4(10.4)	58.2(9.8)	NA[*]			
SWHS	Mainland China	114	78	9	200	57.1(8.0)	58.3(7.7)	54.7(8.4)	58.6(8.4)	83	146	18	32
TLCS	Mainland China	339	49	32	237	57.0(9.7)	57.2(10.1)	57.1(7.4)	57.2(9.8)	NA[*]			
WLCS	Mainland China	79	0	14	343	57.8(10.7)	NA[*]	58.6(12.4)	59.7(7.9)	NA[*]			
YLCS	Mainland China	562	179	330	642	54.4(10.8)	51.9(10.2)	56.5(10.7)	54.9(11.2)	534	598	24	40
Total		6609	4715	743	7457	58.8(11.2)	59.1(11.1)	58.3(11.0)	55.1(13.7)	2162	1986	481	689

[†] CAMSCH: Chinese Academy of Medical Sciences Cancer Hospital Study; FLCS: Fudan Lung Cancer Study; GDS: Guangdong Study; GEL-S: Genes and Environment in Lung Cancer, Singapore study; GELAC: Genetic Epidemiological Study of Lung Adenocarcinoma (in Taiwan); HKS: Hong Kong Study; JLCS: Japanese Female Lung Cancer Collaborative Study; SKLCS: South Korea Multi-Center Lung Cancer Study (includes Seoul National University Study; Korean University Medical Center Study; and Kyungpook National University Hospital Study); SLCS: Shenyang Lung Cancer Study; CNULCS: Chonnam National University Lung Cancer Study; SWHS: Shanghai Women's Health Study; TLCS: Tianjin Lung Cancer Study; WLCS: Wuhan Lung Cancer Study; YLCS: Yunnan Lung Cancer Study. [‡] Ever exposed to family members/husband (SWHS) smoking in the home. [*] NA - not applicable.

Supplementary Table 2. Results from a GWAS among never-smoking Asian females for previously published SNPs and lung cancer

SNP	Chromosome position	Plausible candidate gene(s)	Allele[a]	MAF[b]	Subjects (Control/Case)	OR (95% CI)	P_{trend}
rs4488809	3q28	*TP63*[14, 17]	T|C	0.42|0.46	4493|5450	1.19(1.12-1.26)	4.18×10^{-9}
rs2736100	5p15.33	*TERT*, *hTERT*[11, 12, 14, 17-19]	T|G	0.40|0.48	4543|5505	1.38(1.30-1.47)	4.24×10^{-27}
rs401681	5p15.33	*CLPTM1L*[23]	C|T	0.32|0.30	4452|5382	0.90(0.84-0.96)	7.24×10^{-4}
rs451360	5p15.33	*CLPTM1L*[52]	C|T	0.12|0.10	4405|5222	0.84(0.76-0.92)	2.68×10^{-4}
rs6489769	12p13.33	*RAD52*, *BC039168*[25]	T|C	0.47|0.46	4323|673[c]	1.01(0.89-1.14)	0.86
rs753955	13q12.12	*MIPEP*[17]	T|C	0.33|0.34	4485|5454	1.02(0.95-1.08)	0.63
rs1051730	15q25.1	*CHRNA5*, *CHRNA3*[12, 21, 22]	C|T	0.03|0.03	4543|5509	1.10(0.92-1.30)	0.30
rs12914385	15q25.1	*CHRNA5*,*CHRNA3*,*CHRNB4*[12, 51]	C|T	0.28|0.28	4396|5330	0.99(0.93-1.05)	0.70
rs8034191	15q25.1	*IREB2*, *AGPHD1*[12, 19, 21]	T|C	0.03|0.03	4544|5510	1.10(0.92-1.30)	0.29
rs667282	15q25.1	*CHRNA5*[24]	A|G	0.46|0.47	4485|5449	1.04(0.99-1.11)	0.14
rs17728461	22q12.2	*MTMR3-HORMAD2-LIF*[17]	G|A	0.19|0.19	4454|5385	1.00(0.92-1.07)	0.90
rs36600	22q12.2	*MTMR3*[17]	G|T	0.09|0.09	4493|5453	0.96(0.87-1.06)	0.41
rs1663689 (rs7922444)[d]	10p14	Close to *GATA3*[26]	T|C	0.42|0.41	4490|5453	0.96(0.90-1.02)	0.16
rs2895680	5q32	*PPP2R2B-STK32A-DPYSL3*[26]	T|C	0.31|0.31	4378|5300	0.98(0.92-1.05)	0.61
rs4809957	20q13.2	*CYP24A1*[26]	G|A	0.38|0.39	4395|5320	1.01(0.95-1.08)	0.69
rs247008	5q31.1	*IL3-CSF2-P4HA2*[26]	C|T	0.47|0.48	4536|5506	1.02(0.96-1.08)	0.46
rs9439519 (rs1204094)[e]	1p36.32	*AJAP1-NPHP4*[26]	A|C	0.34|0.34	4489|5453	1.02(0.96-1.08)	0.58
rs7216064	17q24.3	*BPTF*[20]	A|G	0.37|0.33	4422|4066[f]	0.86(0.80-0.92)	6.59×10^{-6}
rs3817963	6p21.3	*BTNL2*[20]	A|G	0.26|0.28	4513|4179[f]	1.11(1.04-1.19)	3.56×10^{-3}

[a]Minor allele listed second; [b]Minor allele frequency; [c]Squamous cell cancer cases only[25]; [d]r^2 of surrogate SNP = 0.86; [e]r^2 of surrogate SNP = 0.73; [f]Adenocarcinoma cancer cases only[20].

Supplementary Table 3. SNPs selected for replication that did not achieve genome-wide significance and lung cancer in a GWAS among never-smoking Asian females

SNP	**Plausible candidate gene(s)**	**Chromosome position**	**Subset**	**Allele**[a]	**MAF**[b] **(Control/Case)**	**Subjects (Control/Case)**	**OR (95% CI)**	P_{trend}
rs10197940	*TNFAIP6*/*RIF1*	2q23.3	Scan	T|C	0.49/0.46	4493/5456	0.88(0.83-0.93)	7.46×10^{-6}
			replication	T|C	0.47/0.46	2870/1082	0.94(0.85-1.05)	0.27
			combined	T|C	0.48/0.46	7363/6538	0.89(0.85-0.94)	5.38×10^{-6}
rs11823347	*NELL1*	11p15.1	Scan	C|A	0.16/0.13	4483/5446	0.81(0.75-0.88)	7.45×10^{-7}
			replication	C|A	0.14/0.15	2894/1094	1.02(0.89-1.18)	0.75
			combined	C|A	0.15/0.13	7377/6540	0.86(0.80-0.92)	3.59×10^{-5}
rs3781453	*FAM53B*	10q26.13	Scan	A|G	0.32/0.34	4491/5455	1.13(1.06-1.20)	1.24×10^{-4}
			replication	A|G	0.34/0.32	2900/1091	0.90(0.81-1.01)	0.063
			combined	A|G	0.33/0.34	7391/6546	1.07(1.02-1.13)	9.5×10^{-3}

[a]Minor allele listed second; [b]Minor allele frequency; [c]For the HLA Class II region, because rs2395185 failed the TaqMan design, we chose to genotype by TaqMan rs28366298, its perfect surrogate (r^2=1.0). The reported P value is based on the meta-analysis of rs2395185 in the scan set and rs28366298 in the TaqMan set.

Supplementary Table 4. SNPs selected for replication that did not achieve genome-wide significance and adeno- and squamous carcinoma of the lung in a GWAS among never-smoking Asian females

SNP	Putative gene	Chromo-some position	Allele[a]	MAF[b]			Adenocarcinoma			Squamous carcinoma			$P_{heterogeneity}$[c]
				1	2	3	Subjects (Control/Case)	OR (95% CI)	P_{trend}	Subjects (Control/Case)	OR (95% CI)	P_{trend}	
rs10197940	*TNFAIP6/ RIF1*	2q23.3	T|C	0.48	0.45	0.47	7019/4662	0.88(0.83-0.93)	6.07×10^{-6}	6702/753	0.89(0.79-0.99)	0.039	0.18
rs11823347	*NELL1*	11p15.1	C|A	0.15	0.13	0.14	7033/4666	0.82(0.76-0.89)	3.03×10^{-6}	6712/753	0.92(0.78-1.08)	0.30	0.15
rs3781453	*FAM53B*	10q26.13	A|G	0.33	0.35	0.32	7047/4671	1.14(1.07-1.21)	1.72×10^{-5}	6727/754	1.02(0.90-1.15)	0.78	0.016

[a] Minor allele listed second; [b] Minor allele frequency; 1: MAF in controls, 2: MAF in adenocarcinoma, 3: MAF in squamous carcinoma; [c] Tested by case-case analysis.

Supplementary Table 5. Bioinformatic characterization of SNPs that are highly correlated with rs7086803 within the associated region on 10q25.2

SNPID	position	r^2 with rs7086803	divergence time	Digital Dnase I hypersensitivity cluster	Weak enhancer	Strong enhancer	Enhancer- and promoter-associated histone mark	Transcription factor ChIP-seq	Human EST	Repeating element	UCSC gene transcript
rs75295815	not found in hg18	1.000									
rs11196080	114,482,568	1.000	167.4 Mya	Y	Y	Y	Y	Y			
rs2419812	114,482,825	1.000	94.4 Mya		Y		Y		Y		
rs1885281	114,482,888	1.000	94.4 Mya		Y		Y		Y		
rs1885282	114,483,087	1.000	29.2 Mya		Y		Y			Y	
rs7920475	114,483,870	1.000	98.8 Mya							Y	
rs10885380	114,484,286	1.000	94.4 Mya							Y	
rs41292624	114,486,707	1.000	98.8 Mya							Y	Y
rs7086803	114,488,466	-	98.8 Mya	Y		Y					
rs7916100	114,490,108	1.000	29.2 Mya			Y					Y
rs7916159	114,490,367	1.000	94.4 Mya			Y					Y
rs6585182	114,490,392	1.000	98.8 Mya			Y					Y
rs7096588	114,490,841	1.000	98.8 Mya		Y	Y	Y				Y
rs7075510	114,491,499	1.000	167.4 Mya		Y	Y	Y				
rs7075671	114,491,675	1.000	167.4 Mya	Y	Y	Y	Y	Y			
rs12217440	114,492,208	1.000	98.8 Mya		Y	Y	Y				
rs7094841	114,492,401	1.000	167.4 Mya		Y	Y	Y				
rs7085406	114,493,486	1.000	none		Y	Y	Y	Y			
rs7086477	114,494,395	1.000	29.2 Mya			Y					
rs7090269	114,494,523	1.000	none			Y					
rs7090171	114,494,587	1.000	29.2 Mya			Y					
rs11196086	114,495,088	1.000	92.4 Mya		Y						
rs11196087	114,496,396	1.000	none						Y		
rs11196088	114,497,723	1.000	94.4 Mya		Y				Y		
rs11196089	114,499,280	0.986	94.4 Mya		Y				Y		

Supplementary Table 6a. QC exclusion threshold

QC group	Exclude threshold			
	Sample heterozygosity	Sample missing rate	Locus missing rate	Locus HWE
CAMSCH _660W	< 0.26 or > 0.30	>0.06	>0.1	< 0.0000001
FLCS _660W	< 0.25 or > 0.30	>0.04	>0.1	< 0.0000001
GELAC _370K	< 0.27 or > 0.31	>0.05	>0.1	< 0.0000001
GELAC _610K	< 0.25 or > 0.29	>0.03	>0.1	< 0.0000001
GDS_660W	< 0.27 or > 0.30	>0.04	>0.1	< 0.0000001
HERPACC _610K†	< 0.24 or > 0.29	>0.01	>0.1	< 0.0000001
HERPACC _660W†	< 0.25 or > 0.29	>0.04	>0.1	< 0.0000001
HKS_660W	< 0.27 or > 0.30	>0.05	>0.1	< 0.0000001
Korea-Univ _660W‡	< 0.27 or > 0.30	>0.05	>0.1	< 0.0000001
Kyungpook-Univ_660W‡	< 0.27 or > 0.30	>0.05	>0.1	< 0.0000001
NCC _660W†	< 0.25 or > 0.30	>0.05	>0.1	< 0.0000001
SLCS _660W	< 0.25 or > 0.30	>0.05	>0.1	< 0.0000001
GEL-S _660W	< 0.25 or > 0.31	>0.03	>0.1	< 0.0000001
SNU _660W‡#	< 0.26 or > 0.30	>0.05	>0.1	< 0.0000001
SWHS _660W	< 0.26 or > 0.30	>0.04	>0.1	< 0.0000001
GELAC _660W	< 0.25 or > 0.30	>0.05	>0.1	< 0.0000001
TLCS _660W	< 0.27 or > 0.30	>0.03	>0.1	< 0.0000001
WLCS _660W	< 0.27 or > 0.30	>0.01	>0.1	< 0.0000001
YLCS _660W	< 0.25 or > 0.30	>0.06	>0.1	< 0.0000001

†JLCS ‡SKLCS #CNULCS included with SNU

Supplementary Table 6b. Summary of excluded samples

QC group	Number of samples excluded			
	Sample heterozygosity	Sample missing rate	Discordant expected dup	Total
CAMSCH _660W	6	18		22*
FLCS _660W	1	27		27*
GELAC _370K	0	0		0
GELAC _610K	2	7		7*
GDS_660W	0	4		4
HERPACC _610K†	0	0		0
HERPACC _660W†	18	15		27*
HKS_660W	12	161		164*
Korea-Univ _660W‡	1	16		16*
Kyungpook-Univ_660W‡	0	96		96
NCC _660W†	2	24		24*
SLCS _660W	3	7	2	10*
GEL-S _660W	3	4		7
SNU _660W‡#	36	152		153*
SWHS _660W	5	28		28*
GELAC _660W	15	39		42*
TLCS _660W	1	4	2	7
WLCS _660W	0	1		1
YLCS _660W	11	122	2	126*
Total	116	725	6	761

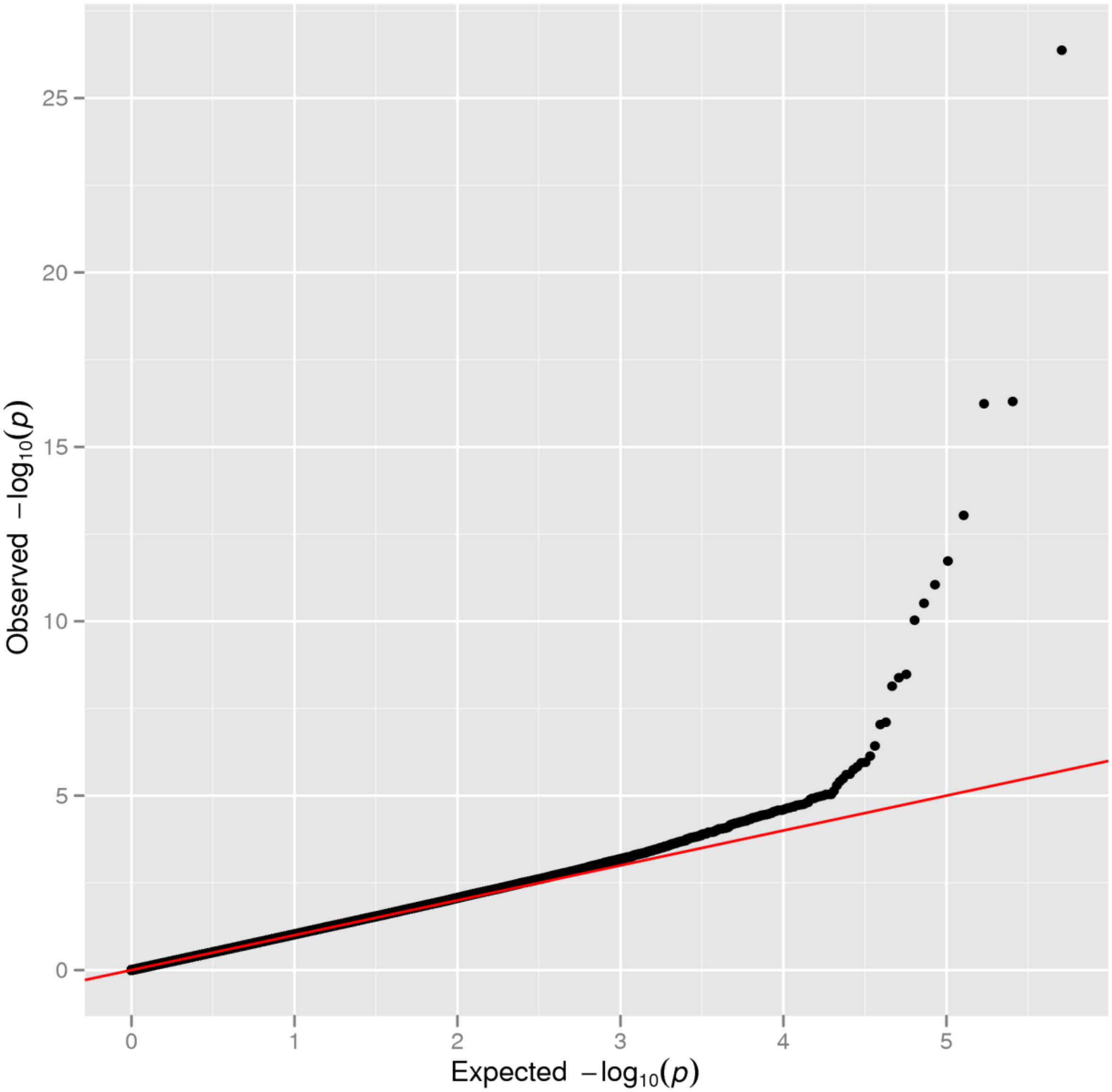
†JLCS ‡SKLCS #CNULCS included with SNU *Total number of unique samples

Supplementary Table 6c. Summary of excluded loci

QC group	Number of excluded Loci	Number of loci failed HWE	Number of remaining Loci
CAMSCH _660W	97054	740	559570
FLCS _660W	97395	458	559511
GELAC _370K	24810	1539	344049
GELAC _610K	29942	891	590068
GDS_660W	96739	456	560169
HERPACC _610K†	30074	410	590417
HERPACC _660W†	100225	132	557007
HKS_660W	111320	167	545877
Korea-Univ _660W‡	129367	15	527982
Kyungpook-Univ_660W‡	239349	31	417984
NCC _660W†	99154	168	558042
SLCS _660W	96785	623	559956
GEL-S _660W	96590	364	560410
SNU _660W‡#	111642	780	544942
SWHS _660W	123234	33	534097
GELAC _660W	101374	648	555342
TLCS _660W	96839	317	560208
WLCS _660W	96453	2	560909
YLCS _660W	110472	790	546102

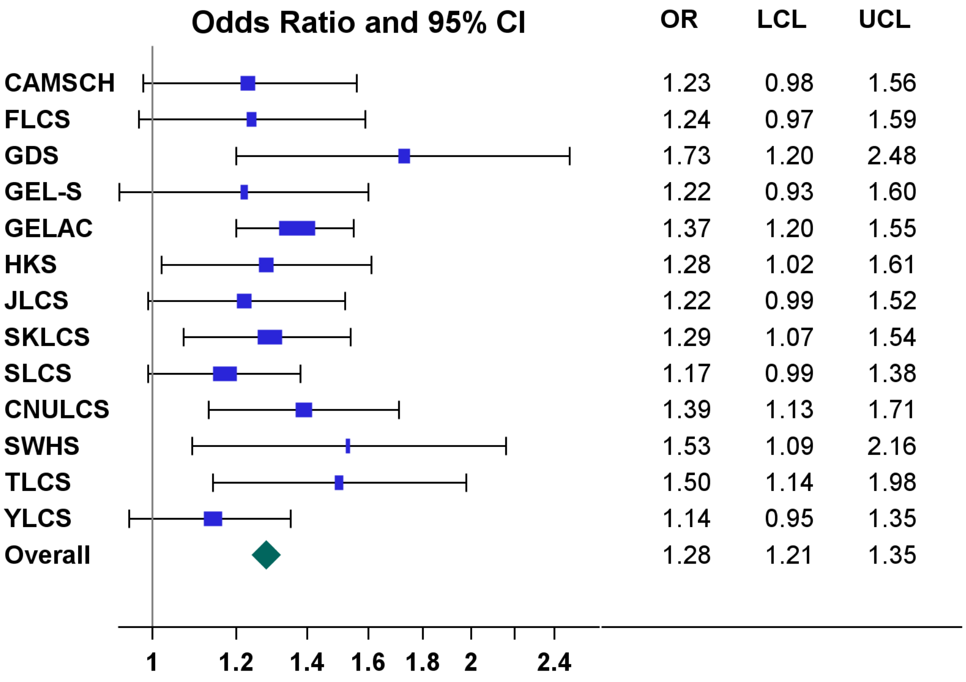
†JLCS ‡SKLCS #CNULCS included with SNU

Supplementary Figure 1. Quantile-Quantile (Q-Q) plot of observed versus expected P values for the GWAS of lung cancer among never-smoking Asian females. Observed P values are compared with the expected uniform distribution.



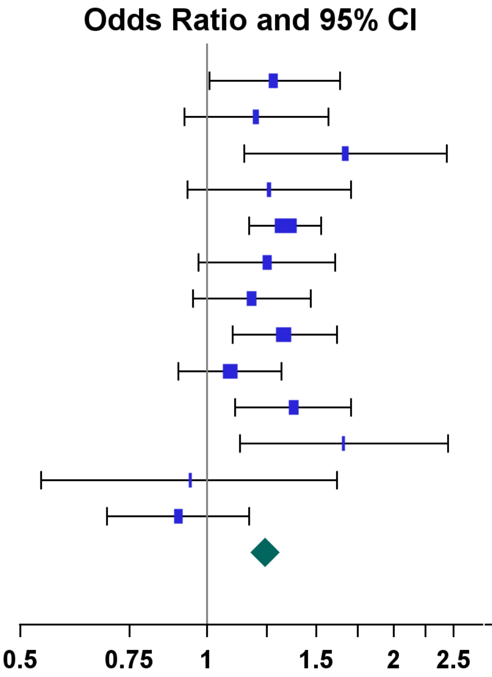
Supplementary Figure 2. Forest plots by study for three new loci. Each of the three new regions are depicted for association with all lung cancer and adenocarcioma: (a,b) *VTI1A*, (c,d) *ROS1/DCBLD1*, (e,f) HLA Class II region. Squares represent the estimated odds ratio (OR) for each study and are proportional to the weight contributed by each study to the summary OR estimates. Lines indicate the 95% confidence interval (CI). Diamonds represent the summary OR estimates and confidence intervals. Study names corresponding to the acronyms shown in the plots and statistical methods are detailed in the Supplementary Note and Supplementary Table 1. Due to limited numbers, WLCS was not included in these plots. Heterogeneity tests for all lung cancer and adenocarcinoma only, respectively: *VTI1A*, P = 0.82, 0.26; *ROS1/DCBLD1*, P = 0.020, 0.99; HLA Class II region, P = 0.062, 0.36. Forest plots are generated using SAS 9.2.

Supplementary Figure 2a. Forest plot for *VTI1A* and all lung cancer



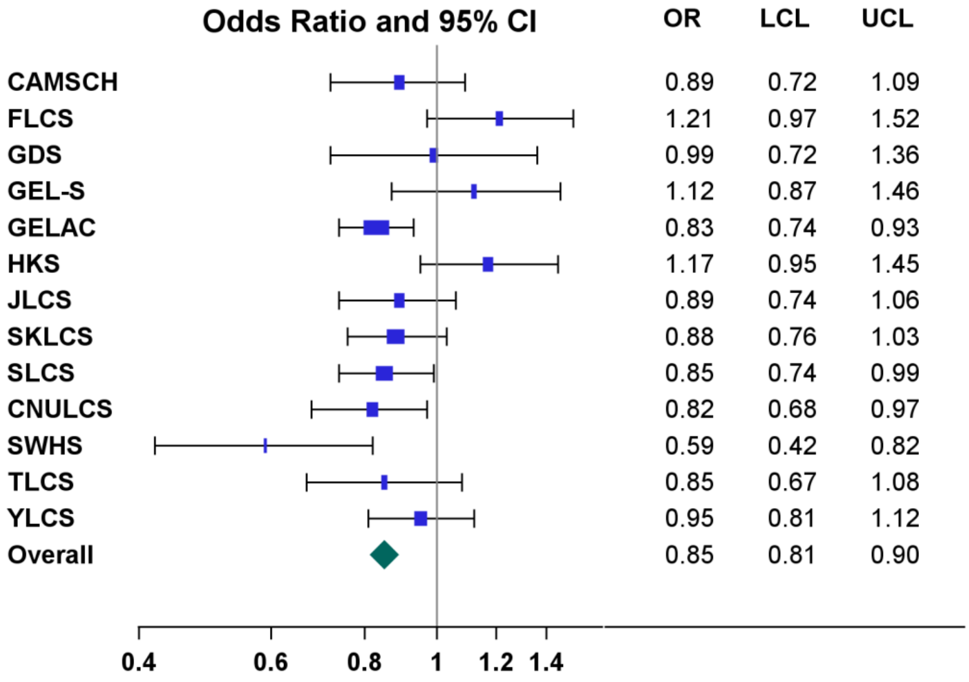
	OR	LCL	UCL
CAMSCH	1.23	0.98	1.56
FLCS	1.24	0.97	1.59
GDS	1.73	1.20	2.48
GEL-S	1.22	0.93	1.60
GELAC	1.37	1.20	1.55
HKS	1.28	1.02	1.61
JLCS	1.22	0.99	1.52
SKLCS	1.29	1.07	1.54
SLCS	1.17	0.99	1.38
CNULCS	1.39	1.13	1.71
SWHS	1.53	1.09	2.16
TLCS	1.50	1.14	1.98
YLCS	1.14	0.95	1.35
Overall	1.28	1.21	1.35

Supplementary Figure 2b. Forest plot for *VTI1A* and adenocarcinoma of the lung



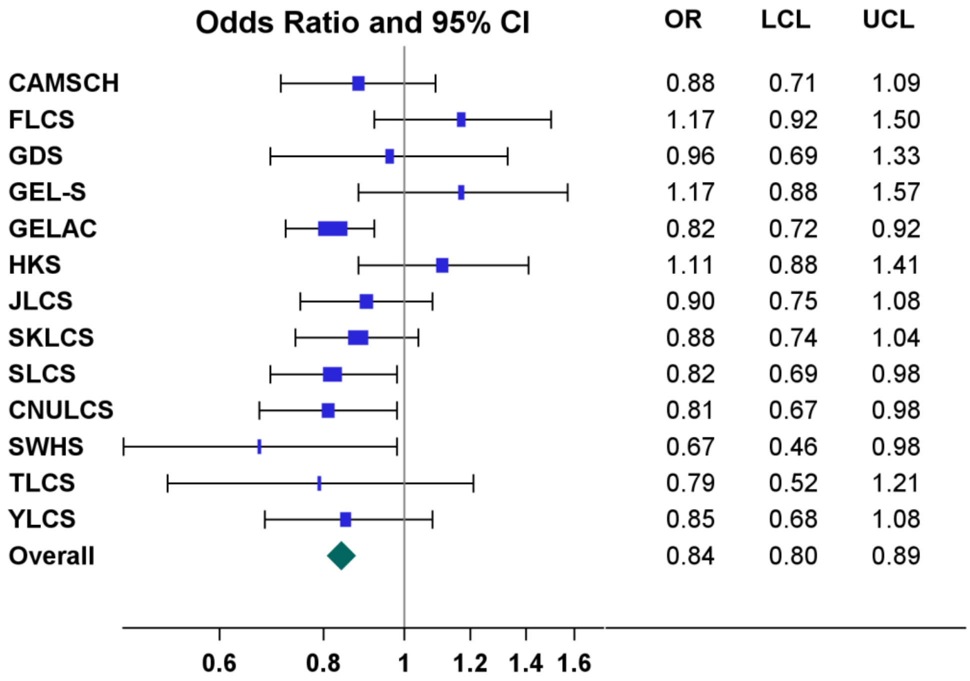
	OR	LCL	UCL
CAMSCH	1.28	1.01	1.64
FLCS	1.20	0.92	1.57
GDS	1.67	1.15	2.44
GEL-S	1.26	0.93	1.71
GELAC	1.34	1.17	1.53
HKS	1.25	0.97	1.61
JLCS	1.18	0.95	1.47
SKLCS	1.33	1.10	1.62
SLCS	1.09	0.90	1.32
CNULCS	1.38	1.11	1.71
SWHS	1.66	1.13	2.45
TLCS	0.94	0.54	1.62
YLCS	0.90	0.69	1.17
Overall	1.24	1.17	1.32

Supplementary Figure 2c. Forest plot for *ROS1/DCBLD1* and all lung cancer

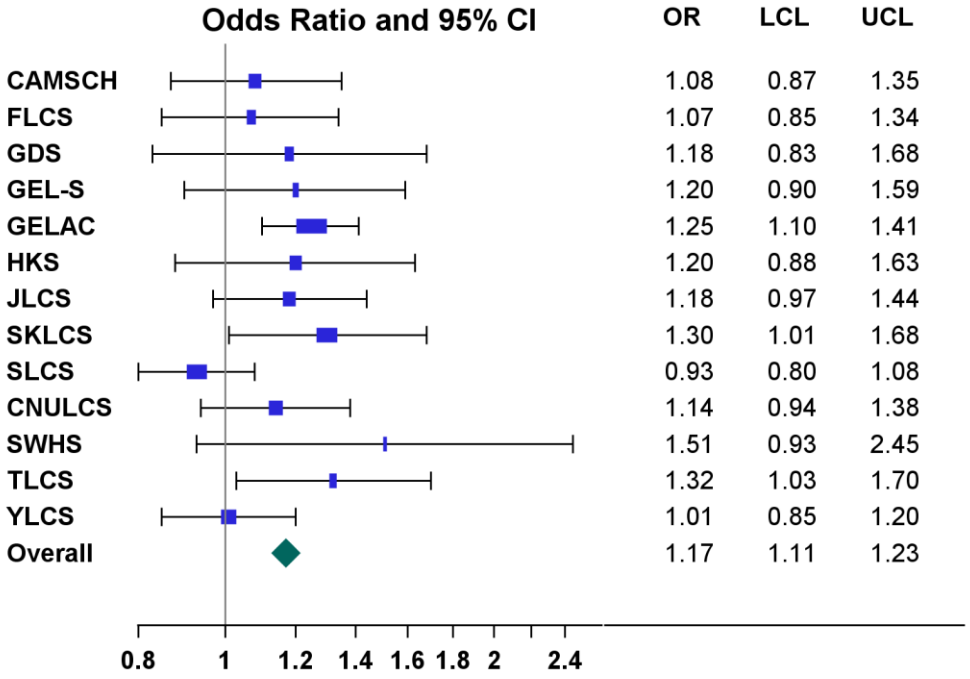


	OR	LCL	UCL
CAMSCH	0.89	0.72	1.09
FLCS	1.21	0.97	1.52
GDS	0.99	0.72	1.36
GEL-S	1.12	0.87	1.46
GELAC	0.83	0.74	0.93
HKS	1.17	0.95	1.45
JLCS	0.89	0.74	1.06
SKLCS	0.88	0.76	1.03
SLCS	0.85	0.74	0.99
CNULCS	0.82	0.68	0.97
SWHS	0.59	0.42	0.82
TLCS	0.85	0.67	1.08
YLCS	0.95	0.81	1.12
Overall	0.85	0.81	0.90

Supplementary Figure 2d. Forest plot for *ROS1/DCBLD1* and adenocarcinoma of the lung

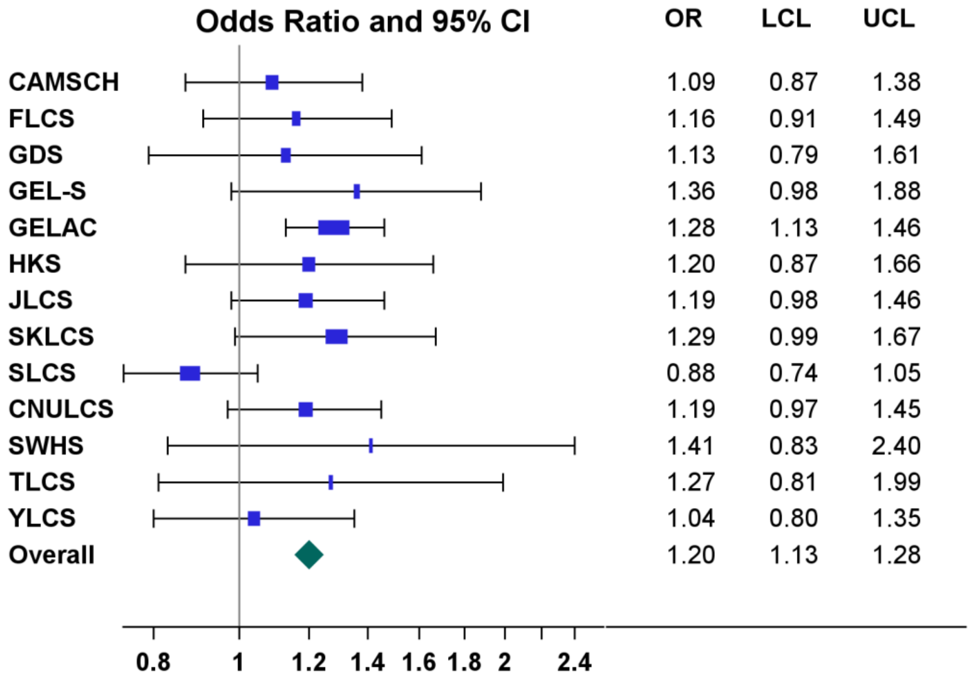


Supplementary Figure 2e. Forest plot for HLA Class II region and all lung cancer



	OR	LCL	UCL
CAMSCH	1.08	0.87	1.35
FLCS	1.07	0.85	1.34
GDS	1.18	0.83	1.68
GEL-S	1.20	0.90	1.59
GELAC	1.25	1.10	1.41
HKS	1.20	0.88	1.63
JLCS	1.18	0.97	1.44
SKLCS	1.30	1.01	1.68
SLCS	0.93	0.80	1.08
CNULCS	1.14	0.94	1.38
SWHS	1.51	0.93	2.45
TLCS	1.32	1.03	1.70
YLCS	1.01	0.85	1.20
Overall	1.17	1.11	1.23

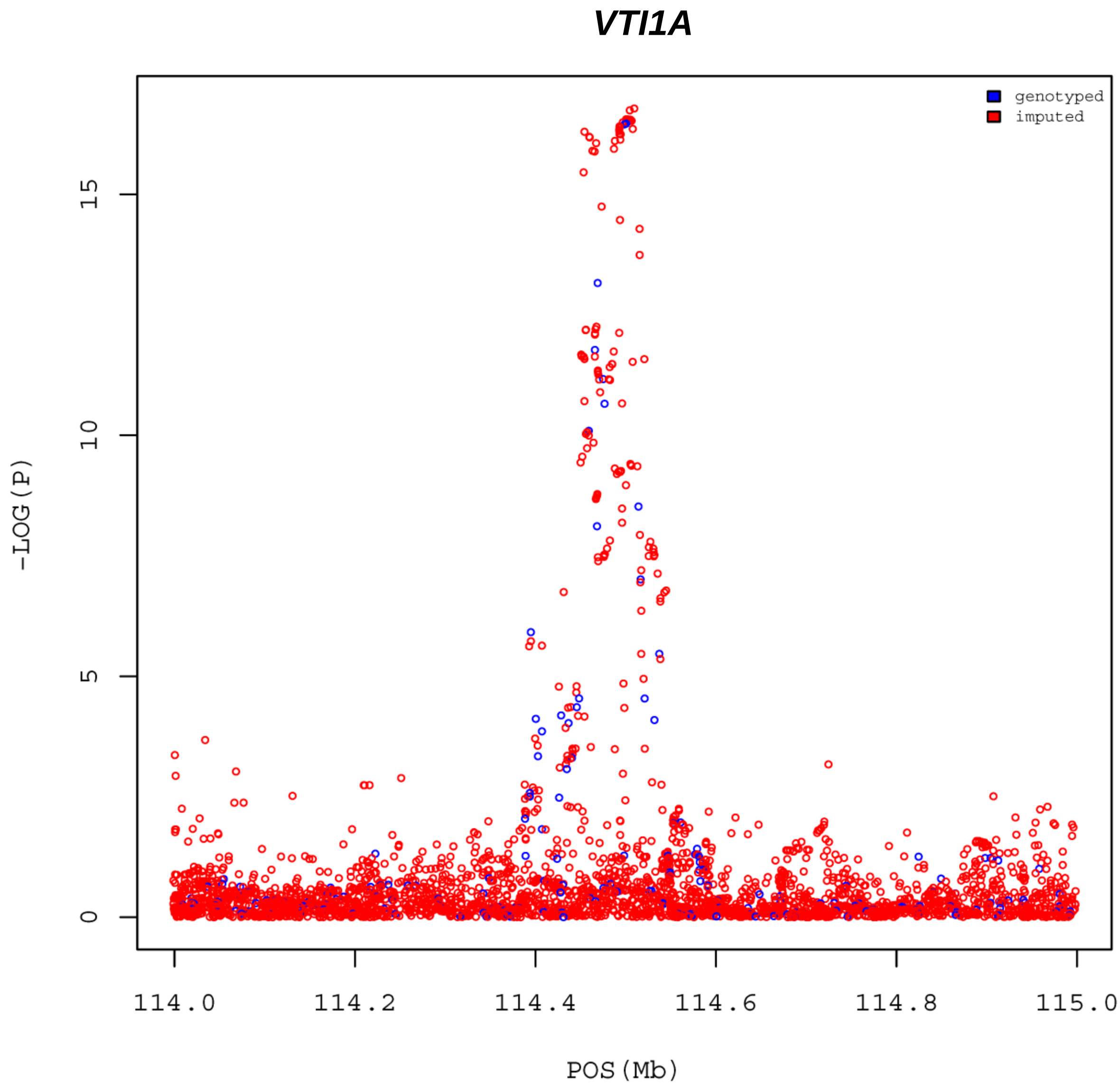
Supplementary Figure 2f. Forest plot for HLA Class II region and adenocarcinoma of the lung



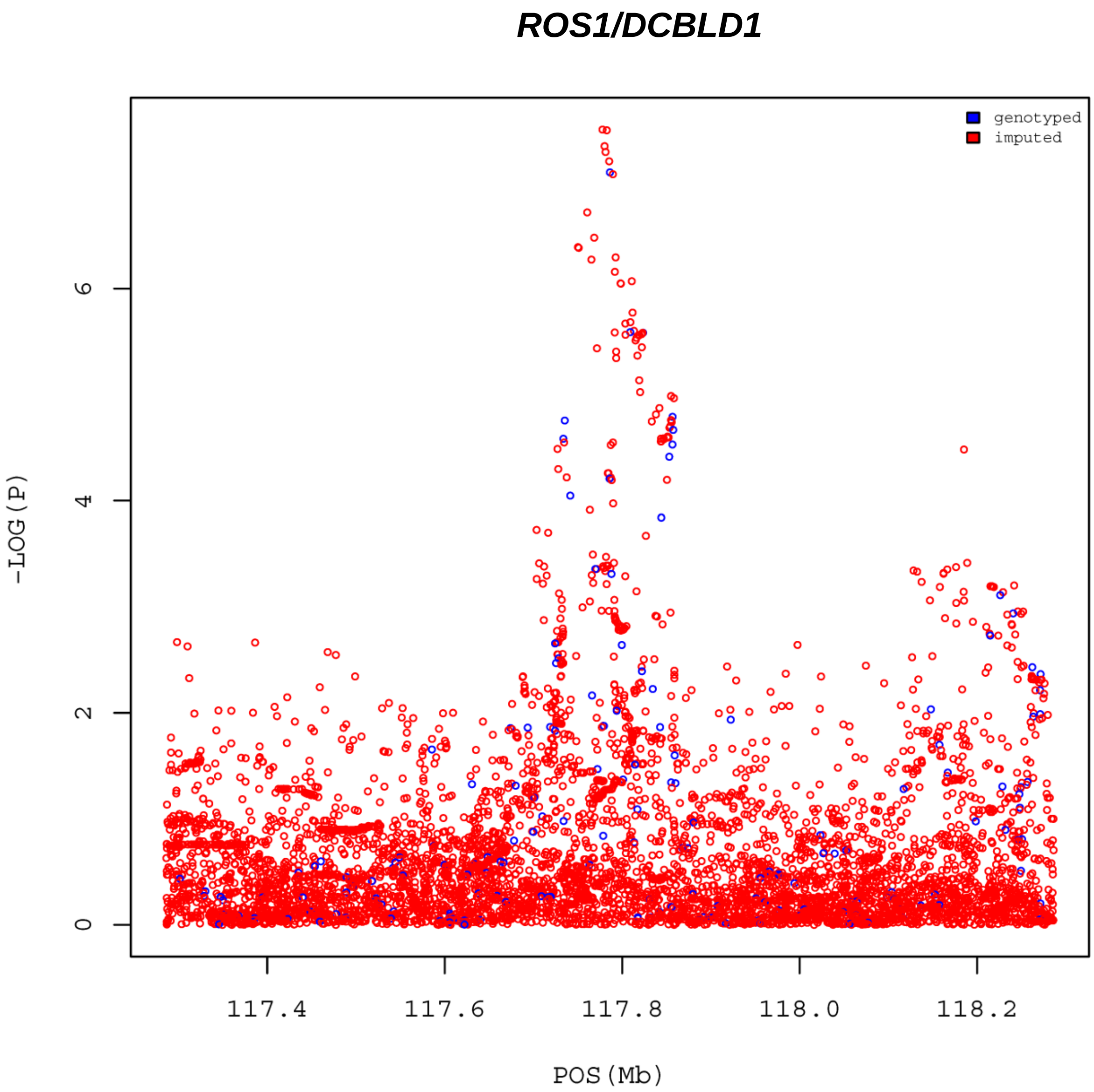
	OR	LCL	UCL
CAMSCH	1.09	0.87	1.38
FLCS	1.16	0.91	1.49
GDS	1.13	0.79	1.61
GEL-S	1.36	0.98	1.88
GELAC	1.28	1.13	1.46
HKS	1.20	0.87	1.66
JLCS	1.19	0.98	1.46
SKLCS	1.29	0.99	1.67
SLCS	0.88	0.74	1.05
CNULCS	1.19	0.97	1.45
SWHS	1.41	0.83	2.40
TLCS	1.27	0.81	1.99
YLCS	1.04	0.80	1.35
Overall	1.20	1.13	1.28

Supplementary Figure 3a, b, c. Imputation analysis for new regions. The $-log_{10}(P)$ (Y axis) for both genotyped SNPs (blue) and imputed SNPs (red) was plotted on the genomic coordinates (X axis; NCBI genome build 37). The imputed signals in each of three regions showed no substantial improvement over the genotyped signals. (a) 10q25.2 region (chr10: 113,998,476-114,998,476); (b) 6q22.2 region (chr6: 117,286,180-118,286,180); and (c) 6p21.32 region (chr6: 31,933,167-32,933,167). Note that the imputation for the HLA locus may not be reliable mainly because it has extended LD as well as being enriched with polymorphisms. HLA-typing will be necessary to unravel the haplotypes involved in this particular association.

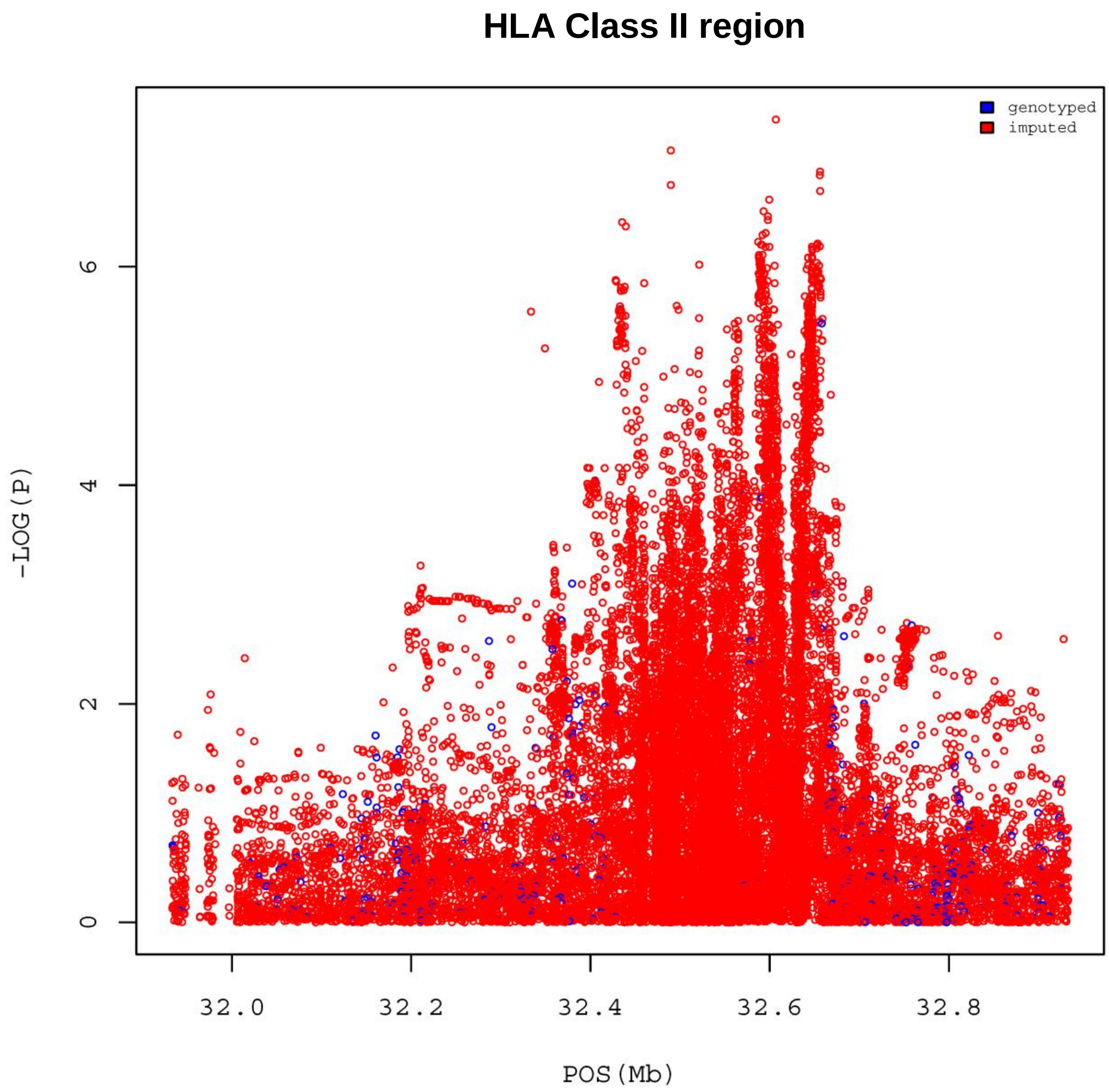
Supplementary Figure 3a. Imputation analysis for new regions



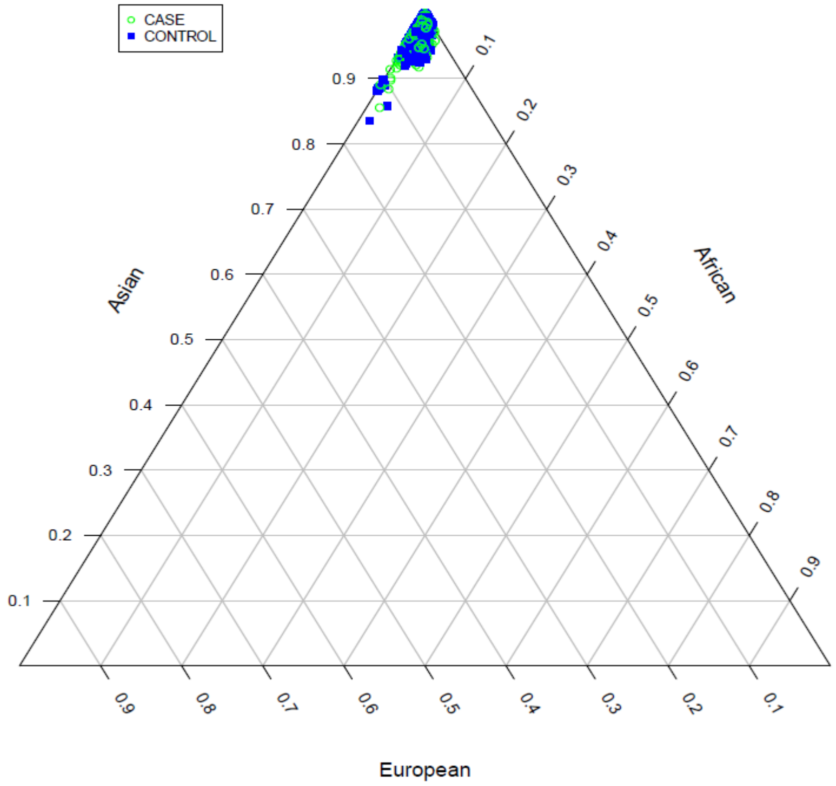
Supplementary Figure 3b. Imputation analysis for new regions



Supplementary Figure 3c. Imputation analysis for new regions

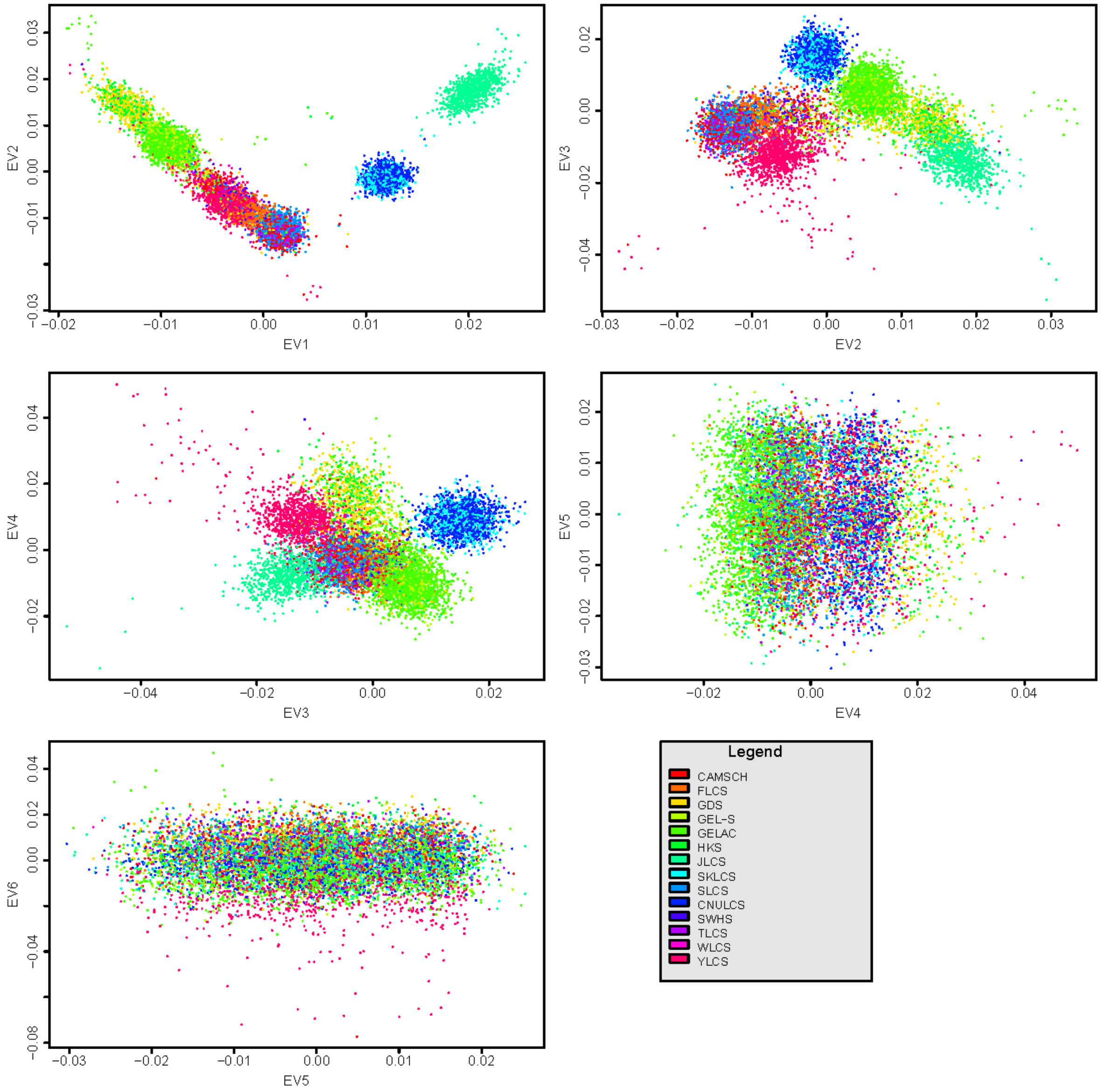


Supplementary Figure 4. Plot of estimated admixture coefficients for individuals included in the GWAS scan. For details, see the **Online Methods**. Three individuals with < 86% Asian ethnicity were removed from further analyses.



Supplementary Figure 5a, b. Plot of top eigenvectors based on the principal components analysis. See details in the **Online Methods**. Three Asian populations (Han Chinese, Korean and Japanese) formed separate clusters based on the first two principal components (PCs). Analysis was adjusted for three eigenvectors (EV1, EV2 and EV4) in the all lung cancer cases analysis.

Supplementary Figure 5a. Plot of top eigenvectors based on the principal components analysis



Supplementary Figure 5b. Plot of top eigenvectors based on the principal components analysis