Human Molecular Genetics, 2012, Vol. 21, No. 22 ***4980–4995***
doi:10.1093/hmg/dds334
Advance Access published on August 16, 2012

Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls

Maria N. Timofeeva[1], Rayjean J. Hung[2], Thorunn Rafnar[3], David C. Christiani[4], John K. Field[5], Heike Bickeböller[6], Angela Risch[7,41], James D. McKay[1], Yufei Wang[8], Juncheng Dai[9], Valerie Gaborieau[1], John McLaughlin[2], Darren Brenner[2], Steven A. Narod[10], Neil E. Caporaso[11], Demetrius Albanes[11], Michael Thun[12], Timothy Eisen[13], H.-Erich Wichmann[14,15], Albert Rosenberger[6], Younghun Han[16], Wei Chen[16], Dakai Zhu[16], Margaret Spitz[18], Xifeng Wu[17], Mala Pande[17], Yang Zhao[4], David Zaridze[19], Neonilia Szeszenia-Dabrowska[20], Jolanta Lissowska[21], Peter Rudnai[22], Eleonora Fabianova[23], Dana Mates[24], Vladimir Bencko[25], Lenka Foretova[26], Vladimir Janout[27], Hans E. Krokan[28], Maiken Elvestad Gabrielsen[28], Frank Skorpen[29], Lars Vatten[30], Inger Njølstad[31], Chu Chen[32], Gary Goodman[32], Mark Lathrop[33], Simone Benhamou[34], Tõnu Vooder[35], Kristjan Välk[36], Mari Nelis[37,38], Andres Metspalu[37], Olaide Raji[5], Ying Chen[5], John Gosney[5], Triantafillos Liloglou[5], Thomas Muley[39,41], Hendrik Dienemann[40,41], Gudmar Thorleifsson[3], Hongbing Shen[9], Kari Stefansson[3], Paul Brennan[1], Christopher I. Amos[16,42,*,†], Richard Houlston[8,†], Maria Teresa Landi[11,†] and for TRICL Research Team[‡]

[1]International Agency for Research on Cancer, Lyon 69008, France, [2]Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, ON, Canada M5G 1X5, [3]DeCODE Genetics, Reykjavik, Iceland, [4]Harvard University School of Public Health, Boston, MA 02115, USA, [5]Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, The University of Liverpool Cancer Research Centre, Institute of Translational Medicine, The University of Liverpool, Liverpool, UK, [6]Department of Genetic Epidemiology, University Medical Center, Georg-August-University Göttingen, 37073 Göttingen, Germany, [7]DKFZ - German Cancer Research Center, Translational Lung Research Centre Heidelberg (TLRC-H), Member of the German Center for Lung Research, 69120 Heidelberg, Germany, [8]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK, [9]Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China, [10]Women's College Research Institute, Toronto, Ontario, Canada, [11]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, U.S. Public Health Service, Bethesda, MD 20892, USA, [12]American Cancer Society, Epidemiology and Surveillance Research, Atlanta, GA 30301, USA, [13]Department of Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK, [14]Helmholtz Center Munich, Institute of Epidemiology I, 85764 Neuherberg, Germany, [15]Chair of Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität Munich, 81377 Munich, Germany, [16]Department of Genetics and [17]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA, [18]Baylor College of Medicine, Houston, TX 77030, USA, [19]Institute of Carcinogenesis,

[*]To whom correspondence should be addressed at: Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, 46 Centerra Parkway Suite 330, Lebanon, NH 03766. Tel: +603-653-6696; Email: Christopher.I.Amos@Dartmouth.edu
[†]These authors contributed equally to this work.
[‡]Transdisciplinary Research in Cancer of the Lung.

Russian N.N. Blokhin Cancer Research Centre, 115478 Moscow, Russia, [20]Department of Epidemiology, Institute of Occupational Medicine, 91348 Lodz, Poland, [21]The M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw 02781, Poland, [22]National Institute of Environmental Health, Budapest 1097, Hungary, [23]Regional Authority of Public Health, Banská Bystrica 97556, Slovak Republic, [24]National Institute of Public Health, Bucharest 050463, Romania, [25]1st Faculty of Medicine, Institute of Hygiene and Epidemiology, Charles University in Prague, 128 00 Prague 2, Czech Republic, [26]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno 65653, Czech Republic, [27]Palacky University, Olomouc 77515, Czech Republic, [28]Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim 7489, Norway, [29]Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine and [30]Department of Public Health and General Practice, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim 7489, Norway, [31]Department of Community Medicine, University of Tromso, Tromso 9037, Norway, [32]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, [33]Centre d'Etude du Polymorphisme Humain (CEPH), Paris 75010, France, [34]INSERM U946, Paris 75010, France, [35]Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia, [36]Competence Centre on Reproductive Medicine and Biology, 50410 Tartu, Estonia, [37]Estonian Genome Center, Institute of Molecular and Cell Biology, Tartu 51010, Estonia, [38]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland, [39]Translational Research Unit and [40]Department of Thoracic Surgery, Thoraxklinik am Universitätsklinikum Heidelberg, Germany, [41]Translational Lung Research Centre Heidelberg (TLRC-H), Member of the German Center for Lung Research, 69126 Heidelberg, Germany and [42]Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH 03750, USA

Received May 4, 2012; Revised July 23, 2012; Accepted August 1, 2012

Recent genome-wide association studies (GWASs) have identified common genetic variants at 5p15.33, 6p21–6p22 and 15q25.1 associated with lung cancer risk. Several other genetic regions including variants of *CHEK2* (22q12), *TP53BP1* (15q15) and *RAD52* (12p13) have been demonstrated to influence lung cancer risk in candidate- or pathway-based analyses. To identify novel risk variants for lung cancer, we performed a meta-analysis of 16 GWASs, totaling 14 900 cases and 29 485 controls of European descent. Our data provided increased support for previously identified risk loci at 5p15 ($P = 7.2 \times 10^{-16}$), 6p21 ($P = 2.3 \times 10^{-14}$) and 15q25 ($P = 2.2 \times 10^{-63}$). Furthermore, we demonstrated histology-specific effects for 5p15, 6p21 and 12p13 loci but not for the 15q25 region. Subgroup analysis also identified a novel disease locus for squamous cell carcinoma at 9p21 (*CDKN2A/p16*^{INK4A}*/p14*^{ARF}*/CDKN2B/p15*^{INK4B}*/ANRIL*; rs1333040, $P = 3.0 \times 10^{-7}$) which was replicated in a series of 5415 Han Chinese ($P = 0.03$; combined analysis, $P = 2.3 \times 10^{-8}$). This large analysis provides additional evidence for the role of inherited genetic susceptibility to lung cancer and insight into biological differences in the development of the different histological types of lung cancer.

INTRODUCTION

Lung cancer is a major cause of cancer death worldwide accounting for over 1 million deaths each year (1). The major lung cancer histologic subtypes (small cell and non-small cell) have different clinicopathological characteristics reflective of differences in carcinogenesis (2).

While lung cancer is largely caused by tobacco smoking, there is increasing evidence for the role of inherited genetic factors in disease etiology (3). Notably, genome-wide association studies (GWASs) of lung cancer have robustly demonstrated that polymorphic variation at 5p15.33 (*TERT/CLPTM1L*), 6p21.33 (*BAT3/MSH5*) and 15q25.1 (*CHRNA5/CHRNA3/CHRNB4*) influences lung cancer risk in European populations (4–9). Additionally, single-nucleotide polymorphisms (SNPs) at 22q12 (10,11) and the 15q15.2 locus containing the TP53-binding protein 1 gene have been associated with lung cancer risk (12–14). Three additional susceptibility regions at 13q12.12, 22q12.2 (15) and 3q28 (16) have been identified in GWASs on Asian populations, but to date these regions have not been implicated in lung cancer risk in individuals of European ancestry.

Given the biological differences across lung cancer subtypes, histology- and smoking-specific associations have been conducted. Analyses have shown that SNP rs2736100 (*TERT*) is primarily associated with adenocarcinoma risk (17,18) and variation in *RAD52* at 12p13 with squamous cell carcinoma risk (19). Significant heterogeneity by smoking status and age of onset has been shown for SNPs at the 15q25 locus harboring nicotinic acetylcholine receptor genes (17).

The statistical power of individual GWAS has been limited by the modest effect sizes of genetic variants, the need to

establish stringent thresholds of statistical significance and financial constraints on the numbers of variants that could be studied. Additionally, due to sample size limitations, few comprehensive histology and smoking history subgroup analyses have been performed in individual GWAS. Meta-analysis of existing GWAS data therefore offers the opportunity to discover additional disease loci harboring common variants associated with lung cancer risk and explore the variability in genetic effects according to disease heterogeneity.

In this study, we conducted a pooled analysis of data from 16 GWASs of lung cancer providing data on 14 900 cases and 29 485 controls of European ancestry. We studied associations by histology, sex, smoking status, age of onset, stage and family history of lung cancer and explored the individual contribution of SNPs in previously identified risk loci. To explore how these genetic findings translate into non-European populations, we evaluated selected SNPs in a Han Chinese study of 2338 lung cancer cases and 3077 controls.

RESULTS

A total of 14 900 lung cancer cases and 29 485 controls of European descent from 16 previously reported lung cancer GWASs undertaken by nine analytical centers were included in the meta-analysis (Table 1, Supplementary Material, Table S1 and Fig. S1). The meta-analysis was primarily based on pooling GWAS summary results from 318 094 SNPs featured on Illumina HumanHap 300 BeadChips arrays. For studies genotyped on HumanHap550 or 610Quad Illumina platforms, an additional 217 914 SNPs were available to inform our analysis.

Some degree of genomic over-dispersion (genomic inflation) is expected under a polygenic model even in the absence of population stratification and other technical artifacts (20), and the meta-analysis showed modest evidence of over-dispersion ($\lambda = 1.10$) for the core 318 094 SNPs typed on Illumina HumanHap 300 BeadChips platform (Fig. 1, Supplementary Material, Fig. S2). Adjustment for a genomic inflation factor of 1.10 in this meta-analysis conservatively reduces the power to detect an association. The λ normalized to 1000 cases and 1000 controls was only 1.005, when the approach proposed by Freedman *et al.* (21) was applied.

SNPs mapping to the previously identified risk loci at 5p15, 6p21 and 15q25 provided the best evidence for an association with lung cancer (Supplementary Material, Tables S2 and S3 and Fig. 1). The strongest association was found for rs1051730, which maps to exon 5 of *CHRNA3* at 15q25 ($P = 2.2 \times 10^{-63}$; Fig. 2G), and rs8034191 ($r^2 = 0.93$, $D' = 1$ between the two SNPs, $P = 9.5 \times 10^{-59}$), which is located in the second intron of the *AGPHD1* gene. Consistent with previous observations (17,18), the rs1051730 association was significant in smokers ($P = 1.8 \times 10^{-59}$) and not in never smokers ($P = 0.06$, Fig. 2, Supplementary Material, Fig. S3). The association also appeared slightly stronger in females than in males [respective odds ratios (ORs) 1.42 and 1.29, $P_{het} = 0.02$] and for late-stage rather than early-stage disease (respective ORs 1.39 and 1.28, $P_{het} = 0.06$).

Thirty-one additional SNPs localizing to 15q25.1 with a varied level of linkage disequilibrium (LD) with rs1051730 (Supplementary Material, Table S4) showed a genome-wide significant association (Supplementary Material, Table S2). The third strongest evidence for association was observed for rs6495309 ($P = 1.1 \times 10^{-32}$), which maps the 3′ downstream of *CHRNB4* and shows weaker correlation with rs1051730 ($r^2 = 0.15$; $D' = 1.00$; Fig. 2F). After adjusting for rs1051730, the effect estimate for rs6495309 was greatly attenuated ($P = 4.0 \times 10^{-5}$), while rs1051730 remained significant ($P = 2.4 \times 10^{-26}$; Fig. 3) when allele dosage for rs6495309 was included into a model. Two intronic variants of *CHRNA5* (rs680244, effect allele T, OR = 0.90, $P = 7.2 \times 10^{-10}$, and rs6495306, effect allele G, OR = 0.91, $P = 1.8 \times 10^{-9}$) changed the direction of effect when controlling for rs1051730 (OR = 1.14 $P = 1.4 \times 10^{-8}$ and OR = 1.13, $P = 4.1 \times 10^{-8}$ for rs680244, effect allele T, and rs6495306, effect allele G, respectively, after controlling for allelic dosage). Conversely, an analysis adjusting for rs6495309 enhanced their effects consistently across studies ($P = 9.2 \times 10^{-31}$ and 5.1×10^{-32} for rs680244 and rs6495306, respectively). No other 15q25.1 variant showed a significant association when allelic dosages for rs1051730 and rs6495309 were included into the statistical model.

After imputation, the most significant association in the meta-analysis of GWAS data from individuals of European ancestry was shown by rs951266 ($P = 2.8 \times 10^{-62}$, Supplementary Material, Fig. S4), which maps to intron 2 of *CHRNA5* and is in LD with both rs1051730 and rs16969968. Rs951266 also showed evidence for an association in the Han Chinese population (MAF = 0.04, $P < 0.01$). Several other rare imputed variants that do not directly correlate with the 15q25 variants identified in GWAS of European descendants ($r^2 < 0.05$ and $D' = 1$) showed association with lung cancer risk in the Han Chinese population (Supplementary Material, Fig. S6). These variants map within or in close proximity to *IREB2*.

As previously reported (6), two independent susceptibility variants, rs2736100 and rs401681, which annotate to *TERT* and *CLPTM1L*, were identified in the 5p15.33 region (Fig. 2A and B). Also consistent with previous findings (17,18), the risk associated with rs2736100 was largely confined to adenocarcinoma ($P = 1.7 \times 10^{-19}$). In contrast, rs401681 influenced the risk of all lung cancer histologies, but had its strongest effect on squamous cell carcinoma (OR = 0.84, $P = 3.7 \times 10^{-11}$) and large-cell carcinoma (OR = 0.78, $P = 0.006$). Both SNPs in the 5p15.33 locus had a stronger effect in never smokers (OR = 1.25 for rs2736100 and OR = 0.80 for rs401681) than in ever smokers (OR = 1.11, $P_{het} = 0.04$ for rs2736100 and OR = 0.88; $P_{het} = 0.11$ for rs401681). The rs2736100 association was stronger in women than in men (respective ORs = 1.21 and 1.12; $P_{het} = 0.05$) and in late-stage versus early-stage disease (respective ORs = 1.19 and 1.07; $P_{het} = 0.05$). Logistic regression including the allelic dosage of two independent SNPs (rs401681 and rs2736100) as covariates showed no support for additional independent associations at 5p15.33 (Fig. 3). Consistent with previous reports (15), rs2736100 and rs401681 both showed an association with lung cancer risk in Han Chinese (Supplementary Material, Fig. S5). In the meta-analysis of imputed genotypes, rs2853677, localizing

Table 1. Studies included in the meta-analysis

Study	Subjects (*n*)		Location	Study design	Illumina genotyping platform	Number of SNPs
	Cases	Controls				
MDACC[a]	1150	1134	Texas, USA	Hospital-based case–control	317 K	312 829
Liverpool Lung Project	543	2501	Liverpool, UK	Population-bases cases, UK Blood Service collections controls (UKBS, WTCCCII[l])	317 K (cases), 1.2 M (UKBS controls)	283 347
ICR-GWA study[b]	1952	2699	UK	Hospital-based cases, 1958 Birth cohort controls (58C, WTCCCII[l])	550 K (cases), 1.2 M (58C controls)	283 347
SLRI[c]/Toronto	331	499	Toronto, Canada	Hospital-based case–control	317 K	314 285
IARC[d] GWAS						312 706
Central Europe	1854	2453	Romania, Hungary, Slovakia, Poland, Russia, Czech Republic	Multicenter hospital-based case–control	317 K, 370Duo	
CARET[e]	394	391	6 US Centers	Cancer Prevention Trial	370Duo	
Estonia	109	851	Estonia	Hospital-based case–control	317 K, 370Duo	
France	143	145	Paris Areas, France	Hospital-based case–control	370Duo	
HUNT2/ Tromso[f]	394	393	Norway	Population-based case–control	370Duo	
DeCODE Genetics	830	11 228	Iceland	Population-based case–control	317 K, 370Duo	290 386
HGF Germany[g]	487	480	Germany	Population-based case–control (<50 years)	550 K	503 381
Harvard	984	970	Massachusetts, USA	Hospital-based case–control	610Quad	543 697
NCI GWAS						506 062
EAGLE[h]	1920	1979	Italy	Population-based case–control	550 K, 610QUAD	
ATBC[i]	1732	1271	Finland	Cohort	550 K, 610QUAD	
PLCO[j]	1380	1817	10 US Centers	Cohort-Cancer Prevention Trial	317 K + 240 S, 550 K, 610QUAD	
CPS-II[k]	697	674	All US States	Cohort	550 K, 610QUAD	
Overall	14 900	29 485				

[a]MD Anderson Cancer Center.
[b]Institute of Cancer Research.
[c]Samuel Lunenfeld Research Institute.
[d]International Agency for Research on Cancer.
[e]Carotene and Retinol Efficacy Trial cohort.
[f]North Trondelag Health Study 2 / Tromsø IV.
[g]Helmholtz-Gemeinschaft Deutscher Forschungszentren Lung Cancer GWAS.
[h]Environment And Genetics in Lung cancer Etiology study.
[i]Alpha-Tocopherol, Beta-Carotene Cancer Prevention study.
[j]Prostate, Lung, Colon, Ovary screening trial.
[k]Cancer Prevention Study II nutrition cohort.
[l]Wellcome Trust Case Control Consortium.

to intron 2 of *TERT*, showed the strongest evidence for an association with adenocarcinoma (OR for the G allele = 1.33; $P = 2.2 \times 10^{-18}$) and rs465498, localizing to intron 10 of *CLPTM1L*, showed the strongest association with lung cancer overall (OR for the A allele = 1.15, $P = 8.2 \times 10^{-18}$; Supplementary Material, Fig. S4). Both variants showed correlations with the previously identified SNPs ($r^2 = 1.0$ and $D' = 1.0$ between genotyped rs401681 and imputed rs465498 at *CLPTM1L* and $r^2 = 0.54$ and $D' = 0.80$ between genotyped rs2736100 and imputed rs2853677 at *TERT*).

The analysis of previously reported lung cancer risk locus at 6p21–6p22, which encompasses *HLA*, is complicated by an extended LD structure (7). The strongest 6p21–6p22 association was shown for rs3117582 ($P = 2.3 \times 10^{-14}$), which maps the 5′ upstream of the DNA repair gene *BAT3*, and is in complete LD with rs3131379 in *MSH5*, a DNA mismatch repair gene. This association was stronger for squamous cell carcinoma (respective ORs for squamous carcinoma and adenocarcinoma, 1.30 and 1.12, $P_{het} = 0.02$; Fig. 2, Supplementary Material, Fig. S3). Logistic regression including the allelic dosage of rs3117582 did not identify any SNPs associated with lung cancer risk with genome-wide significance (Fig. 3). When markers were imputed, the strongest signal for squamous cell carcinoma in this region was observed for two correlated variants ($r^2 = 1.0$, $D' = 1.0$; $r^2 = 0.76$, $D' = 0.93$ with rs3117582 for both SNPs): rs2523546 (effect allele G, OR = 0.76, $P = 1.1 \times 10^{-10}$) and rs2523571 (effect allele T, OR = 0.76, $P = 9.7 \times 10^{-11}$) located in the 3′UTR region of *HLA-B* (Supplementary Material, Fig. S4).

Excluding SNPs mapping to 5p15.33, 6p21–6p22 and 15q25.1, no SNP showed evidence for a genome-wide significant association with lung cancer risk (Fig. 1). Stratification of lung cancer cases by histology did, however, reveal associations with squamous cell carcinoma risk for the previously described risk locus at 12p13 and two potential novel disease loci at 9p21.3 and 2q32.1 ($P < 5.0 \times 10^{-7}$; Figs 1

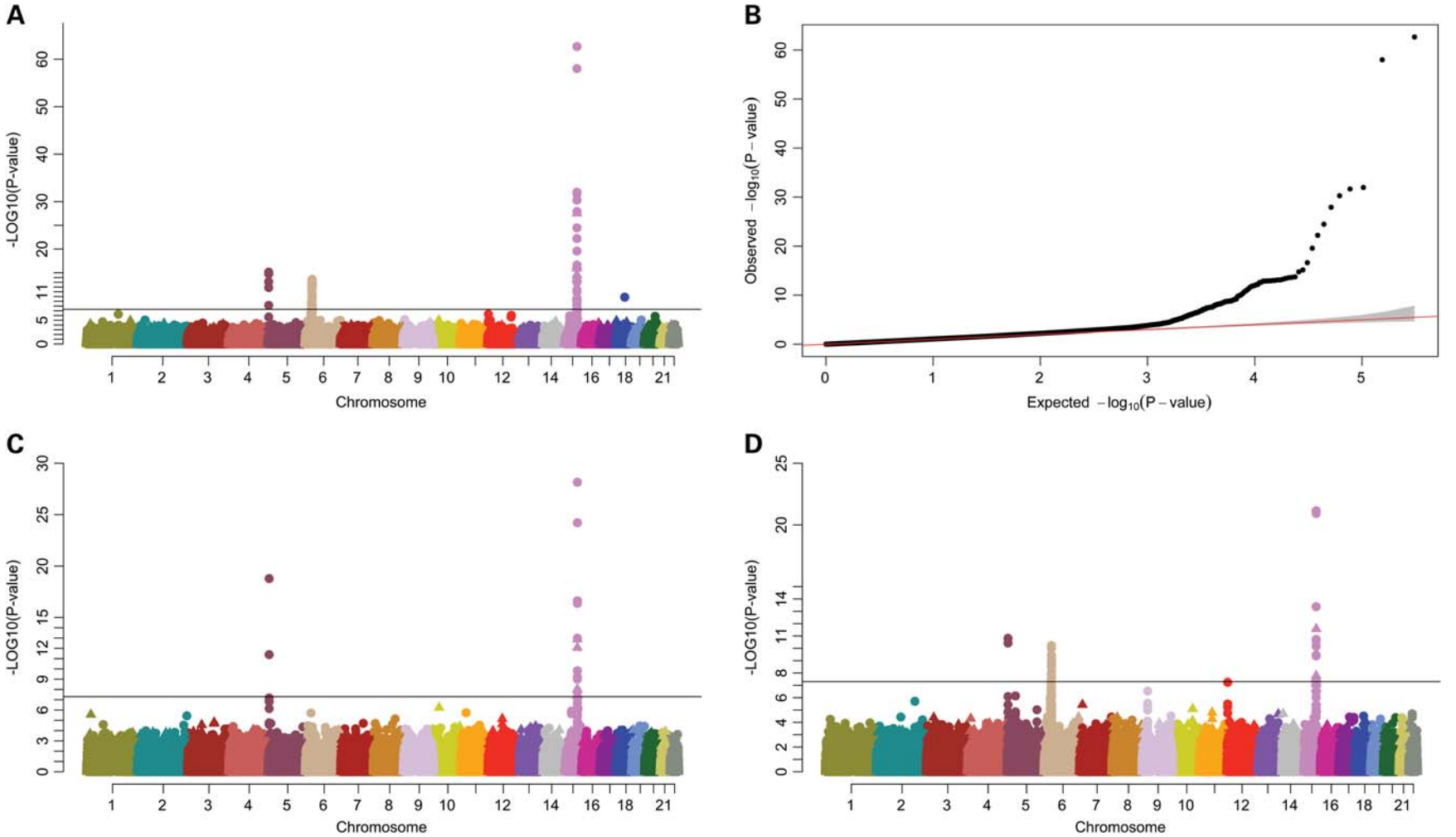


Figure 1. Manhattan and quantile–quantile (Q–Q) plots for the meta-analysis of lung cancer overall and major histologies. Combined ORs and *P*-values were derived from the per-allele model. Core 318 094 SNPs corresponding to the Illumina HumanHap 300 BeadChips array are shown in the Manhattan plots as round-shaped. Additional 217 914 SNPs corresponding to the Illumina HumanHap550 array are shown as triangle-shaped. (**A**) The Manhattan plot of *P*-values for the fixed-effects model for the overall meta-analysis. rs1551821 at 18q21.1 reached genome-wide significance for the fixed effect (effect allele C, OR = 0.81, $P = 6.01 \times 10^{-10}$). However, strong heterogeneity by study ($P_{het} = 3.11 \times 10^{-6}$, $I^2 = 85\%$) driven by two UK studies (OR = 0.90, $P = 0.06$ when the ICR removed), observed deviation from the Hardy–Weinberg equilibrium in the SLRI/Toronto, HGF Germany and MDACC studies and no evidence of association for the correlated SNPs within locus indicated possible chance finding (Supplementary Material, Fig. S2). (**B**) The Q–Q plot for *P*-values in the $-\log_{10}$ scale for the fixed-effects model for the core 318 094 SNPs. The inflation factor for the 90% bottom SNPs (λ) = 1.10. The red line represents the concordance of observed and expected values. The shaded area indicates a 99% concentration band. (**C**) The Manhattan plot of *P*-values for the fixed-effects model for adenocarcinoma histology. The inflation factor for the 90% bottom SNPs (λ) = 1.05. (**D**) The Manhattan plot of *P*-values for the fixed-effects model for squamous cell carcinoma histology. The inflation factor for the 90% bottom SNPs (λ) = 1.04.

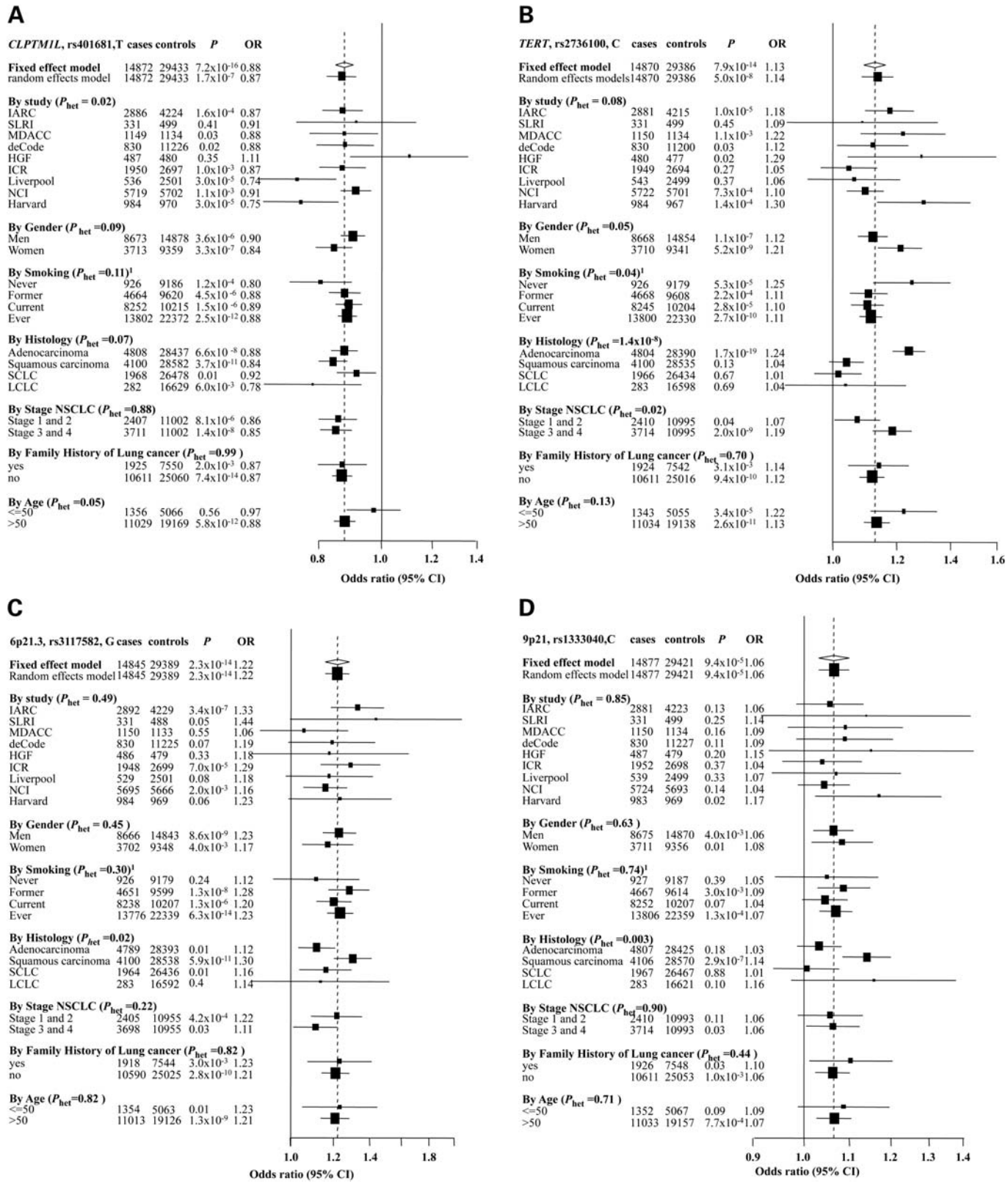
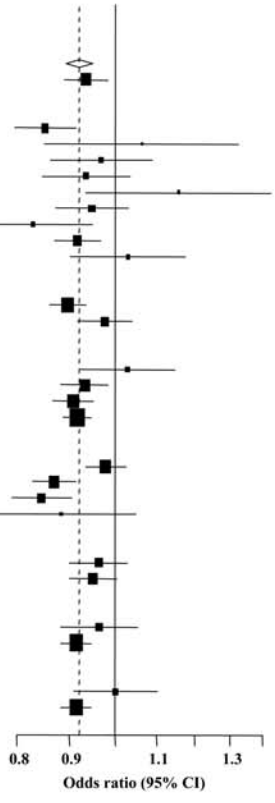


Figure 2. Association between SNPs on 5p15.33, 6p22.3-6p21.31, 9p21.3, 12p13.33 and 15q25.1 and the risk of lung cancer. Combined ORs and 95% CIs were derived from the per-allele model. Except for the ORs for the random-effects model, results for the fixed-effects model are presented. Squares represent ORs; size of the square represents the inverse of the variance of the log ORs; horizontal lines represent 95% CIs; diamonds represent the summary estimate combining the study-specific estimates with a fixed-effects model; solid vertical lines represent OR = 1; dashed vertical lines represent the overall ORs. Results within different strata (histology, age, smoking, gender, family history and stage) are presented for the fixed-effects model. The allele frequency of selected SNPs by study and the case–control status are presented in Supplementary Material, Table S7. [1]Heterogeneity assessed between ever and never smoking groups. NSCLC, non-small-cell lung cancer; SCLC, small-cell lung cancer; LCLC, large-cell lung cancer.

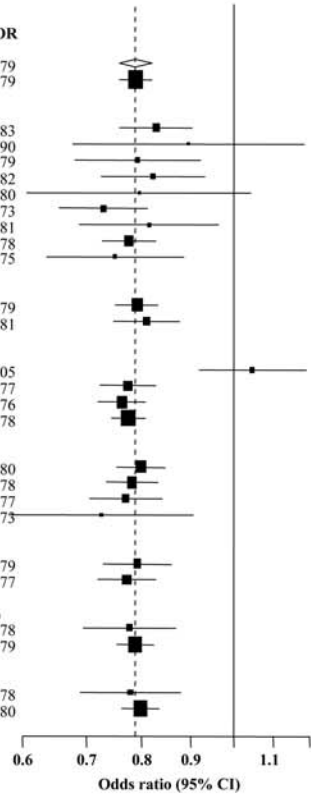
E

12p13, rs10849605,T	cases	controls	*P*	OR
Fixed effect model	14863	29369	5.0×10^{-7}	0.92
Random effects model	14863	29369	0.01	0.94
By study (P_{het} = 0.03)				
IARC	2885	4224	1.0×10^{-5}	0.85
SLRI	331	498	0.59	1.06
MDACC	1148	1131	0.61	0.97
deCode	830	11211	0.22	0.94
HGF	484	478	0.18	1.16
ICR	1945	2694	0.22	0.95
Liverpool	539	2495	0.01	0.83
NCI	5717	5668	2.0×10^{-3}	0.92
Harvard	984	970	0.65	1.03
By Gender (P_{het} =0.03)				
Men	8671	14832	6.0×10^{-7}	0.90
Women	3708	9350	0.5	0.98
By Smoking (P_{het} =0.04)[1]				
Never	925	9157	0.59	1.03
Former	4662	9602	0.02	0.93
Current	8245	10195	7.0×10^{-5}	0.91
Ever	13794	22331	6.0×10^{-7}	0.92
By Histology (P_{het}=0.0002)				
Adenocarcinom	4802	28374	0.4	0.98
Squamous carcinoma	4102	28519	6.0×10^{-8}	0.87
SCLC	1966	26418	2.0×10^{-6}	0.85
LCLC	283	16580	0.16	0.88
By Stage NSCLC (P_{het} =0.77)				
Stage 1 and 2	2405	10961	0.27	0.96
Stage 3 and 4	3713	10961	0.08	0.95
By Family History of Lung cancer (P_{het} =0.27)				
yes	1921	7527	0.43	0.97
no	10606	25010	2.0×10^{-6}	0.92
By Age (P_{het} =0.09)				
<=50	1350	5060	0.97	1.00
>50	11028	19120	2.0×10^{-6}	0.92



F

15q25, rs6495309,T	cases	controls	*P*	OR
Fixed effect model	14893	29465	1.1×10^{-32}	0.79
Random effects model	14893	29465	1.1×10^{-32}	0.79
By study (P_{het} = 0.73)				
IARC	2891	4230	3.0×10^{-5}	0.83
SLRI	331	499	0.44	0.90
MDACC	1150	1133	2.8×10^{-3}	0.79
deCode	830	11227	3.0×10^{-3}	0.82
HGF	487	480	0.1	0.80
ICR	1951	2698	3.1×10^{-9}	0.73
Liverpool	542	2498	0.02	0.81
NCI	5727	5730	2.2×10^{-14}	0.78
Harvard	984	970	7.2×10^{-4}	0.75
By Gender (P_{het} =0.64)				
Men	8686	14910	1.4×10^{-18}	0.79
Women	3714	9361	2.6×10^{-7}	0.81
By Smoking (P_{het} =1.7×10^{-5})[1]				
Never	927	9200	0.49	1.05
Former	4670	9632	3.1×10^{-13}	0.77
Current	8265	10217	1.7×10^{-19}	0.76
Ever	13822	22388	2.8×10^{-32}	0.78
By Histology (P_{het} =0.60)				
Adenocarcinoma	4812	28469	1.1×10^{-13}	0.80
Squamous carcinoma	4105	28614	4.3×10^{-14}	0.78
SCLC	1971	26511	5.2×10^{-9}	0.77
LCLC	283	16656	0.01	0.73
By Stage NSCLC (P_{het} =0.65)				
Stage 1 and 2	2411	11030	5.9×10^{-8}	0.79
Stage 3 and 4	3715	11030	5.4×10^{-13}	0.77
By Family History of Lung cancer (P_{het} =0.82)				
yes	1925	7550	9.1×10^{-6}	0.78
no	10625	25085	3.4×10^{-25}	0.79
By Age (P_{het} =0.71)				
<=50	1354	5068	6.5×10^{-5}	0.78
>50	11045	19201	2.6×10^{-22}	0.80



G

CHRNA3, rs1051730,A	cases	controls	*P*	OR
Fixed effect model	14879	29427	2.2×10^{-63}	1.32
Random effects model	14879	29427	2.2×10^{-63}	1.32
By study (P_{het} = 0.99)				
IARC	2886	4228	4.0×10^{-15}	1.34
SLRI	328	499	0.01	1.33
MDACC	1149	1134	2.1×10^{-6}	1.35
deCode	830	11223	8.9×10^{-6}	1.28
HGF	486	479	0.02	1.28
ICR	1952	2694	1.1×10^{-9}	1.30
Liverpool	541	2499	8.3×10^{-4}	1.27
NCI	5723	5702	2.1×10^{-23}	1.32
Harvard	984	969	3.2×10^{-6}	1.39
By Gender (P_{het} =0.02)				
Men	8675	14881	9.1×10^{-30}	1.29
Women	3711	9355	5.7×10^{-25}	1.42
By Smoking (P_{het} =0.002)[1]				
Never	924	9181	0.06	1.12
Former	4668	9613	8.1×10^{-19}	1.29
Current	8257	10214	3.6×10^{-36}	1.36
Ever	13811	22368	1.8×10^{-59}	1.34
By Histology (P_{het} =0.24)				
Adenocarcinoma	4806	28431	7.0×10^{-29}	1.32
Squamous carcinoma	4106	28576	7.2×10^{-22}	1.29
SCLC	1969	26473	3.4×10^{-14}	1.31
LCLC	283	16623	4.7×10^{-6}	1.50
By Stage NSCLC (P_{het} =0.06)				
Stage 1 and 2	2408	10998	3.7×10^{-12}	1.28
Stage 3 and 4	3714	10998	1.5×10^{-29}	1.39
By Family History of Lung cancer (P_{het} = 0.33)				
yes	1926	7542	1.4×10^{-11}	1.36
no	10614	25053	5.0×10^{-43}	1.30
By Age (P_{het} =0.64)				
<=50	1352	5067	1.2×10^{-9}	1.36
>50	11033	19167	8.1×10^{-48}	1.33

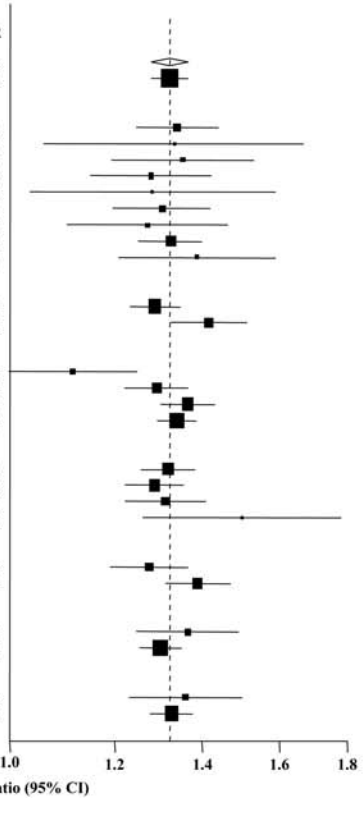


Figure 2. (Continued).

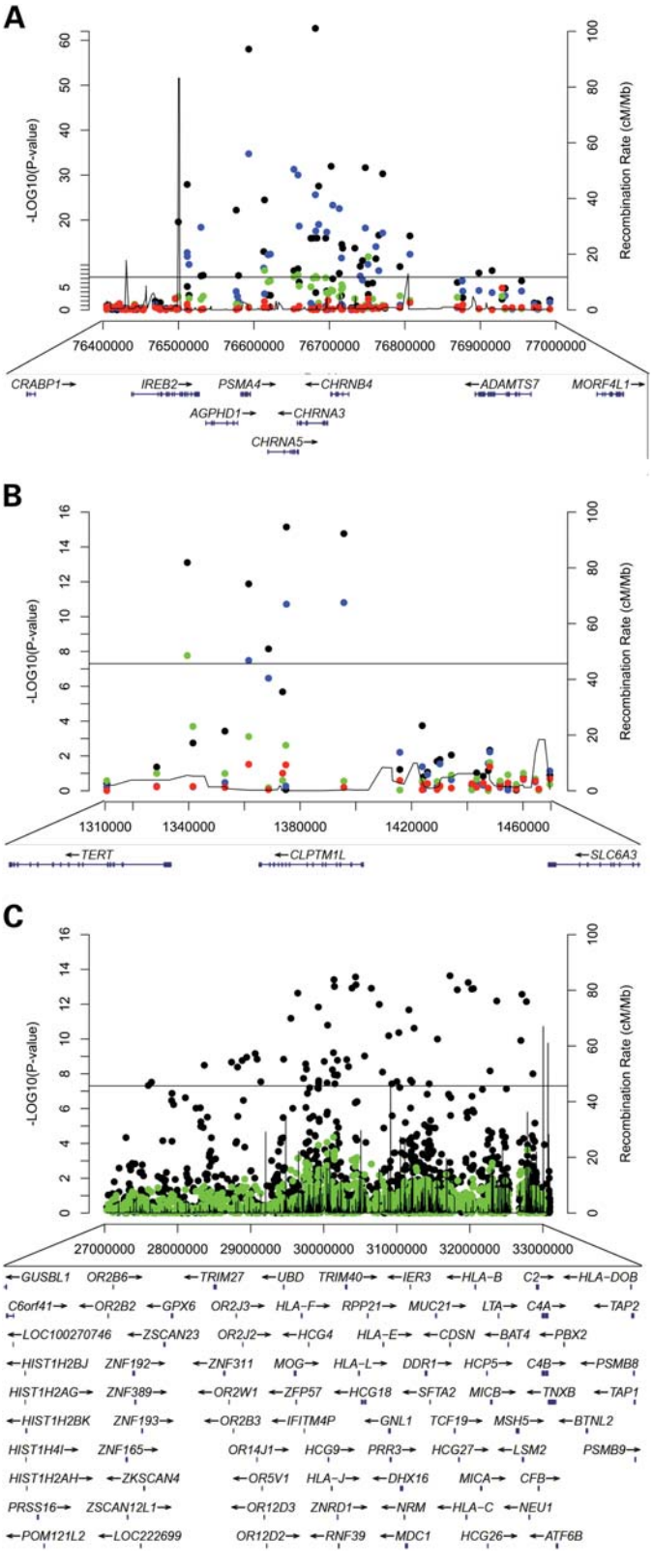


Figure 3. The regional plot of the 15q25, 5p15 and 6p21–6p22 loci after controlling for most significantly associated SNPs within the locus. *P*-values for log-additive association results ($-\log_{10}$) are shown with the recombination rate based on HapMap phase II data. (**A**) 15q25 locus. Black dots, results ($-\log_{10}$ P) for SNPs genotyped within the region. Blue, results after the inclusion of rs6495309 allele dosage as a covariate; green, results after the inclusion of rs1051730 allele dosage as a covariate; red, a model includes allele dosages for both SNPs. rs7173743 showed association ($P = 1.4 \times 10^{-5}$) after controlling for both SNPs with high heterogeneity between studies $I^2 =$ 99.1%. (**B**) 5p15 locus. Black dots, results ($-\log_{10}$ P) for SNPs genotyped within the region. Blue, results after the inclusion of rs2736100 allele dosage as a covariate, *TERT*; green, allele dosage for rs401681 is included as a covariate, *CLPTM1L*; red, allele dosages for both SNPs are included as a covariate. (**C**) 6p21–6p22 locus. Black dots, ($-\log_{10}$ P) for SNPs genotyped within the region; green, allele dosage for rs3117582 is included as a covariate, *BAG6/BAT3*. Two SNPs (rs1003581 and rs130065) reaching genome–wide significance after conditioning on rs3117582 were observed within the same locus with strong heterogeneity by study ($I^2 = 99\%$) suggesting false findings.

and 2, Supplementary Material, Fig. S7). No evidence for association was noted for chromosome X variants in the gender subgroup analysis.

Specifically, rs10849605, which maps within intron 1 of the DNA double-strand repair gene *RAD52* in the 12p13.33 region, was inversely associated with lung cancer risk (effect allele T; OR = 0.92, $P = 5.0 \times 10^{-7}$). This association was particularly strong among smokers (OR = 0.92, $P = 6.0 \times 10^{-7}$) and cases with squamous (OR = 0.87; $P = 6.0 \times 10^{-8}$) and small-cell carcinoma (OR = 0.85, $P = 2.0 \times 10^{-6}$; $P_{het} = 0.0002$ across histologies; Fig. 2E, Supplementary Material, Fig. S3). This variant was not significantly associated with lung cancer risk overall or any histology group in the Han Chinese GWAS (Table 2, Supplementary Material, Fig. S5).

To further explore this region, we performed a meta-analysis of imputed variants from 15 GWASs from eight analytical centers on individuals of European ancestry. In this analysis, rs3748522, which maps to intron 1 of *RAD52* and is in strong LD with rs10849605 ($r^2 = 0.90$, $D' = 1$), provided the strongest association with squamous cell carcinoma (effect allele A, OR = 0.86, $P = 3.4 \times 10^{-8}$; Supplementary Material, Fig. S4).

The strongest evidence for association at 9p21 was shown with rs1333040, which is located ~ 74 kb upstream of *CDKN2B,* within intron 12 of *CDKN1B* antisense *RNA 1* or *ANRIL* (effect allele C, OR = 1.06, $P = 9.4 \times 10^{-5}$; Figs 1 and 2D). A subgroup analysis by histology revealed strong heterogeneity ($P_{het} = 0.003$) with the strongest association for squamous cell cancer (OR = 1.14, $P = 2.9 \times 10^{-7}$). Rs1333040 also showed an association with squamous cell carcinoma in the Han Chinese population ($P = 0.03$, Supplementary Material, Fig. S5). In the combined analysis of all data sets, this association attained genome-wide significance ($P = 2.3 \times 10^{-8}$; Table 2). In an analysis of imputed data across eight studies, the lowest *P*-value was shown for rs1537372 (effect allele G, OR = 1.14, $P = 3.3 \times 10^{-6}$) ($r^2 = 0.60$, $D' = 0.95$ with rs1333040) located in the intron 14 of *CDKN1B-AS/ANRIL* (Supplementary Material, Fig. S4).

The next strongest association with squamous cell carcinoma risk was shown for rs11683501 at 2q32.1 after adjustment for smoking (effect allele G, OR = 1.17, $P = 1.6 \times 10^{-7}$, Supplementary Material, Fig. S7). This SNP is located 3′ downstream of the nucleoporin 35 kDa (*NUP35*) gene (22). Imputation did not identify any stronger association (Supplementary Material, Fig. S4), and rs11683501 did not show an

Table 2. Association of selected SNPs at 9p21 and 12p13 with the risk of lung cancer overall and by histology[a]

Marker	Subgroup	Meta-analysis EAF controls	EAF cases	OR (95% CI)	*P*-value	Han Chinese EAF controls	EAF cases	OR (95% CI)	*P*-value	Overall OR (95% CI)	*P*-value
9p21	Overall	0.41	0.43	1.06 (1.03–1.10)	9.45×10^{-5}	0.30	0.32	1.10 (1.01–1.20)	0.03	1.07 (1.04–1.10)	1.05×10^{-5}
CDKN2B/CDKN2BAS	Adenocarcinoma	0.41	0.41	1.03 (0.99–1.08)	0.18	0.30	0.31	1.07 (0.97–1.19)	0.19	1.04 (0.99–1.09)	0.08
rs1333040[b]	Squamous cell carcinoma	0.41	0.45	1.14 (1.08–1.20)	2.91×10^{-7}	0.30	0.33	1.16 (1.02–1.32)	0.03	1.14 (1.09–1.20)	2.28×10^{-8}
Effect allele C	Small-cell carcinoma	0.40	0.42	1.01 (0.94–1.08)	0.88	0.30	0.28	0.91 (0.72–1.16)	0.46	1.00 (0.93–1.07)	0.95
12p13	Overall	0.49	0.47	0.92 (0.89–0.95)	5.00×10^{-7}	0.68	0.68	0.97 (0.88–1.05)	0.43	0.93 (0.90–0.96)	5.60×10^{-7}
5′UTR RAD52	Adenocarcinoma	0.49	0.46	0.98 (0.94–1.03)	0.40	0.68	0.68	0.96 (0.87–1.07)	0.49	0.98 (0.94–1.02)	0.29
rs10849605	Squamous cell carcinoma	0.49	0.47	0.87 (0.83–0.92)	5.69×10^{-8}	0.68	0.67	0.95 (0.84–1.08)	0.45	0.88 (0.84–0.93)	9.62×10^{-8}
Effect allele T	Small-cell carcinoma	0.48	0.44	0.85 (0.79–0.91)	2.00×10^{-6}	0.68	0.67	0.97 (0.77–1.23)	0.81	0.86 (0.80–0.91)	3.68×10^{-6}

EAF, effect allele frequency.
[a]Results for the fixed effect model is presented.
[b]Results for the Han Chinese population is based on imputed genotypes.

association with risk in the Han Chinese population (Supplementary Material, Fig. S5).

We additionally interrogated variation at 15q15.2, which has been previously identified as a determinant of lung cancer risk (8,12–14,23). In the meta-analysis, rs504417 showed the strongest association, but did not attain genome-wide significance ($P = 1.2 \times 10^{-6}$; Supplementary Material, Fig. S7).

Finally, we evaluated the 3q28, 2q29, 13q12.12 and 22q12.2 regions previously identified in GWASs of Asian populations (15,16,24) as risk factors for lung cancer. None of the SNPs (or their proxies) mapping to these loci showed evidence for an association in European populations (Supplementary Material, Table S5).

DISCUSSION

By pooling summary results from 16 GWASs, we have provided additional evidence for inherited genetic predisposition to lung cancer and have refined associations at the 5p15, 6p21–6p22, 12p13 and 15q25 risk loci. Furthermore, we have shown that 9p21.3 variation is a determinant of squamous cell lung cancer risk.

5p15.33 region

Consistent with previous studies (6,7,17,18), our meta-analysis confirmed two independent signals at 5p15.33 (annotating *TERT* and *CLPTM1L* genes) as determinants of lung cancer risk impacting differentially on lung cancer histology. The rs2736100 variant in *TERT* was principally associated with adenocarcinoma risk and showed stronger effects in women, early-onset disease and never smokers where the proportion of adenocarcinoma cases is generally higher (25–27). Although indirect, the possibility that the association between rs2736100 and adenocarcinoma risk is mediated through an effect on *TERT* is supported by an observation of *TERT* amplification and mRNA overexpression in adenocarcinoma (28), as well as the inhibition of lung adenocarcinoma cell growth promoted by the suppression of hTERT expression (29).

The *CLPTM1L* association appears stronger in squamous cell lung carcinoma and large-cell lung cancer, two histology groups strongly linked to tobacco smoking. This is consistent with the finding that a variant in *CLPTM1L* (rs402710, G) has been associated with high levels of DNA adducts caused by smoking (30).

6p21–6p22 region

A role for the 6p21–6p22 locus in lung cancer development has been previously shown by some (4,7), but not all studies (17,31). This meta-analysis identified 61 SNPs at 6p21–6p22 showing a significant association with lung cancer risk. Most of these SNPs were highly correlated with rs3117582, which had the strongest effect for the squamous cell carcinoma. rs3117582 is located 73 bp 5′ of the gene encoding BCL2-associated athanogene 6 (*BAG6/BAT3*), which belongs to a BAG domain containing a family of proteins that interact with Hsp70/Hsc70 (32). BAT3/BAG6-deficient mice are embryonic lethal with defects in the development of the lung, brain and kidney (33). BAT3/

BAG6 plays an essential role in p53-mediated apoptosis induced by genotoxic stress (34). rs3117582 is in perfect LD with rs3131379, which maps to intron 10 of the DNA mismatch repair mutS homolog 5 (*Escherichia coli*) (*MSH5*) gene. Both *BAT3* and *MSH5* are expressed in lung tissue and are strong potential candidates for being the functional basis for the association (35,36). Since the development of squamous cell lung cancer is strongly influenced by environmental exposure to carcinogens that cause DNA damage, it is highly plausible that genetic variation in the DNA repair mechanism and/or DNA-damage-induced apoptosis would play an etiologic role.

9p21.3 locus

The 9p21.3 region encodes three tumor suppressor genes that play key roles in cell cycle inhibition, senescence and stress-induced apoptosis: $CDKN2A/p16^{INK4A}$ (cyclin-dependent kinase inhibitors 2A), $p14^{ARF}$ (alternative transcript generated by alternative exon 1 of $CDKN2A/p16^{INK4A}$) and $CDKN2B/p15^{INK4B}$ (cyclin-dependent kinase inhibitors 2B) (37). $CDKN2A/p16^{INK4A}$ was originally identified as a melanoma susceptibility gene (38), but is inactivated in many tumors including lung cancer (39–42). 9p21.3 variants associated with lung cancer risk in our study are located 5′ upstream of *CDKN2B*, within the intronic region of the *CDKN2B* antisense RNA (*ANRIL/CDKN2B-AS*). Recent studies have demonstrated 9p21.3 to be a susceptibility locus in many GWASs (43) including on breast cancer (44), glioma (45,46), type 2 diabetes (47–49), endometriosis (50), coronary artery disease (51,52), intracranial aneurysm (53) and glaucoma (44). Several splice variants with varied enhancer activity have been described for *ANRIL* (54), including GQ495921, GQ495919 and GQ495923, which are expressed in lung cancer cell lines (55). Multiple SNPs, including rs1333040 reported here, have been shown to be associated with *ANRIL* mRNA expression in peripheral blood (56). *ANRIL* recruits a polycomb repression complex (PRC2) to silence *CDKN2B* but not *CDKN2A* (54,57,58).

The identified SNP rs1333040 correlates ($0.7 < r^2 < 0.8$) with 24 variants located within or 3′UTR downstream of *CDKN2B-AS*. None of these variants are located within the coding sequence. However, the possibility that the identified variant is tagging a functional SNP located directly within the $CDKN2A/p16^{INK4A}$, $p14^{ARF}$ or $CDKN2B/p15^{INK4B}$ genes cannot be excluded. Further studies are needed to evaluate the effect that the SNPs we identified may have on *ANRIL/CKDN2B-AS*.

12p13.33 locus

The 12p13 risk variants map within the *RAD52* homolog gene which plays a role in DNA double-strand repair and homologous recombination (59,60). A role of the *RAD52* in lung carcinogenesis was originally proposed from a candidate gene study reported by Danoy *et al*. (61), a finding confirmed by a pathway-based analysis using GWAS data from the National Cancer Institute (NCI), UK, and the MD Anderson Cancer Center (MDACC) studies which robustly demonstrated an association for squamous lung cancer (19). The role of *RAD52* in repairing double-strand breaks induced by tobacco smoking is supported by the association being confined to smokers.

15q25.1 region

The present study has confirmed the smoking-related effect of 15q25 variation on lung cancer risk and has provided additional support for the existence of several independent disease loci within the *CHRNA5/CHRNA3/CHRNB4* region. This is consistent with genotyping data which has shown several distinct signals for smoking behavior and lung cancer risk within this region (62–64). Saccone *et al*. (63) described four distinct loci influencing smoking behavior at 15q25 with at least two of them (locus 1 annotated by rs1051730/rs16969968 and locus 3 annotated by rs588765) having independent effects on smoking behavior. The second locus annotated by rs6495308 was more strongly associated with heavy smoking. In contrast, the Oxford-GlaxoSmithKline study reported a secondary locus distinct from rs6495308 (62).

Our current study supports the existence of the two distinct signals defined by rs1051730/rs16969968 and rs6495309/rs6495308/rs2036534. Reciprocal attenuation of the effects for these two signals when allele dosage for an opposite variant is included into a model raises the possibility of an underlying haplotypic effect ($r^2 = 0.17$, $D' = 1.0$ between these two SNPs) or an imperfect correlation with an unknown functional variant. We also observed an effect for rs680244/rs6495306 ($r^2 = 1.0$, $D' = 1.0$ with rs588765 for both) in our meta-analysis, which remained significant at a genome-wide level when controlled for rs6495309 and strongly diminished when controlled for both rs1051730 and rs6495309. This suggests that the rs588765/rs680244/rs6495306 effect on lung cancer risk is not independent. Similar to the earlier observation from Saccone *et al*. (63), these variants had opposite effects when adjusted for rs1051730, which may reflect a haplotypic organization in which the rs1051730 allele increases risk while other associated SNPs decrease the risk.

Impact of variants on squamous cell carcinoma and adenocarcinoma

Our study confirmed a different genetic background for the two major histological subtypes of lung cancer—squamous cell carcinoma and adenocarcinoma. Although the role of the *CHRNA5/CHRNA3/CHRNB4* locus at 15q25 and, to some extent, the *CLPTM1L* locus at 5p15.33 appeared independent from the histology type, all other identified genomic regions showed strong heterogeneity by histology, suggesting different genetic etiologies for these lung cancer subtypes. The significance of cell cycle control (*CDKN2A/ARF/CDKN2B/ANRIL*), DNA damage response and DNA repair genes (*RAD52* and *BAT3/MSH5*) in squamous cell carcinoma is consistent with the notion of a particularly strong effect of smoking on the development of this histological subtype (65) and suggests candidate drug targets that may have clinical utility (66).

The power of the meta-analysis to identify 5p15.33, 6p21–6p22 and 15q25.1 risk SNPs and loci was over 90%, making it unlikely that additional lung cancer susceptibility variants of similar magnitude and allele frequencies can be identified by simply increasing sample size in Europeans. In contrast, the power of our study was limited to detect rarer variants

(i.e. MAFs < 0.05) and common variants of a small effect size (i.e. RR ≤ 1.05) and/or with modest effects confined to a specific histology (Supplementary Material, Fig. S8). The present study was also limited to the genetic variants tagged by the genotyping arrays used. Several novel variants were identified within 5p15.33, 6p21–6p22, 9p21.3, 12p13.33 and 15q25.1 through imputation. The imputed variants correlated with the previously genotyped SNPs in individuals of European descendents, suggesting no additional independent signal within known loci to be identified. However, the replication of imputed variants by direct genotyping would be helpful to completely characterize the strength of effects of these SNPs.

In summary, by pooling results from 16 GWASs, we have been able to comprehensively assay the relationship between common genetic variation and lung cancer risk. Furthermore, we have been able to demonstrate a novel relationship between 9p21.3 variation and squamous cell lung carcinoma. This study provides valuable insights into the pathogenesis of lung cancer, indicating that there is etiological heterogeneity to disease development which is influenced by inherited genetic variation. The identification of additional risk loci is likely to require genotyping larger series using arrays formatted to capture variants poorly tagged by current platforms.

MATERIAL AND METHODS

The study was conducted under the auspices of the Transdisciplinary Research In Cancer of the Lung (TRICL) Research Team, which is a part of the Genetic Associations and MEchanisms in ONcology (GAME-ON) consortium, and associated with the International Lung Cancer Consortium (ILCCO).

Description of studies

The meta-analysis was based on summary data from 16 previously reported lung cancer GWASs undertaken by nine analytical centers providing genotype data on 14 900 lung cancer cases and 29 485 controls of European descent: the MD Anderson Cancer Center lung cancer study (5); cases from the Liverpool Lung Project and control individuals from the UK Blood Service collections (UKBS) (4,67); the UK lung cancer GWAS from the Institute for Cancer Research including lung cancer cases from the Genetic Lung Cancer Predisposition Study (GELCAPS) and controls from the 1958 Birth Cohort (7,68,69); the deCODE Genetics lung cancer study (9); the Helmholtz-Gemeinschaft Deutscher Forschungszentren (HGF) lung cancer GWAS (70); the lung cancer study from Canada (University of Toronto and Samuel Lunenfeld Research Institute) (4); the Harvard lung cancer study (71); the NCI lung cancer GWAS including the Environment and Genetics in Lung Cancer Etiology (EAGLE) study (72), the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) (73), the Prostate, Lung, Colon, Ovary Screening Trial (PLCO) (74) and the Cancer Prevention Study II Nutrition Cohort (CPS-II) (18,75); the IARC lung cancer GWAS (4) including Central Europe GWAS (76), the Carotene and Retinol Efficacy Trial (CARET) cohort lung cancer GWAS (77), the HUNT2/Tromso 4 study (78), lung cancer GWAS from France (79) and the lung cancer study from Estonia (80,81) (Table 1; Supplementary Material, Material and methods). All participants provided informed written consent. All studies were reviewed and approved by institutional ethics review committees at the involved institutions. In each of these studies, SNP genotyping had been performed using Illumina HumanHap 300 BeadChips, HumanHap550 or 610 Quad arrays. Further details about selection criteria, cancer diagnosis, genotyping and quality control in each study are provided in the Supplementary Material, Material and methods. Lung cancer diagnosis in most studies was based on histopathology or cytology but in a minority on clinical history and imaging.

The Chinese lung cancer GWAS included 2338 lung cancer cases and 3077 controls from the Nanjing and Beijing Lung Cancer Studies (15) genotyped using Affymetrix Human SNP Array 6.0 chips (Supplementary Material, Material and methods). The Nanjing and Beijing Lung Cancer Studies provided summary data on the top SNPs for overall lung cancer risk and risk by specific histology. The selected loci were 5p15.33 (1.20–1.61 Mb), 6p22.3–6p21.31 (22.0–36.5 Mb), 15q25.1 (76.1–77.2 Mb), 18q2.3 (40.0–21.5 Mb), 12p13.33 (0.54–1.54 Mb), 2q32.1 (183.4–184.5 Mb) and 9p21.3 (21.66–22.2 Mb) (NCBI Build 36).

Statistical methods

Study-specific analysis of GWAS data

Associations between SNP genotypes and lung cancer risk were evaluated under a log-additive model of inheritance. Additionally, we explored dominant, recessive and co-dominant models. Each study center provided summary statistics from two models: (i) unconditional logistic regression adjusted for sex, age at diagnosis or age at recruitment (5-year age intervals), country/study center where appropriate and significant principal components for population stratification and (ii) additionally adjusted for smoking status coded as categorical variable never/current/former. Analyses stratified by histology (adenocarcinoma, squamous cell carcinoma, large-cell carcinoma and small-cell carcinoma), sex, age at diagnosis for cases or recruitment for controls (≤ 50 and > 50 years), smoking status (current, former, never), tumor stage (I–IV) and family history of lung cancer in a first-degree relative were performed (Supplementary Material, Table S1). Both the UK studies did not contribute data to the smoking analysis, since this information was not available for controls. In addition to the above analyses, each centre provided lung cancer risk estimates for 15q25, 6p21 and 5p15 loci after controlling for allelic dosage for the most significantly associated SNP(s) within the locus. For the 15q25 locus, the statistical model included rs1051730 and/or rs6495309 allelic dosages as covariates; for the 6p21 locus, rs3117582 allelic dosage and for the 5p15 locus allelic, dosages for rs401681 and/or rs2736100.

Prior to undertaking the meta-analysis of all GWAS data sets, we searched for potential errors and biases in data from each case–control series (82). With the exception of the Liverpool Lung Project, quantile–quantile (Q–Q) plots showed that there was minimal inflation of the test statistics, indicating

that substantial cryptic population substructure or differential genotype calling between cases and controls was unlikely in each of the GWASs (Supplementary Material, Fig. S1).

Imputation

To refine the association of the previously reported and newly identified disease loci, we imputed untyped genotypes using Impute2 (83), Mach1 (84,85) or minimac (86) software and HapMap Phase II, Phase III and/or 1000 Genome Project data release 2010-08 or 2010-06 reference genotypes (Supplementary Material, Table S6). The selected loci were 5p15.33 (1.20–1.61 Mb), 6p22.3–6p21.31 (22.0–36.5 Mb), 15q25.1 (76.1–77.2 Mb), 18q2.3 (40.0–21.5 Mb), 12p13.33 (0.54–1.54 Mb), 2q32.1 (183.4–184.5 Mb) and 9p21.3 (21.66–22.2 Mb) (NCBI Build 36). The analytical scheme was similar to the meta-analysis but taking imputation uncertainty into account by using posterior means or allele dosage in logistic regression. Imputed allele dosage for each SNP was tested for association with lung cancer risk using the two models with and without adjustment for smoking as described above. The meta-analysis of imputed genotypes included all studies except the HGF Germany where imputed data were not available. Poorly imputed SNPs defined by an RSQR $<$ 0.30 with MACH1/minimac or an information measure Is $<$ 0.30 with IMPUTE2 were excluded from the analyses (Supplementary Material, Table S6).

Meta-analysis

The meta-analysis was primarily based on pooling GWAS results for the log-additive model of inheritance from 318 094 SNPs featured on Illumina HumanHap 300 BeadChips arrays. For studies genotyped on HumanHap550 or 610Quad Illumina platforms, additional 217 914 SNPs were available to inform our analysis.

Meta-analysis under fixed and random-effects models was conducted. As with individual studies, we examined for the over-dispersion of *P*-values in the meta-analysis by generating Q–Q plots and deriving an inflation factor λ by comparison of observed versus expected *P*-values for the meta-analysis applying the estlambda function within the GenABEL package (87). Cochran's Q statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity were calculated. I^2 values $\geq$75% are considered the characteristic of large heterogeneity (88). To assess the robustness of associations in the meta-analysis, we performed a sensitivity analysis sequentially excluding studies. Wherever removing one study resulted in a $>$10% change of the OR point estimates, we reported results separately (89).

All calculations were performed using PLINK (90) and SAS version 9.2 (SAS Institute Inc., Cary, NC, USA). Q–Q and Manhattan plots were created using an R library GenABEL (87). We used LocusZoom for regional visualization of results (91). Power calculations were performed using QUANTO version 1.2.4 for the main effect of gene and the log-additive model of inheritance stipulating a *P*-value of 5.0×10^{-8} (92).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

WEB ADDRESSES

HapMap Project: http://www.hapmap.org
1000 Genome Project: http://www.1000genomes.org/
MACH1.0: http://www.sph.umich.edu/csg/abecasis/mach/index.html
Impute 2: http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
GenABEL: http://www.genabel.org/packages/GenABEL
SNP Annotation and Proxy Search: http://www.broadinstitute.org/mpg/snap/index.php
TRICL: http://www.u19tricl.org/
R project: www.r-project.org/
PLINK: http://pngu.mgh.harvard.edu/~purcell/plink/
QUANTO: http://hydra.usc.edu/gxe/
AceView: integrative annotation of cDNA-supported genes in human, mouse, rat, worm and Arabidopsis: http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/
International Lung Cancer Consortium: http://ilcco.iarc.fr/
Genetic Association Mechanisms in Oncology (GAME-ON): A Post-Genome Wide Association Initiative: http://epi.grants.cancer.gov/gameon/

ACKNOWLEDGEMENTS

We would like to thank all individuals who participated in this study and who made this study possible.

We gratefully acknowledge the Genetic Associations and MEchanisms in ONcology (GAME-ON) consortium.

We are grateful to clinicians who took part in the GELCAPS consortium and the Liverpool Lung Project (LLP) program. The ICR and LLP studies made use of genotyping data from the WTCCC2; a full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.

We thank Dr H. Dally, Mrs A. Seidel, Dr T. Muley, Dr S. Thiel, Dr H. Wikman, Ms U. von Seydlitz-Kurzbach, Ms D. Bodemer, Dr C. Klappenecker, Mr M. Hoffmann and Mr C. Beynon for help with sample and/or data collection and archiving for the Heidelberg lung study. We are grateful to all patients and staff at the Thoraxklinik-Heidelberg, who participated in the Heidelberg lung cancer study and particularly Prof. Peter Drings for making it possible.

We gratefully acknowledge the KORA study group especially G. Fischer, H. Grallert, N. Klopp, C. Gieger, R. Holle, C. Hanrieder and A. Steinwachs, Institute of Epidemiology, Helmholtz Centre Munich, Neuherberg, Germany, and all individuals, who participated as cases or controls in this study and the KORA Study Center and their co-workers for organizing and conducting the data collection.

We also gratefully acknowledge the LUCY-consortium especially G. Wölke and V. Zietemann for coordinating recruitment, all patients and staff of the participating hospitals: Aurich (Dr Heidi Kleen); Bad Berka (Dr med. R. Bonnet, Klinik für Pneumologie, Zentralklinik Bad Berka GmbH); Bonn (Prof. Ko, Dr Geisen, Innere Medizin I, Johanniter Krankenhaus); Bonn (Dr Stier, Medizinische Poliklinik, Universität Bonn); Bremen (Prof. Dr D. Ukena, Dr Penzl, Zentralkrankenhaus Bremen Ost, Pneumologische Klinik); Chemnitz (PD Dr Schmidt, OA Dr Jielge, Klinikum Chemnitz, Abteilung Innere Medizin); Coswig (Prof. Höffken, Dr Schmidt,

Fachkrankenhaus Coswig); Diekholzen (Dr Hamm, Kreiskrankenhaus Diekholzen, Klinik für Pneumologie); Donaustauf (Prof. Pfeifer, Dr v. Bültzingslöwen, Fr. Schneider, Fachklinik für Atemwegserkrankungen); Essen (Prof. Teschler, Dr Fischer, Ruhrlandklinik—Universitätsklinik, Abt. Pneumologie); Gauting (Prof. Häußinger, Prof. Thetter, Dr Düll, Dr Wagner, Pneumologische Klinik München-Gauting); Gera (CA MR Dr Heil, OÄ Dr Täuscher, OA Dr Lange, II. Medizinische Klinik, Wald-Klinikum Gera); Göttingen (Prof. Trümper, Prof. Griesinger, Dr Overbeck, Abteilung Onkologie, Hämatologie); Göttingen (Prof. Schöndube, Dr Danner, Abteilung Thorax-, Herz- und Gefäßchirurgie); Göttingen/Weende (Dr med. Fleischer, Ev. Krankenhaus Göttingen-Weende e.V., Abteilung Allgemeinchirurgie); Greifenstein (Prof. Morr, Dr M. Degen, Dr Matter, Pneumologische Klinik, Waldhof Elgershausen); Greifswald (Prof. Ewert, Dr Altesellmeier, Universitätsklinik Greifswald, Klinik für Innere Medizin B); Hannover (Prof. Schönhofer, Dr Kohlaußen, Klinikum Hannover Oststadt, Medizinische Klinik II, Pneumologie); Heidelberg (Prof. Drings, Dr Herrmann, Thoraxklinik-Heidelberg GmbH, Abt. Innere Medizin-Onkologie); Hildesheim (Prof. Kaiser, St. Bernward Krankenhaus, Medizinische Klinik II); Homburg (Prof. Sybrecht, OA Dr Gröschel, Dr Mack, Uniklinik des Saarlandes, Innere Medizin V); Immenhausen (Prof. Andreas, Dr Rittmeyer, Fachklinik für Lungenerkrankungen); Köln (Priv. Doz. Dr Stölben, Kliniken der Stadt Köln, Lungenklinik Krankenhaus Merheim); Köln (Prof. Wolf, Dr Staratschek-Jox, Klinikum der Universität Köln, Klinik I für Innere Medizin); Leipzig (Prof. Gillisen, OA Dr Cebulla, Städt. Klinikum St Georg, Robert-Koch-Klinik); Leipzig (Kreymborg, Universitätsklinikum Leipzig, Medizinische Klinik I, Abteilung Pneumologie); Lenglern (Prof. Criée, Dr Körber, Dr Knaack, Ev. Krankenhaus Weende e.V., Standort Lenglern, Abt. Pneumologie; München (Prof. Huber, Dr Borgmeier, Klinikum der LMU-Innenstadt, Abt. Pneumologie); Neustadt a. Harz (Dr Keppler, Schäfer, Evangelisches Fachkrankenhaus für Atemwegserkrankungen); Rotenburg (Prof. Schaberg, Dr Struß, Diakoniekrankenhaus Rotenburg, Lungenklinik Unterstedt); St. Pölten – Österreich (OA Dr M. Wiesholzer, Zentralklinikum St Pölten, I. Medizinische Klinik).

We thank Li Su, Yang Zhao, Geoffrey Liu, John Wain, Rebecca Heist and Kofi Asomaning.

We gratefully acknowledge the Icelandic Cancer Registry (www.krabbameinsskra.is) for assistance in the ascertainment of the Icelandic UBC patients for the deCODE lung cancer study.

Conflict of Interest statement. Some authors involved in this study own stock in DeCODE Genetics.

FUNDING

This study was supported by the grant from the National Institute of Health (NIH) (U19CA148127). The SLRI study was supported by Canadian Cancer Society Research Institute (020214), Ontario Institute of Cancer and Cancer Care Ontario Chair Award to R.H. The ICR study was supported by Cancer Research UK (C1298/A8780 and C1298/A8362—Bobby Moore Fund for Cancer Research UK) and NCRN, HEAL and Sanofi-Aventis. Additional funding was obtained from NIH grants (5R01CA055769, 5R01CA127219, 5R01CA133996, and 5R01CA121197). The Liverpool Lung Project (LLP) was supported by The Roy Castle Lung Cancer Foundation, UK. The ICR and LLP studies made use of genotyping data from the Wellcome Trust Case Control Consortium 2 (WTCCC2); a full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Sample collection for the Heidelberg lung cancer study was in part supported by a grant (70-2919) from the Deutsche Krebshilfe. The work was additionally supported by a Helmholtz-DAAD fellowship (A/07/97379 to M.N.T.) and by the National Institute of Health (USA) (U19CA148127). The KORA Surveys were financed by the GSF, which is funded by the German Federal Ministry of Education, Science, Research and Technology and the State of Bavaria. The LUng Cancer in the Young study (LUCY) was funded in part by the National Genome Research Network (NGFN), the DFG (BI 576/2-1; BI 576/2-2), the Helmholtzgemeinschaft (HGF) and the Federal office for Radiation Protection (BfS: STSch4454). Genotyping was performed in the Genome Analysis Center (GAC) of the Helmholtz Zentrum Muenchen. Support for the Central Europe, HUNT2/Tromsø and CARET genome-wide studies was provided by Institut National du Cancer, France. Support for the HUNT2/Tromsø genome-wide study was also provided by the European Community (Integrated Project DNA repair, LSHG-CT-2005-512113), the Norwegian Cancer Association and the Functional Genomics Programme of Research Council of Norway. Support for the Central Europe study, Czech Republic, was also provided by the European Regional Development Fund and the State Budget of the Czech Republic (RECAMO, CZ.1.05/2.1.00/03.0101). Support for the CARET genome-wide study was also provided by grants from the National Cancer Institute of the U.S. National Institutes of Health (R01 CA111703 and UO1 CA63673) and by funds from the Fred Hutchinson Cancer Research Center. Additional funding for study coordination, genotyping of replication studies and statistical analysis was provided by the US National Cancer Institute (R01 CA092039). The lung cancer GWAS from Estonia was partly supported by a FP7 grant (REGPOT 245536), by the Estonian Government (SF0180142s08), by EU RDF in the frame of Centre of Excellence in Genomics and Estoinian Research Infrastructure's Roadmap and by University of Tartu (SP1GVARENG). The work reported in this paper was partly undertaken during the tenure of a Postdoctoral Fellowship from the IARC (for MNT). The Environment and Genetics in Lung Cancer Etiology (EAGLE), the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) and the Prostate, Lung, Colon, Ovary Screening Trial (PLCO) studies and the genotyping of ATBC, the Cancer Prevention Study II Nutrition Cohort (CPS-II) and part of PLCO were supported by the Intramural Research Program of NIH, NCI, Division of Cancer Epidemiology and Genetics. ATBC was also supported by U.S. Public Health Service contracts (N01-CN-45165, N01-RC-45035 and N01-RC-37004) from the NCI. PLCO was also supported by individual contracts from the NCI to the University of Colorado Denver (NO1-CN-25514), Georgetown University (NO1-CN-25522), Pacific Health Research Institute (NO1-

CN-25515), Henry Ford Health System (NO1-CN-25512), University of Minnesota (NO1-CN-25513), Washington University (NO1-CN-25516), University of Pittsburgh (NO1-CN-25511), University of Utah (NO1-CN-25524), Marshfield Clinic Research Foundation (NO1-CN-25518), University of Alabama at Birmingham (NO1-CN-75022, Westat, Inc. NO1-CN-25476), University of California, Los Angeles (NO1-CN-25404). The Cancer Prevention Study II Nutrition Cohort was supported by the American Cancer Society. The NIH Genes, Environment and Health Initiative (GEI) partly funded DNA extraction and statistical analyses (HG-06-033-NCI-01 and RO1HL091172-01), genotyping at the Johns Hopkins University Center for Inherited Disease Research (U01HG004438 and NIH HHSN268200782096C) and study coordination at the GENEVA Coordination Center (U01 HG004446) for EAGLE and part of PLCO studies. Funding for the MD Anderson Cancer Study was provided by NIH grants (P50 CA70907, R01CA121197, RO1 CA127219, U19 CA148127, RO1 CA55769) and CPRIT grant (RP100443). Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is funded through a federal contract from the National Institutes of Health to The Johns Hopkins University (HHSN268200782096C). The Harvard Lung Cancer Study was funded by Funded by National Institutes of Health (CA074386, CA092824, CA090578). The Nanjing study and Beijing studies were funded by the China National High-Tech Research and Development Program Grant (2009AA022705), the National Key Basic Research Program Grant (2011CB503805) and the National Natural Science Foundation of China (30730080, 30972541, 30901233 and 30872178). Funding to pay the Open Access publication charges for this article was provided by the grant from the National Institute of Health (NIH) (U19CA148127).

REFERENCES

1. Parkin, D.M., Bray, F., Ferlay, J. and Pisani, P. (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
2. Gazdar, A., Franklin, W.A., Brambilla, E., Hainaut, P., Yokota, J. and Harris, C.C. (2004) In Travis, W.D., Brambilla, E., Muller-Hermelink, H.K. and Harris, C.C. (eds), *World Health Organization Classification of Tumors. Pathology and Genetics of Tumors of the Lung, Pleura and Heart*. IARCPress, Lyon, pp. 21–23.
3. Brennan, P., Hainaut, P. and Boffetta, P. (2011) Genetics of lung-cancer susceptibility. *Lancet Oncol.*, **12**, 399–408.
4. Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P. *et al.* (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*, **452**, 633–637.
5. Amos, C.I., Wu, X., Broderick, P., Gorlov, I.P., Gu, J., Eisen, T., Dong, Q., Zhang, Q., Gu, X., Vijayakrishnan, J. *et al.* (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.*, **40**, 616–622.
6. McKay, J.D., Hung, R.J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J. *et al.* (2008) Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.*, **40**, 1404–1406.
7. Wang, Y., Broderick, P., Webb, E., Wu, X., Vijayakrishnan, J., Matakidou, A., Qureshi, M., Dong, Q., Gu, X., Chen, W.V. *et al.* (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.*, **40**, 1407–1409.
8. Broderick, P., Wang, Y., Vijayakrishnan, J., Matakidou, A., Spitz, M.R., Eisen, T., Amos, C.I. and Houlston, R.S. (2009) Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res.*, **69**, 6633–6641.
9. Thorgeirsson, T.E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K.P., Manolescu, A., Thorleifsson, G., Stefansson, H., Ingason, A. *et al.* (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*, **452**, 638–642.
10. Brennan, P., McKay, J., Moore, L., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D. *et al.* (2007) Uncommon CHEK2 mis-sense variant and reduced risk of tobacco-related cancers: case control study. *Hum. Mol. Genet.*, **16**, 1794–1801.
11. Cybulski, C., Masojc, B., Oszutowska, D., Jaworowska, E., Grodzki, T., Waloszczyk, P., Serwatowski, P., Pankowski, J., Huzarski, T., Byrski, T. *et al.* (2008) Constitutional CHEK2 mutations are associated with a decreased risk of lung and laryngeal cancers. *Carcinogenesis*, **29**, 762–765.
12. Rudd, M.F., Webb, E.L., Matakidou, A., Sellick, G.S., Williams, R.D., Bridle, H., Eisen, T. and Houlston, R.S. (2006) Variants in the GH-IGF axis confer susceptibility to lung cancer. *Genome Res.*, **16**, 693–701.
13. Rafnar, T., Sulem, P., Besenbacher, S., Gudbjartsson, D.F., Zanon, C., Gudmundsson, J., Stacey, S.N., Kostic, J.P., Thorgeirsson, T.E., Thorleifsson, G. *et al.* (2011) Genome-wide significant association between a sequence variant at 15q15.2 and lung cancer risk. *Cancer Res.*, **71**, 1356–1361.
14. Truong, T., Sauter, W., McKay, J.D., Hosgood, H.D. III, Gallagher, C., Amos, C.I., Spitz, M., Muscat, J., Lazarus, P., Illig, T. *et al.* (2010) International Lung Cancer Consortium: coordinated association study of 10 potential lung cancer susceptibility variants. *Carcinogenesis*, **31**, 625–633.
15. Hu, Z., Wu, C., Shi, Y., Guo, H., Zhao, X., Yin, Z., Yang, L., Dai, J., Hu, L., Tan, W. *et al.* (2011) 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.
16. Miki, D., Kubo, M., Takahashi, A., Yoon, K.A., Kim, J., Lee, G.K., Zo, J.I., Lee, J.S., Hosono, N., Morizono, T. *et al.* (2010) Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nat. Genet.*, **42**, 893–896.
17. Truong, T., Hung, R.J., Amos, C.I., Wu, X., Bickeboller, H., Rosenberger, A., Sauter, W., Illig, T., Wichmann, H.E., Risch, A. *et al.* (2010) Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J. Natl. Cancer Inst.*, **102**, 959–971.
18. Landi, M.T., Chatterjee, N., Yu, K., Goldin, L.R., Goldstein, A.M., Rotunno, M., Mirabello, L., Jacobs, K., Wheeler, W., Yeager, M. *et al.* (2009) A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.*, **85**, 679–691.
19. Shi, J., Chatterjee, N., Rotunno, M., Wang, Y., Pesatori, A.C., Consonni, D., Li, P., Wheeler, W., Broderick, P., Henrion, M. *et al.* (2012) Inherited variation at chromosome 12p13.33 including RAD52 influences squamous cell lung carcinoma risk. *Cancer Disc.*, **2**, 131–139.
20. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M. *et al.* (2011) Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.*, **19**, 807–812.
21. Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N. *et al.* (2004) Assessing the impact of population stratification on genetic association studies. *Nat. Genet.*, **36**, 388–393.
22. Hawryluk-Gara, L.A., Platani, M., Santarella, R., Wozniak, R.W. and Mattaj, I.W. (2008) Nup53 is required for nuclear envelope and nuclear pore complex assembly. *Mol. Biol. Cell*, **19**, 1753–1762.
23. Kiyohara, C., Horiuchi, T., Miyake, Y., Takayama, K. and Nakanishi, Y. (2010) Cigarette smoking, TP53 Arg72Pro, TP53BP1 Asp353Glu and the risk of lung cancer in a Japanese population. *Oncol. Rep.*, **23**, 1361–1368.
24. Yoon, K.A., Park, J.H., Han, J., Park, S., Lee, G.K., Han, J.Y., Zo, J.I., Kim, J., Lee, J.E., Takahashi, A. *et al.* (2010) A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population. *Hum. Mol. Genet.*, **19**, 4948–4954.
25. Etzel, C.J., Lu, M., Merriman, K., Liu, M., Vaporciyan, A. and Spitz, M.R. (2006) An epidemiologic study of early onset lung cancer. *Lung Cancer*, **52**, 129–134.

26. la Cruz, C.S., Tanoue, L.T. and Matthay, R.A. (2011) Lung cancer: epidemiology, etiology, and prevention. *Clin. Chest Med.*, **32**, 605–644.
27. Subramanian, J., Morgensztern, D., Goodgame, B., Baggstrom, M.Q., Gao, F., Piccirillo, J. and Govindan, R. (2010) Distinctive characteristics of non-small cell lung cancer (NSCLC) in the young: a surveillance, epidemiology, and end results (SEER) analysis. *J. Thorac. Oncol.*, **5**, 23–28.
28. Strazisar, M., Mlakar, V. and Glavac, D. (2009) The expression of COX-2, hTERT, MDM2, LATS2 and S100A2 in different types of non-small cell lung cancer (NSCLC). *Cell. Mol. Biol. Lett.*, **14**, 442–456.
29. Xie, M., Chen, Q., He, S., Li, B. and Hu, C. (2011) Silencing of the human TERT gene by RNAi inhibits A549 lung adenocarcinoma cell growth in vitro and in vivo. *Oncol. Rep.*, **26**, 1019–1027.
30. Zienolddiny, S., Skaug, V., Landvik, N.E., Ryberg, D., Phillips, D.H., Houlston, R. and Haugen, A. (2009) The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. *Carcinogenesis*, **30**, 1368–1371.
31. Jaworowska, E., Trubicka, J., Lener, M.R., Masojc, B., Zlowocka-Perlowska, E., McKay, J.D., Renard, H., Oszutowska, D., Wokolorczyk, D., Lubinski, J. *et al.* (2011) Smoking related cancers and loci at chromosomes 15q25, 5p15, 6p22.1 and 6p21.33 in the Polish population. *PLoS One*, **6**, e25057.
32. Takayama, S. and Reed, J.C. (2001) Molecular chaperone targeting and regulation by BAG family proteins. *Nat. Cell Biol.*, **3**, E237–E241.
33. Desmots, F., Russell, H.R., Lee, Y., Boyd, K. and McKinnon, P.J. (2005) The reaper-binding protein scythe modulates apoptosis and proliferation during mammalian development. *Mol. Cell Biol.*, **25**, 10329–10337.
34. Sasaki, T., Gan, E.C., Wakeham, A., Kornbluth, S., Mak, T.W. and Okada, H. (2007) HLA-B-associated transcript 3 (Bat3)/Scythe is essential for p300-mediated acetylation of p53. *Genes Dev.*, **21**, 848–861.
35. Parkinson, H., Sarkans, U., Kolesnikov, N., Abeygunawardena, N., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Holloway, E. *et al.* (2011) ArrayExpress update–an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **39**, D1002–D1004.
36. Thierry-Mieg, D. and Thierry-Mieg, J. (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.*, **7**(Suppl. 1), S12–S14.
37. Gil, J. and Peters, G. (2006) Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nat. Rev. Mol. Cell Biol.*, **7**, 667–677.
38. Hussussian, C.J., Struewing, J.P., Goldstein, A.M., Higgins, P.A., Ally, D.S., Sheahan, M.D., Clark, W.H. Jr, Tucker, M.A. and Dracopoli, N.C. (1994) Germline p16 mutations in familial melanoma. *Nat. Genet.*, **8**, 15–21.
39. Hamada, K., Kohno, T., Kawanishi, M., Ohwada, S. and Yokota, J. (1998) Association of CDKN2A(p16)/CDKN2B(p15) alterations and homozygous chromosome arm 9p deletions in human lung carcinoma. Genes Chromosomes. Cancer, **22**, 232–240.
40. Washimi, O., Nagatake, M., Osada, H., Ueda, R., Koshikawa, T., Seki, T., Takahashi, T. and Takahashi, T. (1995) In vivo occurrence of p16 (MTS1) and p15 (MTS2) alterations preferentially in non-small cell lung cancers. *Cancer Res.*, **55**, 514–517.
41. Xiao, S., Li, D., Corson, J.M., Vijg, J. and Fletcher, J.A. (1995) Codeletion of p15 and p16 genes in primary non-small cell lung carcinoma. *Cancer Res.*, **55**, 2968–2971.
42. Hawes, S.E., Stern, J.E., Feng, Q., Wiens, L.W., Rasey, J.S., Lu, H., Kiviat, N.B. and Vesselle, H. (2010) DNA hypermethylation of tumors from non-small cell lung cancer (NSCLC) patients is associated with gender and histologic type. *Lung Cancer*, **69**, 172–179.
43. Pasmant, E., Sabbagh, A., Vidaud, M. and Bieche, I. (2011) ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.*, **25**, 444–448.
44. Burdon, K.P., Macgregor, S., Hewitt, A.W., Sharma, S., Chidlow, G., Mills, R.A., Danoy, P., Casson, R., Viswanathan, A.C., Liu, J.Z. *et al.* (2011) Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1. *Nat. Genet.*, **43**, 574–578.
45. Wrensch, M., Jenkins, R.B., Chang, J.S., Yeh, R.F., Xiao, Y., Decker, P.A., Ballman, K.V., Berger, M., Buckner, J.C., Chang, S. *et al.* (2009) Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat. Genet.*, **41**, 905–908.
46. Shete, S., Hosking, F.J., Robertson, L.B., Dobbins, S.E., Sanson, M., Malmer, B., Simon, M., Marie, Y., Boisselier, B., Delattre, J.Y. *et al.* (2009) Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.*, **41**, 899–904.
47. Takeuchi, F., Serizawa, M., Yamamoto, K., Fujisawa, T., Nakashima, E., Ohnaka, K., Ikegami, H., Sugiyama, T., Katsuya, T., Miyagishi, M. *et al.* (2009) Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes*, **58**, 1690–1699.
48. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
49. Parra, E.J., Below, J.E., Krithika, S., Valladares, A., Barta, J.L., Cox, N.J., Hanis, C.L., Wacher, N., Garcia-Mena, J., Hu, P. *et al.* (2011) Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. *Diabetologia*, **54**, 2038–2046.
50. Uno, S., Zembutsu, H., Hirasawa, A., Takahashi, A., Kubo, M., Akahane, T., Aoki, D., Kamatani, N., Hirata, K. and Nakamura, Y. (2010) A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat. Genet.*, **42**, 707–710.
51. Takeuchi, F., Yokota, M., Yamamoto, K., Nakashima, E., Katsuya, T., Asano, H., Isono, M., Nabika, T., Sugiyama, T., Fujioka, A. *et al.* (2012) Genome-wide association study of coronary artery disease in the Japanese. *Eur. J. Hum. Genet.*, **20**, 333–340.
52. Schunkert, H., Konig, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C. *et al.* (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.*, **43**, 333–338.
53. Bilguvar, K., Yasuno, K., Niemela, M., Ruigrok, Y.M., von Und Zu, F.M., van Duijn, C.M., van den Berg, L.H., Mane, S., Mason, C.E., Choi, M. *et al.* (2008) Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat. Genet.*, **40**, 1472–1477.
54. Jarinova, O., Stewart, A.F., Roberts, R., Wells, G., Lau, P., Naing, T., Buerki, C., McLean, B.W., Cook, R.C., Parker, J.S. and McPherson, R. (2009) Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler. Thromb. Vasc. Biol.*, **29**, 1671–1677.
55. Folkersen, L., Kyriakou, T., Goel, A., Peden, J., Malarstig, A., Paulsson-Berne, G., Hamsten, A., Hugh, W., Franco-Cereceda, A., Gabrielsen, A. and Eriksson, P. (2009) Relationship between CAD risk genotype in the chromosome 9p21 locus and gene expression. Identification of eight new ANRIL splice variants. *PLoS One*, **4**, e7677.
56. Cunnington, M.S., Santibanez, K.M., Mayosi, B.M., Burn, J. and Keavney, B. (2010) Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet.*, **6**, e1000899.
57. Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M. and Xiong, Y. (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene*, **30**, 1956–1962.
58. Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P. and Cui, H. (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, **451**, 202–206.
59. Mortensen, U.H., Lisby, M. and Rothstein, R. (2009) Rad52. *Curr. Biol.*, **19**, R676–R677.
60. Khanna, K.K. and Jackson, S.P. (2001) DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet.*, **27**, 247–254.
61. Danoy, P., Michiels, S., Dessen, P., Pignat, C., Boulet, T., Monet, M., Bouchardy, C., Lathrop, M., Sarasin, A. and Benhamou, S. (2008) Variants in DNA double-strand break repair and DNA damage-response genes and susceptibility to lung and head and neck cancers. *Int. J. Cancer*, **123**, 457–463.
62. Liu, J.Z., Tozzi, F., Waterworth, D.M., Pillai, S.G., Muglia, P., Middleton, L., Berrettini, W., Knouff, C.W., Yuan, X., Waeber, G. *et al.* (2010) Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.*, **42**, 436–440.
63. Saccone, N.L., Culverhouse, R.C., Schwantes-An, T.H., Cannon, D.S., Chen, X., Cichon, S., Giegling, I., Han, S., Han, Y., Keskitalo-Vuokko, K. *et al.* (2010) Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.*, **6**, e1001053.

64. Saccone, N.L., Schwantes-An, T.H., Wang, J.C., Grucza, R.A., Breslau, N., Hatsukami, D., Johnson, E.O., Rice, J.P., Goate, A.M. and Bierut, L.J. (2010) Multiple cholinergic nicotinic receptor genes affect nicotine dependence risk in African and European Americans. *Genes Brain Behav.*, **9**, 741–750.
65. Pesch, B., Kendzia, B., Gustavsson, P., Jockel, K.H., Johnen, G., Pohlabeln, H., Olsson, A., Ahrens, W., Gross, I.M., Bruske, I. *et al.* (2012) Cigarette smoking and lung cancer-relative risk estimates for the major histological types from a pooled analysis of case-control studies. *Int. J. Cancer*, **131**, 1210–1219.
66. Postel-Vinay, S., Vanhecke, E., Olaussen, K.A., Lord, C.J., Ashworth, A. and Soria, J.C. (2012) The potential of exploiting DNA-repair defects for optimizing lung cancer treatment. *Nat. Rev. Clin. Oncol.*, **9**, 144–155.
67. Field, J.K., Smith, D.L., Duffy, S. and Cassidy, A. (2005) The Liverpool Lung Project research protocol. *Int. J. Oncol.*, **27**, 1633–1645.
68. Eisen, T., Matakidou, A. and Houlston, R. (2008) Identification of low penetrance alleles for lung cancer: the GEnetic Lung CAncer Predisposition Study (GELCAPS). *BMC Cancer*, **8**, 244.
69. Power, C. and Elliott, J. (2006) Cohort profile: 1958 British birth cohort (National Child Development Study). *Int. J. Epidemiol.*, **35**, 34–41.
70. Sauter, W., Rosenberger, A., Beckmann, L., Kropp, S., Mittelstrass, K., Timofeeva, M., Wolke, G., Steinwachs, A., Scheiner, D., Meese, E. *et al.* (2008) Matrix metalloproteinase 1 (MMP1) is associated with early-onset lung cancer. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 1127–1135.
71. Su, L., Zhou, W., Asomaning, K., Lin, X., Wain, J.C., Lynch, T.J., Liu, G. and Christiani, D.C. (2006) Genotypes and haplotypes of matrix metalloproteinase 1, 3 and 12 genes and the risk of lung cancer. *Carcinogenesis*, **27**, 1024–1029.
72. Landi, M.T., Consonni, D., Rotunno, M., Bergen, A.W., Goldstein, A.M., Lubin, J.H., Goldin, L., Alavanja, M., Morgan, G., Subar, A.F. *et al.* (2008) Environment And Genetics in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. *BMC Public Health*, **8**, 203.
73. 1994) The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann. Epidemiol.*, **4**, 1–10.
74. Hayes, R.B., Sigurdson, A., Moore, L., Peters, U., Huang, W.Y., Pinsky, P., Reding, D., Gelmann, E.P., Rothman, N., Pfeiffer, R.M. *et al.* (2005) Methods for etiologic and early marker investigations in the PLCO trial. *Mutat. Res.*, **592**, 147–154.
75. Calle, E.E., Rodriguez, C., Jacobs, E.J., Almon, M.L., Chao, A., McCullough, M.L., Feigelson, H.S. and Thun, M.J. (2002) The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer*, **94**, 2490–2501.
76. Scelo, G., Constantinescu, V., Csiki, I., Zaridze, D., Szeszenia-Dabrowska, N., Rudnai, P., Lissowska, J., Fabianova, E., Cassidy, A., Slamova, A. *et al.* (2004) Occupational exposure to vinyl chloride, acrylonitrile and styrene and lung cancer risk (europe). *Cancer Causes Control*, **15**, 445–452.
77. Omenn, G.S., Goodman, G., Thornquist, M., Grizzle, J., Rosenstock, L., Barnhart, S., Balmes, J., Cherniack, M.G., Cullen, M.R. and Glass, A. (1994) The beta-carotene and retinol efficacy trial (CARET) for chemoprevention of lung cancer in high risk populations: smokers and asbestos-exposed workers. *Cancer Res.*, **54**, 2038s–2043s.
78. Holmen, J., Midthjell, K., Krüger, O., Langhammer, A., Holmen, T.L., Bratberg, G.H., Vatten, L. and Lund-Larsen, P.G. (2003) The Nord-Trøndelag Health Study 1995–97 (HUNT 2):Objectives, contents, methods and participation. *Norsk Epidemiologi*, **13**, 19–32.
79. Feyler, A., Voho, A., Bouchardy, C., Kuokkanen, K., Dayer, P., Hirvonen, A. and Benhamou, S. (2002) Point: myeloperoxidase -463G –> a polymorphism and lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1550–1554.
80. Nelis, M., Esko, T., Magi, R., Zimprich, F., Zimprich, A., Toncheva, D., Karachanak, S., Piskackova, T., Balascak, I., Peltonen, L. *et al.* (2009) Genetic structure of Europeans: a view from the North-East. *PLoS One*, **4**, e5472.
81. Valk, K., Vooder, T., Kolde, R., Reintam, M.A., Petzold, C., Vilo, J. and Metspalu, A. (2010) Gene expression profiles of non-small cell lung cancer: survival prediction and new biomarkers. *Oncology*, **79**, 283–292.
82. Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H.E. *et al.* (2005) Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.*, **37**, 1243–1246.
83. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
84. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
85. Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009) Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, **10**, 387–406, 387–406.
86. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, 44, 955–959.
87. Aulchenko, Y.S., Ripke, S., Isaacs, A. and van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
88. Higgins, J.P., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *Br Med J*, **327**, 557–560.
89. Altshuler, D., Daly, M.J. and Lander, E.S. (2008) Genetic mapping in human disease. *Science*, **322**, 881–888.
90. Purcell, S. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
91. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
92. Gauderman, W.J. (2002) Sample size requirements for matched case-control studies of gene-environment interaction. *Stat. Med.*, **21**, 35–50.