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

Maria N. Timofeeva¹, Rayjean J. Hung², Thorunn Rafnar³, David C. Christiani⁴, John K. Field⁵, Heike Bickeböller⁶, Angela Risch⁷,⁴¹, James D. McKay², Yufei Wang⁸, Juncheng Dai⁹, Valerie Gaborieau¹, John McLaughlin², Darren Brenner², Steven A. Narod¹⁰, Neil E. Caporaso¹¹, Demetrius Albanes¹¹, Michael Thun¹², Timothy Eisen¹³, H.-Erich Wichmann¹⁴,¹⁵, Albert Rosenberger⁶, Younghun Han¹⁶, Wei Chen¹⁶, Dakai Zhu¹⁶, Margaret Spitz¹⁸, Xifeng Wu¹⁷, Mala Pande¹⁷, Yang Zhao⁴, David Zaridze¹⁹, Neonilia Szeszenia-Dabrowska²⁰, Jolanta Lissowska²¹, Peter Rudnai²², Eleonora Fabianova²³, Dana Mates²⁴, Vladimir Bencko²⁵, Lenka Foretova²⁶, Vladimir Janout²⁷, Hans E. Krokan²⁸, Maiken Elvestad Gabrielsen²⁸, Frank Skorpen²⁹, Lars Vatten³⁰, Inger Njølstad³¹, Chu Chen³², Gary Goodman³², Mark Lathrop³³, Simone Benhamou³⁴, Tönu Voore³⁵, Kristjan Válk³⁶, Mari Nelis³⁷,³⁸, Andres Metspalu³⁷, Olaide Raji³⁵, Ying Chen³⁵, John Gosney⁵, Triantafillos Liloglou⁵, Thomas Muley³⁹,⁴¹, Hendrik Dienemann⁴⁰,⁴¹, Godmar Thorleifsson³, Hongbing Shen⁹, Kari Stefansson³, Paul Brennan¹, Christopher I. Amos¹⁶,⁴²,*,†, Richard Houlston⁸,†, Maria Teresa Landi¹¹,† and for TRICL Research Team³

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

© The Author 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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 GWAS 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 A 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 ($P = 1.8 \times 10^{-55}$) 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 CHRN4 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 ($P = 4.0 \times 10^{-7}$), while rs1051730 remained significant ($P = 2.4 \times 10^{-26}$, Fig. 3) when allele dosage for rs6495309 was included into a model. Two intrinsic 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 allele dosage). Conversely, an analysis adjusting for rs6495309 enhanced their effects consistently across studies ($P = 9.2 \times 10^{-31}$ and $5.1 \times 10^{-32}$ for rs680244 and rs6495306, respectively). No other 15q25.1 variant showed a significant association when allele 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 ($P = 2.8 \times 10^{-82}$, 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
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 loci 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 associated with lung cancer risk with genome-wide significance (Fig. 3). When markers were imputed, the strongest signal was shown for rs3117582 (\( P = 8.2 \times 10^{-18} \)) and rs2523571 (effect allele G, OR = 0.76, \( P = 1.1 \times 10^{-10} \)) 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 loci 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 500 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 × 10^{-10}). However, strong heterogeneity by study (P_{het} = 3.11 × 10^{-26}, 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. Heterogeneity assessed between ever and never smoking groups. NSCLC, non-small-cell lung cancer; SCLC, small-cell lung cancer; LCLC, large-cell lung cancer.
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 (-log10 P) are shown with the recombination rate based on HapMap phase II data. (A) 15q25 locus. Black dots, results (-log10 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 × 10^-8) after controlling for both SNPs with high heterogeneity between studies (Phet = 99.1%). (B) 5p15 locus. Black dots, results (-log10 P) for SNPs genotyped within the region. Blue, results after the inclusion of rs2736100 allele dosage as a covariate. (C) 6p21–6p22 locus. Black dots, results (-log10 P) for SNPs genotyped within the region; green, allele dosage for rs401681 is included as a covariate, CLPTM1L; red, allele dosages for both SNPs are included as a covariate. (D) 15q25 locus. Black dots, results (-log10 P) for SNPs genotyped within the region; green, allele dosage for rs6495309 is included as a covariate; red, allele dosage for both SNPs is included as a covariate.

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 × 10^-7). This association was particularly strong among smokers (OR = 0.92, P = 6.0 × 10^-7) and cases with squamous (OR = 0.87; P = 6.0 × 10^-8) and small-cell carcinoma (OR = 0.85, P = 2.0 × 10^-6; Phet = 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 the DNA repair gene RAD52, provided the strongest association with squamous cell carcinoma (effect allele A, OR = 0.86, P = 3.4 × 10^-8; Supplementary Material, Fig. S4).

The strongest evidence for association at 9p21 was shown with rs1333040, which is located ≈ 3′ to the CYLD gene (22). Rs1333040 also showed an association with squamous cell cancer (effect allele A, OR = 0.86, P = 2.9 × 10^-10). Rs1333040 is located ≈ 1.5 Mb upstream of CDKN1B, which is in strong LD with rs1333040 (r^2 = 0.96, D′ = 1). In an analysis of imputed data from eight studies, the lowest P-value was shown for rs1333040 (effect allele C, OR = 1.06, P = 9.4 × 10^-8) located in the intron between CDKN1B and 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 × 10^-7, Supplementary Material, Fig. S7). This SNP is located 3′ downstream of the NIP3 gene (22). Imputation did not identify any stronger association (Supplementary Material, Fig. S4), and rs11683501 did not show an
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).
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/p16INK4a (cyclin-dependent kinase inhibitors 2A), p14ARF (alternative transcript generated by alternative exon 1 of CDKN2A/p16INK4a) and CDKN2B/p15INK4B (cyclin-dependent kinase inhibitors 2B) (37). CDKN2A/p16INK4a 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 (ANKL/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 < r² < 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/p16INK4a, p14ARF or CDKN2B/p15INK4B genes cannot be excluded. Further studies are needed to evaluate the effect that the SNPs we identified may have on ANRIL/CDKN2B-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/CRBN4 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 haplotype effect (r² = 0.17, D′ = 1.0 between these two SNPs) with an imperfect correlation with an unknown functional variant. We also observed an effect for rs680244/rs6495306 (r² = 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 haplotype 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/CRBN4 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/ANKL), 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 Prediction 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), 18q23 (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 R^2 < 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 ≥ 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).

**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
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/
International Lung Cancer Consortium: http://ilcco.iarc.fr/

**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, 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. Bülzingslöwen, Fr. Schneider, Fachklinik für Atemwegserkrankungen); Essen (Prof. Teschter, Dr Fischer, Ruhrlandklinik—Universitätsklinik, Abt. Pneumologie); Gauting (Prof. Häußlinger, Prof. Thetter, Dr Dühl, Dr Wagner, Pneumologische Klinik München-Gauting); Gera (CA MR Dr Heil, OA Dr Täuscher, OA Dr Lange, II. Medizinische Klinik, Wald-Klinikum Gera); Göttinngen (Prof. Trümper, Prof. Griesinger, Dr Overbeck, Abteilung Onkologie, Hämatologie); Göttinngen (Prof. Schöndube, Dr Danner, Abteilung Thorax-, Herz- und Gefäßchirurgie); Güttlingen/Weende (Dr med. Fleischer, Ev. Krankenhaus Güttlingen-Weende e.V., Abteilung Allgemeinchirurgie); Greifenstein (Prof. Morr, Dr M. Degen, Dr Matter, Pneumologische Klinik, Waldhof Elgershausen); Greifswald (Prof. Ewert, Prof. 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); Immnenhausen (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. Gillissen, OA Dr Cebulla, Städt. Klinikum St Georg, Robert-Koch-Klinik); Leipzig (Kreyimborg, Universitätsklinikum Leipzig, Medizinische Klinik I, Abteilung Pneumologie); Lengern (Prof. Criée, Dr Körber, Dr Knaack, Ev. Krankenhaus Weene e.V., Standort Lengern, 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 Kosi 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 LiUng 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 Helmholtz Gemeinschaft (HGF) and the Federal office for Radiation Protection (BfS: STSch4454). Genotyping was performed in the Genome Analysis Center (GAC) of the Helmholtz Zentrum München. 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 U01 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 Estonian 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 (N01-CN-25514), Georgetown University (N01-CN-25522), Pacific Health Research Institute (N01-
REFERENCES


