Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer

Yufei Wang,1,61 James D McKay,2,61,62 Thorunn Rafnar,3 Zhaoming Wang,4 Maria N Timofeeva,2 Peter Broderick,1 Xuchen Zong,5 Marina Laplana,6 Yongyue Wei,7 Younghun Han,8 Amy Lloyd,1 Manon Delahaye-Soudeix,2 Daniel Chubb,1,7 William Wheeler,8,9 Nilanjan Chatterjee,4 Guðmar Thorleifsson,3 Patrick Sulem,3 Geoffrey Liu,10 Rudolf Kaaks,11,12 Marc Henrion,1 Ben Kinnersley,1 Maxime Vallée,2 Florence Le Calvez-Kelm,2 Victoria L Stevens,13 Susan M Gapstur,13 Wei V Chen,14 David Zaridze,15 Neonilia Szeszenia-Dabrowska,16 Jolanta Lissowska,17 Peter Rudnai,18 Eleonora Fabianova,19 Dana Mates,20 Vladimir Bencko,21 Lenka Foretova,22 Vladimir Janout,23 Hans E Krokan,24 Maiken Elvestad Gabrielsen,24 Frank Skorpen,25 Lars Vatten,26 Inger Njølstad,27 Chu Chen,28 Gary Goodman,28 Simone Benhamou,29 Tonu Voode,30 Kristjan Välk,21 Mari Nelis,32,33 Andres Metspalu,32 Marcin Lener,34 Jan Lubinski,34,35 Mattias Johansson,35 Paolo Vineis,35,36 Antonio Agudo,37 Francoise Clavel-Chapelon,38–40, H Bas Bueno-de-Mesquita,35,41,42 Dimitrios Trichopoulou,43–45, Kay-Tee Khaw,46,47 Mikhail Johansson,47 Elisabete Weiderpass,48–51, Anne Tjønneland,52,53 Eli Riboli,53,54 Mark Ladbrooke,55 Ghislaine Scelo,56 Demetrius Albanes,57 Neil E Caporaso,58 Yuntao Song,59,60 Jian Gu,54,55 Xifeng Wu,54,61 Margaret R Spitz,55 Hendrik Diemmann,12,62,63 Albert Rosenberg,57 Li Su,7 Athena Matakidou,58,59 Timothy Eisen,59,60,64 Kari Stefansson,3,65 Angela Risch,6,12,66 Stephen J Chanock,4,67 David C Christiani,7 Rayjean J Hung,5 Paul Brennan,5 Maria Teresa Landi,61,62,67 Richard S Houlston,1,61,62,68 & Christopher I Amos,6,68

We conducted imputation to the 1000 Genomes Project of four genome-wide association studies of lung cancer in populations of European ancestry (11,348 cases and 15,861 controls) and genotyped an additional 10,246 cases and 38,295 controls for follow-up. We identified large-effect genome-wide associations for squamous lung cancer with the rare variants BRCA2 p.Lys326X (rs11571833, odds ratio (OR) = 2.47, P = 4.74 × 10−20) and CHEK2 p.Ile157Thr (rs17879961, OR = 0.38, P = 1.27 × 10−13). We also showed an association between common variation at 3q28 (TP63, rs13314271, OR = 1.13, P = 7.22 × 10−10) and lung adenocarcinoma that had been previously reported only in Asians. These findings provide further evidence for inherited genetic susceptibility to lung cancer and its biological basis. Additionally, we further demonstrated that imputation can identify rare disease-causing variants with substantive effects on cancer risk from preexisting genome-wide association study data.

Lung cancer causes over 1 million deaths each year worldwide. Although primarily caused by tobacco smoking, studies have also implicated inherited genetic factors in the etiology of lung cancer; notably, genome-wide association studies (GWAS) in Europeans have consistently identified polymorphic variation at 15q25 (CHRNA5-CHRNA3-CHRNB4), 5p15.33 (TERT-CLPTM1) and 6p21.33 (BAG6) as determinants of lung cancer risk. Additionally, susceptibility loci for lung cancer at 3q28, 6q22.2, 13q12.12, 10q25.2 and 22q11.2 in Asians have been identified through GWAS.

Non-small cell lung cancer (NSCLC) is the most common lung cancer histology, comprised primarily of adenocarcinoma (AD) and squamous cell carcinoma (SQ). These lung cancer histologies have different molecular characteristics that reflect differences in etiology and carcinogenesis. Perhaps not surprisingly, there is variability in the genetic effects on lung cancer risk by histology, with subtype-specific associations at 5p15.33 (TERT-CLPTM1) for AD and 6p21.33 associations for SQ than for AD.

To identify additional lung cancer susceptibility loci, we conducted a meta-analysis of four lung cancer GWAS in populations of European ancestry: the MD Anderson Cancer Center (MDACC) GWAS, the Institute of Cancer Research (ICR) GWAS, the National Cancer Institute (NCI) GWAS and the International Agency for Research on Cancer (IARC) GWAS (Online Methods), which were genotyped using Illumina HumanHap 317, 317+240s, 370Duo, 550, 610 or 1M arrays (Supplementary Table 1). After filtering, the studies provided genotypes on 11,348 cases and 15,861 controls (Supplementary Table 1). Before undertaking meta-analysis of the GWAS data, we searched for potential errors and biases in the data sets. Quantile-quantile (Q-Q) plots of genome-wide association test statistics showed minimal inflation,
rendering substantial cryptic population substructure or differential genotype calling between cases and controls unlikely (λ = 1.01–1.05; Supplementary Fig. 1). To bring genotype data obtained from different arrays into a common platform and recover untyped genotypes, we imputed >10 million SNPs using 1000 Genomes Project data as the reference. Q-Q plots for all SNPs and those restricted to rare SNPs (minor allele frequency (MAF) <1%) after imputation did not show evidence of substantive overdispersion introduced by imputation (λ = 0.99–1.06 and λ = 0.82–1.05, respectively; Supplementary Fig. 1).

Pooling data from each GWAS, we derived joint ORs and 95% confidence intervals (CIs) under a fixed-effects model for each SNP and the associated per-allele P values. To explore variability in associations according to tumor histology, we derived ORs for all lung cancer, AD and SQ.

Our meta-analysis identified 50 SNPs that showed evidence of association with lung cancer, AD or SQ (P < 5.0 × 10^-6; Fig. 1) at loci not reported previously in Europeans (Fig. 1). We evaluated 1-Mb regions encompassing these 50 SNPs for association through in silico replication in the Harvard15 and deCODE16 series. Nine of the SNPs within these 50 regions showed support for an association (combined P < 5.0 × 10^-7). We attempted genotyping of these nine SNPs in four additional series: the Heidelberg–European Prospective Investigation into Cancer and Nutrition (EPIC), ICR, IARC and Toronto replications (Supplementary Table 2b and Online Methods). rs185577307 could not be genotyped because of repetitive sequence. Collectively, genotypes were available from 21,594 cases and 54,156 controls, providing 80% power to detect a variant with MAF of 0.01 and conferring a relative risk of ≥1.5. In the combined analysis of all GWAS plus replication series data, SNPs mapping to 13q13.1 and 22q12.1; the associations remained statistically significant after adjustment for multiple testing (P < 3.0 × 10^-6; Fig. 2 and Supplementary Table 3). A stronger role for BRCA2 in SQ etiology than in AD etiology is reflected in the higher observed mutational frequency in the respective lung cancers (~6% and 1% (refs. 18,19)). Thr9976 was recently shown to confer a 1.26-fold increased breast cancer risk20 and has been suggested previously as a risk factor for esophageal and pancreatic cancers21,22. We found no evidence for an association between Thr9976 and lung cancer risk in nonsmokers using directly genotyped samples (Supplementary Table 2); however, these cases comprised <10% of each cohort, and therefore our power to demonstrate a relationship was limited. Previous analyses of families carrying highly penetrant BRCA2 mutations have found either no evidence for any excess risk or a reduced risk of lung cancer in carriers23,24.

A possible explanation for these observations is that members of the families studied tended to smoke less than the general population24.

The RAD51–BRCA2 interaction is pivotal for BRCA2-mediated double strand–break repair, and exon 27 of BRCA2 encodes one of the highly conserved RAD51 binding domains: homozygous deletion of exon 27 in mice confers susceptibility to tumors, including lung cancer25. Thr9976 leads to the loss of the C-terminal domain of BRCA2, inviting speculation that the SNP is functional. Although the deleted region is distal to the RAD51 binding domain and an impact on nuclear localization is unknown26,27, the nearby BRCA2 p.Thr3387Ala alteration interrupts CHK2 phosphorylation and abrogates BRCA2-CHK2-RAD51–mediated recombination repair28. Alternatively, the association might be a consequence of linkage disequilibrium (LD) with another BRCA2 mutation. Studies of families with breast cancer of northern European ancestry show that the BRCA2 c.6275delTT and c.4889C>G mutations, which are highly...
penetrant for breast and ovarian cancer, originated on a p.Lys3326X haplotype. To gain further insight into a probable genetic basis of the 13q13.1 lung cancer association, we sequenced germline DNA from 70 individuals with lung cancer who carried c.9976A>T from the UK Genetic Lung Cancer Predisposition Study for the c.6275delTT and c.4889C>G mutations; we did not find c.6275delTT or c.4889C>G in any of these individuals. Similarly, sequencing the coding region of BRCA2 identified no clearly pathogenic mutations among 13 individuals from the 1958 British Birth Cohort (58BC), 11 individuals with lung cancer from IARC or 24 individuals with lung cancer carrying Thr9976 from TCGA. In Iceland, Thr9976 is not correlated with the founder BRCA2 mutation resulting in p.256_257del (c.999del5), which greatly increases the risk of breast and ovarian cancer. Paradoxically, whereas Thr9976 is a risk factor for lung cancer, in this population this SNP is not associated with risk of breast or ovarian cancer (Supplementary Table 5). Although in vitro studies

Figure 2 Plots of the ORs of lung cancer associated with 13q13.1 (rs11571833 and rs56084662), 22q12.1 (rs17879961) and 3q28 (rs13314271) risk loci. (a–l) All lung cancer based on 21,594 lung cancer cases and 54,156 controls (a–d), SQ based on 6,477 SQ cases and 53,333 controls (e–h) and AD based on 7,031 AD cases and 53,189 controls (i–l). The studies are weighted according to the inverse of the variance of the log of the OR calculated by unconditional logistic regression. Horizontal lines indicate the 95% CIs. Boxes are the OR point estimates, and the area of the box is proportional to the weight of the study. Diamonds and broken lines indicate the overall summary estimate derived under a fixed-effects (FE) model, with the CI given by the width. Unbroken vertical lines show the null value (OR = 1.0).
Figure 3 Regional plots of associations at susceptibility loci for SQ and AD. (a–c) Association results and recombination rates for the 13q13.1 in SQ (a), 22q12.1 in SQ (b) and 3q28 in AD (c). The SQ-related plots (a, b) were based on 3,275 SQ cases and 15,038 controls from the discovery phase; the AD-related plot (c) was based on 3,442 AD cases and 14,894 controls from the discovery phase. Association results of both genotyped (circles) and imputed (diamonds) SNPs in the GWAS samples and recombination rates for each locus are shown. For each plot, $-\log_{10} P$ values (y axes) of the SNPs are shown according to their chromosomal positions (x axes). The top genotyped SNP in each combined analysis is indicated by a large diamond and is labeled by its rsID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP: white ($r^2 = 0$) through to dark red ($r^2 = 1.0$). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association. Genes have been redrawn to show the relative positions; therefore, maps are not to physical scale.

have failed to demonstrate that p.Lys3326X affects DNA repair\(^\text{30}\), our findings raise the possibility that p.Lys3326X may have a direct effect on lung cancer risk. The fact that somatic mutation of BRCA2 is not associated with p.Lys3326X carrier status\(^\text{15}\) (Supplementary Table 6a) suggests that any impact the SNP has on lung cancer risk is mediated through alternative mechanisms.

The relationship at 22q12.1 between the rs17879961 (c.470T>C) and SQ in the combined series (OR = 0.38, $P = 1.27 \times 10^{-13}$) validates an association that has been reported previously\(^\text{31,32}\) (Fig. 2 and Supplementary Tables 3 and 4). The frequency of rs17879961 varies markedly between populations: it has a MAF of ~5% in eastern Europeans (for example, individuals in the IARC series) but is almost monomorphic in most northern Europeans. This likely accounts for the failure to demonstrate a significant relationship in the ICR, MDACC, Toronto and deCODE series, which comprise largely western European populations (Fig. 2 and Supplementary Table 3). rs17879961 is responsible for the missense mutation in CHEK2 resulting in p.Ile157Thr; CHEK2 is a cell cycle–control gene encoding a pluripotent kinase that can cause arrest or apoptosis in response to DNA damage. Acquired mutation of CHEK2 is rarely seen in lung cancer, and the CHEK2 p.Ile157Thr alteration does not appear to correlate with mutation (Supplementary Table 6a), raising the possibility that carrier status per se influences cancer risk. The p.Ile157Thr substitution lies in a functionally important domain of CHEK2 and causes reduced or abolished binding of principal substrates. Although Cys470 increases breast cancer risk\(^\text{33}\), here Cys470 was associated with reduced lung cancer risk. A mechanism for the paradoxical associations is not immediately apparent. However, CHEK2 can have opposite effects on damaged stem cells, retarding stem cell division until DNA damage is repaired or activating apoptosis if damage cannot be repaired. Although speculative, in the presence of continued DNA damage to squamous epithelia by tobacco smoke, the normal stem cell defenses involving CHEK2 might be attenuated by a reduction in CHEK2 activity as a result of p.Ile157Thr\(^\text{34}\). Concordant with such a model is our observation of a paradoxically increased lung cancer risk in nonsmokers ($P = 0.05$) and in correlated subgroups of AD and women, although this increase was based on small numbers (Supplementary Table 2).

The association between variation at 3q28 marked by rs13314271 and lung cancer risk was restricted to AD (OR = 1.13, $P = 7.22 \times 10^{-10}$, Fig. 2 and Supplementary Table 3). rs13314271 maps within intron 1 of TP63 (Fig. 3). Variation at TP63 defined by the intron 1 SNP rs4488809, which is in complete LD with rs13314271 ($r^2 = 1.00$, $D’ = 1.00$), is associated with AD in Asians\(^\text{8}\). Our findings provide robust evidence for the generalizability of a relationship between 3q28 variation and AD. We found a weak association between rs13314271 and lung cancer risk in nonsmokers ($P = 0.03$; Supplementary Table 2b). TP63 is a member of the tumor suppressor TP53 gene family, which is pivotal in cellular differentiation and responsiveness to cellular stress\(^\text{34,35}\). Exposure of cells to DNA damage leads to induction of TP63, and both isoforms have the ability to transactivate TP53 target genes, thereby affecting cellular responsiveness to DNA damage\(^\text{36}\). Although rs13314271 does not map to an evolutionary conserved region, rs7636839, which is correlated with rs13314271 and rs4488809 ($r^2 = 1.0$), does map to an evolutionary conserved region and has predicted enhancer activity (Supplementary Table 6b). Moreover, rs4488809 has been shown to be an expression quantitative trait locus for TP63 in lung tissue\(^\text{37}\). Although the mechanism by which 3q28 variation affects AD development is unknown, accumulation of DNA damage and a lack of response to genotoxic stress are recognized to contribute to lung carcinogenesis; hence, loss of repair fidelity as a consequence of differential TP63 expression is likely deleterious.

There was no association between rs11571833, rs17879961 and rs13314271 genotypes and cigarette consumption on the basis of smoking information on 43,693 Icelandic subjects (Supplementary Table 7), which is in contrast to the association of 15q25 and risk of lung cancer. Although there is some overlap, distinct DNA lesions are ostensibly repaired by different DNA repair pathways. Histology-specific relationships seen implicate the BRCA2-CHEK2-RAD52 double strand–break repair and homologous recombination pathways as a determinant of SQ and defective TP53 and TERT apoptosis–telomerase regulation as a basis of AD risk.
In conclusion, our findings provide further evidence for inherited genetic susceptibility to lung cancer and underscore the importance of searching for histology-specific risk variants. Our data also provide an important proof of principle that 1000 Genomes imputation can be used to detect new, low-frequency, large-effect associations, thereby extending the utility of preexisting GWAS data. Notably, this study facilitated the identification of BRCA2 Thr9976, which is the strongest genetic association in lung cancer reported so far. For a smoker carrying this variant (2% of the population), the risk of developing lung cancer is approximately doubled, which may have implications for identifying high-risk ever-smoking subjects for lung cancer screening. Additionally, future study of the effects of PARP inhibition in smokers with lung cancer carrying BRCA2 Thr9976 may be warranted.

URLS. R suite, http://www.r-project.org/; 1000 Genomes Project, http://www.1000genomes.org/; SNAP, http://www.broadinstitute.org/mpg/snap/; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute_v2.html; MACH, http://www.sph.umich.edu/csg/abecasis/MACH/; Minimac, http://genome.sph.umich.edu/wiki/Minimac/; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snpptest/snptest.html; ProbABEL, http://www.genabel.org/packages/ProbABEL/; mach2dat, http://genome.sph.umich.edu/wiki/Mach2mdat:_Association_with_MACH_output; Wellcome Trust Case Control Consortium, http://www.wtccc.org.uk/; RegulomeDB, http://regulome.stanford.edu/; HaploReg v2, http://www.broadinstitute.org/mammals/haploreg/haploreg.php; Transdisciplinary Research In Cancer of the Lung (TRICL), http://u19tricl.org/; Genetic Associations and Mechanisms in Oncology (GAME-ON) consortium, http://epi.grants.cancer.gov/gameon/; International Lung Cancer Consortium (ILCCO), http://ilcco.iarc.fr/; Icelandic Cancer Registry, http://www.krabbameinsskra.is/; Genome Analysis Toolkit (GATK), http://www.broadinstitute.org/gatk/; The Cancer Genome Atlas (TCGA), http://cancergenome.nih.gov/; Leiden Open Variation Database (LOVD), http://www.lonv.nl/3/home/; Breast Cancer IARC database, http://brca.iarc.fr/.

METHODS

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

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all individuals who participated in this study. We are also grateful to the patients, clinicians and allied health care professions. We thank Z. Chen and K. Boyle for sample handling and data management of the Toronto study, and L. Admas and L.R. Zhang for field recruitment. We thank L. Su, Y. Zhao, G. Liu, J. Wain, R. Heist and K. Asomaning for providing computing support at MDACC. We thank G. Thomas and Synergy Lyon Cancer (Lyon France) for high performance computing support and J. Olivier and A. Chabrier for IARC’s PGM frame of Centre of Excellence in Genomics and Estonian Research Infrastructure’s Roadmap and the University of Tartu (SPIVARENG) and an IARC Postdoctoral Fellowship (M.N.T.). Work at the NCI was supported by the Intramural Research Program of the NCI, the NCI, US Public Health Service contracts N01-RC-45139, N01-RC-45136, N01-RC-45137, N01-RC-45140, N01-RC-45141, N01-CN-25151, N01-CN-25512, N01-CN-25513, N01-CN-25516, N01-CN-25511, N01-CN-25524, N01-CN-25518, N01-CN-75022, N01-CN-25476 and N01-CN-25404), the American Cancer Society, the NIH Genes, Environment and Health Initiative in part by HG-06-033-NCI-01 and RO1HL091172-01, genotyping at the Johns Hopkins University Incenter for Inherited Disease Research (U01HG004438 and NIH HHSN2652007028966C) and study coordination at the GENEVA Coordination Center (U01HG004466). Work was also supported by NIH grants RO1CA121197, R01CA127219, U19CA148127 and RO1CA55769 and a Cancer Prevention Research Institute of Texas grant (RP100443). Genotyping was provided by the Center for Inherited Disease Research (CIDR). Work performed at Harvard was supported by the NIH (CA074386, CA092824 and CA090578). The Icelandic study was supported in part by NIH DA17932.

AUTHOR CONTRIBUTIONS

R.S.H. and Y. Wang conceived the study and provided overall project management and drafted the NCI (in the UK, Y. Wang performed statistics and bioinformatics of UK data and conducted all meta-analyses; additional support was provided by M.H.; P. Broderick oversaw genotyping and sequencing; A.L. and B.K. performed genotyping and Sanger sequencing; A. Matakidou, T.E. and R.S.H. were responsible for the development and operation of the Genetic Lung Cancer Predisposition Study (GELCAPS); and D.C. and P. Broderick performed next-generation sequencing. At IARC, J.D.M. and P. Brennan provided overall project management; M.N.T., M.D.-S., V.G. and M.V. performed statistics and bioinformatics of IARC data and conducted meta-analysis; J.D.M. and E.L.C.-K. oversaw genotyping and sequencing; and G.S., D.Z., N.S., D.J., J. Lisowski, P.R., E.E., D.M., Y.B., L.F., V.I., H.E.K., M.E.G., E.S., L.V., I.N., C.C., G.G., M. Lathrop, S.B., T.V., K.V., M.N., A. Metspalu, M. Lathrop, J. Lubiński, Mattias Johannson, P.V., A.A., F.C.-C., H.B.-d.-M., D.T., K.-T., Mikkel Johannson, E.W., A.T., R.K. and E.R. provided samples and data. For the Dartmouth and MDACC component, C.L.A. provided overall project management, obtained funding and contributed to statistical analyses; W.V.C. performed imputation analysis; Y.H. performed statistical analyses; and M.R.S. oversaw sample collection and development of the epidemiological studies. M.R.S. was also responsible for collecting samples that are a part of this research. X.W. provided ongoing support for the research protocol and supported large laboratory management of samples. Y.T. and J.G. performed genotyping. At the NCI, M.T.L. was responsible for the overall project and managed the Environment and Genetics in Lung Cancer Etiology (EAGLE) study; N.E.C. managed the Prostate, Lung, Colon, Ovary Screening Trial (PLCO) study; D.A. managed the α-Tocopherol, β-Carotene Cancer Prevention Study (ATBC); S.M.G. and V.L.S. managed the Cancer Prevention Study II Nutrition Cohort (CPS-II) study; N.C. and W.W. performed statistical analyses; Z.W. performed genotyping and imputation analysis; and S.J.C. oversaw genotyping and imputation analysis. At decode, T.R. and K.S. were responsible for the development and operation of deCODE’s lung cancer study; and G.T. and P.S. contributed to the imputations and statistical analysis of the Icelandic data. At Harvard, D.C.C. was responsible for the overall conduct of the project; L.S. was responsible for sample management, genotyping and laboratory quality control; and Y. Wei performed data management and statistical analyses. For the Heidelberg-EPIC replication, M. Laplana managed DNA samples and performed genotyping; A. Rosenberger managed genotype and phenotype information; A. Risch supervised genotyping and data analysis; and R.K., A. Risch and H.D. conceived and managed studies that contributed samples. For the Toronto replication, R.J.H. and G.L. provided.
overall supervision of the study conduct, including study design, field recruitment, genotyping and statistical analysis; and X.Z. performed the statistical analysis.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/ reprints/index.html.

1. Ferlay, J. et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer 127, 2893–2917 (2010).
2. Hung, R.J. et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452, 633–637 (2008).
3. Amos, C.I. et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat. Genet. 40, 616–622 (2008).
4. Thorgeirsson, T.E. et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452, 638–642 (2008).
5. McKay, J.D. et al. Lung cancer susceptibility locus at 5p13.33. Nat. Genet. 40, 1404–1406 (2008).
6. Wang, Y. et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat. Genet. 41, 382–385 (2010).
7. Hu, Z. et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat. Genet. 43, 792–796 (2011).
8. Mike, J. et al. Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. Nat. Genet. 42, 893–896 (2010).
9. Lan, Q. et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nat. Genet. 44, 1330–1335 (2012).
10. Travis, W.D. et al. International association for the study of lung cancer/American Thoracic Society/European Respiratory Society: international multidisciplinary classification of lung adenocarcinoma: executive summary. Proc. Am. Thorac. Soc. 8, 381–385 (2011).
11. Broncken, P. et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. Cancer Res. 69, 6633–6641 (2009).
12. Landi, M.T. et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am. J. Hum. Genet. 83, 790–799 (2008).
13. Timofeeva, M.N. et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14,900 cases and 29,485 controls. Cancer Res. 65, 417–426 (2005).
14. Brennan, P. et al. Uncommon CHEK2 missense variant and reduced risk of tobacco-related cancers: case-control study. Hum. Mol. Genet. 16, 1794–1801 (2007).
15. Dubyaki, C. et al. Constitutional CHEK2 mutations are associated with a decreased risk of lung and laryngeal cancers. Carcinogenesis 29, 762–765 (2008).
16. Han, F.F., Guo, C.L. & Liu, L.H. The effect of CHEK2 variant 11S7 on cancer susceptibility: evidence from a meta-analysis. DNA Cell Biol. 32, 325–335 (2013).
17. Flores, E.R. The roles of p63 and p73 in cancer. Cell Cycle 6, 300–304 (2007).
18. Tomlinson, I.P. et al. Germline TP53 mutations and the risk of esophageal squamous cell carcinoma. Oncogene 27, 1290–1296 (2008).
19. Martin, S.T. et al. Increased prevalence of the BRCA2 polyomorphism stop-mutation K3326X among individuals with familial pancreatic cancer. Oncogene 24, 3652–3656 (2005).
20. Breast Cancer Linkage Consortium. Cancer risks in BRCA2 mutation carriers. J. Natl. Cancer Inst. 91, 1310–1316 (1999).
21. Akbari, M.R. et al. Germline BRCA2 mutations and the risk of esophageal squamous cell carcinoma. Oncogene 27, 519–525 (2012).
22. Michaelidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45, 353–361 (2013).
23. Merck, K. et al. Nuclear localization signals of the BRCA2 protein. Biochem. Biophys. Res. Commun. 270, 171–175 (2000).
24. Mieghem, P.V. et al. The checkpoint kinases Chk1 and Chk2 regulate the functional associations between hBRCA2 and Rad51 in response to DNA damage. Oncogene 17, 3979–3985 (2008).
25. Mazzocca, A. et al. A polyomorphism stop codon in BRCA2. Nat. Genet. 14, 253–254 (1996).
26. Wu, K. et al. Functional evaluation and cancer risk assessment of BRCA2 unclassified variants. Cancer Res. 65, 3825–3829 (2005).
27. Mankowski, L. et al. Germline TP53 mutations and the risk of hereditary breast cancer. J. Natl. Cancer Inst. 99, 1677–1680 (2007).
28. Pettinato, A. et al. Properties of the six isoforms of p63: p53-like regulation in response to genotoxic stress and cross talk with ΔNp73. Carcinogenesis 29, 273–281 (2008).
29. Hao, K. et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. Nat. Genet. 40, 314–320 (2012).

NATURE GENETICS VOLUME 46 | NUMBER 7 | JULY 2014
ONLINE METHODS

Studies. 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 is associated with the International Lung Cancer Consortium (ILCOCO). Tumors from patients were classified as AD, SQ, large-cell carcinoma (LCC), mixed adenocarcinoma (MADSQ) and other NSCLC histologies following either the International Classification of Diseases for Oncology (ICD-O) or WHO coding. Tumors with overlapping histologies were classified as mixed.

Ethics. All participants provided informed written consent. All studies were reviewed and approved by institutional ethics review committees at the involved institutions.

GWAS. The meta-analysis was based on data from four previously reported lung cancer GWAS of European populations: the MDACC GWAS, the ICR GWAS, the NCI GWAS and the IARC GWAS. In each of the studies, SNP genotyping had been performed using Illumina HumanHap 317, 317+2405, 370, 550, 610 or 1M arrays (Supplementary Table 1).

IARC GWAS. The IARC GWAS comprised 3,062 lung cancer cases and 4,455 controls derived from five case-control studies: (i) the Carotene and Retinol Efficacy Trial (CARET) cohort, (ii) the Central Europe multicenter hospital-based case-control study, (iii) the hospital-based case-control study from France, (iv) the hospital-based case-control lung cancer study from Estonia, and (v) the population-based HUNT2/Tromsø IV lung cancer studies. Patient and control DNAs were derived from EDTA–venous blood samples. The patients with lung cancer were classified according to ICD-O-3: SQ: 8070/3, 8071/3, 8072/3, 8074/3; AD: 8140/3, 8250/3, 8260/3, 8310/3, 8480/3, 8560/3, 8251/3, 8490/3, 8570/3, 8574/3; with tumors with overlapping histologies being classified as mixed. After applying standardized quality-control procedures, 2,533 cases and 3,791 controls were included in the current analysis (Supplementary Table 1).

NCI GWAS. Details of the NCI GWAS have been reported previously. Briefly, the study comprised samples from four series: (i) the Environment and Genetics in Lung cancer Etiology (EAGLE) study, a population-based case-control study of 2,100 lung cancer cases and 2,120 healthy controls enrolled in Italy between 2002 and 2005 (ref. 44), in which cancers were classified according to the ICD-O coding for histology and grading and histology of ~10% of tumors was confirmed by an independent pathologist from the NCI; (ii) the hospital-based case-control study from France, (iii) the hospital-based case-control lung cancer study from Estonia, and (v) the population-based HUNT2/Tromsø IV lung cancer studies. Patient and control DNAs were derived from EDTA–venous blood samples. The patients with lung cancer were classified according to ICD-O-3: SQ: 8070/3, 8071/3, 8072/3, 8074/3; AD: 8140/3, 8250/3, 8260/3, 8310/3, 8480/3, 8560/3, 8251/3, 8490/3, 8570/3, 8574/3; with tumors with overlapping histologies being classified as mixed. After applying standardized quality-control procedures, 2,533 cases and 3,791 controls were included in the current analysis (Supplementary Table 1).

ICR GWAS. The ICR GWAS comprised 1,952 cases (1,166 male; mean age at diagnosis 57 years, s.d. 6 years) with pathologically confirmed lung cancer ascertainment through the Genetic Lung Cancer Predisposition Study (GELCAPS) conducted between March 1999 and July 2004 (ref. 48). All cases were British residents and were self-reported to be of European ancestry. To ensure that data and samples were collected from bona fide lung cancer cases and avoid issues of bias from survivorship, only incident cases with histologically or cytologically (if not AD) confirmed primary disease were ascertained. Tumors from patients were classified according to ICD-O-3: specifically, SQ: 8070/3, 8071/3, 8072/3, 8074/3; AD: 8140/3, 8250/3, 8260/3, 8310/3, 8480/3, 8560/3, 8251/3, 8490/3, 8570/3, 8574/3; with tumors with overlapping histologies being classified as mixed. Patient DNA was derived from EDTA–venous blood samples using conventional methodologies. Genotype frequencies were compared with publicly accessible data generated by the UK Wellcome Trust Case-Control Consortium 2 (WTCCC2) study49 of individuals from the 1958 British Birth Cohort (58BC), and blood service was typed using Illumina Human 1M-Duo Custom_v1 Array BeadChips.

Quality control of GWAS data sets. Standard quality control was performed on all scans, excluding individuals with low call rate (<90%) and extremely high or low heterozygosity (P < 1.0 × 10⁻⁷), as well as all individuals evaluated to be of non-European ancestry (using the HapMap version 2 CEU, JPT/CHB and YRI populations as a reference; Supplementary Table 1). For apparent first-degree relative pairs, we removed the control from a case-control pair; otherwise, we excluded the individual with the lower call rate.

Replication series. To validate promising associations from the meta-analysis, we made use of in silico data and imputed genotypes from Harvard and deCODE GWAS data sets together with data from the direct-genotyping Heidelberg-EPIC, ICR, IARC and Toronto replication series.

Harvard. For the Harvard Lung Cancer Susceptibility Study, details of participant recruitment have been described previously. Replication was based on data derived from 1,000 cases and 1,000 controls genotyped using Illumina HumanHap610–Quad arrays. Cases were patients aged >18 years with newly diagnosed, histologically confirmed primary NSCLC. Controls were healthy non–blood related family members and friends of patients with cancer or with cardiovascular conditions undergoing surgery. The histological classification of lung tumors was performed by two staff pulmonary pathologists at Massachusetts General Hospital according to ICD-O-3: specifically, AD: 8140/3, 8250/3, 8260/3, 8310/3, 8480/3 8560/3, LCC: 8012/3, 8031/3; SQ: 8070/3, 8071/3, 8072/3, 8074/3; other NSCLC: 8010/3, 8020/3, 8021/3, 8032/3, 8230/3. Unqualified samples were excluded if they fit the following quality-control criteria: (i) overall genotype completion rates <95%; (ii) gender discrepancies; (iii) unexpected duplicates or probable relatives (based on a pairwise identity-by-state value of PI_HAT in PLINK >0.185); (iv) heterozygosity rates >6 times the s.d. from the mean; or (v) individuals evaluated to be of non-European ancestry (using HapMap release 23 including the JPT, CEPH, CEU and YRI populations as a reference). Unqualified SNPs were excluded when they fit the following quality-control criteria: (i) SNPs were not mapped on autosomes; (ii) SNPs had a call rate <95% in all GWAS samples; (iii) SNPs had MAF <0.01; or (iv) the genotype distributions of SNPs deviated from those expected by Hardy-Weinberg equilibrium.
(P < 1.0 × 10⁻⁶). After applying these prespecified quality controls, genotype data were available for 984 cases and 970 controls.

decode. The Icelandic lung cancer study has been described previously⁴. The primary source of information on the Icelandic lung cancer cases is the Icelandic Cancer Registry (ICAR), which covers the entire population of Iceland (http://www.cancerregistry.is/krabbameinska/indexen.jsp?id=summary). The sources of data in the ICAR are all pathology and hematology laboratories and all hospital departments and health care facilities in the country. ICAR registration is based on the ICD system and includes information on histology (systemized nomenclature of medicine, SNOMED). ICAR registration also uses the ICD-O system, which takes histology diagnosis into account. Over 94% of diagnoses in the ICAR have histological confirmation. Briefly, according to the ICAR, a total of 4,252 patients were diagnosed with lung cancer from January 1, 1955 to December 31, 2010. Recruitment of both prevalent and incident cases was initiated in 1998, the recruitment is ongoing and DNA samples from lung cancer cases are subjected to whole-genome genotyping as they are collected. The controls used in this study consisted of individuals from other GWAS that were age and sex matched to cases, with no individual disease group accounting for >10% of all controls. Samples were assayed with the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips at deCODE genetics. SNPs were excluded if they had (i) a yield <95%, (ii) MAF < 1% in the population, (iii) deviation from Hardy-Weinberg equilibrium (HWE; P < 10⁻⁶), (iv) inheritance error rate (>0.001) or (v) if there was a substantial difference in allele frequency between chip types (in which case the SNP was removed from a single chip type if that resolved the difference, but if it did not then the SNP was removed from all chip types). All samples with a call rate of <97% were removed from the analysis. The Icelandic sample set is drawn from the Icelandic population, which is a small homogeneous founder population with almost no detectable population substructure. Thus, there was no need to adjust for such substructure in the association analysis. In addition, the comprehensive Icelandic genealogy database allowed us to exclude individuals not of Icelandic origin from the analysis. SNP genotypes were phased using the method of long-range phasing⁵¹; for the HumanHap series of chips, 304,937 SNPs were used for long-range phasing, whereas for the Omni series of chips, 564,196 SNPs were used. An initial imputation step was carried out on each chip series separately to create a single harmonized, long-range phased genotype data set consisting of 707,525 SNPs for 95,085 Icelandic individuals. Two sets of genotypes were imputed into this data set with methods previously described⁵¹: (i) genotypes for about 38 million variants using the 1000 Genomes phase I integrated variant set (v3) as training set and (ii) genotypes for about 34 million variants identified in 2,230 whole genome–sequenced Icelanders. The first set of imputed genotypes was used for replicating the association with variants in the 3p15.33, 2p21 and 12q13.33 regions using IMPUTE (v2.1.1)⁵¹ to perform the case-control analysis. The second set was used when testing the relationship between the p.Lys3326X and c.999del5 genotypes and risk of different cancer types in the Icelandic population using a method that allowed including individuals that had not been chip typed but for whom genotype probabilities were imputed using methods of familial imputation⁵¹.

Heidelberg-EPIC. This study comprised 1,253 Heidelberg-EPIC controls and 1,362 lung cancer cases from the Heidelberg lung cancer study recruited between 1994 and 1998 and between 1996 and 2007, respectively. Details of the Heidelberg-EPIC controls and the Heidelberg lung cancer study have been described previously⁵⁴,⁵⁵. All subjects were aged 18 years or older, and information on lifestyle risk factors and medical and family history was collected through interviews based on standardized questionnaires. The EPIC Lung and the Heidelberg-EPIC studies were performed independently with no sample overlap with those analyzed as part of the IARC replication series. Histological classification of tumors was obtained from pathology reports, where it was recorded by a staff pulmonary pathologist according to WHO guidelines. Blood samples from patients with malignant lung disease categorized as follows were included: AD, SCLC, NSCLC, LCC, carcinoid, mixed lung tumors or mixed without SCLC. The above-described EPIC Lung and Heidelberg-EPIC studies were performed independently with no sample overlap. Genotypes for SNPs showed no significant departure from HWE, with the exception of rs13314271 in cases.

ICR replication. This study comprised 2,448 cases (1,664 male; mean age at diagnosis 71.8 years, s.d. 6.7 years) with pathologically confirmed lung cancer ascertained through GELCAPS⁴⁹ and 2,989 controls (1,469 male; mean age at sampling 60.6 years, s.d. 12.0 years) collected through the National Study of Colorectal Cancer Genetics with no personal history of malignancy. Cases were subclassified into histological subtypes based on ICD coding as described above (in the section detailing the ICR GWAS). Both cases and controls were British residents and had self-reported European ancestry. The genotype distributions of genotypes for each of the SNPs typed in replication showed no significant departure from HWE.

IARC replication. This analysis comprised three studies: (i) EPIC Lung⁵,²⁷, a nested case-control study performed within the EPIC (European Prospective Investigation into Cancer and Nutrition) prospective cohort totaling 1,119 lung cancer cases and 2,546 controls (matched one or two to cases for age, sex, center and time of recruitment) selected from 8 of the 10 countries participating in EPIC (Sweden, Netherlands, UK, France, Germany, Spain, Italy and Norway); (ii) the Szczecin case-control study⁵², a consecutive series of 849 incident lung cancer cases ascertained from the outpatient oncology clinic in the regional hospital of Szczecin between 2004 and 2007 (the 1,072 controls were individuals without diagnosed cancer or family history of cancer matched to cases by sex, age and region recruited by general medical practitioners); and (iii) Moscow L2. 1,081 newly diagnosed lung cancer cases and 2,119 controls recruited from three hospitals within the Moscow area of Russia between 2007 and 2011. Information on lifestyle risk factors and medical and family history was collected from subjects by interview using a standard questionnaire. Cases were subclassified into histological subtypes based on ICD-O-3 coding as described above (in the section detailing the IARC GWAS). The distributions of genotypes for each of the SNPs typed in replication showed no departure from HWE in each country or study series.

Toronto. This study was conducted in the greater Toronto area from 2008 to 2013. Lung cancer cases were recruited at the hospitals in the network of the University of Toronto. Controls were selected randomly from individuals registered in the family medicine clinics databases and were frequency matched with cases on age and sex. All subjects were interviewed, and information on lifestyle risk factors, occupational history and medical and family history was collected using a standard questionnaire. Tumors were centrally reviewed by the reference pathologist (a member of the International Association for the Study of Lung Cancer (IASLC) committee) and a second pathologist in the University Health Network. If the reviews conflicted, a consensus was arrived at after discussion. Coding of histology was based on 2001 WHO/IASLC. After applying standardized quality control procedures and restricting the data to participants with self-reported European ancestry, data and samples were available on 1,084 cases and 966 controls. The genotype distributions of genotypes for each of the SNPs typed in replication showed no significant departure from HWE.

Replication genotyping. Genotyping of rs1519542, rs13314271, rs55731496, rs149423192, rs4592420, rs11571833, rs56084662 and rs17879961 was performed using competitive allele-specific PCR KASPar chemistry (LCG, Hertfordshire, UK; UK replication sequen), Sequenom (Sequenom, Inc., San Diego, US; Toronto replication and Heidelberg-EPIC replication (rs1519542, rs55731496, rs149423192, rs4592420, rs11571833, rs56084662 and rs17879961)) or TaqMan (Carlsbad, CA; IARC replication series and Heidelberg-EPIC replication (rs13314271)). All primers, probes and conditions used are available on request. Call rates for SNP genotypes were >95% in each of the replication series.

To ensure the quality of genotyping in all assays, at least two negative controls and 1–10% duplicates (showing a concordance of >99%) were genotyped at each center. To exclude technical artifacts in genotyping, at the ICR and IARC we performed cross-platform validation of 96 samples and sequenced a set of 96 randomly selected samples from each case and control series to confirm genotyping accuracy. Assays were found to be performing robustly; concordance was >99%.

Statistical and bioinformatic analyses. Data were imputed for all scans for over 10 million SNPs using data from the 1000 Genomes Project (phase 1
Gene expression profiles of non-small cell lung cancer: survival

The Genetic structure of Europeans: a view from the North-East.

Relationship between genotypes and smoking.

Sequence analysis of BRCA2 in constitutional DNA. At the ICR, targeted sequencing for the BRCA2 mutations c.6275delTT and c.4889C>G was performed using Illumina TruSeq capture technology (Illumina, Inc, San Diego). Sequencing was based on exome sequencing data generated using Illumina TruSeq capture technology (Illumina, Inc, San Diego). Sequencing for the BRCA2 was based on exome sequencing data.

LD metrics were calculated in PLINK using 1000 Genomes data and plotted on the basis of the distribution of CIs defined by Gabriel et al.67.

Relationship between genotypes and smoking. To examine the relationship between rs11571833 (BRCA2 p.Lys3326X), rs17879961 (CHEK2 p.Ile157Thr) and rs13314271 (TP63) genotype and cigarette consumption (cigarettes per day)66, we used data on 43,693 Icelandic subjects (including 34,850 chip-typed individuals).

Sequence analysis of BRCA2 in constitutional DNA. At the ICR, targeted sequencing for the BRCA2 mutations c.6275delTT and c.4889C>G was performed by Sanger implemented on an ABI3700 analyzer ( Applied Biosystems; Carlsbad, CA) or by an RSQR < 0.30 with MaCH or an information measure Is < 0.40 with IMPUTE2 (ref. 3), ProbABEL62, MaCH2dat v.1.12 (ref. 58) or the gln function in R. Principle components generated using common SNPs were included in the analysis to limit the effects of cryptic population stratification that might cause inflation of test statistics. The association between each SNP and lung cancer risk was assessed by Cochran-Armitage trend test. The adequacy of the case-control matching and possibility of differential genotyping of cases and controls were formally evaluated using Q-Q plots of test statistics. Meta-analysis was undertaken using inverse-variance approaches. The inflation factor λ was based on the 90% least-significant directly typed SNPs62. ORs and associated 95% CIs were calculated by unconditional logistic regression using R (v.2.6). Stata v.10 (State College, Texas, US) and PLINK (v1.06) software. Cochran's Q statistic to test for heterogeneity and the P statistic to quantify the proportion of the total variation due to heterogeneity were calculated65. P values ≥75% are considered to be characteristic of large heterogeneity.65 Additionally, analyses stratified by histology, sex, age and smoking status (current, former or never) were performed. All statistical tests were two sided.

The fidelity of imputation as assessed by the concordance between imputed and directly genotyped SNPs was examined in a subset of samples from the UK GWAS, MDACC GWAS, IARC GWAS and NCI GWAS discovery series (Supplementary Table 2).

LD metrics were calculated in PLINK using 1000 Genomes data and plotted using SNAPE9. LD blocks were defined on the basis of HapMap recombination rate (cM/Mb) as defined using the Oxford recombination hotspots and on the basis of the distribution of CIs defined by Gabriel et al.67.

Analysis of TCGA data. The exomes of 243 individuals with lung SQ and 338 individuals with lung AD in TCGA (Project Number #3230) were analyzed at IARC using an in-house pipeline based on the GATK tool kit. At IARC, Qiagen Generead (SABiosciences/Qiagen Hilde, Germany) was used to amplify the coding region of BRCA2 using Ionsuite software. Sequence changes were referenced to the Leiden Open Variation Database (LOVD2) and the BReast Cancer IARC database.

Copy number variation. Copy number variation was assessed from Human SNP Array 6.0 data. We retrieved level 3 TCGA data comprising normalized log, ratios of the fluorescence intensities between the target sample and a reference sample. We included only tumor-normal paired data in our analysis. We considered a log, ratio ≤0.5 as reflecting loss and a log, ratio >0.5 as reflecting gain. Annotation was performed by adding the genes contained in each of the remaining segments using Ensembl databases.

38. Omenn, G.S. et al. The β-carotene and retinol efficacy trial (CARET) for chemoprevention of lung cancer in high risk populations: smokers and asbestos-exposed workers. Cancer Res. 54 (suppl. 7), 2038s–2043s (1994).
39. Scigli, G. et al. Occupational exposure to vinyl chloride, acrylonitrile and styrene and lung cancer risk (Europe). Cancer Causes Control 15, 445–452 (2004).
40. Feyler, A. et al. Point: myeloperoxidase –463G>A polymorphism and lung cancer risk. Cancer Epidemiol. Biomarkers Prev. 11, 1550–1554 (2002).
41. Neils, M. et al. Genetic structure of Europeans: a view from the North-East. PLoS ONE 4, e5472 (2009).
42. Vålk, K. et al. Gene expression profiles of non-small cell lung cancer: survival of several human diseases and other traits. Nature 497, 283–292 (2010).
43. Holmen, J. et al. The Nord-Trondelag Health Study 1995–97 (HUNT2): objectives, contents, methods and participation. Norsk Epidemiologi 13, 1932 (2003).
44. Landi, M.T. et al. Environment And Genotypes in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. BMC Public Health 8, 203 (2008).
45. ATBC Cancer Prevention Study Group. The α-tocopherol, β-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. Ann. Epidemiol. 4, 1–10 (1994).
46. Hayes, R.B. et al. Methods for haplotype and allele frequency estimation in a large sample of members of the pedigrees in the pedigrones-a family study of early markers in the PLCO trial. Mutat. Res. 592, 147–154 (2005).
47. Calle, E.E. et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer 94, 2490–2501 (2002).
48. Eiser, T., Makikado, A. & Houlston, R. Identification of low penetration alleles for lung cancer: the Genetic Lung Cancer Predisposition Study (GELCAPS). BMC Cancer 8, 244 (2008).
49. Welcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678 (2007).
50. Su, L. et al. Genotypes and haplotypes of matrix metalloproteinase 1, 3 and 12 gene families and the risk of lung cancer. Cancer genomics 27, 1004–1029 (2006).
51. Kong, A. et al. Parental origin of sequence variants associated with complex diseases. Nature 462, 868–874 (2009).
52. Styrkarsdottir, U. et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. Nature 497, 283–292 (2010).
53. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
54. Boeing, H., Wahrendorf, J. & Becker, N. EPIC-Germany—a source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. Ann. Nutr. Metab. 43, 195–204 (2009).
55. Dally, H. et al. The CYP544A1 allele increases risk for small cell lung cancer: effect of gender and smoking dose. Pharmacogenetics 13, 607–618 (2003).
56. Penegar, S. et al. National study of colorectal cancer genetics. Br. J. Cancer 97, 1305–1309 (2007).
57. Timofeeva, M.N. et al. Genetic polymorphisms in 15q25 and 19q13 loci, cotinine level, and the risk of lung cancer in EPIC. Cancer Epidemiol. Biomarkers Prev. 20, 2250–2261 (2011).
58. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. 34, 816–834 (2010).
59. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44, 955–999 (2012).
60. Zeggini, E. et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat. Genet. 40, 638–645 (2008).
61. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11, 499–511 (2010).
62. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 11, 134 (2010).
63. Clayton, D.G. et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat. Genet. 37, 1243–1246 (2005).
64. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
65. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. Br. Med. J. 327, 557–560 (2003).
66. Johnson, A.D. et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24, 2938–2939 (2008).
67. Gabriel, S.B. et al. The structure of haplotype blocks in the human genome. Science 296, 2225–2229 (2002).
68. Thorgeirsson, T.E. et al. Sequence variants at CHRNA3/CHRNA6 and CYP2A6 affect smoking behavior. Nat. Genet. 42, 448–453 (2010).
Corrigendum: Rare variants of large effect in \textit{BRCA2} and \textit{CHEK2} affect risk of lung cancer

Yufei Wang, James D McKay, Thorunn Rafnar, Zhaoming Wang, Maria N Timofeeva, Peter Broderick, Xuchen Zong, Marina Laplana, Yongyue Wei, Younghun Han, Amy Lloyd, Manon Delahaye-Sourdeix, Daniel Chubb, Valerie Gaborieau, William Wheeler, Nilanjan Chatterjee, Gudmar Thorleifsson, Patrick Sulem, Geoffrey Liu, Rudolf Kaaks, Marc Henrion, Ben Kinnersley, Maxime Vallée, Florence Le Calvez-Kelm, Victoria I. Stevens, Susan M Gapstur, Wei V Chen, David Zaridze, Neonilia Szeshenia-Dabrowska, Jolanta Lissowska, Peter Rudnai, Eleonora Fabjanova, Dana Mates, Vladimir Bencko, Lenka Foretova, Vladimir Janout, Hans E Krokan, Maiken Elvestad Gabrielsen, Frank Skorpen, Lars Vatten, Inger Njolstad, Chu Chen, Gary Goodman, Simone Benhamou, Tonu Vooder, Kristjan Välk, Mari Neli, Marts Metspalu, Marcin Lener, Jan Lubiński, Mattias Johansson, Paolo Vineis, Antonio Agudo, Francoise Clavel-Chapelon, H Bas Bueno-de-Mesquita, Dimitrios Trichopoulos, Kay-Tee Khaw, Mikael Johansson, Elisabete Weiderpass, Anne Tjonneland, Elio Riboli, Mark Lathrop, Ghislaine Scelo, Demetrius Albanes, Neil E Caporaso, Yuanqing Ye, Jian Gu, Xifeng Wu, Margaret R Spitz, Hendrik Dienemann, Albert Rosenberger, Li Su, Athena Matakidou, Timothy Eisen, Kari Stefansson, Angela Risch, Stephen J Chanock, David C Christiani, Rayjean J Hung, Paul Brennan, Maria Teresa Landi, Richard S Houlston & Christopher I Amos

\textit{Nat. Genet.} 46, 736–741 (2014); published online 1 June 2014; corrected after print 23 January 2017

In the version of this article initially published, the name of author Florence Le Calvez-Kelm appeared incorrectly as Florence LeCalvez-Kelm. The error has been corrected in the HTML and PDF versions of the article.