Published OnlineFirst November 11, 2011; DOI:10.1158/0008-5472.CAN-11-1403

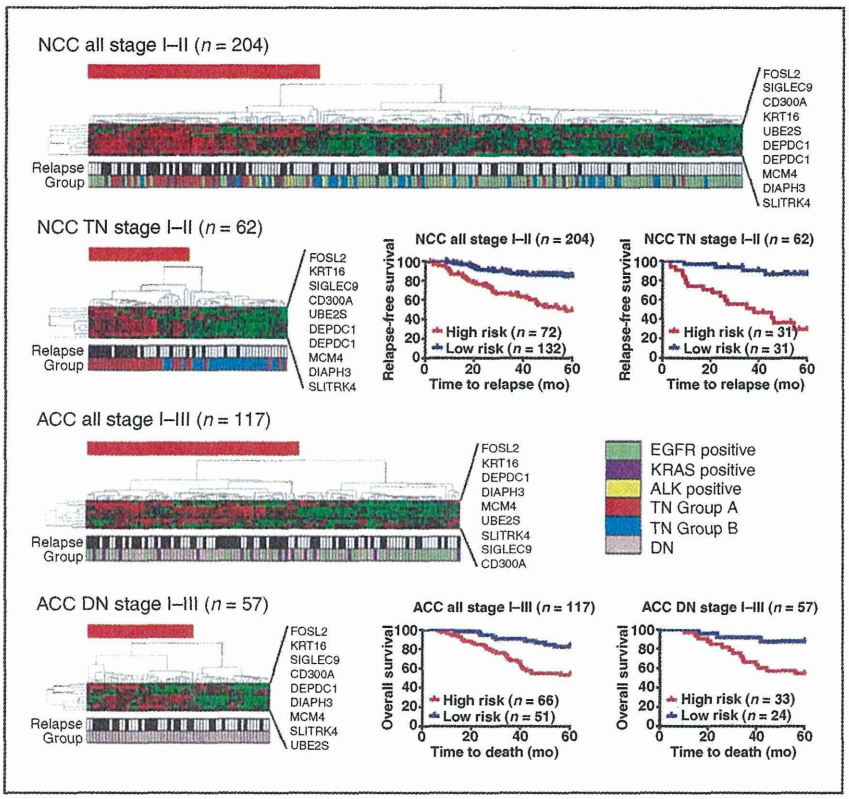

Figure 3. Unsupervised hierarchical clustering based on the expression of a set of 9 genes. All 204 stage I–II adenocarcinomas and 62 triple-negative (TN) stage I–II adenocarcinomas of the National Cancer Center (NCC) data set subjected to survival analysis were analyzed, and a cluster with higher expression of these genes than the other cluster was recognized as a high-risk group (red bar). Results of 117 adenocarcinomas, including 57 double-negative (DN) adenocarcinomas, of the Aichi Cancer Center (ACC) data set are shown below.

level of a total of 54,675 probes." Then, 11 of the 174 genes were further selected as being associated with prognosis of patients with triple-negative adenocarcinomas. Therefore, higher expression of several genes among the 11 genes was predicted to be associated with poorer prognosis, even when all adenocarcinoma cases, including *EGFR*-positive, *KRAS*-positive, and *ALK*-positive adenocarcinomas were analyzed together. Furthermore, triple-negative adenocarcinomas with poor prognosis would be separated into a high-risk group classified with this procedure. For this reason, we next analyzed all 204 adenocarcinoma cases. Among the 11 genes with 12 probes, 9 genes with 10 probes showed significant associations with both RFS and OS in all 204 adenocarcinoma cases and also in 162 stage I adenocarcinoma cases. *LOC152225* and *KIF19* were excluded because of no significant associations in stage I adenocarcinoma cases. As predicted, higher expression of the 9 genes was correlated with poorer prognosis in the analysis of RFS and OS among 204 stages I and II cases and also among 162 stage I cases.

The result strongly indicated that unsupervised hierarchical clustering using this 10 probe set (9 genes) would separate the patients into high-risk and low-risk groups for prognosis and that all group A triple-negative adenocarcinoma patients with poor prognosis would be classified into the high-risk group (Fig. 3 and Supplementary Table S3). As expected, expression profiling of these 9 genes successfully separated the 204 patients into high-risk and low-risk groups with significantly different RFS (HR = 3.79, 95% CI = 2.19–6.55, P = 1.9E-06) as well as OS (HR = 5.72, 95% CI = 2.53–12.87, P = 2.5E-05). Furthermore, if 62 triple-negative cases only were separated with these 9 genes, HRs for both RFS and OS were much higher than those with separation of all the 204 cases. All the relapsed cases in group A were separated into the high-risk group in the analyses of both cases (all the 204 cases and the 62 triple-negative cases only), supporting that triple-negative adenocarcinomas cases with poor prognosis can be selected as a high-risk group from all the adenocarcinoma cases by expression profiling of these 9 genes (Fig. 3). This profiling further separated 162 stage I cases as well as 46 stage I triple-negative adenocarcinoma cases into high-risk and low-risk groups with significantly different RFS as well as OS (Supplementary Fig. S5 and Supplementary Table S3). Again, HRs for both RFS and OS were much higher in triple-negative adenocarcinoma cases than in all adenocarcinoma cases. Accordingly, high levels of expression in these 9 genes were concluded to be distinct characteristics of triple-negative adenocarcinomas with poor prognosis.

Validation of associations using independent expression profiling data

To validate the present findings using the data of other cohorts, we searched for expression profiling data with

Published OnlineFirst November 11, 2011; DOI:10.1158/0008-5472.CAN-11-1403

mutation data of the *EGFR*, *KRAS*, and *ALK* genes in various databases. However, there has been no cohort in which expression profiles specifically in triple-negative adenocarcinomas were analyzed. Therefore, unsupervised hierarchical clustering using these 9 genes was done on a cohort of 117 Japanese lung adenocarcinoma cases because expression profile data as well as *EGFR/KRAS* mutation data were available only in this cohort (32). This study included 57 adenocarcinoma cases without *EGFR* and *KRAS* mutations. Although a different array platform was used, the data for all the 9 genes were available for clustering. These cases were separated into 2 groups of 33 cases and 24 cases (Fig. 3). OS of the 33 cases was significantly shorter than that of the 24 cases (HR = 3.17, 95% CI = 1.17–8.63, $P = 2.4E-02$; Supplementary Table S3). As with our cohort, the high-risk group showed a significantly higher HR of 2.73, even when all the 117 cases were analyzed together. Although *ALK* mutation data were not available for this cohort, the results strongly supported that expression profiling of the 9 genes would be highly informative for prediction of prognosis of lung adenocarcinoma patients, in particular patients with *EGFR*- and *KRAS*-negative adenocarcinomas.

Associations of *DEPDC1* expression with prognosis of NSCLC patients

Associations of gene expression with prognosis in various cancers are available from the PrognoScan database (22). Therefore, associations of expression of these 9 genes with prognosis of NSCLC patients were examined in 7 other cohorts (Table 4). Notably, *DEPDC1* expression was positively associated with poor prognosis in 4 of the 7 cohorts; MSK, Nagoya, Duke, and Seoul. The results strongly indicated that *DEPDC1* expression can be a novel prognostic marker for patients with NSCLC. Representative data showing the association of *DEPDC1* expression with prognosis in 204 adenocarcinoma patients obtained from the minimum *P* value approach are shown in Supplementary Fig. S6. Associations of *DEPDC1* expression with RFS and OS were validated by quantitative RT-PCR analysis of 204 stages I and II cases and also of 162 stage I cases (Supplementary Fig. S3).

FOSL2 expression was associated with prognosis in 3 of the 7 cohorts, whereas *MCM4*, *CD300A*, and *UBE2S* expression was associated in 1 cohort, respectively (Table 4).

Discussion

In this study, we attempted to characterize *ALK*-positive adenocarcinomas and triple-negative adenocarcinomas by genome-wide expression profiling. For this purpose, we selected a set of genes that are not transcriptionally activated in any *EGFR*-positive and *KRAS*-positive adenocarcinomas, and obtained 2 pieces of unique evidence. One is that *ALK*-positive adenocarcinomas show unique expression profiles in comparison with any other types of adenocarcinomas. The other is that there is a group of patients with extremely poor prognosis among triple-negative adenocarcinomas. This group, herein designated as group A, of patients showed much worse prognoses than patients with *EGFR*, *KRAS*, or *ALK* mutations and also than the other group, group B, of patients with triple-negative adenocarcinomas.

ALK-positive adenocarcinomas are sensitive to ALK TKIs with an overall response rate of 55% (8). Therefore, for the clinical application of ALK-targeted therapy, it is indispensable to develop a simple and reliable method for detection of *ALK* rearrangements in lung adenocarcinomas. Here, we showed that *ALK* expression is exclusively high only in *ALK*-positive adenocarcinomas and that several other genes, including *GRIN2A*, are overexpressed together with *ALK* specifically in *ALK*-positive adenocarcinomas. Therefore, GRIN2A can be a biomarker for detection of *ALK*-positive adenocarcinomas. *GRIN2A* encodes an *N*-methyl-D-aspartate (NMDA) receptor, which is a neurotransmitter-gated ion channel involved in regulation of synaptic function in the central nervous system (33). It was noted that the *GRIN2A* gene was recently reported to be frequently mutated in melanoma (34). Therefore, although the biological significance of *GRIN2A* upregulation in *ALK*-positive adenocarcinomas remains unclear, *GRIN2A* expression may play some important role in the phenotype unique to *ALK*-positive adenocarcinomas. Expression profiles unique to *ALK*-positive adenocarcinomas, shown here, will be also informative to improve clinical detection of *ALK* rearrangements.

Group A cases were discriminated by expression profiling of 9 genes among stage I–II cases who received complete surgical resection of tumors. Therefore, this gene set will be applicable as biomarkers to select lung adenocarcinoma patients who will benefit from adjuvant therapy after surgery, in particular to select them among patients with triple-negative adenocarcinomas. For this reason, combined analyses of this expression profiling with mutational analyses of the *EGFR*, *KRAS*, and *ALK* genes will be appropriate to pick out triple-negative adenocarcinoma patients with poor prognosis from all the adenocarcinoma patients. Molecular targeting drugs against triple-negative adenocarcinomas are not available at present; therefore, genes upregulated in group A cases will also be applicable as targets for therapy. *DEPDC1* was previously identified as being upregulated in bladder cancer and breast cancer (35–37). Because *DEPDC1* expression was hardly detectable in any normal tissues except testis, it has been considered as a cancer/testis antigen and also as a promising target of therapeutic drugs (35, 36). This study showed that *DEPDC1* is preferentially expressed in triple-negative adenocarcinomas with poor prognosis. In the PrognoScan database, *DEPDC1* expression is shown to be positively associated with poor prognosis in bladder cancer, multiple myeloma, breast cancer, glioma, and melanoma. Therefore, *DEPDC1* could be a novel target for diagnosis as well as therapy in various cancers, including lung adenocarcinoma.

Identification of genetic alterations that occur specifically in group A cases will be also of great importance for the development of target therapy for stages I and II lung adenocarcinoma patients with poor outcomes. Group A cases include males and ever-smokers as a majority (Table 1); therefore, group A cases were likely to carry several genetic alterations induced by tobacco carcinogens leading to poor outcomes. Identification of genetic alterations in

Published OnlineFirst November 11, 2011; DOI:10.1158/0008-5472.CAN-11-1403

group A adenocarcinomas will further facilitate the development of targeted therapies for lung adenocarcinomas with poor prognosis.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Dr. Teruhiko Yoshida and Ms. Sachiyo Mimaki for their efforts in expression profiling.

Grant Support

This work was supported in part by grants-in-aid from the Ministry of Health, Labor and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control and from the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO: 10–41). K. Shiraishi was an awardee of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 23, 2011; revised October 19, 2011; accepted October 21, 2011; published OnlineFirst November 11, 2011.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893–917.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108.
3. Colvy T, Noguchi M, Henschke C, Vazquez M, Geisinger K, Yokose T, et al. Adenocarcinoma. In:Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. World Health Organization classification of tumors: pathology and genetics; tumours of the lung, pleura, thymus and heart. IARC Press: Lyon, France; 2004. p. 35–44.
4. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. Lancet Oncol 2011;12:175–80.
5. Herbst RS, Heymach JV, Lippman SM: Lung cancer. N Engl J Med 2008;359:1367–80.
6. Janku F, Stewart DJ, Kurzrock R. Targeted therapy in non-small-cell lung cancer–is it becoming a reality? Nat Rev Clin Oncol 2010;7:401–14.
7. Bronte G, Rizzo S, La Paglia L, Adamo V, Siragusa S, Ficorella C, et al. 2288; Driver mutations and differential sensitivity to targeted therapies: a new approach to the treatment of lung adenocarcinoma. Cancer Treat Rev 2010;36 Suppl 3:S21–9.
8. Gerber DE, Minna JD. ALK inhibition for non-small cell lung cancer: from discovery to therapy in record time. Cancer Cell 2010;18:548–51.
9. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers–a different disease. Nat Rev Cancer 2007;7:778–90.
10. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 2009;361:958–67.
11. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al.: Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009;361:947–57.
12. Konstantinopoulos PA, Karamouzis MV, Papavassiliou AG. Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets. Nat Rev Drug Discov 2007;6:541–55.
13. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561–6.
14. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 2006;24:1679–88.
15. Angulo B, Suarez-Gauthier A, Lopez-Rios F, Medina PP, Conde E, Tang M, et al. Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK3CA overexpression by gene amplification. J Pathol 2008;214:347–56.
16. Motoi N, Szoke J, Riely GJ, Seshan VE, Kris MG, Rusch VW, et al. Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. Am J Surg Pathol 2008;32:810–27.
17. Meyerson M, Carbone D. Genomic and proteomic profiling of lung cancers: lung cancer classification in the age of targeted therapy. J Clin Oncol 2005;23:3219–26.
18. Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. World Health Organization classification of tumors: pathology and genetics; tumours of the lung, pleura, thymus and heart. IARC Press: Lyon, France; 2004. p. 1–344.
19. Saito M, Schetter AJ, Mollerup S, Kohno T, Skaug V, Bowman ED, et al. The association of microRNA expression with prognosis and progression in early-stage, non-small cell lung adenocarcinoma: a retrospective analysis of three cohorts. Clin Cancer Res 2011;17:1875–82.
20. Robins JM, Gail MH, Lubin JH. More on "Biased selection of controls for case-control analyses of cohort studies". Biometrics 1986;42:293–9.
21. Richardson DB. An incidence density sampling program for nested case-control analyses. Occup Environ Med 2004;61:e59.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–8. Available from: http://rana.lbl.gov/eisen/.
23. Takano T, Ohe Y, Sakamoto H, Tsuta K, Matsuno Y, Tateishi U, et al. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. J Clin Oncol 2005;23:6829–37.
24. Takano T, Ohe Y, Tsuta K, Fukui T, Sakamoto H, Yoshida T, et al. Epidermal growth factor receptor mutation detection using high-resolution melting analysis predicts outcomes in patients with advanced non small cell lung cancer treated with gefitinib. Clin Cancer Res 2007;13:5385–90.
25. Takeuchi K, Choi YL, Soda M, Inamura K, Togashi Y, Hatano S, et al. Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. Clin Cancer Res 2008;14:6618–24.
26. Mizuno H, Kitada K, Nakai K, Sarai A. 2288; PrognoScan: a new database for meta-analysis of the prognostic value of genes. 2288; BMC Med Genomics 2009;2:18. Available from: http://gibk21.bio.kyutech.ac.jp/PrognoScan/index.html.
27. Mitsudomi T. Advances in target therapy for lung cancer. Jpn J Clin Oncol 2010;40:101–6.
28. Suda K, Tomizawa K, Mitsudomi T. Biological and clinical significance of KRAS mutations in lung cancer: an oncogenic driver that contrasts with EGFR mutation. Cancer Metastasis Rev 2010;29:49–60.
29. Zhang X, Zhang S, Yang X, Yang J, Zhou Q, Yin L, et al. Fusion of EML4 and ALK is associated with development of lung adenocarcinomas lacking EGFR and KRAS mutations and is correlated with ALK expression. Mol Cancer 2010;9:188.
30. Boland JM, Erdogan S, Vasmatzis G, Yang P, Tillmans LS, Johnson MR, et al. Anaplastic lymphoma kinase immunoreactivity correlates with ALK gene rearrangement and transcriptional up-regulation in non-small cell lung carcinomas. Hum Pathol 2009;40:1152–8.
31. Mino-Kenudson M, Chirieac LR, Law K, Hornick JL, Lindeman N, Mark EJ, et al. A novel, highly sensitive antibody allows for the routine detection of ALK-rearranged lung adenocarcinomas by standard immunohistochemistry. Clin Cancer Res 2010;16:1561–71.

Published OnlineFirst November 11, 2011; DOI:10.1158/0008-5472.CAN-11-1403

32. Tomida S, Takeuchi T, Shimada Y, Arima C, Matsuo K, Mitsudomi T, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. J Clin Oncol 2009;27:2793–9.
33. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, et al. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet 2010;42:1021–6.
34. Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, et al. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. Nat Genet 2011;43:442–6.
35. Kanehira M, Harada Y, Takata R, Shuin T, Miki T, Fujioka T, et al. Involvement of upregulation of DEPDC1 (DEP domain containing 1) in bladder carcinogenesis. Oncogene 2007;26:6448–55.
36. Harada Y, Kanehira M, Fujisawa Y, Takata R, Shuin T, Miki T, et al. Cell-permeable peptide DEPDC1-ZNF224 interferes with transcriptional repression and oncogenicity in bladder cancer cells. Cancer Res 2010; 70:5829–39.
37. Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, Kemmner W. Identification of early molecular markers for breast cancer. Mol Cancer 2011;10:15.

nature genetics

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

Qing Lan[1,68], Chao A Hsiung[2,68], Keitaro Matsuo[3,68], Yun-Chul Hong[4,68], Adeline Seow[5,68], Zhaoming Wang[6,68], H Dean Hosgood III[1,7,68], Kexin Chen[8,68], Jiu-Cun Wang[9,10,68], Nilanjan Chatterjee[1], Wei Hu[1], Maria Pik Wong[11], Wei Zheng[12], Neil Caporaso[1], Jae Yong Park[13], Chien-Jen Chen[14], Yeul Hong Kim[15], Young Tae Kim[16], Maria Teresa Landi[1], Hongbing Shen[17,18], Charles Lawrence[19], Laurie Burdett[6], Meredith Yeager[6], Jeffrey Yuenger[6], Kevin B Jacobs[6], I-Shou Chang[20], Tetsuya Mitsudomi[21], Hee Nam Kim[22], Gee-Chen Chang[23,24], Bryan A Bassig[1,25], Margaret Tucker[1], Fusheng Wei[26], Zhihua Yin[27], Chen Wu[28,29], She-Juan An[30], Biyun Qian[8], Victor Ho Fun Lee[31], Daru Lu[9,10], Jianjun Liu[32,33], Hyo-Sung Jeon[34], Chin-Fu Hsiao[2], Jae Sook Sung[15], Jin Hee Kim[35], Yu-Tang Gao[36], Ying-Huang Tsai[37], Yoo Jin Jung[16], Huan Guo[38], Zhibin Hu[17,18], Amy Hutchinson[6], Wen-Chang Wang[2], Robert Klein[39], Charles C Chung[1], In-Jae Oh[40,41], Kuan-Yu Chen[42], Sonja I Berndt[1], Xingzhou He[43], Wei Wu[27], Jiang Chang[28,29], Xu-Chao Zhang[30], Ming-Shyan Huang[44], Hong Zheng[8], Junwen Wang[45,46], Xueying Zhao[9,10], Yuqing Li[32], Jin Eun Choi[34], Wu-Chou Su[47], Kyong Hwa Park[15], Sook Whan Sung[48], Xiao-Ou Shu[12], Yuh-Min Chen[23,49], Li Liu[50], Chang Hyun Kang[16], Lingmin Hu[17,18], Chung-Hsing Chen[20], William Pao[51], Young-Chul Kim[40,41], Tsung-Ying Yang[24], Jun Xu[52], Peng Guan[27], Wen Tan[28,29], Jian Su[30], Chih-Liang Wang[53], Haixin Li[8], Alan Dart Loon Sihoe[54], Zhenhong Zhao[9,10], Ying Chen[5], Yi Young Choi[34], Jen-Yu Hung[44], Jun Suk Kim[55], Ho-Il Yoon[56], Qiuyin Cai[12], Chien-Chung Lin[47], In Kyu Park[16], Ping Xu[57], Jing Dong[17,18], Christopher Kim[1], Qincheng He[27], Reury-Perng Perng[49], Takashi Kohno[58], Sun-Seog Kweon[59,60], Chih-Yi Chen[61], Roel Vermeulen[62], Junjie Wu[9,10], Wei-Yen Lim[5], Kun-Chieh Chen[24], Wong-Ho Chow[1], Bu-Tian Ji[1], John K C Chan[63], Minjie Chu[17,18], Yao-Jen Li[14], Jun Yokota[64], Jihua Li[65], Hongyan Chen[9,10], Yong-Bing Xiang[36], Chong-Jen Yu[42], Hideo Kunitoh[66], Guoping Wu[26], Li Jin[9,10], Yen-Li Lo[2], Kouya Shiraishi[58], Ying-Hsiang Chen[2], Hsien-Chih Lin[2], Tangchun Wu[38,69], Yi-Long Wu[30,69], Pan-Chyr Yang[67,69], Baosen Zhou[27,69], Min-Ho Shin[60,69], Joseph F Fraumeni Jr[1,69], Dongxin Lin[28,29,69], Stephen J Chanock[1,69] & Nathaniel Rothman[1,69]

To identify common genetic variants that contribute to lung cancer susceptibility, we conducted a multistage genome-wide association study of lung cancer in Asian women who never smoked. We scanned 5,510 never-smoking female lung cancer cases and 4,544 controls drawn from 14 studies from mainland China, South Korea, Japan, Singapore, Taiwan and Hong Kong. We genotyped the most promising variants (associated at $P < 5 \times 10^{-6}$) in an additional 1,099 cases and 2,913 controls. We identified three new susceptibility loci at 10q25.2 (rs7086803, $P = 3.54 \times 10^{-18}$), 6q22.2 (rs9387478, $P = 4.14 \times 10^{-10}$) and 6p21.32 (rs2395185, $P = 9.51 \times 10^{-9}$). We also confirmed associations reported for loci at 5p15.33 and 3q28 and a recently reported finding at 17q24.3. We observed no evidence of association for lung cancer at 15q25 in never-smoking women in Asia, providing strong evidence that this locus is not associated with lung cancer independent of smoking.

It is estimated that 25% of lung cancer cases arise in individuals who never smoked. Lung cancer in never smokers ranks as the seventh most common cause of cancer death worldwide[1]. A number of observations suggest that the molecular pathogenesis of lung cancer differs by smoking status. Differences have been reported by smoking status in cellular and molecular carcinogenic pathways, distinct profiles of oncogenic mutations (for example, in *EGFR*) and response to targeted therapy[2,3]. Compared to lung cancer in smokers, cases in never smokers are more likely to arise in women at a younger age, and there is a greater proportion of cases with the adenocarcinoma histology subtype[3]. Epidemiological studies of lung cancer in never smokers have shown that the incidence of lung cancer in women is particularly high in Asia[4], which is partially attributed to exposure to environmental tobacco smoke, combustion products from indoor heating and cooking fuel, and cooking oil fumes[4–10].

A full list of affiliations appears at the end of the paper.

Received 31 May; accepted 5 October; published online 11 November 2012; doi:10.1038/ng.2456

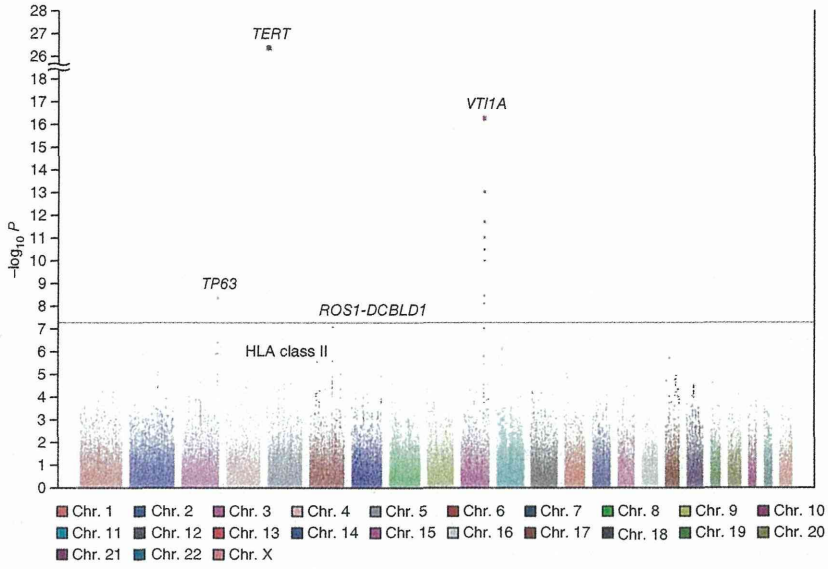

Figure 1 Association results from a GWAS of never-smoking women in Asia. Manhattan plot based on *P* values derived from 1-degree-of-freedom tests of genotype trend effect in unconditional logistic regression analysis adjusted for study, age and three eigenvectors in a GWAS of lung cancer in never-smoking Asian females, including 5,510 lung cancer cases and 4,544 controls. The *x* axis represents chromosomal location, and the *y* axis shows *P* values on a negative logarithmic scale. The red horizontal line represents the genome-wide significance threshold of $P = 5 \times 10^{-8}$. Labeled are two previously associated loci (*TERT* at 5p15.33 and *TP63* at 3q28) together with three newly identified loci (*VTI1A* on chromosome 10 and *ROS1-DCBLD1* and the HLA class II region on chromosome 6).

To gain insight into the etiology of lung cancer in never-smoking women, we formed the Female Lung Cancer Consortium in Asia (FLCCA), which includes studies drawn from mainland China, South Korea, Japan, Singapore, Taiwan and Hong Kong. Previously, we published the first genome-wide association study (GWAS) of lung cancer in never-smoking Asian women, including 584 cases and 585 controls with large-scale replication, reporting an association at 5p15.33 near the *TERT* gene[11]; in this study, it was also notable that the estimated effect of the associated locus was greater in nonsmoking Asian women than the reported effect size observed in primarily smokers of European ancestry[12]. We also confirmed an association signal in *TP63* at 3q28 (ref. 13), replicating the report from a GWAS conducted in Japan[14].

To identify new susceptibility loci in Asian never-smoking women, we conducted a lung cancer GWAS in 14 studies (13 case-control studies and 1 cohort study; **Supplementary Table 1** and **Supplementary Note**). Samples were scanned at six centers (Online Methods): the US National Cancer Institute (NCI) Cancer Genomic Research (CGR) Laboratory, the Genome Institute of Singapore, the Memorial Sloan-Kettering Cancer Center (MSKCC), GeneTech Biotech in Taiwan, Gene-Square Biotech in Beijing and deCODE Genetics in Iceland. After stringent quality control analysis of genotypes (Online Methods), we combined data sets for 5,510 lung cancer cases and 4,544 controls using a previously described clustering algorithm[15]. The primary analysis was performed using logistic regression for genotype trend effect (with 1 degree of freedom) adjusted for study center, age and three eigenvectors (on the basis of principal-components analysis). A comparison of the observed and expected *P* values in the quantile-quantile plot showed an enrichment of observed signals with small *P* values compared to the null distribution of no association, with little evidence for genomic inflation (unscaled $\lambda = 1.014$, $\lambda_{1000} = 1.003$; **Supplementary Fig. 1**)[16].

The overall association results are shown in a Manhattan plot, in which we observed both new and known loci that exceeded the threshold for genome-wide significance ($P < 5 \times 10^{-8}$; **Fig. 1**). We observed association at two previously established loci, rs2736100 at 5p15.33 (refs. 11,12,14,17–19) and rs4488809 at 3q28 (refs. 13,14). We also observed support for association of a recently reported locus marked by rs7216064 at 17q24.3 (ref. 20) (**Supplementary Table 2**). Notably, there was no evidence for association across the 15q25 region, which has been associated with smoking-related lung cancer[12,19,21–24]. We did not observe strong association signals for other loci reported in either European[25] or Asian[17,26] populations (**Supplementary Table 2**).

In our primary scan, we observed one new locus at 10q25.2, marked by rs7086803, that substantially exceeded the threshold for genome-wide significance (odds ratio (OR) = 1.32, 95% confidence interval (CI) = 1.24–1.41; $P = 5.04 \times 10^{-17}$) (**Fig. 1** and **Table 1**). We developed assays to genotype 13 SNPs associated at $P < 5 \times 10^{-6}$ in the initial scan, using analysis of all cases or the most common subtype in nonsmokers, adenocarcinoma. We genotyped 1,099 new cases and 2,913 controls drawn from the same studies as in the initial scan. In a combined analysis of 6,609 cases and 7,457 controls, 3 new loci achieved associations at genome-wide significance (**Table 1**): 10q25.2 (rs7086803: OR = 1.28, 95% CI = 1.21–1.35; $P = 3.54 \times 10^{-18}$), 6q22.2 (rs9387478: OR = 0.85, 95% CI = 0.81–0.90; $P = 4.14 \times 10^{-10}$)

Table 1 New loci associated with lung cancer in a GWAS of never-smoking Asian females

SNP	Plausible candidate gene(s)	Chromosome position	Subset	Allele[a]	MAF[b] Control	MAF[b] Case	Subjects Control	Subjects Case	OR (95% CI)	P_{trend}
rs7086803	*VTI1A*	10q25.2	GWAS	G/A	0.26	0.32	4,492	5,457	1.32 (1.24–1.41)	5.04×10^{-17}
			Replication	G/A	0.27	0.31	2,887	1,085	1.23 (1.10–1.37)	3.36×10^{-4}
			Combined	G/A	0.27	0.31	7,379	6,542	1.28 (1.21–1.35)	3.54×10^{-18}
rs9387478	*ROS1*, *DCBLD1*	6q22.2	GWAS	C/A	0.50	0.46	4,542	5,510	0.85 (0.81–0.90)	7.79×10^{-8}
			Replication	C/A	0.49	0.47	2,891	1,091	0.92 (0.83–1.01)	0.088
			Combined	C/A	0.50	0.46	7,433	6,601	0.85 (0.81–0.90)	4.14×10^{-10}
rs2395185[c] (rs28366298)	HLA class II region	6p21.32	GWAS	G/T	0.35	0.38	4,541	5,504	1.16 (1.09–1.23)	2.60×10^{-6}
			Replication	A/C	0.37	0.42	2,880	1,008	1.20 (1.08–1.33)	7.93×10^{-4}
			Combined	Meta			7,421	6,512	1.17 (1.11–1.23)	9.51×10^{-9}

[a]Minor allele listed second. [b]Minor allele frequency. [c]For the HLA class II region, because a TaqMan assay could not be designed for rs2395185, we instead genotyped rs28366298, its perfect surrogate ($r^2 = 1.0$), by TaqMan. The reported *P* value is based on meta-analysis of the rs2395185 results in the GWAS and the rs28366298 results in the TaqMan set.

and 6p21.32 (rs2395185: OR = 1.17, 95% CI = 1.11–1.23; $P = 9.51 \times 10^{-9}$) (**Fig. 2**, **Table 1**, **Supplementary Fig. 2** and **Supplementary Tables 3** and **4**).

Analysis by histological subtype of lung cancer showed that both the 6q22.2 (rs9387478) and 6p21.32 (rs2395185) loci were associated with adenocarcinoma only, which comprised 71% of cases (**Table 2**). The estimated effects were consistent across studies (**Supplementary Fig. 2**). We note that rs7086803 showed a somewhat larger effect for squamous carcinoma compared to adenocarcinoma (**Table 2**), but, as the number of squamous carcinoma cases analyzed was small, we consider this a preliminary observation requiring independent replication.

To explore the relationship between these three regions and lung cancer in populations of European ancestry, we analyzed data from a previously reported GWAS of 5,718 lung cancer cases and 5,739 controls, including men and women who were primarily ever smokers[12]. We found no evidence for association at the three newly associated loci. In a subanalysis of 350 never-smoker cases and 1,379 never-smoker controls drawn from this study, we observed some evidence of association for rs2395185 (M.T.L., unpublished data), but larger studies are warranted.

We imputed SNPs catalogued in the 1000 Genomes Project March 2012 release and the Division of Cancer Epidemiology and Genetics

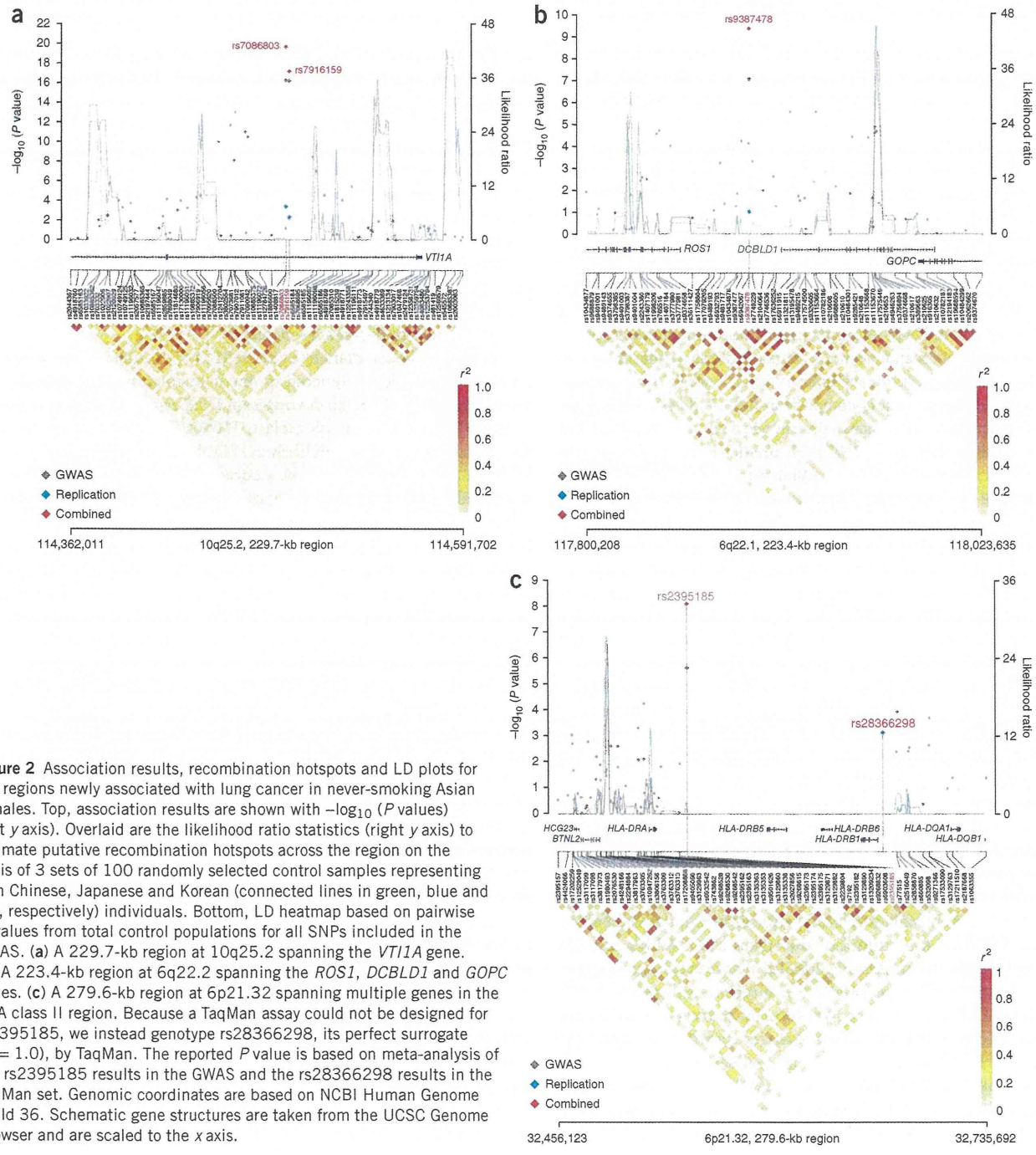

Figure 2 Association results, recombination hotspots and LD plots for the regions newly associated with lung cancer in never-smoking Asian females. Top, association results are shown with $-\log_{10}$ (*P* values) (left *y* axis). Overlaid are the likelihood ratio statistics (right *y* axis) to estimate putative recombination hotspots across the region on the basis of 3 sets of 100 randomly selected control samples representing Han Chinese, Japanese and Korean (connected lines in green, blue and red, respectively) individuals. Bottom, LD heatmap based on pairwise r^2 values from total control populations for all SNPs included in the GWAS. (**a**) A 229.7-kb region at 10q25.2 spanning the *VTI1A* gene. (**b**) A 223.4-kb region at 6q22.2 spanning the *ROS1*, *DCBLD1* and *GOPC* genes. (**c**) A 279.6-kb region at 6p21.32 spanning multiple genes in the HLA class II region. Because a TaqMan assay could not be designed for rs2395185, we instead genotype rs28366298, its perfect surrogate ($r^2 = 1.0$), by TaqMan. The reported *P* value is based on meta-analysis of the rs2395185 results in the GWAS and the rs28366298 results in the TaqMan set. Genomic coordinates are based on NCBI Human Genome Build 36. Schematic gene structures are taken from the UCSC Genome Browser and are scaled to the *x* axis.

Table 2 New loci associated with adenocarcinoma and squamous carcinoma of the lung in a GWAS of never-smoking Asian females

SNP	Putative gene	Chromosome position	Allele[a]	MAF[b] 1	2	3	Adenocarcinoma Subjects Control	Case	OR (95% CI)	P_{trend}	Squamous carcinoma Subjects Control	Case	OR (95% CI)	P_{trend}	$P_{heterogeneity}$[c]
rs7086803	*VTI1A*	10q25.2	G/A	0.27	0.31	0.34	7,035	4,666	1.24 (1.17–1.32)	1.19×10^{-11}	6,714	756	1.36 (1.21–1.54)	7.11×10^{-7}	0.014
rs9387478	*ROS1*, *DCBLD1*	6q22.2	C/A	0.50	0.46	0.48	7,089	4,726	0.84 (0.80–0.89)	1.55×10^{-9}	6,768	755	0.90 (0.81–1.01)	0.078	0.060
rs2395185[d] (rs28366298)	HLA class II region	6p21.32	Meta				7,390	4,696	1.20 (1.13–1.28)	9.47×10^{-10}	7,211	742	1.05 (0.93–1.18)	0.42	0.56

[a]Minor allele listed second. [b]Minor allele frequency. 1, MAF in controls; 2, MAF in adenocarcinoma; 3, MAF in squamous carcinoma. [c]Tested by case-case analysis. [d]For the HLA class II region, because a TaqMan assay could not be designed for rs2395185, we instead genotyped rs28366298, its perfect surrogate ($r^2 = 1.0$), by TaqMan. The reported *P* value is based on meta-analysis of the rs2395185 results in the GWAS and the rs28366298 results in the TaqMan set.

Imputation Reference Set version 1 (ref. 27) using the IMPUTE2 program[28] across a 1-Mb region centered on the index SNP (Online Methods). For the two regions outside of the human leukocyte antigen (HLA) region, the association analysis did not identify new signals that were substantially stronger than those found for the genotyped SNPs (**Supplementary Fig. 3a,b**). Although there seem to be stronger signals in the imputed data for the HLA class II region (**Supplementary Fig. 3c**), HLA typing will be necessary to unravel the specific haplotypes involved.

At the 6q22 locus, six SNPs were highly correlated with rs9387478 (r^2 = 0.99–1.00). Two SNPs, rs9387478 and rs6937083 (pairwise $r^2 = 1$), were observed within a region defined by the Encyclopedia of DNA Elements (ENCODE) as containing both chromatin state segmentation and enhancer- and promoter-associated histone marks. Although the evidence for evolutionary conservation is weak (that is, a cross-species sequence alignment comparison indicated conservation at the site of ~29.2 million years since divergence from a common ancestor), rs6937083 falls within an ENCODE-predicted transcription factor–binding site and an exon of the AceView-predicted gene, *DCBLD1*. The architecture of the region on chromosome 10q25 is more complicated because there are 23 perfectly correlated SNPs ($r^2 = 1$) and 1 highly correlated SNP ($r^2 = 0.99$). All localize to intron 7 or the UTR of one transcript of the *VTI1A* gene (encoding vesicle transport through interaction with t-SNAREs homolog 1A (yeast)). Sixteen fall within putatively functional regions, defined as ENCODE DNase I hypersensitivity clusters, chromatin state segmentation, the UTR of *VTI1A*, ENCODE enhancer- and promoter-associated histone marks and/or highly conserved (that is, a cross-species sequence alignment comparison indicated conservation at the site of 300 million years since divergence from a common ancestor) regions (**Supplementary Table 5**). rs11196080 is noteworthy because many of the functionally predicted areas converge on this SNP, making this a high-priority variant for functional follow-up studies.

The strongest new association signal, rs7086803 at 10q25.2, maps to intron 7 of the *VTI1A* gene, which has been implicated in lung carcinogenesis. Loss of VTI1A activity has been reported to reduce high-frequency spontaneous neurotransmitter release[29] and rapid progressive neurodegeneration in the peripheral ganglia[30]. VTI1A is also involved in Acrp30-containing vesicles in adipocytes, and lower amounts of VTI1A in cultured adipocytes can inhibit adiponectin secretion[31]. Lower amounts of adiponectin have previously been associated with advanced lung cancer[31,32]. A recent study reported recurrent *VTI1A-TCF7L2* fusions in colorectal cancers, and a colorectal carcinoma cell line with the fusion gene was shown to be dependent on *VTI1A-TCF7L2* for anchorage-independent growth[33].

The rs9387478 SNP at 6q22.2 is located in an interval that contains two candidate genes: *DCBLD1* (encoding discoidin, CUB and LCCL domain containing 1) and *ROS1* (encoding the ROS proto-oncogene receptor tyrosine kinase). ROS1 functions as both an integral membrane protein and a receptor tyrosine kinase[34]. Expression of *Ros1* is specifically increased in lung cancer tissue in mouse models, and *ROS1* expression levels are higher in non-small cell lung cancer (NSCLC)[35]. *ROS1* fusions in lung adenocarcinoma and NSCLC, particularly in Asian never smokers, have been identified as drivers of oncogenesis[36–38]. *ROS1* rearrangements were found to be more common in lung adenocarcinomas from never smokers and younger affected individuals[39]. There is limited evidence concerning the functional role of the protein encoded by *DCBLD1*; a related gene at 3q12.2, *DCBLD2* (encoding discoidin, CUB and LCCL domain containing 2; also known as *CLCP1*) regulates cellular proliferation and invasion and may have an important role in cancer metastasis[40–42].

The third locus, marked by rs2395185 at 6p21.3, is located within 20 kb of *HLA-DRA* (encoding major histocompatibility complex, class II, DRα) and 52 kb downstream of *HLA-DRB5* (encoding major histocompatibility complex, class II, DRβ5). There was no evidence for strong linkage disequilibrium (LD) between this SNP and other SNPs at 6p21.32 reported to be associated with lung cancer[17,23]. There was little LD with a recently reported SNP at 6p21.3, rs3817963, which was associated with lung cancer in a Japanese population[20]; the r^2 in Han Chinese and Japanese HapMap samples was 0.18 and 0.10, respectively, and *D* was 0.57 and 0.43, respectively. These data suggest that our locus probably represents a new HLA class II–related finding for nonsmoking lung cancer susceptibility. Further mapping across the complex HLA region is required to characterize the specific susceptibility alleles or haplotypes involved in nonsmoking lung cancer risk. We also note that rs2395185 has been previously associated with ulcerative colitis[43], Hodgkin lymphoma[44] and type 1 diabetes[45].

In previous GWAS of lung cancer, in which a majority of cases were smokers, SNPs across a region at 15q25 have been associated with lung cancer risk[12,19,21–24]. However, studies of smoking-related behavior have also identified variants at 15q25, raising the possibility that the variants previously identified by GWAS for lung cancer could mediate risk through effects on tobacco use[46]. We previously genotyped additional SNPs across 15q25 in Asian studies and observed no evidence of association with lung cancer in never-smoking Asian females[11]. Notably, in our current, larger study, there was no evidence for association with lung cancer at 15q25 in the never-smoking population overall or in the major subtypes. These data provide strong evidence that this locus is not associated with lung cancer independent of smoking in never-smoking females in Asia, which contrasts with the results from a smaller Asian study[24] but is consistent with previous reports from smaller studies conducted in populations of European ancestry[12,47,48].

We investigated the relationship between our new loci and known environmental exposures. The association between exposure to

environmental tobacco smoke in the home and adenocarcinoma in the five studies with data available yielded an OR of 1.36 ($P = 1.2 \times 10^{-4}$) in an analysis of 1,770 cases and 2,675 controls, consistent with previous reports[8]. The effect of environmental tobacco smoke was stronger for subjects with the GG genotype at rs2395185, with OR = 1.78 ($P = 1.15 \times 10^{-5}$), compared to subjects with the GT or TT genotypes, OR = 1.16 ($P = 0.15$), with $P_{interaction} = 0.002$. The association between the T allele at rs2395185 and risk of adenocarcinoma in subjects with and without exposure to environmental tobacco smoke yielded OR = 1.13 ($P = 0.031$) and OR = 1.43 ($P = 5.6 \times 10^{-4}$), respectively, with $P_{interaction} = 0.037$. There was no evidence of interaction with the other two new loci reported here.

In summary, we conducted a GWAS of lung cancer in never-smoking females in Asia and identified three new susceptibility loci at 10q25.2, 6q22.2 and 6p21.32. We also confirmed associations with two previously reported regions at 5p15.3 and 3q28 and a recently reported locus at 17q24.3. It is notable that our strongest finding at 10q25.2 has not been reported previously in lung cancer GWAS. This observation suggests that the etiology of lung cancer in never smokers in Asia may have unique genetic characteristics. This is consistent with the distinct pattern of environmental risk factors that have been causally linked to lung cancer in never-smoking females in Asia[4–8,10] and the distinct molecular phenotypes of lung cancer in never smokers[2,3]. Further work is warranted to map the new regions. Functional work is required to identify the variants that directly account for the underlying association, as well as to study how the genetic variants interact with established environmental risk factors, including environmental tobacco smoke, cooking fumes and fuel use, in never-smoking females in Asia.

URLs. CGF, http://cgf.nci.nih.gov/; GLU, http://code.google.com/p/glu-genetics/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm; Structure, http://pritch.bsd.uchicago.edu/structure.html; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; liftOver, http://hgdownload.cse.ucsc.edu/downloads.html; SAS v9.2 (used to generate forest plots), http://support.sas.com/kb/43/855.html.

METHODS

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

Accession codes. The CGEMS data portal provides access to individual-level data for investigators from certified scientific institutions after approval of their submitted Data Access Request.

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

ACKNOWLEDGMENTS

We thank J.-J. Yang, X.-N. Yang, Q. Zhou, W.-B. Guo, S.-L. Chen, Y. Huang, Z. Xie, J.-G. Chen, H.-H. Yan, K. Tajima, Y. Yatabe, T. Hida, K.-L. Chuah, A. Ng, P. Eng, S.-S. Leong, M.-K. Ang, E. Lim, T.-K. Lim, M. Teh, W.-T. Poh and A. Tee. The overall GWAS project was supported by the intramural program of the US National Institutes of Health/National Cancer Institute. A list of support provided to individual studies is provided in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

Q.L., N.R., S.J.C., D. Lin, C.A.H., Y.-C.H., K.M., A.S., H.D.H., J.Y.P., C.-J.C., Y.H.K., Y.T.K., C.L., Y.-L.W., P.-C.Y., B.Z., M.-H.S., J.F.F., K.C., W.Z., T.W., H.S., I.-S.C., D. Lu, N. Caporaso, W.P., R.K., J. Liu, M.T.L., N. Chatterjee, M.T. and M.Y. organized and designed the study. S.J.C., D. Lin, R.K., J. Liu, C.A.H., K.M., T.W., L.B., M.Y., J. Yuenger, Z.Y., C.W., H.G., A.H., W.W., Y.L., W.P., H.-C.L. and B.Z. conducted and supervised the genotyping of samples. Z.W., K.B.J., N.R., Q.L., S.J.C., N. Chatterjee, C.A.H., H.D.H., W.H., M.Y., I.-S.C., C.-F.H., W.-C.W., C.C.C., S.I.B., C.-H.C., R.V. and Y.-H.C. contributed to the design and execution of statistical analysis. Q.L., N.R., S.J.C., Z.W., W.H., C.C.C., C.A.H., K.M., Y.-C.H., A.S., H.D.H., N. Chatterjee, N. Caporaso, C.L., M.Y., B.A.B., M.T., S.-J.A., S.I.B., M.T.L., C.K., R.V., Y.-L.W., J.F.F. and I.-S.C. wrote the first draft of the manuscript. C.A.H., Q.L., B.Z., Y.-C.H., K.M., A.S., K.C., J.-C.W., M.P.W., W.Z., J.Y.P., W.H., C.-J.C., Y.H.K., Y.T.K., T.W., H.S., I.-S.C., T.M., H.N.K., F.W., Z.Y., C.W., S.-J.A., G.-C.C., B.Q., V.H.F.L., D. Lu, H.-S.J., J.S.S., J.H.K., Y.-T.G., Y.-H.T., Y.J.J., H.G., Z.H., I.-J.O., K.-Y.C., X.H., W.W., J.C., X.-C.Z., M.-S.H., H.Z., J. Wang, X.Z., J.E.C., W.-C.S., K.H.P., S.W.S., X.-O.S., Y.-M.C., L.L., C.H.K., L.H., Y.-C.K., T.-Y.Y., J.X., P.G., W.T., J.S., C.-L.W., H.L., A.D.L.S., Z.Z., Y.C., Y.Y.C., J.-Y.H., J.S.K., H.-I.Y., Q.C., C.-C.L., I.K.P., P.X., J.D., Q.H., R.-P.P., T.K., S.-S.K., C.-Y.C., R.V., J. Wu, W.-Y.L., K.-C.C., W.-H.C., B.-T.J., J.K.C.C., M.C., Y.-J.L., J. Yokota, J. Li, H.C., Y.-B.X., C.-J.Y., H.K., G.W., L.J., Y.-L.L., K.S., Y.-L.W., P.-C.Y., M.-H.S., J.F.F., D. Lin, S.J.C. and N.R. conducted the epidemiological studies and contributed samples to the GWAS and/or follow-up genotyping. All authors contributed to the writing of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2456.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Sun, S., Schiller, J.H. & Gazdar, A.F. Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer* **7**, 778–790 (2007).
2. Sun, Y. *et al.* Lung adenocarcinoma from East Asian never-smokers is a disease largely defined by targetable oncogenic mutant kinases. *J. Clin. Oncol.* **28**, 4616–4620 (2010).
3. Rudin, C.M. *et al.* Lung cancer in never smokers: molecular profiles and therapeutic implications. *Clin. Cancer Res.* **15**, 5646–5661 (2009).
4. Thun, M.J. *et al.* Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS Med.* **5**, e185 (2008).
5. Gao, Y.T. *et al.* Lung cancer among Chinese women. *Int. J. Cancer* **40**, 604–609 (1987).
6. Gu, D. *et al.* Cigarette smoking and exposure to environmental tobacco smoke in China: the international collaborative study of cardiovascular disease in Asia. *Am. J. Public Health* **94**, 1972–1976 (2004).
7. Lan, Q., Chapman, R.S., Schreinemachers, D.M., Tian, L. & He, X. Household stove improvement and risk of lung cancer in Xuanwei, China. *J. Natl. Cancer Inst.* **94**, 826–835 (2002).
8. Couraud, S., Zalcman, G., Milleron, B., Morin, F. & Souquet, P.J. Lung cancer in never smokers—a review. *Eur. J. Cancer* **48**, 1299–1311 (2012).
9. Samet, J.M. *et al.* Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clin. Cancer Res.* **15**, 5626–5645 (2009).
10. Lo, Y.L. *et al.* Risk factors for primary lung cancer among never smokers by gender in a matched case-control study. *Cancer Causes Control* published online, doi:10.1007/s10552-012-9994-x (22 May 2012).
11. Hsiung, C.A. *et al.* The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS Genet.* **6**, e1001051 (2010).
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.* **85**, 679–691 (2009).
13. Hosgood, H.D. III *et al.* Genetic variant in *TP63* on locus 3q28 is associated with risk of lung adenocarcinoma among never-smoking females in Asia. *Hum. Genet.* **131**, 1197–1203 (2012).
14. Miki, D. *et al.* Variation in *TP63* is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nat. Genet.* **42**, 893–896 (2010).
15. Amundadottir, L. *et al.* Genome-wide association study identifies variants in the *ABO* locus associated with susceptibility to pancreatic cancer. *Nat. Genet.* **41**, 986–990 (2009).
16. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
17. 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).
18. McKay, J.D. *et al.* Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.* **40**, 1404–1406 (2008).
19. Truong, T. *et al.* Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J. Natl. Cancer Inst.* **102**, 959–971 (2010).
20. Shiraishi, K. *et al.* A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nat. Genet.* **44**, 900–903 (2012).
21. 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).
22. 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).

23. Wang, Y. *et al.* Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.* **40**, 1407–1409 (2008).
24. Wu, C. *et al.* Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. *Cancer Res.* **69**, 5065–5072 (2009).
25. Shi, J. *et al.* Inherited variation at chromosome 12p13.33, including *RAD52*, influences the risk of squamous cell lung carcinoma. *Cancer Discov.* **2**, 131–139 (2012).
26. Dong, J. *et al.* Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nat. Genet.* **44**, 895–899 (2012).
27. Wang, Z. *et al.* Improved imputation of common and uncommon SNPs with a new reference set. *Nat. Genet.* **44**, 6–7 (2012).
28. 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).
29. Ramirez, D.M., Khvotchev, M., Trauterman, B. & Kavalali, E.T. Vti1a identifies a vesicle pool that preferentially recycles at rest and maintains spontaneous neurotransmission. *Neuron* **73**, 121–134 (2012).
30. Kunwar, A.J. *et al.* Lack of the endosomal SNAREs vti1a and vti1b led to significant impairments in neuronal development. *Proc. Natl. Acad. Sci. USA* **108**, 2575–2580 (2011).
31. Bose, A. *et al.* The v-SNARE Vti1a regulates insulin-stimulated glucose transport and Acrp30 secretion in 3T3-L1 adipocytes. *J. Biol. Chem.* **280**, 36946–36951 (2005).
32. Petridou, E.T. *et al.* Circulating adiponectin levels and expression of adiponectin receptors in relation to lung cancer: two case-control studies. *Oncology* **73**, 261–269 (2007).
33. Bass, A.J. *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent *VTI1A-TCF7L2* fusion. *Nat. Genet.* **43**, 964–968 (2011).
34. Lemmon, M.A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
35. Acquaviva, J., Wong, R. & Charest, A. The multifaceted roles of the receptor tyrosine kinase ROS in development and cancer. *Biochim. Biophys. Acta* **1795**, 37–52 (2009).
36. Li, C. *et al.* Spectrum of oncogenic driver mutations in lung adenocarcinomas from East Asian never smokers. *PLoS ONE* **6**, e28204 (2011).
37. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 (2007).
38. Takeuchi, K. *et al. RET*, *ROS1* and *ALK* fusions in lung cancer. *Nat. Med.* **18**, 378–381 (2012).
39. Bergethon, K. *et al. ROS1* rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* **30**, 863–870 (2012).
40. Kim, M. *et al.* Epigenetic down-regulation and suppressive role of *DCBLD2* in gastric cancer cell proliferation and invasion. *Mol. Cancer Res.* **6**, 222–230 (2008).
41. Koshikawa, K. *et al.* Significant up-regulation of a novel gene, *CLCP1*, in a highly metastatic lung cancer subline as well as in lung cancers *in vivo*. *Oncogene* **21**, 2822–2828 (2002).
42. Nagai, H. *et al.* CLCP1 interacts with semaphorin 4B and regulates motility of lung cancer cells. *Oncogene* **26**, 4025–4031 (2007).
43. Silverberg, M.S. *et al.* Ulcerative colitis–risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat. Genet.* **41**, 216–220 (2009).
44. Urayama, K.Y. *et al.* Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status–defined subgroups. *J. Natl. Cancer Inst.* **104**, 240–253 (2012).
45. Nakanishi, K. & Shima, Y. Capture of type 1 diabetes–susceptible HLA DR-DQ haplotypes in Japanese subjects using a tag single nucleotide polymorphism. *Diabetes Care* **33**, 162–164 (2010).
46. Chanock, S.J. & Hunter, D.J. Genomics: when the smoke clears. *Nature* **452**, 537–538 (2008).
47. Spitz, M.R., Amos, C.I., Dong, Q., Lin, J. & Wu, X. The *CHRNA5-A3* region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *J. Natl. Cancer Inst.* **100**, 1552–1556 (2008).
48. Wang, Y., Broderick, P., Matakidou, A., Eisen, T. & Houlston, R.S. Chromosome 15q25 (*CHRNA3-CHRNA5*) variation impacts indirectly on lung cancer risk. *PLoS ONE* **6**, e19085 (2011).

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA. [2]Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan. [3]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. [4]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. [5]Saw Swee Hock School of Public Health, National University of Singapore, Singapore. [6]Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA. [7]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, USA. [8]Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China. [9]Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China. [10]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China. [11]Department of Pathology, Li Ka Shing (LKS) Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [12]Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Institute for Medicine and Public Health, Vanderbilt University, Nashville, Tennessee, USA. [13]Lung Cancer Center, Kyungpook National University Medical Center, Daegu, Republic of Korea. [14]Genomic Research Center, Taipei, Taiwan. [15]Department of Internal Medicine, Division of Oncology/Hematology, College of Medicine, Korea University Anam Hospital, Seoul, Republic of Korea. [16]Department of Thoracic and Cardiovascular Surgery, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea. [17]Ministry of Education Key Laboratory of Modern Toxicology, Nanjing Medical University, Nanjing, China. [18]Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Nanjing Medical University, Nanjing, China. [19]Westat, Rockville, Maryland, USA. [20]National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan. [21]Division of Thoracic Surgery, Kinki University School of Medicine, Sayama, Japan. [22]Genome Research Center for Hematopoietic Diseases, Chonnam National University Hwasun Hospital, Hwasun-eup, Republic of Korea. [23]Department of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan. [24]Division of Chest Medicine, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [25]Division of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA. [26]China National Environmental Monitoring Center, Beijing, China. [27]Department of Epidemiology, School of Public Health, China Medical University, Shenyang, China. [28]Department of Etiology & Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [29]State Key Laboratory of Molecular Oncology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [30]Guangdong Lung Cancer Institute, Medical Research Center and Cancer Center of Guangdong General Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China. [31]Department of Clinical Oncology, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [32]Department of Human Genetics, Genome Institute of Singapore, Singapore. [33]School of Life Sciences, Anhui Medical University, Hefei, China. [34]Cancer Research Center, Kyungpook National University Medical Center, Daegu, Republic of Korea. [35]Institute of Environmental Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea. [36]Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. [37]Department of Respiratory Therapy, Chang Gung Memorial Hospital, Chiayi, Taiwan. [38]Institute of Occupational Medicine and Ministry of Education Key Laboratory for Environment and Health, School of Public Health, Huazhong University of Science and Technology, Wuhan, China. [39]Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [40]Lung and Esophageal Cancer Clinic, Chonnam National University Hwasun Hospital, Hwasun-eup, Republic of Korea. [41]Department of Internal Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea. [42]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. [43]Chinese Center for Disease Control and Prevention, Beijing, China. [44]Department of Internal Medicine, Kaohsiung Medical University Hospital, School of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan. [45]Department of Biochemistry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [46]Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [47]Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan. [48]Department of Thoracic and Cardiovascular Surgery, Seoul National University Bundang Hospital, Seongnam, Republic of Korea. [49]Chest Department, Taipei Veterans General Hospital, Taipei, Taiwan. [50]Cancer Center, Union Hospital, Huazhong University of Science and Technology, Wuhan, China. [51]Division of Hematology and Oncology, Vanderbilt University Medical Center, Nashville, Tennessee, USA. [52]School of Public Health, The University of Hong Kong, Hong Kong, China. [53]Department of Pulmonary and Critical Care, Chang Gung Memorial Hospital, Taoyuan, Taiwan. [54]Department of Surgery, Division of Cardiothoracic Surgery, Queen Mary Hospital, Hong Kong, China. [55]Department of Internal Medicine, Division of Medical Oncology, College of Medicine, Korea University Guro Hospital, Seoul, Republic of Korea. [56]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea. [57]Department of Oncology, Wuhan Iron and Steel Corporation Staff Worker Hospital, Wuhan, China. [58]Division of Genome Biology, National Cancer Center Research Institute, Tokyo, Japan. [59]Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun-eup, Republic of Korea. [60]Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea. [61]Cancer Center, China Medical University and Hospital, Taichung, Taiwan. [62]Division of Environmental Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands. [63]Department of Pathology, Queen Elizabeth Hospital, Hong Kong, China. [64]Division of Multistep Carcinogenesis, National Cancer Center Research Institute, Tokyo, Japan. [65]Qujing Center for Diseases Control and Prevention, Sanjiangdadao, Qujing, China. [66]Department of Respiratory Medicine, Mitsui Memorial Hospital, Tokyo, Japan. [67]Department of Internal Medicine, National Taiwan University College of Medicine, Taipei, Taiwan. [68]These authors contributed equally to this work. [69]These authors jointly directed this work. Correspondence should be addressed to Q.L. (qingl@mail.nih.gov).