Genetic Modifiers of Progression-Free Survival in Never-Smoking Lung Adenocarcinoma Patients Treated with First-Line Tyrosine Kinase Inhibitors

I-Shou Chang^{1*}, Shih Sheng Jiang^{1*}, James Chih-Hsin Yang^{2,3*}, Wu-Chou Su⁴, Li-Hsin Chien⁵, Chin-Fu Hsiao^{5,6}, Jih-Hsiang Lee², Chih-Yi Chen^{7,8}, Chung-Hsing Chen¹, Gee-Chen Chang^{9,10}, Zhaoming Wang¹¹, Fang-Yi Lo⁵, Kuan-Yu Chen¹², Wen-Chang Wang^{5,13}, Yuh-Min Chen^{14,15}, Ming-Shyan Huang¹⁶, Ying-Huang Tsai¹⁷, Yu-Chun Su⁵, Wan-Shan Hsieh⁵, Wen-Chi Shih⁵, Shwn-Huey Shieh^{18,19}, Tsung-Ying Yang¹⁰, Qing Lan²⁰, Nathaniel Rothman²⁰, Chien-Jen Chen²¹, Stephen J. Chanock²⁰, Pan-Chyr Yang²², and Chao A. Hsiung⁵

¹National Institute of Cancer Research, ⁵Institute of Population Health Sciences, and ⁶Taiwan Lung Cancer Tissue/Specimen Information Resource Center, National Health Research Institutes, Zhunan, Taiwan; ²Department of Oncology and ²²Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan; ³Graduate Institute of Oncology and Cancer Research Center, College of Medicine and ¹²Division of Pulmonary Medicine, Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan; ⁴Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan; ⁷Institute of Medicine and ⁸Division of Thoracic Surgery, Department of Surgery, Chung Shan Medical University Hospital, Taichung, Taiwan; ⁹Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan; ¹⁰Division of Chest Medicine, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan; ¹¹Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, and ²⁰Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; ¹³The Ph.D. Program for Translational Medicine, College of Medical Science and Technology and ¹⁵College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan; ¹⁴Department of Chest Medicine, Taipei Veterans General Hospital, Taipei, Taiwan; ¹⁶Department of Internal Medicine, Kaohsiung Medical University Hospital, School of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan; ¹⁷Division of Pulmonary and Critical Care Medicine, Chiayi Chang Gung Memorial Hospital, Chiayi, Taiwan; ¹⁸Department of Health Services Administration and ¹⁹Department of Nursing, China Medical University, Taichung, Taiwan; and ²¹Genomic Research Center, Academia Sinica, Taipei, Taiwan

Abstract

Rationale: Patients with non-small cell lung cancer (NSCLC) with mutated epidermal growth factor receptor (*EGFR*) are relatively sensitive to *EGFR*-tyrosine kinase inhibitor (TKI) treatment and have longer progression-free survival (PFS) when treated with *EGFR*-TKI compared with platinum-based chemotherapy. However, many patients with advanced NSCLC who have mutated *EGFR* do not respond to first-line *EGFR*-TKI treatment and still have shorter PFS.

Objectives: The aim of this study was to identify genetic variants associated with PFS among patients with lung adenocarcinoma who were treated with first-line *EGFR*-TKIs.

Methods: A genome-wide association study on PFS was performed in never-smoking women diagnosed with lung adenocarcinoma and who were treated with first-line *EGFR*-TKIs (n = 128). Significant single-nucleotide polymorphisms (SNPs) were selected for follow-up association analysis (n = 198) and for replication assay in another independent cohort (n = 153).

Measurements and Main Results: We identified SNPs at 4q12 associated with PFS at genome-wide significance ($P < 10^{-8}$) and with an estimated hazard ratio of more than 4. This association was also replicated in a larger but similar cohort and in an independent NSCLC cohort. Follow-up functional analyses showed that these SNPs were associated with the expression of *EGFR*, which encodes the TKI target, and with a nearby gene neuromedin-U, which encodes a G protein-coupled receptor ligand known to be involved in the progression of NSCLC. Considering these as possible prognostic biomarkers for the treatment of patients with late-stage lung cancer, we found that these SNPs were not associated with *EGFR* mutation status or with polymorphism of the Bcl2-interacting mediator of cell death gene.

Conclusions: Genetic variants in 4q12 merit further investigation to assess their potential as pharmacogenomic predictors for and to understand the biology underlying its influence on PFS in patients treated with TKI therapy.

Keywords: lung neoplasms; molecular targeted therapy; never-smokers; genome-wide association study; single-nucleotide polymorphism

(Received in original form February 13, 2016; accepted in final form September 23, 2016)

*These authors contributed equally to this work.

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

Am J Respir Crit Care Med Vol 195, Iss 5, pp 663–673, Mar 1, 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201602-0300OC on September 26, 2016

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Patients with non-small cell lung cancer (NSCLC) with mutated *EGFR* are relatively sensitive to epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitor (TKI) treatment and have longer progression-free survival (PFS) when treated with EGFR-TKIs compared with platinum-based chemotherapy. However, many patients with advanced NSCLC who have mutated *EGFR* do not respond to first-line EGFR-TKI treatment and still have shorter PFS. The underlying mechanism of such a phenomenon remains unclear.

What This Study Adds to the

Field: Single-nucleotide polymorphisms (SNPs) at chromosome 4q12 were shown to be associated at a genome significance level with PFS of patients with lung adenocarcinoma treated with TKIs. Interestingly, these SNPs are associated with the expression of EGFR signaling pathway genes, including *EGFR*, which encodes the target of TKIs. More intriguingly, these SNPs are also associated with the expression of the gene *neuromedin-U*, which encodes a G protein-coupled receptor ligand known to be involved in the progression of NSCLC. These associations increase the likelihood that these SNPs may be potential biomarkers that may be useful in the treatment of patients with lung adenocarcinoma, and that one may establish the mechanism through which these SNPs modify the effect of TKIs on PFS.

Non-small cell lung cancer (NSCLC) is the most common form of lung cancer and is the leading cause of cancer-related death in many countries. Approximately two-thirds of patients with NSCLC are diagnosed at an advanced stage (1). Until recently, conventional treatment with cytotoxic chemotherapy provided limited efficacy and had significant toxicity (2). In 2004, three studies reported that tumors with epidermal growth factor receptor (*EGFR*) mutations were highly sensitive to a targeting agent, a tyrosine kinase inhibitor (TKI), which is known as EGFR-TKI (3–5). Subsequent phase III randomized clinical trials reported longer median progression-free survival (PFS) in patients with mutated *EGFR* treated with first-line EGFR-TKIs (gefitinib, erlotinib, and afatinib) compared with those treated with standard platinum-based chemotherapy (6–12). Testing for *EGFR* mutation has since become a standard of therapy in patients with advanced NSCLC (13). Our account of targeted therapy for NSCLC is brief; more comprehensive reviews have been published recently (14, 15).

Despite these promising results, a cumulative review of the literature has indicated that, overall, first-line EGFR-TKI treatment in patients with advanced *EGFR*-mutated NSCLC has yielded mixed results, and many patients continue to experience short PFS. Among patients with *EGFR*-mutated lung adenocarcinoma (ADC), some of those treated with a first-line TKI exhibited a shorter PFS compared with the median PFS reported for those treated with first-line standard chemotherapy (6). These results prompted a search for new biomarkers that could identify subgroups of patients who might benefit from the treatment.

In view of the concept that germ-line genetic variants can strongly influence both the pharmacokinetic and pharmacodynamic profiles of administered agents, and of the example that a genome-scale association between solute carrier organic anion transporter family member 1B1 (*SLCO1B1*) and statin-induced myopathy was established in a genome-wide association study (GWAS) using only 85 cases and 90 controls (16), we performed a GWAS to search for common variants associated with the PFS. In addition to replicating the initial GWAS findings that reached genome-wide significance, we conducted functional analyses to investigate the plausibility of this association further and to obtain clues regarding the biological mechanism responsible for this association (17). We note that a recent GWAS reported pharmacogenomics loci for thoracic diseases that reached genome-scale significance (18).

We examined whether the associated single-nucleotide polymorphisms (SNPs) are expression quantitative trait loci (eQTL) that correlate with the expression of *EGFR*, B-Raf proto-oncogene, serine/threonine kinase (*BRAF*), Erb-B2 receptor tyrosine kinase 2 (*ERBB2*), and growth factor receptor-bound protein 2 (*GRB2*), whose encoded products interact physically with the EGFR. We also examined whether these SNPs correlate with the expression of neighboring genes and if any of the *cis*-eQTL genes are involved in tumor progression (19). We used both in-house and public domain gene expression data derived from clinical specimens to examine the clinical relevance of the implicated genes (Figure 1). We believed that if these associations were consistent among themselves, we would gain confidence in

Supported by Taiwan National Science Council (NSC; now Ministry of Science and Technology), Department of Health (DOH; now Ministry of Health and Welfare), National Health Research Institutes, and in part by grants from the National Research Program for Biopharmaceuticals in Taiwan (DOH 101-TD-PB-111-TM015, DOH 102-TD-PB-111-TM024, Taiwan Lung Cancer Tissue/Specimen Information Resource Center (NSC 101-2325-B-400-020), and by Taiwan Bioinformatics Institute Core Facility (NSC 100-2319-B-400-001).

Author contributions: Jointly supervised the study: C.A.H., I.-S.C., S.S.J., J.C.-H.Y., C.-J.C., and P.-C.Y. Conceived and designed the experiment: C.A.H., I.-S.C., S.S.J., J.C.-H.Y., W.-C. Su, K.-Y.C., Y.-H.T., Q.L., N.R., C.-J.C., S.J.C., and P.-C.Y. Performed the experiments: S.S.J., J.C.-H.Y., J.-H.L., Z.W., F.-Y.L., and W.-S.H. Contributed to the design and execution of statistical analysis: C.A.H., I.-S.C., J.C.-H.Y., W.-C. Su, L.-H.C., C.-F.H., C.-Y.C., C.-H.C., G.-C.C., Z.W., F.-Y.L., K.-Y.C., W.-C.W., Y.-M.C., M.-S.H., Y.-H.T., Y.-C.S., W.-C. Shih, Q.L., N.R., C.-J.C., S.J.C., and P.-C.Y. Contributed to the bioinformatic analysis: I.-S.C., S.S.J., L.-H.C., and C.-H.C. Contributed reagents and materials: C.A.H., J.C.-H.Y., W.-C. Su, J.-H.L., C.-Y.C., G.-C.C., K.-Y.C., Y.-M.C., M.-S.H., Y.-H.T., S.-H.S., T.-Y.Y., Q.L., N.R., C.-J.C., S.J.C., and P.-C.Y. Wrote the first draft of the manuscript: C.A.H., I.-S.C., S.S.J., and S.J.C. Conducted the epidemiological studies: C.A.H., I.-S.C., J.C.-H.Y., W.-C. Su, L.-H.C., C.-Y.C., G.-C.C., K.-Y.C., Y.-M.C., M.-S.H., Y.-H.T., T.-Y.Y., C.-J.C., and P.-C.Y. Contributed to the recruitment of subjects: C.A.H., J.C.-H.Y., W.-C. Su, J.-H.L., C.-Y.C., G.-C.C., K.-Y.C., Y.-M.C., M.-S.H., Y.-H.T., S.-H.S., T.-Y.Y., C.-J.C., and P.-C.Y. All authors reviewed and approved the manuscript.

Correspondence and requests for reprints should be addressed to Chao A. Hsiung, Ph.D., Institute of Population Health Sciences, National Health Research Institutes, 35 Keyan Road, Zhunan, 35053, Taiwan. E-mail: hsiung@nhri.org.tw

Figure 1. Enhancing the biological plausibility of the association between single-nucleotide polymorphisms (SNPs) and progression-free survival (PFS) by expression quantitative trait loci (eQTL) analyses and associations between gene expression and clinical phenotypes. *BRAF* = B-Raf proto-oncogene, serine/threonine kinase; *CEP135* = centrosomal protein 135; *CLOCK* = clock circadian regulator; *EGFR* = epidermal growth factor receptor; *ERBB2* = Erb-B2 receptor tyrosine kinase 2; *GELAC* = Genetic Epidemiological Study of Lung Adenocarcinoma; *GRB2* = growth factor receptor-bound protein 2; *GWAS* = genome-wide association study; *NMU* = neuromedin-U; *NTU NSCLC* = National Taiwan University non-small cell lung cancer; *PDCL2* = phosphatidylcholine-specific phospholipase 2; *SRD5A3* = steroid 5 α -reductase 3; *TMEM165* = transmembrane protein 165.

the biological plausibility of the findings of this study. Considering the possibility of using these SNPs as prognosis biomarkers, we also examined whether these variants correlated with the *EGFR* mutation status, with the *Bcl2*-interacting mediator of cell death (*BIM*) polymorphism (20), the deletion of which predicts a shorter PFS and with PFS after treatment with standard chemotherapy.

Methods

Cohorts and Subcohorts

Cohorts and subcohorts of this study are shown in Figure 2. All patients with lung

ADC in GELAC (Genetic Epidemiological Study of Lung Adenocarcinoma) (see online supplement) who were treated with first-line *EGFR*-TKI therapy, who did not undergo resection, and who were recruited before August 23, 2010 were included in this study (Table 1); this cohort is termed TALL (n = 198). All subjects were followed until July 25, 2012. Never-smokers in TALL formed the subcohort TNS. Females in TNS formed the subcohort TFNS. Patients in TFNS and in our earlier GWAS (21, 22) formed the TGFNS cohort.

This study also used the PGFNS cohort, which consisted of never-smoking female patients with late-stage lung ADC who were included in our GWAS (21, 22) and

who were treated with first-line, platinum-based doublets, without resection. The recruitment and follow-up periods for PGFNS were similar to those reported for TGFNS.

Three-Step Genetic Association Study of PFS

The first step of our genetic association study of PFS was the standard GWAS based on TGFNS (n = 128). One region containing all the SNPs associated with PFS at genome-wide significance from this step was determined. We then obtained imputed genotype data at SNPs in this region for subjects in TGFNS. The second step examined the association between PFS and all the imputed and genotyped SNPs in this region. Based on the second step genetic association study, we selected SNPs for subsequent studies. SNPs associated with PFS at genome-wide significance and located in transcribed sequences were prioritized, and the collection of the selected SNPs spanned the region determined in the first step.

Eventually, five SNPs were chosen and genotyped in every patient in TALL using the MassARRAY (Sequenom Inc., San Diego, CA) system at the Genotyping Core of the Taiwan National Research Program for Genomic Medicine (NRPGM) or TaqMan (Applied Biosystems, Foster City, CA) at the Taiwan National Health Research Institutes (NHRI). The call rates of these five SNPs were between 96% and 100%. rs476184 and rs17725110 had already been genotyped with the Illumina Human 660W Quad BeadChip (Illumina Inc., San Diego, CA), and their concordance rates were 100% and 98.4%, respectively.

National Taiwan University NSCLC Validation Cohort

The National Taiwan University (NTU) NSCLC cohort had 153 patients with advanced NSCLC from two prospective studies that started in 2006 and 2008 (ClinicalTrials.gov Identifiers: NCT00173875 and NCT00977509, respectively). Patients received *EGFR*-TKIs as first-line therapy and were followed regularly until June 31, 2013. Use of blood samples and clinical data was approved by the Research Ethics Committee of the NTU Hospital.

Genotyping of the five SNPs, which were selected as previously described using the GELAC cohort, of the subjects in NTU NSCLC group was performed using the TaqMan assay. The characteristics of the

Figure 2. Cohorts and subcohorts used in this study. Cohorts from which association of a genome-wide level of significance ($P < 10^{-8}$) was generated or from which an association was successfully replicated ($P < 0.05$) ADC = adenocarcinoma; ALL-TA = all lung ADC patients from NTU NSCLC; ALL-TN = all NTU NSCLC patients; CMUH = China Medical University Hospital; EGFR-TKI = epidermal growth factor receptor–tyrosine kinase inhibitor; eQTL = expression quantitative trait loci; FNS-TA = female in NS-TA; FNS-TN = female in NS-TN; GELAC = Genetic Epidemiological Study of Lung Adenocarcinoma; GWAS = genome-wide association study; LCTC = Lung Cancer Tissue Cohort; LCTCSMB = patients in LCTC who have single-nucleotide polymorphism array data, *EGFR* mutation status and *Bcl2*-interacting mediator of cell death (*BIM*) polymorphism data; LCTCNS = never-smoking patients with lung ADC in LCTCSMB who have gene expression array data; NS-TA = never-smokers in ALL-TA; NS-TA-M = *EGFR* mutation positive in NS-TA; NS-TN = never-smokers in ALL-TN; NS-TN-M = *EGFR* mutation positive in NS-TN; NTU NSCLC = National Taiwan University non-small cell lung cancer; PGFNS = female never-smoking patients with late-stage lung ADC who were in our GWAS (21, 22) and treated with first-line platinum based doublets, without resection; TALL = all the patients with lung ADC who were treated with first-line EGFR-TKI, without resection; TFNS = female never-smokers in TALL; TGFNS = patients in TFNS and in our earlier GWAS (21, 22); TNS = never-smokers in TALL.

subjects in NTU NSCLC group are given in Table E1 in the online supplement. We noted that six patients belonged to both the NTU NSCLC and TALL groups of the GELAC cohorts; however, the main results of this study remain unaltered based on the association study without these six patients.

Tissue Cohort for eQTL Study

We gathered a lung cancer tissue cohort (LCTC) of patients with lung cancer who were treated surgically at the China Medical University Hospital, Taiwan. The LCTC was supported by the NRPGM, Taiwan, and approved by the institutional review boards

of the NHRI and of the China Medical University. We obtained informed consent from each subject in the LCTC and collected their tumor tissue, adjacent nontumor tissue, blood, and clinical information. In some of the LCTC subjects, microarray gene expression experiments were conducted using the Illumina WG-DASL HumanRef-8 v3 or HumanHT-12 v4 BeadChip (Illumina Inc.). Genome-scale genotype data based on buffy coat DNA were also obtained for some of these subjects using the Illumina Human 660W Quad BeadChip. Finally, we obtained the genome-scale expression and genotypic data of 115 never-smoking

patients with lung ADC, which we refer to as LCTCNS. The subcohort had 124 subjects, for whom we obtained SNP array data, *EGFR* mutation status, and *BIM* genotype data, was termed LCTCSMB. LCTCSMB contained LCTCNS.

The eQTL study was based on the LCTCNS group and was performed for only the five previously selected SNPs. The genotypes at three of the SNPs were obtained from imputation.

Allelic Expression Imbalance

Because three of the five selected SNPs are in transcribed regions, we conducted allele

Table 1. Basic Characteristics of the Genetic Epidemiological Study of Lung Adenocarcinoma Cohorts

Cohorts	TALL (n = 198)	TNS (n = 160)	TFNS (n = 147)	PGFNS (n = 360)
Age(yr)				
Median	65	65	64	57
Range	33–94	33–87	33–87	24–80
Sex, n (%)				
Male	43 (21.7)	13 (8.1)	0	0
Female	155 (78.3)	147 (91.9)	147 (100)	360 (100)
Smoking history, n (%)				
Never-smokers	160 (80.8)	160 (100)	147 (100)	360 (100)
Smokers	38 (19.2)	0	0	0
ECOG performance status, n (%)				
0	37 (18.7)	32 (20.0)	27 (18.4)	54 (15.0)
1	91 (46.0)	68 (42.5)	61 (41.5)	238 (66.1)
2	38 (19.2)	32 (20.0)	31 (21.1)	35 (9.7)
3	14 (7.1)	12 (7.5)	12 (8.2)	4 (1.1)
4	4 (2.0)	3 (1.9)	3 (2.0)	2 (0.6)
Unknown	14 (7.1)	13 (8.1)	13 (8.8)	27 (7.5)
Follow-up time from treatment to progression or censored time*				
Median, d	167	162	155	218
Range, d	5–1750	5–1750	5–1160	1–2972
Disease stage at diagnosis, n (%)				
IA	0	0	0	1 (0.3)
IB	0	0	0	1 (0.3)
IIA	0	0	0	1 (0.3)
IIB	1 (0.5)	0	0	1 (0.3)
IIIA	8 (4.0)	4 (2.5)	4 (2.7)	10 (2.8)
IIIB	30 (15.2)	22 (13.8)	19 (12.9)	53 (14.7)
IV	159 (80.3)	134 (83.8)	124 (84.4)	284 (78.9)
Unknown	0	0	0	9 (2.5)
First-line EGFR-TKI therapy, n (%)				
Gefitinib (250 mg daily)	178 (89.9)	147 (91.9)	137 (93.2)	
Erlotinib (150 mg daily)	18 (9.1)	12 (7.5)	9 (6.1)	
Erlotinib (100 mg daily)	2 (1.0)	1 (0.6)	1 (0.7)	
First-line platinum-based, n (%)				
Cisplatin + gemzar				221 (61.4)
Cisplatin + taxotere				28 (7.8)
Cisplatin + taxol				30 (8.3)
Cisplatin + navelbine				58 (16.1)
Cisplatin + alimta				6 (1.7)
Carboplatin + gemzar				2 (0.6)
Carboplatin + taxotere				1 (0.3)
Carboplatin + taxol				13 (3.6)
Carboplatin + navelbine				1 (0.3)

Definition of abbreviations: ECOG = Eastern Cooperative Oncology Group; EGFR-TKI = epidermal growth factor receptor-tyrosine kinase inhibitor; PGFNS = female never-smoking patients with late-stage lung adenocarcinoma who were in our genome-wide association study (21, 22) and treated with first-line platinum based doublets, without resection; TALL = all the patients with lung adenocarcinoma who were treated with first-line EGFR-TKIs, without resection; TFNS = female never-smokers in TALL; TNS = never-smokers in TALL.

*That the median for the patients treated with platinum-based doublets seemed to be longer needs some explanations. First, this table shows that compared with the groups treated with TKI, the group treated with platinum-based doublets was approximately 7 years younger and was healthier in terms of ECOG performance status. Second, a subset of the group treated with platinum-based doublets were selected to match the age and ECOG performance status of the group treated with TKIs; comparison of the progression-free survival between these two groups indicated no significant difference between them, using either a log-rank test or a two-sample test for right censored data.

specific expression studies for them. Using adjacent nontumor tissues, we studied the allelic expression imbalance (AEI) for rs576732, which is located in the 3'-untranslated region (UTR) of steroid 5

α-reductase 3 (*SRD5A3*); for rs1801260, which is located in the 3'-UTR of clock circadian regulator (*CLOCK*); and for rs3805383, which is a synonymous variant of *NMU*.

For each of these experiments, all heterozygous subjects in LCTCNS were selected, and their genomic DNA (gDNA) and complementary RNA (cRNA) were subjected to pyrosequencing using a standard protocol (23). cRNA was reverse transcribed to cDNA using the SuperScript III kit (Life Technologies, Foster City, CA) and random hexamers as primers. Polymerase chain reaction and sequencing primers were designed using PyroMark Assay Design 2.0 software (Qiagen, Crawley, UK) (see Table E2).

Other details of AEI experiments are available as Methods in the online supplement.

Statistical Analysis

The association between a SNP and PFS was assessed using Cox's regression model with age and additively coded genotype as covariates. For the eQTL study, we applied a linear regression model to each SNP-gene probe pair, with the adjusted logarithmic expression level being the response variable and additively coded genotype as the covariate. The adjusted logarithmic expression levels were the residuals obtained from a linear regression performed after normalization and surrogate variable analysis, described in HIGH-THROUGHPUT DATA GENERATION AND PREPROCESSING in the online supplement. We used a probe to examine whether the expression level at this probe in the tumor tissue differed from that in the adjacent nontumor tissue, and performed paired *t* tests based on the adjusted logarithmic expression level to report the *P* value. The expression data used in this study were deposited in Gene Expression Omnibus (accession number GSE46539). The associations between the expression of a gene and time-to-recurrence or relapse for data in the public domain were studied using Cox's regression model, without adjusting the expression levels.

The association between an SNP and *EGFR* mutation status or *BIM* polymorphism was similarly studied, with the exception that logistic regression models were used.

Results

Genetic Association

We conducted a discovery GWAS based on 128 female never-smoking patients with lung ADC who were treated with first-line

EGFR-TKI (without resection) and who were included in the GELAC component of the previous GWAS of female lung ADC susceptibility (21, 22) (Table 1). After the application of rigorous quality control metrics, we analyzed 418,580 SNPs in these 128 patients with lung ADC. The analyses were based on Cox's regression model, which was adjusted for age only, because Eastern Cooperative Oncology Group performance status and tumor stage were not significant covariates. The Q-Q plot and the Manhattan plot are shown in Figure E1 in the online supplement.

Nine correlated SNPs ($r^2 > 0.87$) located at 4q12 were significantly associated with PFS; all had P values $< 10^{-7}$, and seven had P values $< 10^{-8}$. Interestingly, we observed that most had elevated estimated hazard ratios (≥ 5). Table E3 shows all the significant associations from the GWAS, including the additional SNPs identified by imputation in this region. Five SNPs (rs576732, rs476184, rs1801260, rs17725110, and rs3805383) located at 4q12 were chosen for follow-up analyses in the

TALL cohort ($n = 198$) of all the patients with lung ADC who were treated with first-line EGFR-TKIs, without resection, in the GELAC cohort. The analysis showed that rs576732, rs476184, and rs1801260 were associated with PFS at the genome-wide level, with P values of 6.3×10^{-8} , 8.5×10^{-9} , and 1.6×10^{-8} , respectively. The P values for two other SNPs, rs17725110 and rs3805383, approached genome-wide significance. Thus, the results of the collective analyses point toward a locus that contributes to PFS after EGFR-TKI therapy. Table 2 reports the results of this genetic association study based on three sub-cohorts of GELAC. Figure E2 provides the recombination rate and linkage disequilibrium plot in this region.

For each of these five SNPs, we also used the log-rank test to compare PFS in the three groups defined according to the number of risk alleles and found that all yielded similar results. For example, rs476184 (see Figure E3) was associated with a significant difference in PFS ($P = 1.3 \times 10^{-14}$).

Because most of the patients were never-smokers, it is plausible that the effect

pertains to only this subcategory of patients with lung cancer. The results were validated only in the never-smokers in the NTU NSCLC cohort, as described in METHODS and Table E1.

Independent replication of the association between PFS and the 4q12 variants was pursued in 153 patients with advanced NSCLC drawn from the NTU NSCLC cohort. These patient cases showed that the EGFR mutation status predicted longer PFS and that the BIM deletion polymorphism predicted shorter survival in late-stage NSCLC (24, 25).

Table 3 shows the replication study based on the NTU NSCLC cohort, in which we validated the association in the never-smokers in the GELAC cohort. In particular, it shows that the P values for rs476184 were 0.01 and 0.02 among never-smoking patients with lung ADC and never-smoking patients with NSCLC, respectively. The data listed in Table 3 also suggest the presence of a signal based on the smaller set of never-smokers with an EGFR mutation. The results obtained using the log-rank test for the NTU NSCLC cohort provide further analytical support of the involvement of the 4q12 locus (see Figure E4).

Because six patients in the NTU NSCLC group were included already in GELAC, we also conducted a replication study without these six patients and report the results in Table E4. The results in Table E4 are similar to those in Table 3.

Associations between SNPs, Gene Expressions, and Survival

We examined the association between SNPs at 4q12 and the expression of EGFR, BRAF, ERBB2, and GRB2, which are not on chromosome 4, and the expression of genes located within 500 kb from any of these SNPs. Using Illumina's expression chip, we considered the eight nearest genes, centrosomal protein 135 (CEP135), CLOCK, exocyst complex component 1 (EXOC1), kinase insert domain receptor (KDR), NMU, phosphatidylinositol-3-OH kinase class II (PI3K), and transmembrane protein 165 (TMEM165). We first examined whether these 12 genes were associated with tumor status and the time-to-recurrence of tumors.

Data were obtained from the tumor and adjacent nontumor tissues of 115 never-smoking patients with lung ADC from the in-house LTCNS group (see METHODS). Table E5a shows that, except for CEP135 and CLOCK, the expression level of each of the other 10 genes in lung ADC tissues

Table 2. Association between SNPs at 4q12 and PFS Based among TFNS, TNS, and All Patients with Lung ADC Treated with First-Line TKI in GELAC Cohort

SNP	Chr	Position*	Allele†	MAF‡	P Value§	Hazard Ratio§ (95% CI)
rs576732	4	55931894	G/A	0.08	4.6×10^{-7} 5.8×10^{-8} 6.3×10^{-8}	3.72 (2.23–6.20) 3.97 (2.41–6.52) 3.54 (2.24–5.59)
rs476184	4	55941621	T/C	0.0718	3.2×10^{-8} 3.1×10^{-8} 8.5×10^{-9}	4.58 (2.67–7.85) 4.90 (2.90–8.28) 4.05 (2.51–6.51)
rs1801260	4	55996126	T/C	0.0704	2.8×10^{-8} 5.7×10^{-9} 1.6×10^{-8}	4.62 (2.69–7.94) 4.91 (2.88–8.39) 4.04 (2.49–6.55)
rs17725110	4	56166779	A/G	0.0631	4.1×10^{-6} 5.4×10^{-7} 7.7×10^{-7}	3.64 (2.10–6.31) 3.92 (2.30–6.69) 3.49 (2.13–5.73)
rs3805383	4	56170095	C/T	0.0541	3.5×10^{-4} 8.0×10^{-5} 4.7×10^{-5}	2.88 (1.61–5.14) 3.09 (1.76–5.42) 2.90 (1.74–4.84)

Definition of abbreviations: ADC = adenocarcinoma; Chr = chromosome; CI = confidence interval; EGFR = epidermal growth factor receptor; GELAC = Genetic Epidemiological Study of Lung Adenocarcinoma; MAF = minor allele frequency; PFS = progression-free survival; SNP = single-nucleotide polymorphism; TALL = all the patients with lung adenocarcinoma who were treated with first-line EGFR-TKIs, without resection; TKI = tyrosine kinase inhibitor; TFNS = female never-smokers in TALL; TNS = never-smokers in TALL.

*The positions of the SNPs were taken from human NCBI (National Center for Biotechnology Information) dbSNP build 36 (h18).

†Minor allele listed second.

‡MAF, based on TALL (all of the 198 subjects).

§ P values and hazard ratios for each SNP were obtained from fitting a Cox's regression model for each of the cohorts—TFNS, TNS, and TALL (all)—adjusted for age. The three coordinates in each cell were ordered to give corresponding values for TFNS, TNS, and TALL. Characteristics of TFNS, TNS, and TALL are given in Table 1.

||The largest hazard ratios appeared in the group of never-smokers.

Table 3. Replication of the Associations between the SNPs at 4q12 and PFS among Never-Smokers, Based on NTU NSCLC

SNP	NSCLC			ADC		
	MAF*	P Value ^{†‡}	Hazard Ratio [†] (95% CI)	MAF*	P Value ^{†‡}	Hazard Ratio ^{†‡} (95% CI)
rs576732	0.09	2.06×10^{-1}	1.49 (0.80–2.76)	0.10	9.15×10^{-2}	1.72 (0.92–3.21)
		2.01×10^{-2}	1.88 (1.10–3.20)		1.10×10^{-2}	2.01 (1.17–3.46)
		2.12×10^{-1}	1.33 (0.85–2.09)		1.48×10^{-1}	1.4 (0.89–2.21)
		3.49×10^{-2}	2.34 (1.06–5.14)		1.89×10^{-2}	2.63 (1.17–5.88)
		4.70×10^{-1}	1.32 (0.63–2.76)		3.64×10^{-1}	1.42 (0.67–3.00)
rs476184	0.09	2.06×10^{-1}	1.49 (0.80–2.76)	0.10	9.15×10^{-2}	1.72 (0.92–3.21)
		2.01×10^{-2}	1.88 (1.10–3.20)		1.10×10^{-2}	2.01 (1.17–3.46)
		2.33×10^{-1}	1.31 (0.84–2.06)		1.67×10^{-1}	1.38 (0.87–2.17)
		3.49×10^{-2}	2.34 (1.06–5.14)		1.89×10^{-2}	2.63 (1.17–5.88)
		4.70×10^{-1}	1.32 (0.63–2.76)		3.64×10^{-1}	1.42 (0.67–3.00)
rs1801260	0.09	2.06×10^{-1}	1.49 (0.80–2.76)	0.10	9.15×10^{-2}	1.72 (0.92–3.21)
		2.01×10^{-2}	1.88 (1.10–3.20)		1.10×10^{-2}	2.01 (1.17–3.46)
		2.33×10^{-1}	1.31 (0.84–2.06)		1.67×10^{-1}	1.38 (0.87–2.17)
		3.49×10^{-2}	2.34 (1.06–5.14)		1.89×10^{-2}	2.63 (1.17–5.88)
		4.70×10^{-1}	1.32 (0.63–2.76)		3.64×10^{-1}	1.42 (0.67–3.00)
rs17725110	0.09	4.31×10^{-1}	1.25 (0.72–2.18)	0.09	2.66×10^{-1}	1.37 (0.79–2.37)
		5.20×10^{-2}	1.58 (1.00–2.49)		3.46×10^{-2}	1.64 (1.04–2.59)
		3.05×10^{-1}	1.25 (0.82–1.90)		2.32×10^{-1}	1.29 (0.85–1.97)
		2.45×10^{-1}	1.45 (0.77–2.73)		1.99×10^{-1}	1.51 (0.80–2.84)
		7.70×10^{-1}	1.10 (0.58–2.09)		6.83×10^{-1}	1.14 (0.60–2.18)
rs3805383	0.09	3.80×10^{-1}	1.26 (0.76–2.09)	0.10	2.29×10^{-1}	1.36 (0.82–2.26)
		5.66×10^{-2}	1.51 (0.99–2.32)		3.77×10^{-2}	1.57 (1.03–2.41)
		2.89×10^{-1}	1.24 (0.83–1.84)		2.19×10^{-1}	1.28 (0.86–1.91)
		2.45×10^{-1}	1.45 (0.77–2.73)		1.99×10^{-1}	1.51 (0.80–2.84)
		7.70×10^{-1}	1.10 (0.58–2.09)		6.83×10^{-1}	1.14 (0.60–2.18)

Definition of abbreviations: ADC = adenocarcinoma; ALL-TA = all lung ADC patients from NTU NSCLC; ALL-TA-M = *EGFR* mutation positive patients in ALL-TA; ALL-TN = all NTU NSCLC patients; ALL-TN-M = *EGFR* mutation positive patients in ALL-TN; CI = confidence interval; *EGFR* = epidermal growth factor receptor; FNS-TA = female in NS-TA; FNS-TN = female in NS-TN; MAF = minor allele frequency; NS-TA = never-smokers in ALL-TA; NS-TA-M = *EGFR* mutation positive in NS-TA; NS-TN = never-smokers in ALL-TN; NS-TN-M = *EGFR* mutation positive in NS-TN; NTU NSCLC = National Taiwan University non-small cell lung cancer; PFS = progression-free survival; SNP = single-nucleotide polymorphism; TKI = tyrosine kinase inhibitor.

*MAF, based on all the NSCLC and all of the patients with lung ADC from the NTU NSCLC cohorts, respectively.

[†]P values and hazard ratios were obtained from the same Cox's model as those in Table 2. The five coordinates for each SNP in the left part contain results for FNS-TN, NS-TN, ALL-TN, NS-TN-M, and ALL-TN-M; those in the right part contain results for FNS-TA, NS-TA, ALL-TA, NS-TA-M, and ALL-TA-M. These subcohorts are defined in Tables E1a-d in the online supplement. For example, NS-TN consists of all the 113 never-smoking patients in NTU NSCLC; NS-TN-M consists of all of the 54 never-smoking patients in NTU NSCLC whose *EGFR* mutation status were tested and were positive; NS-TA consists of all of the 105 patients with ADC in NS-TN; and NS-TA-M consists of all the 50 patients with ADC in NS-TN-M.

[‡]P < 0.05 are in bold; they hold for all the never-smoking subcohorts for SNPs rs576732, rs476184, rs1801260. For the other two SNPs, they are generally also small.

differed from that in the adjacent nontumor tissues. Table E5b shows that, except for *CEP135*, the expression of each of the remaining genes differed between the tumor tissues and adjacent nontumor tissues in at least one of the five microarray data sets that were derived from clinical lung tissue samples in the public domain (see METHODS). Among these genes, *NMU*, *SRD5A3*, *EGFR*, *KDR*, and *BRAF* showed the largest-fold change in expression level according to Table E5a. Table E5c shows that except for *EXOC1*, the expression of all of the other genes was associated with the time-to-recurrence or relapse of tumors in patients with lung ADC or patients with NSCLC on the microarray data sets and time-to-recurrence or relapse information from GSE31210, GSE10245, and GSE8894 (see

online supplement). The number of significant correlations observed suggests that it was unlikely that they all occurred by chance.

Using the data from 115 patients from the LCTCNS group, we report the eQTL results in four tables: Tables 4, 5, and 6 for rs3805383 and Table E6 for rs1801260, rs17725110, rs576732, and rs476184. The expression of all of these genes, except for *CEP135*, *KDR*, and *TMEM165*, were associated to varying degrees with the five SNPs, and the three types of associations were compatible when they were statistically significant. For example, the data in Tables 2 to 6 and Table E5c together indicate that the increase in the number of minor alleles at rs3805383 correlated with the decrease in PFS and with the increase in *EGFR* expression in the tumor tissue;

the latter correlated with the decrease in the time-to-recurrence or relapse, although the mRNA probes differed between the different platforms. Similar phenomena were observed to varying degrees for *BRAF*, *ERBB2*, *GRB2*, *CLOCK*, *NMU*, *PDCL2*, and *SRD5A3* in both tumor and nontumor tissues. Among these eight genes, the expression of *CLOCK*, *NMU*, and *SRD5A3* continued to be associated with some of these SNPs, including rs3805383, based on the eQTL analyses of the 274 nontumor lung tissues from mainly non-Asian patients in the Genotype-Tissue Expression (GTEx) study (26) (Tables 4, 5, 6, and E6).

One of the mechanisms that can account for how *cis*-eQTLs work is that these DNA sequence variations are

Table 4. Expression Quantitative Trait Locus Analysis for rs3805383

Sample	BRAF: ILMN_1652472			CLOCK: ILMN_1682399			EGFR			ERBB2		
	ILMN_1652472			ILMN_1682399			ILMN_175535			ILMN_1717902		
	β	P Value		β	P Value		β	P Value		β	P Value	
All (<i>n</i> = 115)	0.118	0.038		0.000	0.998		0.279	0.040		−0.087	0.087	
T	0.031	0.438		−0.078	0.237		0.049	0.609		−0.072	0.152	
N												
EGFR mut ⁺ (<i>n</i> = 72)	0.198	0.008		−0.024	0.815		0.398	0.016		−0.112	0.098	
T	−0.007	0.883		−0.188	0.015		0.037	0.777		−0.038	0.559	
N												
L858R (<i>n</i> = 46)	0.339	0.000		−0.004	0.980		0.755	0.001		−0.177	0.046	
T	−0.030	0.643		−0.216	0.020		0.022	0.886		−0.016	0.862	
N												
GTEx	ENSG 00000157764.7	0.38		ENSG 00000134852.10	0.036		ENSG 00000146648.11	0.1		ENSG 00000141736.9	0.3	
	−0.03			−0.12			0.07			−0.042		
Sample	EXOC1: ILMN_1745583			GRB2: ILMN_1748797			NMU: ILMN_2162253			PDCL2: ILMN_1762409		
	ILMN_1745583			ILMN_1748797			ILMN_2162253			ILMN_1762409		
	β	P Value		β	P Value		β	P Value		β	P Value	
All (<i>n</i> = 115)	0.108	0.100		0.060	0.094		−0.071	0.774		0.149	0.102	
T	0.142	0.013		−0.016	0.637		0.283	0.051		−0.090	0.106	
N												
EGFR mut ⁺ (<i>n</i> = 72)	0.094	0.287		0.065	0.166		0.300	0.349		0.269	0.027	
T	0.112	0.108		−0.010	0.821		0.297	0.102		−0.105	0.155	
N												
L858R (<i>n</i> = 46)	0.199	0.077		0.036	0.517		0.725	0.058		0.208	0.143	
T	0.161	0.061		−0.046	0.437		0.515	0.044		−0.121	0.215	
N												
GTEx	ENSG 00000090989.13	0.79		ENSG 00000177885.9	0.58		ENSG 00000109255.7	0.24		ENSG 00000128039.6	0.06	
	0.01			−0.025			0.024			−0.16		

Definition of abbreviations: BRAF = B-Raf proto-oncogene, serine/threonine kinase; CLOCK = clock circadian regulator; EGFR = epidermal growth factor receptor; EGFR mut⁺ = EGFR mutation positive; ENSG = prefix of ID of Ensembl genes; ERBB2 = ErbB2 receptor tyrosine kinase 2; EXOC1 = exocyst complex component 1; GRB2 = growth factor receptor-bound protein 2; GTEx = genotype-tissue expression; ILMN = prefix of ID of genes' probe designed by Illumina; L858R = never-smoking patients with lung adenocarcinoma in Lung Cancer Tissue Cohort who have the single-nucleotide polymorphism array data, EGFR mutation status data, genotype data of the Bcl2-interacting mediator of cell death gene, and gene expression array data; N = nontumor; NMU = neuromedin-U; PDCL2 = phosphodiesterase 2; SMD5A3 = steroid 5 α -reductase 3; T = tumor.

The first column describes sample information, including sample size and tissue status (tumor or nontumor). Each of the other columns relates to one gene. The column heads give the gene names and the expression probes. The expression results are given based on all of the 115 samples in L858R, the subset of 72 samples with a common mutation (exon 19 deletion or L858R), and the subset of 46 samples with the L858R mutation. For each sample, results are given based on tumor samples and on adjacent nontumor tissues; tumor tissues and nontumor tissues are paired. A probe is included if the association between any of the five single-nucleotide polymorphisms and the expression in any of the six tissue types has a *P* value < 0.05. Results in the last rows were obtained from GTEx portal on May 18, 2016, using 274 nontumor lung tissue. GENCODE ID is shown for each gene.

P < 0.05 are in bold.

Table 5. Mean Levels by Genotype of rs3805383 for the Expression Quantitative Trait Locus Analysis Reported in Table 4 by Its Raw Data from the Experiments

Sample	Tissue	No. Risk Allele	BRAF:		CLOCK:		EGFR		ERBB2		EXOC1:		GRB2:		NMU:		PDCL2:		SRD5A3:	
			ILMN_1652472	ILMN_1682399	ILMN_1755535	ILMN_1696521	ILMN_1717902	ILMN_1728761	ILMN_2352131	ILMN_1745583	ILMN_1748797	ILMN_2162253	ILMN_1762409	ILMN_1678435						
All (n = 115)	T	0	10.930 (1.449)	9.124 (0.977)	7.043 (1.357)	10.869 (1.581)	6.300 (0.618)	6.546 (1.025)	13.214 (1.047)	10.753 (1.25)	11.018 (1.136)	9.338 (1.678)	5.916 (0.588)	9.191 (1.966)						
		1	11.257 (1.338)	9.389 (1.009)	7.446 (1.71)	11.281 (1.283)	6.328 (0.709)	6.712 (1.021)	13.572 (0.938)	10.859 (1.267)	11.117 (1.083)	9.145 (1.754)	5.963 (0.984)	9.515 (2.053)						
	2	13.839 (0)	9.071 (0)	8.114 (0)	13.674 (0)	5.766 (0)	6.434 (0)	14.720 (0)	11.294 (0)	11.913 (0)	13.281 (0)	6.412 (0)	12.767 (0)							
	N	0	10.690 (1.271)	9.093 (0.96)	8.006 (1.373)	10.659 (1.342)	6.248 (0.661)	6.368 (1.037)	13.04 (0.878)	11.294 (1.155)	11.097 (0.986)	7.32 (1.086)	5.725 (0.63)	8.833 (1.736)						
EGFR mut ⁺ (n = 72)	T	1	11.010 (1.195)	7.936 (1.069)	7.213 (1.61)	10.844 (1.181)	6.396 (0.637)	6.444 (0.835)	14.032 (0.841)	11.093 (1.15)	11.348 (0.878)	7.648 (1.138)	5.835 (0.528)	9.115 (1.796)						
		2	12.086 (0)	7.986 (0)	6.229 (0)	12.418 (0)	5.894 (0)	6.113 (0)	14.042 (0)	10.87 (0)	11.346 (0)	8.748 (0)	5.058 (0)	10.939 (0)						
	0	10.789 (1.405)	9.02 (0.874)	7.03 (1.303)	10.927 (1.475)	6.25 (0.639)	6.46 (0.99)	13.18 (1.046)	10.607 (1.283)	10.832 (1.128)	9.804 (1.663)	5.838 (0.539)	8.979 (1.878)							
	1	11.584 (1.405)	9.2 (0.874)	7.315 (1.621)	11.662 (1.4)	6.33 (0.479)	6.585 (0.886)	13.806 (0.884)	10.825 (1.274)	11.222 (0.962)	9.919 (1.383)	6.086 (1.145)	8.036 (2.391)							
L858R (n = 46)	T	2	13.839 (0)	9.071 (0)	8.114 (0)	13.674 (0)	5.766 (0)	6.434 (0)	14.72 (0)	11.294 (0)	11.913 (0)	13.281 (0)	6.412 (0)	12.767 (0)						
		1	10.524 (1.269)	8.016 (0.958)	6.71 (1.313)	10.491 (1.356)	6.177 (0.616)	6.231 (0.863)	12.952 (0.871)	10.542 (1.099)	10.909 (0.983)	7.252 (1.082)	5.66 (0.586)	8.615 (1.718)						
	N	0	11.088 (1.196)	8.371 (0.898)	6.867 (1.371)	11.045 (1.024)	6.377 (0.38)	6.376 (0.589)	13.345 (0.9)	10.917 (0.945)	11.329 (0.783)	7.762 (1.089)	5.837 (0.422)	9.271 (1.832)						
	2	12.086 (0)	7.986 (0)	6.229 (0)	12.418 (0)	5.894 (0)	6.113 (0)	14.042 (0)	10.87 (0)	11.346 (0)	8.748 (0)	5.058 (0)	10.939 (0)							
	T	0	11.084 (1.636)	9.159 (1.177)	7.247 (1.543)	11.147 (1.738)	6.396 (0.698)	6.712 (1.115)	13.314 (1.204)	10.842 (1.404)	11.055 (1.22)	9.586 (1.681)	5.94 (0.567)	9.432 (2.037)						
		1	11.764 (1.379)	9.176 (0.769)	7.308 (1.209)	12.935 (1.559)	6.324 (0.281)	6.566 (0.643)	14.003 (0.859)	10.787 (1.662)	11.155 (1.066)	10.467 (1.259)	5.722 (0.645)	10.365 (2.977)						
	N	0	13.839 (0)	9.071 (0)	8.114 (0)	13.674 (0)	5.766 (0)	6.434 (0)	14.72 (0)	11.294 (0)	11.913 (0)	13.281 (0)	6.412 (0)	12.767 (0)						
	2	10.808 (1.424)	9.11 (1.123)	6.936 (1.507)	10.757 (1.198)	6.317 (0.646)	6.4 (0.966)	13.138 (0.913)	10.757 (1.198)	11.078 (1.111)	7.539 (1.147)	5.750 (0.617)	9.027 (0.897)							
	N	1	11.169 (0.803)	9.059 (0.615)	6.609 (0.639)	11.22 (0.799)	6.482 (0.426)	6.443 (0.369)	14.046 (0.736)	10.719 (0.803)	11.251 (0.53)	8.963 (0.635)	5.859 (0.344)	9.451 (0.62)						
		2	12.086 (0)	7.986 (0)	6.229 (0)	12.418 (0)	5.894 (0)	6.113 (0)	14.042 (0)	10.870 (0)	11.346 (0)	8.748 (0)	5.058 (0)	10.939 (0)						

Definition of abbreviations: BRAF = B-Raf proto-oncogene, serine/threonine kinase; CLOCK = clock circadian regulator; EGFR = epidermal growth factor receptor; EGFR mut⁺ = EGFR mutation positive; ERBB2 = Erb-B2 receptor tyrosine kinase 2; EXOC1 = exocyst complex component 1; GRB2 = growth factor receptor-bound protein 2; ILMN = prefix of ID of genes' probe designed by Illumina; N = nontumor; NMU = neuromedin-U; PDCL2 = phosphodiesterase 3; T = tumor.

The genotype is coded by the number of minor alleles. The detail of the adjustment is described in high-throughput data generation and preprocessing (see the online supplement). The first and second columns describe the sample type and tissue type, respectively. The third column gives the genotype information. Each of the remaining columns gives the mean levels and the SD in the parentheses. *P* values < 0.05 are in bold.

Table 6. Mean Levels by Genotype of rs3805383 for the Expression Quantitative Trait Locus Analysis Reported in Table 4 by the Adjusted Logarithmic Expression Level

Sample	Tissue	No. Risk Allele	BRAF:		CLOCK:		EGFR		ERBB2		EXOC1:		GRB2:		NMU:		PDCL2:		SRD5A3:	
			ILMN_1652472	ILMN_1682399	ILMN_1755535	ILMN_1696521	ILMN_1717902	ILMN_1728761	ILMN_2352131	ILMN_1745583	ILMN_1748797	ILMN_2162253	ILMN_1762409	ILMN_1678435						
All (n = 115)	T	0	0.113 (0.218)	0.021 (0.303)	0.122 (0.511)	0.083 (0.436)	0.027 (0.206)	0.051 (0.259)	0.068 (0.173)	0.036 (0.26)	0.034 (0.145)	0.753 (0.959)	0.062 (0.314)	0.194 (0.373)						
		1	0.166 (0.221)	0.03 (0.256)	0.212 (0.571)	0.118 (0.434)	-0.024 (0.139)	0.228 (0.314)	0.142 (0.125)	0.135 (0.256)	0.084 (0.12)	0.383 (0.967)	0.125 (0.572)	0.251 (0.382)						
	N	2	0.766 (0)	-0.035 (0)	1.9 (0)	0.544 (0)	-0.386 (0)	0.241 (0)	0.46 (0)	0.308 (0)	0.224 (0)	2.544 (0)	0.917 (0)	1.106 (0)						
		0	-0.128 (0.152)	-0.021 (0.251)	-0.158 (0.384)	-0.098 (0.252)	-0.011 (0.203)	-0.058 (0.265)	-0.088 (0.086)	-0.071 (0.232)	-0.042 (0.137)	-0.786 (0.569)	-0.081 (0.216)	-0.219 (0.233)						
EGFR mut ⁺ (n = 72)	T	2	-0.075 (0.182)	-0.053 (0.306)	-0.167 (0.357)	-0.053 (0.205)	-0.075 (0.154)	-0.172 (0.228)	-0.024 (0.12)	0.055 (0.148)	-0.04 (0.092)	-0.515 (0.584)	-0.133 (0.237)	-0.209 (0.207)						
		0	-0.209 (0)	-0.471 (0)	0.318 (0)	-0.067 (0)	-0.204 (0)	-0.158 (0)	-0.04 (0)	0.32 (0)	-0.194 (0)	-0.141 (0)	-0.506 (0)	-0.112 (0)						
	N	1	0.117 (0.223)	-0.007 (0.325)	0.175 (0.481)	0.163 (0.361)	0.028 (0.218)	0.059 (0.272)	0.067 (0.16)	0.016 (0.272)	0.034 (0.151)	0.78 (1.001)	0.042 (0.298)	0.181 (0.385)						
		0	0.237 (0.248)	-0.037 (0.234)	0.288 (0.579)	0.245 (0.505)	-0.026 (0.081)	0.17 (0.28)	0.189 (0.083)	0.077 (0.305)	0.081 (0.078)	0.723 (0.862)	0.208 (0.805)	0.391 (0.389)						
L858R (n = 46)	T	2	0.766 (0)	-0.035 (0)	1.9 (0)	0.544 (0)	-0.386 (0)	0.241 (0)	0.46 (0)	0.308 (0)	0.224 (0)	2.544 (0)	0.917 (0)	1.106 (0)						
		0	-0.132 (0.152)	-0.02 (0.224)	-0.14 (0.411)	-0.109 (0.256)	-0.014 (0.206)	-0.074 (0.249)	-0.03 (0.095)	-0.08 (0.222)	-0.059 (0.135)	-0.806 (0.566)	-0.088 (0.224)	-0.215 (0.236)						
	N	1	-0.121 (0.183)	-0.185 (0.363)	0.318 (0)	-0.038 (0.115)	-0.016 (0.147)	-0.109 (0.266)	-0.037 (0.115)	-0.023 (0.119)	-0.033 (0.101)	-0.532 (0.569)	-0.13 (0.281)	-0.276 (0.162)						
		2	-0.209 (0)	-0.471 (0)	0.318 (0)	-0.067 (0)	-0.204 (0)	-0.158 (0)	-0.04 (0)	0.32 (0)	-0.194 (0)	-0.141 (0)	-0.506 (0)	-0.112 (0)						
L858R (n = 46)	T	0	0.464 (0.172)	-0.03 (0.353)	0.15 (0.551)	0.135 (0.365)	0.04 (0.225)	0.096 (0.279)	0.052 (0.171)	0.016 (0.274)	0.059 (0.145)	0.782 (0.966)	0.039 (0.338)	0.159 (0.361)						
		1	0.464 (0.172)	-0.035 (0.202)	0.735 (0.408)	0.651 (0.517)	-0.086 (0.067)	0.335 (0.295)	0.265 (0.032)	0.291 (0.392)	0.03 (0.037)	1.286 (0.596)	-0.08 (0.377)	0.583 (0.349)						
	N	2	0.766 (0)	-0.035 (0)	1.9 (0)	0.544 (0)	-0.386 (0)	0.241 (0)	0.46 (0)	0.308 (0)	0.224 (0)	2.544 (0)	0.917 (0)	1.106 (0)						
		0	-0.128 (0.17)	-0.047 (0.233)	-0.115 (0.381)	-0.135 (0.273)	-0.01 (0.234)	-0.101 (0.227)	-0.08 (0.103)	-0.116 (0.216)	-0.062 (0.152)	-0.823 (0.633)	-0.081 (0.243)	-0.239 (0.258)						
	N	1	-0.145 (0.074)	-0.269 (0.131)	-0.369 (0.288)	-0.035 (0.129)	0.089 (0.069)	0.058 (0.345)	-0.042 (0.019)	-0.038 (0.119)	-0.08 (0.118)	-0.061 (0.573)	-0.072 (0.247)	-0.252 (0.112)						
		2	-0.209 (0)	-0.471 (0)	0.318 (0)	-0.067 (0)	-0.204 (0)	-0.158 (0)	-0.04 (0)	0.32 (0)	-0.194 (0)	-0.141 (0)	-0.506 (0)	-0.112 (0)						

Definition of abbreviations: BRAF = B-Raf proto-oncogene, serine/threonine kinase; CLOCK = clock circadian regulator; EGFR = epidermal growth factor receptor; EGFR mut⁺ = EGFR mutation positive; ERBB2 = Erb-B2 receptor tyrosine kinase 2; EXOC1 = exocyst complex component 1; GRB2 = growth factor receptor-bound protein 2; ILMN = prefix of ID of genes' probe designed by Illumina; N = nontumor; NMU = neuromedin-U; PDCL2 = phosphodiesterase 3; T = tumor.

The genotype is coded by the number of minor alleles. The detail of the adjustment is described in high-throughput data generation and preprocessing (see the online supplement). The first and second columns describe the sample type and tissue type, respectively. The third column gives the genotype information. Each of the remaining columns gives the mean levels and the SD in the parentheses. *P* values < 0.05 are in bold.

regulatory polymorphisms that alter the level of expression of nearby transcripts. If such variants are synonymous and located in the exon of particular genes, then AEI might be observed. We observed an AEI at rs3805383, which is a synonymous SNP located in the exonic region of *NMU* (see Figure E5a). Using the imputed genotypic data, we identified and pyrosequenced all 14 subjects in the LCTCNS group who were heterozygous at rs3805383. Using allele-specific expression ratios (ASERs), we found an AEI in 4 of the 14 subjects; their ASERs were 1.48, 2.40, 1.52, and 0.57; thus, they met the established criterion for an AEI (23). Figure E5b depicts the pyrograms of representative subjects. We also measured the ASER for each subject in the LCTCNS group who were heterozygous at rs576732 or rs1801260. Although we did not find AEIs according to this criterion, the ASERs for rs576732 provided some suggestive evidence; Table E7 reports the results.

Absence of Correlation with *EGFR* Mutation Status, *BIM* Polymorphism, and PFS after Standard Chemotherapy

Individuals with *EGFR*-mutated NSCLC harboring the *BIM* deletion polymorphism may experience a poorer response to *EGFR*-TKIs compared with those without the polymorphism (20, 24). Our reported variants at 4q12 were not associated with either *EGFR* mutation status or *BIM* polymorphism. Using the LCTCSMB group (n = 124), in which *EGFR* mutation status, *BIM* polymorphism, and the SNP array data were available for each patient, we found no association with either the genotype data at rs476184 and rs17725110 or with the imputed genotype data at rs576732, rs1801260, and rs3805383 (see Table E8). Table E8 also shows that the genetic variants did not correlate with *EGFR* mutation status in the NTU NSCLC cohort. Because the *EGFR* common mutations (exon 19 deletion or L858R mutation) and wild-type status accounted for approximately 61% and 30%, respectively, of the subjects in the LCTCSMB group (see Table E9), the correlation presented in Table E8 was restricted to subjects who were either *EGFR* wild-type or had the *EGFR* common mutations.

We evaluated the association between the 4q12 variants and PFS based on PGFNS (n = 360), which consisted of female never-smoking patients with lung ADC who were

previously studied in a reported GWAS (21, 22) and who were treated with first-line, platinum-based doublets but without resection. Table E10 shows a lack of association between the genotypes at these loci and the PFS among the PGFNS group based on Cox's regression model adjusted for age. Similar results were obtained using the log-rank test for each of these SNPs (see Figure E3b). These results suggest that it might be beneficial to explore first-line, platinum-based doublets instead of *EGFR*-TKIs as a possible treatment for female patients with late-stage lung ADC who are carrying a risk allele at these SNPs.

Using the ENCyclopedia of DNA Element (27) data, we found that several of these SNPs may be involved in the regulation of gene expression (see Table E11). The fact that rs3805383 is located in predicted enhancer region with flanking enhancer and promoter associated histone mark (H3K4Me1 and H3K4Me3) (Figure E5a) strongly suggests that rs3805383 is a regulatory SNP. This is not only in line with the findings of this study but also points to directions for future mechanistic studies.

Discussion

Our study showed that germ-line variants at 4q12 were associated with PFS among never-smoking patients with late-stage lung ADC who were treated with first-line *EGFR*-TKIs. This association reached genome-scale significance in the GELAC cohort and was supported by a genetic study in an independent cohort, NTU NSCLC, in which this association was also established among a smaller group of never-smoking patients and patients with the *EGFR* mutation. We also found associations between these SNPs and the expression of *BRAF*, *EGFR*, *ERBB2*, *GRB2*, *CLOCK*, *NMU*, *PDCL2*, and *SRD5A3* in lung tissues, which was also associated with the time-to-recurrence or relapse, and tumor status. In particular, the correlation between rs3805383 and the expression of *CLOCK*, *NMU*, and *SRD5A3* also appeared in the 274 nontumor lung tissues of mostly non-Asian patients in the GTEx study. We found an AEI of *NMU* at rs3805383.

Interestingly, the mechanistic involvement of *NMU* in lung cancer was reported previously (28). In that report, GHSR1b and NTSR1, which are two G-protein coupled receptors, were identified as *NMU* receptors in lung cancer cells.

Furthermore, NTSR1 was identified as a potential prognostic marker of lung ADC; NTSR1 activation participates in lung cancer progression (29). In contrast, *PDCL2* encodes a member of the phosphoinositide-like protein family that regulates G-protein signaling or chaperone-assisted protein folding (30).

The preceding functional analyses seem to support the biological plausibility of the association between SNPs in 4q12 and PFS. The consistency shown in these association studies reinforces each of the scientific findings. It seems that *BRAF*, *EGFR*, *NMU*, and *SRD5A3* deserve the most attention because their expression levels showed the largest-fold change between tumor and nontumor tissues and had the most significant correlation with genotype at rs3805383, based on our own data.

Because never-smoking and the *EGFR* mutation seems to be related, and patients with these characteristics respond favorably to *EGFR*-TKIs (31), it would be interesting to determine which variants at 4q12 are associated with PFS in patients with the *EGFR* mutation. This and the association between these variants and *EGFR* expression suggest this locus could modify the effect of *EGFR*-TKIs on PFS in patients with lung ADC.

We also found that these variants were not associated with *EGFR* mutation status or with *BIM* polymorphism, nor with PFS in never-smoking patients with late-stage lung ADC who were treated with first-line, platinum-based doublets. This suggests that it may be worthwhile to examine whether patients with the *EGFR* mutation who carry a minor allele at these SNPs might respond more favorably to first-line, platinum-based doublets rather than to first-line TKIs.

Our results warrant further confirmatory studies before transitioning to clinical studies. Because the minor allele frequencies of these variants were much higher in Japanese and European populations (see Table E12), if confirmed, our results might have important implications for clinical practice. Further laboratory studies are needed to examine the validity of these results and to further the understanding of the biological mechanisms responsible for the effect of the 4q12 locus. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank Ko-Wei Chang, Wen-Tsen Fang, and Fang-Yu Tsai for technical assistance.

References

- Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin* 2009;59:225–249.
- Buter J, Giaccone G. Medical treatment of non-small-cell lung cancer. *Ann Oncol* 2005;16 Suppl 2:ii229–ii232.
- Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–2139.
- Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500.
- Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, *et al.* EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci USA* 2004;101:13306–13311.
- Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpawaravong P, Han B, Margono B, Ichinose Y, *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;361:947–957.
- Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, Gemma A, Harada M, Yoshizawa H, Kinoshita I, *et al.*; North-East Japan Study Group. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* 2010;362:2380–2388.
- Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, Seto T, Satouchi M, Tada H, Hirashima T, *et al.*; West Japan Oncology Group. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol* 2010;11:121–128.
- Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, Zhang S, Wang J, Zhou S, Ren S, *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011;12:735–742.
- Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, *et al.*; Spanish Lung Cancer Group in collaboration with Groupe Français de Pneumo-Cancérologie and Associazione Italiana Oncologia Toracica. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EORTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol* 2012;13:239–246.
- Sequist LV, Yang JC, Yamamoto N, O’Byrne K, Hirsh V, Mok T, Geater SL, Orlov S, Tsai CM, Boyer M, *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013;31:3327–3334.
- Wu YL, Zhou C, Hu CP, Feng J, Lu S, Huang Y, Li W, Hou M, Shi JH, Lee KY, *et al.* Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15:213–222.
- Keedy VL, Temin S, Somerfield MR, Beasley MB, Johnson DH, McShane LM, Milton DT, Strawn JR, Wakelee HA, Giaccone G. American Society of Clinical Oncology provisional clinical opinion: epidermal growth factor receptor (EGFR) mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. *J Clin Oncol* 2011;29:2121–2127.
- Jett JR, Carr LL. Targeted therapy for non-small cell lung cancer. *Am J Respir Crit Care Med* 2013;188:907–912.
- Spira A, Halmos B, Powell CA. Update in lung cancer 2014. *Am J Respir Crit Care Med* 2015;192:283–294.
- Wilke RA, Ramsey LB, Johnson SG, Maxwell WD, McLeod HL, Voora D, Krauss RM, Roden DM, Feng Q, Cooper-Dehoff RM, *et al.*; Clinical Pharmacogenomics Implementation Consortium (CPIC). The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin Pharmacol Ther* 2012;92:112–117.
- Wheeler HE, Maitland ML, Dolan ME, Cox NJ, Ratain MJ. Cancer pharmacogenomics: strategies and challenges. *Nat Rev Genet* 2013;14:23–34.
- Israel E, Lasky-Su J, Markezich A, Damask A, Szefer SJ, Schuermann B, Klanderman B, Sylvia J, Kazani S, Wu R, *et al.*; SHARP Investigators. Genome-wide association study of short-acting β_2 -agonists. A novel genome-wide significant locus on chromosome 2 near ASB3. *Am J Respir Crit Care Med* 2015;191:530–537.
- Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011;43:513–518.
- Ng KP, Hillmer AM, Chuah CT, Juan WC, Ko TK, Teo AS, Ariyaratne PN, Takahashi N, Sawada K, Fei Y, *et al.* A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. *Nat Med* 2012;18:521–528.
- Lan Q, Hsiung CA, Matsuo K, Hong YC, Seow A, Wang Z, Hosgood HD III, Chen K, Wang JC, Chatterjee N, *et al.* Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. *Nat Genet* 2012;44:1330–1335.
- Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS, Chatterjee N, Brennan P, Wu C, Zheng W, *et al.* The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS Genet* 2010;6: pii: e1001051.
- Wang H, Elbein SC. Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol Biol* 2007;373:157–176.
- Lee JH, Lin YL, Hsu WH, Chen HY, Chang YC, Yu CJ, Shih JY, Lin CC, Chen KY, Ho CC, *et al.* Bcl-2-like protein 11 deletion polymorphism predicts survival in advanced non-small-cell lung cancer. *J Thorac Oncol* 2014;9:1385–1392.
- Yang CH, Yu CJ, Shih JY, Chang YC, Hu FC, Tsai MC, Chen KY, Lin ZZ, Huang CJ, Shun CT, *et al.* Specific EGFR mutations predict treatment outcome of stage IIIB/IV patients with chemotherapy-naïve non-small-cell lung cancer receiving first-line gefitinib monotherapy. *J Clin Oncol* 2008;26:2745–2753.
- GTEX Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580–585.
- The Encode Project Consortium. A user’s guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 2011; 9: e1001046.
- Takahashi K, Furukawa C, Takano A, Ishikawa N, Kato T, Hayama S, Suzuki C, Yasui W, Inai K, Sone S, *et al.* The neuromedin U-growth hormone secretagogue receptor 1b/neurotensin receptor 1 oncogenic signaling pathway as a therapeutic target for lung cancer. *Cancer Res* 2006;66:9408–9419.
- Alifano M, Souzae F, Dupouy S, Camilleri-Broet S, Younes M, Ahmed-Zaid SM, Takahashi T, Cancellieri A, Damiani S, Baron M, *et al.* Neurotensin receptor 1 determines the outcome of non-small cell lung cancer. *Clin Cancer Res* 2010;16:4401–4410.
- Willardson BM, Howlett AC. Function of phosphatidylcholine-like proteins in G protein signaling and chaperone-assisted protein folding. *Cell Signal* 2007;19:2417–2427.
- Pham D, Kris MG, Riely GJ, Sarkaria IS, McDonough T, Chuai S, Venkatraman ES, Miller VA, Ladanyi M, Pao W, *et al.* Use of cigarette-smoking history to estimate the likelihood of mutations in epidermal growth factor receptor gene exons 19 and 21 in lung adenocarcinomas. *J Clin Oncol* 2006;24:1700–1704.