ARTICLES

Clinical and Biological Features Associated With Epidermal Growth Factor Receptor Gene Mutations in Lung Cancers

Hisayuki Shigematsu, Li Lin, Takao Takahashi, Masaharu Nomura, Makoto Suzuki, Ignacio I. Wistuba, Kwun M. Fong, Huei Lee, Shinichi Toyooka, Nobuyoshi Shimizu, Takehiko Fujisawa, Ziding Feng, Jack A. Roth, Joachim Herz, John D. Minna, Adi F. Gazdar

Background: Mutations in the tyrosine kinase (TK) domain of the epidermal growth factor receptor (EGFR) gene in lung cancers are associated with increased sensitivity of these cancers to drugs that inhibit EGFR kinase activity. However, the role of such mutations in the pathogenesis of lung cancers is unclear. *Methods:* We sequenced exons 18–21 of the EGFR TK domain from genomic DNA isolated from 617 non-small-cell lung cancers (NSCLCs) and 524 normal lung tissue samples from the same patients and 36 neuroendocrine lung tumors collected from patients in Japan, Taiwan, the United States, and Australia and from 243 other epithelial cancers. Mutation status was compared with clinicopathologic features and with the presence of mutations in KRAS, a gene in the EGFR signaling pathway that is also frequently mutated in lung cancers. All statistical tests were two sided. *Results:* We detected a total of 134 EGFR TK domain mutations in 130 (21%) of the 617 NSCLCs but not in any of the other carcinomas, nor in nonmalignant lung tissue from the same patients. In NSCLC patients, EGFR TK domain mutations were statistically significantly more frequent in never smokers than ever smokers (51% versus 10%), in adenocarcinomas versus cancer of other histologies (40% versus 3%), in patients of East Asian ethnicity versus other ethnicities (30% versus 8%), and in females versus males (42% versus 14%; all *P* < .001). EGFR TK domain mutation status was not associated with patient age at diagnosis, clinical stage, the presence of bronchioloalveolar histologic features, or overall survival. The EGFR TK domain mutations we detected were of three common types: in-frame deletions in exon 19, single missense mutations in exon 21, and in-frame duplications/insertions in exon 20. Rare missense mutations were also detected in exons 18, 20, and 21. KRAS gene mutations were present in 50 (8%) of the 617 NSCLCs but not in any tumors with an EGFR TK domain mutation. *Conclusions:* Mutations in either the EGFR TK domain or the KRAS gene can lead to lung cancer pathogenesis. EGFR TK domain mutations are the first molecular change known to occur specifically in never smokers. [J Natl Cancer Inst 2005;97:339–46]

Despite improvements in diagnostic and therapeutic approaches, lung cancer is the major cause of cancer-related deaths *(1,2)*. Non–small-cell lung cancer (NSCLC), the major form of lung cancer, is classified into three histologic types: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Adenocarcinoma is the most frequent histologic type of NSCLC in both sexes in many parts of the world. Although tobacco smoking is the cause of most lung cancers, it is less strongly associated with adenocarcinoma than with other histologic types *(3)*. Lung cancer is also characterized by the accumulation of multiple genetic and/or epigenetic alterations, including those that result in the activation of oncogenes and the inactivation of tumor suppressor genes *(4–6)*. A better understanding of the molecular mechanism by which these alterations affect lung cancer pathogenesis would provide new and more effective strategies for chemoprevention, early diagnosis, and targeted treatment *(7)*.

Deregulation of protein kinase activity is common in malignancies *(8,9)* and has led to the development of therapies that target these oncogenes *(10)*. One such therapy, which has been widely used for the treatment of NSCLC, is gefitinib (Iressa, ZD1839), a small-molecule tyrosine kinase (TK) inhibitor that inhibits the protein kinase activity of epidermal growth factor receptor (EGFR), which is highly expressed in many epithelial cancers, including lung cancers *(11)*. Although results of several preclinical studies and clinical trials have reported mixed results *(12–15)*, the mechanisms of the antitumor effect and drug sensitivity of TK inhibitors have not been fully established because neither the expression nor the phosphorylation status of EGFR was associated with patient response *(16)*. Nonetheless, some patients have dramatic and durable responses to such therapy.

Reports of lung cancers bearing mutations in the EGFR gene have generated considerable interest because such mutations are associated with an increased sensitivity to gefitinib therapy *(17–19)*. All of the mutations detected are located within the TK domain of the EGFR gene, where gefitinib competes with adenosine triphosphate (ATP) for binding to the protein; cells containing these mutations are responsive to the principal ligand (i.e., EGF) and show increased sensitivity to gefitinib in vitro *(17)*. The TK domain EGFR mutations are more frequently found in NSCLCs with adenocarcinoma histology than in those with other histologies, in females than in males, in patients from Japan than in patients from the United States, and in never smokers than in current or former smokers *(17–19),* the same subpopulations that have the highest response rates to gefitinib *(14,15,20,21)*.

To examine the role of EGFR TK domain mutations in lung cancer pathogenesis, we searched for EGFR TK domain mutations in genomic DNA isolated from primary lung tumors from Japan, Taiwan, the United States, and Australia. Our goal was to examine the associations between EGFR TK domain mutation status and patient sex, age at diagnosis, tumor histology, clinical stage, smoking history, and ethnicity. We also examined EGFR TK domain mutations in genomic DNA from nonmalignant lung tissue from many of the patients and from other epithelial cancer samples. EGFR signaling influences multiple downstream pathways, including Ras/Raf/mitogen-activated protein kinase, JAK-STAT cytokine, and phosphatidylinositol 3′-kinase (PI3K)/Akt pathways, which affect cell proliferation, survival, and apoptosis. Several genes in the RAS, RAF, and phosphatidylinositol 3′-kinase families have been found to be mutated in lung cancer, with KRAS gene mutations (especially in codons 12 and 13) especially frequent *(8,9,22)*. DNA sequences of human papillomavirus (HPV) strains 16 and 18 have been reported to be associated with lung cancers from Taiwan and are associated with the female sex and never smoking status *(23)*. Thus, we also examined the relationship between EGFR gene mutation status and KRAS gene mutation status or the presence of HPV DNA sequences.

PATIENTS AND METHODS

Tumor Samples

We obtained 617 NSCLC tumors from patients undergoing surgical resection at Chiba University and Okayama University (Chiba and Okayama, Japan, respectively; n = 263), Veterans General Hospital (Taichung, Taiwan; n = 93), MD Anderson Cancer Center (Houston, TX; n = 160), and Prince Charles Hospital (Brisbane, Australia; n = 101). Tumors were collected at the time of surgical resection and kept frozen at −80°C. All patients from Japan and Taiwan were of East Asian ethnicity. Among patients from the United States, 139 were white, eight were Hispanic, seven were black, and four were East Asian; ethnicity was not known for two patients. One of the Australian patients was East Asian; the rest were white. For the 524 patients from Japan, the United States, and Australia, we also obtained a sample of the corresponding nonmalignant lung tissue from a site located far from the tumor. Most of the tumors were unselected (i.e., collected sequentially; n = 519); the remaining tumors were selected from among patients with well-documented smoking histories from the United States (n = 80) and Australia (n = 18). Six primary small-cell lung cancers (SCLCs) were obtained from patients undergoing treatment at Chiba University, and 25

bronchial carcinoids and 5 large-cell neuroendocrine carcinomas (LCNECs) were obtained from patients undergoing treatment at MD Anderson Cancer Center. A total of 243 epithelial carcinomas arising at sites other than the lung were obtained from patients undergoing treatment at the hospitals affiliated with University of Texas Southwestern Medical Center, Dallas, TX (prostate, bladder, breast, and colorectal cancer) or from Catholic University, Santiago, Chile (gallbladder cancer).

Institutional Review Board permission and patient written informed consent were obtained at each collection site. Clinical information, including patient sex, age at diagnosis, tumor histology, clinical stage, and smoking history, were available for all patients, and valid survival data were available for 436 patients (182 from Japan, 159 from the United States, and 95 from Australia). Clinical staging of lung cancers was performed using the revised International System for Staging Lung Cancer *(24)*. All available pathology slides of adenocarcinomas from the U.S. NSCLCs (n = 97) were reviewed and reevaluated for bronchioloalveolar carcinoma (BAC) subtypes according to the World Health Organization (WHO) classification of lung cancers (which defines BAC as a true noninvasive cancer without stromal, vascular, or pleural invasion) *(25)*. The slides were also scored for the presence of lepidic growth (a feature of BAC) in increments of 10% (with 100% representing pure, or true, BAC tumors). The Taiwanese patients had been previously examined for the presence of HPV types 16 and 18 DNA *(23)*.

DNA Extraction and Sequencing

Genomic DNA was isolated from primary tumor samples by overnight digestion with sodium dodecyl sulfate and proteinase K (Life Technologies Inc., Rockville, MD) at 37°C, followed by standard phenol-chloroform (1 volume:1 volume) extraction and ethanol precipitation.

Intron-based polymerase chain reaction (PCR) primers were used to amplify the seven exons comprising the entire TK domain of the EGFR gene. The primers were as follows (forward and reverse, respectively): exon 18 (5′-AGCATGGTGAGGGCTGAGGTGAC-3′ and 5′-ATATACAGCTTGCAAGGACTCTGG-3′), exon 19 (5′-CCAGATCACTGGGCAGCATGTGGCACC-3′ and 5′-AGCAGGGTCTAGAGCAGAGCAGCTGCC-3′), exon 20 (5′-GATCGCATTCATGCGTCTTCACC-3′ and 5′-TTGCTATCCCAGGAGCGCAGACC-3′), exon 21 (5′-TCAGAGCCTGGCATGAACATGACCCTG-3′ and 5′-GGTCCCTGGTGTCAGGAAAATGCTGG-3′), exon 22 (5′-AATTAGGTCCAGAGTGAGTTAAC-3′ and 5′-ACTTGCATGTCAGAGGATATAATG-3′), exon 23 (5′-CATCAAGAAACAGTAACCAGTAATG-3′ and 5′-AAGGCCTCAGCTGTTTGGCTAAG-3′), and exon 24 (5′-TTGACTGGAAGTGTCGCATCACC-3′ and 5′-CATGTGACAGAACACAGTGACATG-3′). All PCR assays were carried out in a 25-μL volume that contained 100 ng of genomic DNA and 1.25 units of HotStarTaq DNA polymerase (QIAGEN Inc., Valencia, CA). DNA was amplified for 33 cycles at 95°C for 30 seconds, 65°C for 30 seconds, and 72°C for 45 seconds, followed by a 7-minute extension at 72°C. The sequences of the intron-based PCR primers for used to amplify exon 2 of the KRAS gene were as follows (forward and reverse, respectively): 5′-GTATTAACCTTATGTGTGACA-3′ and 5′-GTCCTGCACCAGTAATATGC-3′. DNA was amplified for 33 cycles at 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a 7-minute extension at 72°C. All PCR products were incubated

with exonuclease I and shrimp alkaline phosphatase (Amersham Biosciences Corp., Piscataway, NJ) according to the manufacturer's instructions and then sequenced directly using the Applied Biosystems PRISM dye terminator cycle sequencing method (Perkin-Elmer Corp., Foster City, CA). All sequence variants were confirmed by sequencing the products of independent PCR amplifications in both directions.

Statistical Analyses

We used chi-square tests and Fisher's exact tests (when there were fewer than five expected counts in the contingency table) to assess the relationship between EGFR gene mutations and each of the potentially influential factors, including ethnicity, sex, smoking status, and histologic subtype. We used logistic regression models, with EGFR gene mutation status as the outcome, to examine the effects of ethnicity, sex, smoking status, and histologic subtype, with adjustment for each factor. Kaplan-Meier curves were drawn for the two groups of patients (i.e., those with and without the mutations), and the difference in overall survival between the two groups was investigated by using a Cox proportional hazards model with adjustment for other survival-related risk factors, age, ethnicity, and histological subtype, given the mutation groups were time independent. The agreement between EGFR and KRAS gene mutation status was tested, and the kappa coefficient was determined. The agreement between mutation status for each of the four studied exons of the TK domain and ethnicity, sex, smoking status, and histologic type was tested among unselected patients with EGFR mutations. All statistical tests were two sided, and P values less than .05 were considered statistically significant.

RESULTS

Preliminary sequence analysis of the entire TK domain (exons 18–24) of the EGFR gene of 96 unselected NSCLC samples indicated that all mutations were located within exons 18–21, which encode the N lobe and part of the C lobe of EGFR (Fig. 1). Our subsequent analyses were limited to these four exons. We detected mutations only in DNA extracted from NSCLC samples; no mutations were detected in DNA extracted from neuroendocrine lung tumors (SCLCs, LCNECs, and bronchial carcinoids) or prostate, bladder, breast, colorectal, or gallbladder carcinomas. EGFR TK domain mutations were detected in 130 (21%) of the 617 NSCLCs and in 120 (23%) of the 519 unselected NSCLCs (Table 1). Of the 120 unselected tumors with EGFR TK domain mutations, two had multiple mutations (see below). The frequency of mutations was greater for patients of East Asian ethnicity versus those of other ethnicities (30% versus 8%, $P < .001$), females versus males (42% versus 14%, $P < .001$), never smokers versus ever smokers (51% versus 10%, $P < .001$), and for adenocarcinomas versus other histologies (40% versus 3%, $P < .001$).

Among the adenocarcinoma patients from the United States or Australia, four (80%) of five patients of East Asian ethnicity had an EGFR TK domain mutation, whereas only 12 (16%) of 75 patients of other ethnicities (mainly white) had an EGFR TK domain mutation ($P = .005$). Of the four EGFR TK domain mutation-positive adenocarcinomas in patients of East Asian ethnicity, three were from males and two were from never smokers. Logistic regression models confirmed that age and EGFR TK domain mutation status were independent. Because the distribu-
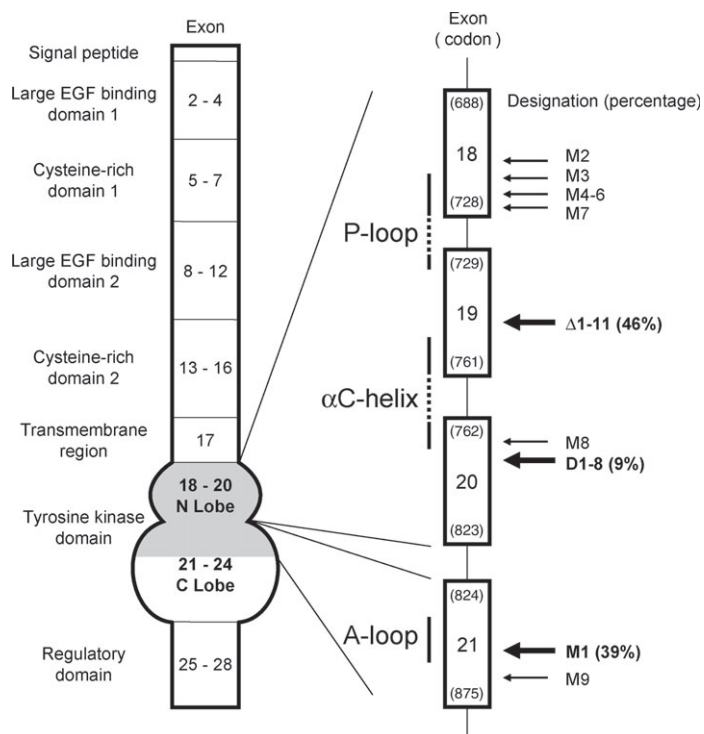
Fig. 1. Locations and types of the 134 epidermal growth factor receptor (EGFR) gene mutations detected in lung cancers. The structure of the EGFR gene is shown at left, and the locations and types of the mutations in the tyrosine kinase (TK) domain are shown at right. All mutations were located within exons 18–21, which encode the N lobe and part of the C lobe of EGFR (shaded area of the gene on the left, which is presented in magnified form on the right). Three major types of mutations (shown in **bold**) formed 94% of the 134 mutations detected and consisted of deletions in exon 19 immediately 5′ of the αC-helix (11 types, labeled Δ1–Δ11), duplications and/or insertions in exon 20 immediately 3′ of the αC-helix (eight types labeled D1–D8), and a single-point mutation, L858R (labeled M1), in the A-loop. The remaining 6% of mutations consisted of missense mutations in the P-loop in exon 18 (six types labeled M2–M7), in the αC-helix in exon 20 (a single type labeled M8), or in the A-loop in exon 21 (a single type labeled M9).

tion of age at diagnosis was not statistically significant different between patients with and without EGFR TK domain mutations, we did not adjust for age in the logistic regression analyses. We examined associations between the variables after adjustment for each other, with EGFR TK domain mutation status as the outcome and all the factors as explanatory variables, and found no interactions.

Even when we confined our analysis to the subgroup with the highest frequencies of mutations, never smokers with adenocarcinomas (both selected and unselected cases combined, n = 157), the EGFR TK domain mutation frequency was statistically significantly higher for patients from Japan and Taiwan than for those from the United States and Australia (64% versus 36%), both before and after adjustment for sex ($P = .003$ and .004, respectively). For the 160 patients from the United States for whom we had detailed smoking data, 3% of current smokers, 8% of former smokers, and 20% of never smokers had EGFR TK domain mutations. Results of a test for trend revealed that smoking status was statistically significantly associated with the presence of EGFR gene mutations (adjusted $P_{trend} = .02$). There were no statistically significant differences in mutational patterns with respect to sex, smoking status, or ethnicity.

Overall, we identified a total 134 EGFR TK domain mutations among 130 tumors. In 95 of the 130 patients for whom adjacent

Table 1. Univariate analysis of the relationship between EGFR TK domain mutation positivity and sex, smoking status, and histologic subtype by country or ethnicity among 519 unselected NSCLC patients*

Country or ethnicity† (no.)	No. with mutation (%)	Sex (no.)	No. with mutation (%)	P‡	Smoking status (no.)	No. with mutation (%)	P‡	Histologic subtype§ (no.)	No. with mutation (%)	P‡
Japan (263)	71 (27)	Male (183)	30 (16)	<.001	Never smoker (78)	47 (60)	<.001	Adenocarcinoma (154)	67 (44)	<.001
		Female (80)	41 (51)		Smoker (185)	24 (13)		Other (109)	4 (4)	
Taiwan (93)	32 (34)	Male (64)	14 (22)	<.001	Never smoker (55)	27 (49)	<.001	Adenocarcinoma (55)	31 (56)	<.001
		Female (29)	18 (62)		Smoker (38)	5 (13)		Other (38)	1 (3)	
United States (80)	11 (14)	Male (43)	3 (7)	.116	Never smoker (26)	7 (27)	.043	Adenocarcinoma (44)	11 (25)	<.001
		Female(37)	8 (22)		Smoker (54)	4 (7)		Other (36)	0 (0)	
Australia (83)	6 (7)	Male (58)	1 (2)	.013	Never smoker (7)	4 (57)	<.001	Adenocarcinoma (36)	5 (14)	.081
		Female (25)	5 (20)		Smoker (76)	2 (3)		Other (47)	1 (2)	
East Asian (361)	107 (30)	Male (251)	47 (19)	<.001	Never smoker (135)	76 (56)	<.001	Adenocarcinoma (214)	102 (48)	<.001
		Female (110)	60 (55)		Smoker (226)	31 (14)		Other (147)	5 (3)	
Other ethnicities (158)	13 (8)	Male (97)	1 (1)	<.001	Never smoker (31)	9 (29)	<.001	Adenocarcinoma (75)	12 (16)	<.001
		Female (61)	12 (20)		Smoker (127)	4 (3)		Other (83)	1 (1)	
Total (519)	120 (23)	Male (348)	48 (14)	<.001	Never smoker (166)	85 (51)	<.001	Adenocarcinoma (289)	114 (40)	<.001
		Female (171)	72 (42)		Smoker (353)	35 (10)		Other (230)	6 (3)	

*EGFR = epidermal growth factor receptor; NSCLC = non–small-cell lung cancer.

†Two-sided *P* value (chi-square test) for testing the relationship between EGFR TK domain mutation status and country (Japan and Taiwan versus United States and Australia) or ethnicity (East Asian versus other) was less than .001.

‡Two-sided (chi-square test).

§Includes squamous cell, adeno-squamous cell, and large cell carcinoma.

nonmalignant lung tissue was available, mutations were absent from the nonmalignant tissue, indicating that the mutations were somatic in origin. The mutations consisted of three different types—in-frame deletions, single-nucleotide substitutions, and in-frame duplications/insertions—and all mutations were located within or near functionally important sites in the receptor (Table 2; Fig. 1). In-frame deletions in exon 19 (11 types, labeled Δ1–Δ11), involving three to seven codons centered around the uniformly deleted codons 747 to 749 (Leu-Arg-Glu sequence), accounted for 62 (46%) of the mutations detected. The in-frame deletions were occasionally accompanied by missense mutations at the carboxyl-terminal amino acid position flanking the deletion. Missense mutations (n = 60, 45% of total mutations) in exons 18, 20, or 21 (nine types, labeled M1–M9) were the second-most-common mutation, especially mutation L858R (M1) in exon 21 (n = 52, 39% of total mutations). In-frame duplications and/or insertions of one to three codons in exon 20 (eight types, labeled D1–D8), involving amino acids 770 to 776, accounted for 12 (9%) of the mutations. We also detected one silent mutation in exon 18 at codon 718. Three tumors had multiple mutations (one tumor had two mutations, and two tumors had three mutations each); each of the remaining tumors had a single mutation.

It has been reported that EGFR TK domain mutations are common among adenocarcinomas of the BAC subtype *(17,19)*. Therefore, two pathologists (I.I.W. and A.F.G.) who are familiar with the WHO classification of lung cancers *(25)* reviewed the pathology slides for 97 adenocarcinomas from the United States. Seven of these adenocarcinomas (7%) were pure BAC tumors, and none of them had EGFR TK domain mutations. Seventeen adenocarcinomas (18%) were classified as having any BAC features (including those classified as pure BAC tumors). BAC features were present in four (27%) of 15 adenocarcinomas with EGFR TK domain mutations and in 13 (16%) of 82 adenocarcinomas without EGFR TK domain mutations (Table 3). There were no statistically significant differences between EGFR TK domain mutation frequencies and the presence or percentage of BAC features (*P* = .29).

We examined the sequence electropherograms for clues about the relative ratios of the mutant and wild-type alleles in individual tumors. We focused on the electropherograms of tumors that had in-frame deletions or duplications/insertions because their interpretation was facilitated by the fact that the entire sequence downstream of those mutations differed from that of the wild-type allele. Tumor specimens invariably contain a considerable percentage of nonmalignant cells; we therefore interpreted electropherograms in which the peaks corresponding to the mutant sequence were equal to or greater than those of the corresponding wild-type sequence as indicating gene amplification of the mutant allele. Using these criteria, we found that approximately 40% of the electropherograms suggested gene amplification of the mutant allele (Fig. 2).

Because KRAS gene mutations in lung cancers are, like EGFR TK domain gene mutations, limited almost exclusively to adenocarcinomas *(22)*, we also analyzed the adenocarcinomas for mutations in codons 12 and 13 of the KRAS gene (the codons mutated in the majority of lung cancers). We found that KRAS gene mutations were present in 50 (8%) of the 617 NSCLCs and were more frequently found in adenocarcinomas than in other histologic types (12% versus 2%; *P* < .001), in ever smokers than in never smokers (10% versus 4%; *P* =.01), and in patients from the United States and Australia than in patients from Japan and Taiwan (12% versus 5%; *P* = .001). There was no statistically significant association between the frequency of KRAS gene mutations and sex (7% in males versus 10% in females; *P* = .4). None of the tumors we examined had mutations in both the EGFR TK domain and the KRAS gene.

Among the unselected (i.e., sequentially collected) adenocarcinoma patients, the EGFR TK domain mutation rates were similar for early-stage disease (i.e., stages I and II; 80 [38%] of 208 cases were mutation positive) and advanced disease (i.e., stages III and IV; 34 [44%] of 78 were mutation positive) (*P* = .5). After we adjusted for other factors, including age at diagnosis, there was no statistically significant difference in overall survival between patients whose tumors did and did not have EGFR TK domain mutations (*P* = .5, Wald's chi-square test, Cox proportional hazards

Table 2. Analysis of the 134 mutations detected in the tyrosine kinase domain of the epidermal growth factor receptor gene*

Type of mutation	Exon	Designation	Nucleotide number and sequence (5'→3')	Amino acid change	No. with mutation (%)
In-frame deletion	19	Δ1	2235–2249 del GGAATTAAGAGAAGC	E746–A750 del	27 (20)
		Δ2	2236–2250 del GAATTAAGAGAAGCA	E746–A750 del	15 (11)
		Δ3	2240–2254 del TAAGAGAAGCAACAT	L747–T751 del	4 (3)
		Δ4	2239–2247 del TTAAGAGAA, 2248 G>C	L747–E749 del P ins	3 (2)
		Δ5	2239–2250 del TTAAGAGAAGCA, 2251 A>C	L747–T750 del P ins	3 (2)
		Δ6	2240–2257 del TAAGAGAAGCAACATCTC	L747–S752 del S ins	3 (2)
		Δ7	2237–2254 del AATTAAGAGAAGCAACAT, 2255 C>T	E746–T751 del V ins	2 (1.5)
		Δ8	2239–2256 del TTAAGAGAAGCAACATCT	L747–S752 del	2 (1.5)
		Δ9	2235–2252 del GGAATTAAGAGAAGCAAC, 2254 T>A, 2255 C>T	E746–T751 del I ins	1 (1)
		Δ10	2237–2251 del AATTAAGAGAAGCAA, 2252 C>T	E746–A750 del V ins	1 (1)
		Δ11	2239–2256 del TTAAGAGAAGCAACATCT, 2258 C>A	L747–S752 del Q ins	1 (1)
Total					62 (46)
Single-nucleotide substitution†	21	M1	2573 T>G	L858R	52 (39)
	18	M2	2126 A>T	E709V	1 (1)
	18	M3	2144 T>G	I715S	1 (1)
	18	M4	2155 G>T	G719C	1 (1)
	18	M5	2155 G>A	G719S	1 (1
	18	M6	2156 G>C	G719A	1 (1)
	18	M7	2159 C>T	S720F	1 (1)
	20	M8	2303 G>T	S768I	1 (1)
	21	M9	2582 T>G	L861Q	1 (1)
Total					60 (45)
In-frame duplication and/or insertion	20	D1	2308–2316 ins GCCAGCGTG	ASV770-772 ins	4 (3)
		D2	2320–2322 ins CAC	H774 ins	2 (1.5)
		D3	2311–2313 ins GGT	G771 ins	1 (1)
		D4	2308–2313 ins TGCGTG	CV770-771 ins	1 (1)
		D5	2317–2222 ins AACCCC, 2223 C>T	NP773-774 ins, H775Y	1 (1)
		D6	2320–2325 ins CCCCAC	PH774-775 ins	1 (1)
		D7	2320–2328 ins AACCCCCAC	NPH774-776 ins	1 (1)
		D8	2323–2328 ins CACGTG, 2322 G>C	HV775-776 ins	1 (1)
Total					12 (9)

*del = deletion; ins = insertion.

†Three patients had multiple mutations as follows: Japan, patient no. 316, mutations at G719S and S768I; Australia, patient no. 108, mutations at I715S, S720F, and L861Q; Australia, patient no. 364, mutations at E709V, L718L (silent), and G719C.

model; Fig. 3). We also analyzed survival data for patients whose tumors had either of the two most common EGFR TK domain mutations (the L858R missense mutation in exon 21 and an in-frame deletion in exon 19). Compared with patients with no EGFR TK domain mutations, those with deletions in exon 19 had worse survival and those with the L858R mutation had better survival, but in neither case was the difference statistically significant.

Of the 93 Taiwanese patients, 58 had been previously tested for the presence of high-risk strains 16 and 18 of HPV *(23)*.

Table 3. EGFR TK domain mutation status and bronchioloalveolar carcinoma (BAC) features among 97 adenocarcinomas from the United States

EGFR gene mutation status (no.)	No. of pure BACs* (%)	No. with any BAC features† (%)	Mean % with BAC features (range)
Positive (15)	0 (0)	4 (27)	30 (20–50)
Negative (82)	7 (9)	13 (16)	75 (30–100)
Total (97)	7 (7)	17 (18)	65 (20–100)

*True noninvasive tumors that have lepidic growth features as defined by the WHO classification *(25)*.

†Tumors with any BAC features including pure BAC tumors (i.e., 100% BAC features).

Among the tested patients, 32 had tumors (55%) that tested positive for HPV DNA, and as previously reported *(23)*, 75% of the HPV DNA-positive tumors were from females and 67% were from never smokers. We found that 12 of the 58 patients with known HPV DNA status had an EGFR TK domain mutation. The concordance between the two groups (i.e., both molecular changes present versus neither present) was 53% (chi-square test; $P = .42$).
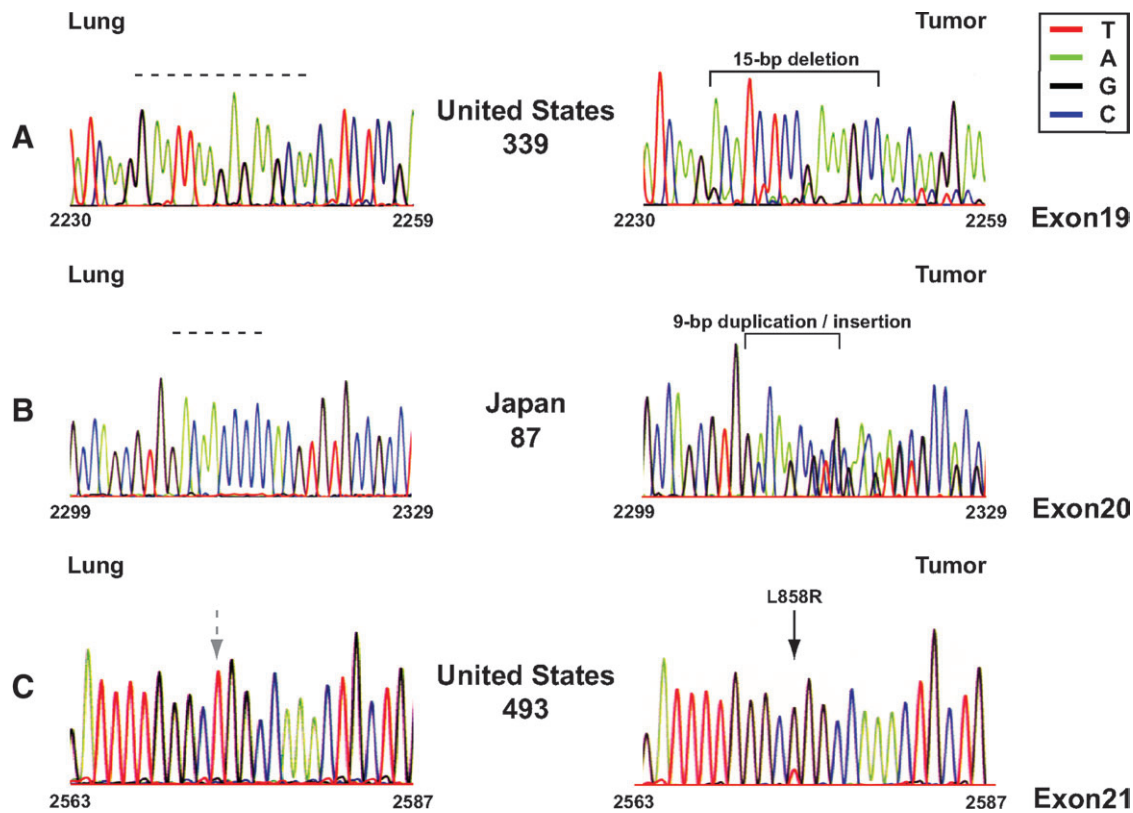
DISCUSSION

A detailed study of mutation patterns within the EGFR gene and their associations with clinicopathologic features of lung cancers is essential because preliminary data have suggested that lung cancer patients whose tumors have EGFR TK domain mutations respond better than those without such mutations to drugs that target the TK domain of the gene *(17–19)*. We tested a large number of primary lung and other cancers for EGFR TK domain mutations and explored the relationship between EGFR TK domain mutation status and multiple parameters associated with lung carcinogenesis, including KRAS gene mutation status and the presence of HPV DNA. We detected EGFR TK domain mutations only in lung cancers of the NSCLC type and not in

Fig. 2. DNA sequence electropherograms illustrating the three major types of tyrosine kinase (TK) domain mutations detected in the epidermal growth factor receptor (EGFR) gene. Exons 18–21 of the TK domain of the gene were amplified individually using intron-based primer pairs, and the resultant amplicons were sequenced. Only the mutation-containing portions of the electropherograms are illustrated; each of the four nucleotides was labeled with unique fluorescent dye. Numbers below electropherograms indicate nucleotide positions. Each panel consists of tumor tissue (right side) and corresponding nonmalignant lung tissue (left side). The tumor designations and country of origin are indicated (center of each panel), as is the involved exon (right side). In all three panels, a mutation is present in the tumor tissue while the corresponding nonmalignant lung tissue contains the wild-type sequence. **A**) A 15-bp deletion (Δ2) in exon 19 (sample number, United States 339), indicated by a **bracket** in the tumor tissue (right) and the corresponding wild-type sequence indicated by a **dashed line** (left). **B**) A 9-bp duplication and/or insertion (D1) in exon 20 (sample number, Japan 87), indicated by a **bracket** in the tumor tissue (right) and the corresponding wild-type sequence indicated by a **dashed line** (left). **C**) L858R (M1) missense mutation in exon 21 (sample number, United States 493). The mutant sequence (right) is indicated by a **solid arrow**, while the wild-type sequence (left) is indicated by a **dashed arrow**.

neuroendocrine lung tumors or in cancers (mainly adenocarcinomas) arising at other sites, including those of breast, colorectal, and prostate origins. Preliminary examination of a subset of the tumors indicated that mutations were limited to the first four exons (exons 18–21) of the TK domain. Three types of mutations constituted 94% of the mutations that we detected: deletions in exon 19, duplications and/or insertions in exon 20, and a single-point mutation in exon 21. The remaining 6% of the mutations we detected consisted of rare missense mutations, mainly in exon 18 but also in exons 20 and 21.

We identified 28 distinct mutations and classified them according to type. The three different mutation types—in-frame deletions, single-nucleotide substitutions, and in-frame duplications and/or insertions—all targeted structures around the ATP-binding cleft of EGFR (which is also the docking site of the small-molecule EGFR kinase inhibitors) *(26),* including the phosphate-binding loop (P-loop), the αC-helix, and the activation loop (A-loop). The L858R (M1) single-nucleotide substitution mutation, which is located near the conserved Asp-Phe-Gly sequence, stabilizes the A-loop *(27)*. The other two common mutation types were located on either side of the αC-helix in the N lobe (Fig. 1), which controls the angle of the ATP-binding pocket. We hypothesize that mutations on either side of the αC-helix (deletions in exon 19 and duplications and/or insertions in exon 20) result in similar conformational changes in EGFR that cause a shift in the helical axis that results in the narrowing of the ATP-binding cleft, which leads to increased gene expression and TK inhibitor sensitivity *(28)*. At least two of the common mutation types (deletions in exon 19 and the L858R missense mutation) are associated with an increase in the amount and duration of ligand-dependent activation, which explains the much greater sensitivity of cells bearing these mutations to EGFR kinase inhibitors *(17)*. Most of the rare mutations target the P-loop. It is interesting that the major mutations always occurred as single mutations in individual tumors, whereas in individual tumors, the rare mutations usually occurred as multiple mutations. These results suggest that the tumorigenic effects of the minor mutations may not be as powerful as those of the major mutations. Additional investigations are required to elucidate the relationships among each type of mutation, TK inhibitor sensitivity, and tumorigenesis.

Our examination of the electropherograms revealed that EGFR TK domain mutations often occurred in a setting of allelic imbalance in which the mutant allele was in excess of the wild-type allele. Polysomy or amplification of the EGFR gene occurs in many cancers, including NSCLC, perhaps relatively early during multistage pathogenesis *(29,30)*. Our findings suggest that specific mutations as well as the increased copy number of the
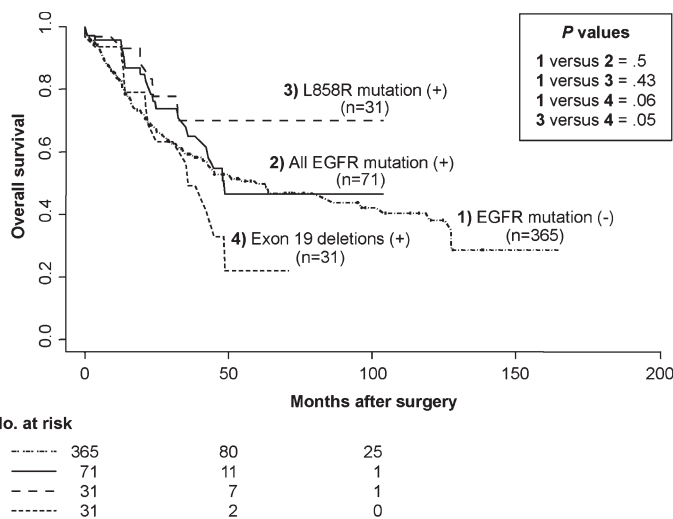
Fig. 3. Kaplan–Meier survival curves for non-small-cell lung cancer patients who underwent curative intent resections and did not receive gefitinib or other epidermal growth factor receptor (EGFR)-targeted therapy. There were no statistically significant differences in the survival curves between patients with (+) or without (−) EGFR TK domain mutations. *P* values (Wald's chi-square test) are two sided. Group 1 = tumors without EGFR TK domain mutations; group 2 = tumors with any EGFR TK domain mutation; group 3 = tumors with the L858R mutation; group 4 = tumors with exon 19 deletions.

mutant allele may play a role in lung cancer pathogenesis (and perhaps in the response to EGFR-targeted therapy).

Among the unselected patients, there were statistically significantly higher mutation frequencies for those with adenocarcinomas versus other histologies, for never smokers versus ever smokers, for patients from East Asian ethnicity versus other ethnicities, and for females versus males. Of the relatively few adenocarcinomas that arose in patients of East Asian ethnicity who were from United States or Australia, 80% had EGFR TK domain mutations, suggesting that ethnicity was associated with mutation rate, irrespective of geographic location. Results of multivariate analyses confirmed that these factors (i.e., sex, smoking status, ethnicity, and histologic subtype) were independent. Our results are consistent with those of previous reports (17–19) and with findings associated with response to gefitinib (14,15). There was no relationship between EGFR TK domain mutation status and either clinical stage or patient survival (in the absence of EGFR kinase inhibitor therapy).

It has been reported that tumors with BAC features have better responses to gefitinib than tumors without such features (20). However, we found no association between EGFR gene mutation status and the BAC subtype of adenocarcinoma, which we defined according to the strict noninvasive criteria as stated by the WHO classification of lung tumors (25). Many pathologists do not use the strict WHO criteria; instead, they use terms such as "adenocarcinomas with BAC features" (which would be termed "adenocarcinomas with mixed subtypes" by the WHO classification) or "bronchioloalveolar carcinoma and its variants" (20). However, even when we applied such a liberal definition for the presence of BAC-like features mixed with other adenocarcinoma subtypes to the tumors in our study, there were no statistically significant differences between EGFR TK domain mutation frequencies and the presence or percentage of BAC features. Accurate assessment of BAC features requires examination of the entire tumor block because BAC features are more prominent at the edges of adenocarcinomas of mixed subtypes than elsewhere

in the tumor. Thus, the limited tumor material typically assessed by reference pathologists for large multi-institutional studies could possibly lead to misleading classifications.

DNA sequences of the high-risk HPV types 16 and 18 have been detected in lung cancers from patients in Taiwan (23) and, less frequently, from patients in other parts of the world. In Taiwan, these HPV DNA sequences are more frequent in lung cancers that arise in female never smokers than in males or in ever smokers and those cancers are usually of the adenocarcinoma type. Thus, the distribution of HPV DNA sequences in lung cancers has similarities to the distribution of EGFR TK domain mutations. However, we found no relationship between the presence of high-risk HPV DNA sequences and EGFR TK domain mutations in our Taiwanese patients.

Both EGFR TK domain and KRAS gene mutations are relatively frequent in NSCLC, especially in adenocarcinomas (31). We found that KRAS gene mutations were more frequent in ever smokers than in never smokers and in patients from Western countries than in patients from East Asian countries. Although mutations in either the EFGR TK domain or the KRAS gene were present in 47% of lung adenocarcinomas, no tumors contained mutations in both genes. Mutant forms of EGFR protein activate multiple downstream signaling pathways, including the RAS, JAK-STAT, and Akt pathways (32–34). Our findings suggest that activation of either the EGFR or RAS signaling pathways has similar effects on lung carcinogenesis.

Mutations in the TP53 gene are also relatively frequent in many cancer types, including lung cancers (35). Most KRAS and TP53 gene mutations in lung cancers are G-to-T transversions, molecular events that are believed to be linked to exposure to tobacco smoke carcinogens (35–37). Lung cancers that arise in never smokers rarely have KRAS gene mutations, and their TP53 gene mutations are seldom G-to-T transversions (35,38), suggesting that these cancers arise in response to exposure to carcinogens other than those present in tobacco smoke.

Although we and others have identified at least 28 different mutations in the TK domain of EGFR, the vast majority of which can be grouped into three major types (Table 2), to date, only the L858R missense mutation in exon 21 and deletions in exon 19 have been proven to be activating mutations (32). Although the activating status of the third major type of mutation (duplications and/or insertions in exon 20) has not been determined, we presume that all mutations of the major types are activating because they always occurred singly, whereas the minor mutations may not be individually sufficient for activation because they often occurred as multiple mutations in individual tumors. However, this hypothesis will be tested in clinical and in vitro studies. In addition, it is possible that other factors such as gene amplification play a role in tumor pathogenesis and response to TK inhibitors.

The findings presented herein support the hypothesis (28) that at least two distinct molecular pathways are involved in the pathogenesis of lung adenocarcinomas, one involving EGFR TK domain mutations and the other involving KRAS gene mutations. The different mutational spectra of the KRAS, EGFR, and TP53 genes and the presence of HPV DNA in lung cancers arising in ever and never smokers suggest that exposure to carcinogens in environmental tobacco smoke may not be the major pathogenic factor involved in the origin of lung cancers in never smokers but that an as-yet-unidentified carcinogen(s) plays an important role. The fact that mutations in the TK domain of the EGFR gene are

more frequent in lung cancers from patients of East Asian ethnicity than in patients of other ethnicities and have not been identified in other human carcinomas suggests that genetic susceptibility to the hypothetical carcinogen(s) may be greater in some subpopulations than in others. In summary, our findings suggest that mutations in the EGFR TK domain occur in a subset of lung adenocarcinomas, whereas KRAS gene mutations occur in a different subset. EGFR TK domain mutations are the first molecular change known to occur specifically in never smokers.

REFERENCES

(1) Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, et al.; American Cancer Society. Cancer statistics, 2004. CA Cancer J Clin 2004;54: 8–29.

(2) Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J, et al.; Eastern Cooperative Oncology Group. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med 2002;346:92–8.

(3) Patel JD, Bach PB, Kris MG. Lung cancer in US women: a contemporary epidemic. JAMA 2004;291:1763–8.

(4) Fong KM, Sekido Y, Minna JD. Molecular pathogenesis of lung cancer. J Thorac Cardiovasc Surg 1999;118:1136–52.

(5) Gazdar AF. The molecular and cellular basis of human lung cancer. Anticancer Res 1994;14:261–7.

(6) Sekido Y, Fong KM, Minna JD. Progress in understanding the molecular pathogenesis of human lung cancer. Biochim Biophys Acta 1998;1378: F21–59.

(7) Minna JD, Gazdar AF, Sprang SR, Herz J. Cancer. A bull's eye for targeted lung cancer therapy. Science 2004;304:1458–61.

(8) Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature 2002;417:949–54.

(9) Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, et al. High frequency of mutations of the PIK3CA gene in human cancers. Science 2004;304:554.

(10) Noble ME, Endicott JA, Johnson LN. Protein kinase inhibitors: insights into drug design from structure. Science 2004;303:1800–5.

(11) Cohen MH, Williams GA, Sridhara R, Chen G, McGuinn WD Jr., Morse D, et al. United States Food and Drug Administration Drug Approval summary: Gefitinib (ZD1839; Iressa) tablets. Clin Cancer Res 2004;10:1212–8.

(12) Wakeling AE, Guy SP, Woodburn JR, Ashton SE, Curry BJ, Barker AJ, Gibson KH. ZD1839 (Iressa): an orally active inhibitor of epidermal growth factor signaling with potential for cancer therapy. Cancer Res 2002;62: 5749–54.

(13) Miller VA, Johnson DH, Krug LM, Pizzo B, Tyson L, Perez W, et al. Pilot trial of the epidermal growth factor receptor tyrosine kinase inhibitor gefitinib plus carboplatin and paclitaxel in patients with stage IIIB or IV non-small-cell lung cancer. J Clin Oncol 2003;21:2094–100.

(14) Fukuoka M, Yano S, Giaccone G, Tamura T, Nakagawa K, Douillard JY, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial). J Clin Oncol 2003;21:2237–46.

(15) Kris MG, Natale RB, Herbst RS, Lynch TJ Jr., Prager D, Belani CP, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. JAMA 2003;290:2149–58.

(16) Suzuki T, Nakagawa T, Endo H, Mitsudomi T, Masuda A, Yatabe Y, et al. The sensitivity of lung cancer cell lines to the EGFR-selective tyrosine kinase inhibitor ZD1839 ('Iressa') is not related to the expression of EGFR or HER-2 or to K-ras gene status. Lung Cancer 2003;42:35–41.

(17) Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, 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;350:2129–39.

(18) Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.

(19) Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, 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;101:13306–11.

(20) Miller VA, Kris MG, Shah N, Patel J, Azzoli C, Gomez J, et al. Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. J Clin Oncol 2004;22:1103–9.

(21) Janne PA, Gurubhagavatula S, Yeap BY, Lucca J, Ostler P, Skarin AT, et al. Outcomes of patients with advanced non-small cell lung cancer treated with gefitinib (ZD1839, "Iressa") on an expanded access study. Lung Cancer 2004;44:221–30.

(22) Ahrendt SA, Decker PA, Alawi EA, Zhu YR, Sanchez-Cespedes M, Yang SC, et al. Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. Cancer 2001;92:1525–30.

(23) Cheng YW, Chiou HL, Sheu GT, Hsieh LL, Chen JT, Chen CY, et al. The association of human papillomavirus 16/18 infection with lung cancer among nonsmoking Taiwanese women. Cancer Res 2001;61:2799–803.

(24) Mountain CF. Revisions in the International System for Staging Lung Cancer. Chest 1997;111:1710–7.

(25) Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E. Histological typing of lung and pleural tumours. 3rd ed. New York (NY): Springer-Verlag; 1999.

(26) Kim DW, Lu B, Hallahan DE. Receptor tyrosine kinase inhibitors as anti-angiogenic agents. Curr Opin Investig Drugs 2004;5:597–604.

(27) Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell 2002;109:275–82.

(28) Gazdar AF, Shigematsu H, Herz J, Minna JD. Mutations and addiction to EGFR: the Achilles 'heal' of lung cancers? Trends Mol Med 2004;10: 481–6.

(29) Hirsch FR, Franklin WA, Veve R, Varella-Garcia M, Bunn PA Jr. HER2/neu expression in malignant lung tumors. Semin Oncol 2002;29:51–8.

(30) Hirsch FR, Scagliotti GV, Langer CJ, Varella-Garcia M, Franklin WA. Epidermal growth factor family of receptors in preneoplasia and lung cancer: perspectives for targeted therapies. Lung Cancer 2003;41:S29–42.

(31) Slebos RJ, Hruban RH, Dalesio O, Mooi WJ, Offerhaus GJ, Rodenhuis S. Relationship between K-ras oncogene activation and smoking in adenocarcinoma of the human lung. J Natl Cancer Inst 1991;83:1024–7.

(32) Sordella R, Bell DW, Haber DA, Settleman J. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. Science 2004;305:1163–7.

(33) Barnes CJ, Kumar R. Epidermal growth factor receptor family tyrosine kinases as signal integrators and therapeutic targets. Cancer Metastasis Rev 2003;22:301–7.

(34) Gschwind A, Fischer OM, Ullrich A. The discovery of receptor tyrosine kinases: targets for cancer therapy. Nat Rev Cancer 2004;4:361–70.

(35) Robles AI, Linke SP, Harris CC. The p53 network in lung carcinogenesis. Oncogene 2002;21:6898–907.

(36) Gealy R, Zhang L, Siegfried JM, Luketich JD, Keohavong P. Comparison of mutations in the p53 and K-ras genes in lung carcinomas from smoking and nonsmoking women. Cancer Epidemiol Biomarkers Prev 1999;8:297–302.

(37) Vahakangas KH, Bennett WP, Castren K, Welsh JA, Khan MA, Blomeke B, et al. p53 and K-ras mutations in lung cancers from former and never-smoking women. Cancer Res 2001;61:4350–6.

(38) Toyooka S, Tsuda T, Gazdar AF. The TP 53 gene, tobacco exposure, and lung cancer. Hum Mutat 2003;21:229–39.

NOTES