Human Cancer Biology

Clinical Cancer Research

Molecular Epidemiology of *EGFR* and *KRAS* Mutations in 3,026 Lung Adenocarcinomas: Higher Susceptibility of Women to Smoking-Related *KRAS*-Mutant Cancers

Snjezana Dogan[1], Ronglai Shen[2], Daphne C. Ang[1], Melissa L. Johnson[3], Sandra P. D'Angelo[3], Paul K. Paik[3], Edyta B. Brzostowski[1], Gregory J. Riely[3], Mark G. Kris[3], Maureen F. Zakowski[1], and Marc Ladanyi[1,4]

Abstract

Purpose: The molecular epidemiology of most *EGFR* and *KRAS* mutations in lung cancer remains unclear.

Experimental Design: We genotyped 3,026 lung adenocarcinomas for the major *EGFR* (exon 19 deletions and L858R) and *KRAS* (G12, G13) mutations and examined correlations with demographic, clinical, and smoking history data.

Results: *EGFR* mutations were found in 43% of never smokers and in 11% of smokers. *KRAS* mutations occurred in 34% of smokers and in 6% of never smokers. In patients with smoking histories up to 10 pack-years, *EGFR* predominated over *KRAS*. Among former smokers with lung cancer, multivariate analysis showed that, independent of pack-years, increasing smoking-free years raise the likelihood of *EGFR* mutation. Never smokers were more likely than smokers to have *KRAS* G > A transition mutation (mostly G12D; 58% vs. 20%, $P = 0.0001$). *KRAS* G12C, the most common G > T transversion mutation in smokers, was more frequent in women ($P = 0.007$) and these women were younger than men with the same mutation (median 65 vs. 69, $P = 0.0008$) and had smoked less.

Conclusions: The distinct types of *KRAS* mutations in smokers versus never smokers suggest that most *KRAS*-mutant lung cancers in never smokers are not due to second-hand smoke exposure. The higher frequency of *KRAS* G12C in women, their younger age, and lesser smoking history together support a heightened susceptibility to tobacco carcinogens. *Clin Cancer Res; 18(22); 6169–77. ©2012 AACR.*

Introduction

EGFR or *KRAS* mutations are present in almost 50% of lung adenocarcinomas in Caucasian patients. More than 90% of *EGFR* mutations are small in frame deletions in exon 19 and L858R missense mutation in exon 21 (1). These mutations are associated with responsiveness to tyrosine kinase inhibitors (TKI) therapy (2–4). *EGFR* mutations are more frequently found in women, Asians, and in never smokers (5, 6). There is an inverse relationship between duration and intensity of cigarette smoking and frequency of *EGFR* mutations suggesting that smoking history has predictive value for the presence of *EGFR* mutations (7, 8).

Although *KRAS* mutations were identified in lung cancer more than 2 decades ago (9, 10), the clinical importance of *KRAS* mutation status became apparent only relatively recently, as lung adenocarcinomas harboring *KRAS* mutations were found to show lack of response to *EGFR* TKI therapy (11, 12). *KRAS*-mutated lung cancers are prognostically unfavorable when compared with *EGFR*-mutated (13–16). In more than 95% of cases, *KRAS* missense mutations are found in codons 12 and 13 (17). Unlike *EGFR* mutations, *KRAS* mutations show no sex predilection, are more frequent in white populations than Asians, and most patients are former or current cigarette smokers (18, 19). *KRAS* mutations known to be smoking-associated (G12C, G12V) are transversion mutations (G > T and G > C), whereas *KRAS* transitions mutations (G > A) are more common in lung adenocarcinomas from patients without any smoking history (20, 21).

Even though the distinctive distribution of *EGFR* and *KRAS* mutations in relation to ethnicity, sex, and smoking history suggests that patient characteristics have a significant predictive value for the presence of these mutations, the etiology of most mutations arising in never smokers remains unknown. In this study, we hypothesized that correlations between demographic, epidemiologic and clinical data, and types of *EGFR* and *KRAS* mutations could

Authors' Affiliations: Departments of [1]Pathology, [2]Epidemiology and Biostatistics, [3]Medicine, and [4]Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Marc Ladanyi, Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065. Phone: 212-639-6369; Fax: 212-717-3515; E-mail: ladanyim@mskcc.org

doi: 10.1158/1078-0432.CCR-11-3265

Translational Relevance

To clarify the molecular epidemiology of *EGFR* and *KRAS* mutations in lung adenocarcinoma, we examined tumor genotyping data in 3,026 patients in relation to demographic, clinical, and smoking history data. In addition to the expected reciprocal associations of *EGFR* and *KRAS* mutations with smoking history, this showed that 11% of smokers had *EGFR*-mutated tumors and 6% of never smokers had *KRAS*-mutated tumors. Pack-years of smoking were predictive for *EGFR* and *KRAS* mutations but even in the context of a nomogram, it is difficult to identify a significant subset of smokers with an *EGFR* mutation likelihood of less than 1%, and therefore, our data do not support excluding any patient subset from *EGFR* testing. The distinct types of *KRAS* mutations in smokers versus never smokers suggest that most *KRAS*-mutant lung cancers in never smokers are not because of second-hand smoke exposure. The higher frequency of *KRAS* G12C in women, their younger age, and lesser smoking history support a heightened susceptibility to tobacco carcinogens.

provide a better insight into specific etiology and/or biology of these mutations. Therefore, we took the advantage of our large clinical dataset and conducted an in-depth retrospective analysis of more than 3,000 consecutive lung adenocarcinoma cases subjected to routine testing for *EGFR* and *KRAS* mutations over a 5-year period.

Materials and Methods

Clinical samples/patients

From September 2004 to December 2009, 3,026 lung adenocarcinomas (including 2 adenosquamous and 1 large cell carcinoma with adenocarcinoma component) were consecutively received and clinically tested for the presence of *EGFR* exon 19 deletion and exon 21 L858R mutation. In January 2006, testing for *KRAS* mutations (codons 12 and 13) was introduced for all cases negative for *EGFR* mutation, and 2,529 cases were received after that time. Cases with more than 1 tumor were included if: all the tumors were either mutation negative, harbored the same mutation, or if 1 tumor harbored *EGFR* or *KRAS* mutation and the other(s) was (were) mutation negative. Twenty-three patients with more than 1 tumor harboring different *KRAS* or *EGFR* mutations were excluded from the study; some of these have been reported separately (22). Clinical samples submitted for molecular testing included surgically resected tumor samples, biopsies, and cytology specimens. Clinical data were collected with the approval of Institutional Review Board of Memorial Sloan-Kettering Cancer Center (MSKCC, New York). Stage designated as IIIB/IV included stages IIIB, IV, and multifocal bronchioloalveolar carcinoma. Smoking status was defined as never smokers (< 100 lifetime cigarettes), former smokers (quit >1 year before diagnosis), or current smokers (still smoking, or quit <1 year before diagnosis). Pack-years of smoking was defined as average number of cigarettes per day/20 $\times$ years smoking.

Mutation detection

DNA was extracted using a kit (DNeasy; Qiagen) from frozen tumor tissue or formalin-fixed paraffin embedded tumor tissue. If necessary, manual microdissection of paraffin sections was done to ensure at least 50% tumor content. *EGFR* mutations were detected by sensitive PCR-specific assays as previously described (23). *KRAS* mutations were detected by PCR sequencing of exon 2 as described (11). In limited volume tumor samples, presence of an exuberant inflammatory response or extensive fibrosis, PCR was conducted with addition of locked nucleic acid oligonucleotide to favor the amplification of mutated allele, if present (24).

Statistical analysis

Cases were divided into 3 groups based on mutation status: *EGFR*-mutated, *KRAS*-mutated, or wild-type for *EGFR*/*KRAS*. The associations were tested between the mutation groups and the demographic or clinical characteristics, and the smoking status using Fisher exact test or unpaired *t* test. A *P* value <0.01 was considered significant. The Bonferroni method was used to control for family-wise error rate. Univariate and multivariate logistic regression analyses were used to test the association of smoking-free years and pack-years of smoking with *EGFR* and *KRAS* mutational status.

Nomogram development and validation

A nomogram was generated for the likelihood of *EGFR* mutation among Caucasian smokers based on the following logistic regression model: EGFR $\sim \beta_0 + \beta_1$ smoke-free-years $+ \beta_2$pack-years $+ \beta_3$gender $+ \beta_4$age $+ \beta_5$age^2. The quadratic term allows a U-shape pattern of the age association with the mutation status. All analyses were conducted using the R package Design and Hmisc. An independent data set was used for validation (25); specifically, we used 375 adenocarcinoma patients who were Caucasian smokers from the Boston cohort included in the study by Girard and colleagues (25) as the validation cohort.

Results

Table 1 summarizes characteristics of patients with lung cancer with *EGFR* and *KRAS* mutations. Our lung adenocarcinoma patient population was predominantly female (1,898/3,026, 62.7%) and this was consistent in each year from 2006 to 2009 (1,624/2,620, 62%) reflecting the routine reflex *EGFR*/*KRAS* testing that was initiated in 2006 (26). Only 13% of the cases (406/3,026) were submitted for testing before 2006 and these showed a slightly higher female to male ratio (274/406, 67.5%) presumably reflecting some referral bias. Of 3,026 cases tested clinically for the 2 major *EGFR* mutations, 593 (20%) were mutated, including 347 exon 19 deletions (59%) and 246 L858R mutations

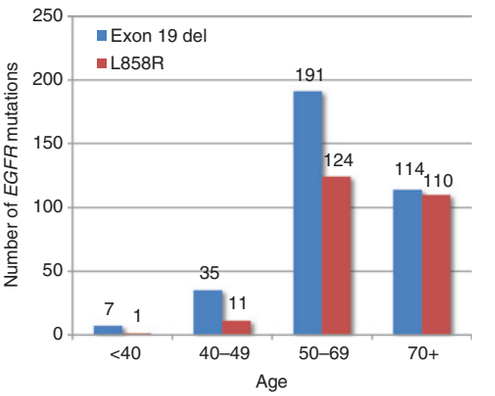
Table 1. Demographic and clinical characteristics of patients according to the *EGFR* and *KRAS* mutation status

	EGFR mutations				*KRAS* mutations			
	All patients ($N = 3026$)	Patients with EGFR mutations ($N = 593$)	All 3026 patients % (95% CI)	*P*	All patients ($N = 2529$)	Patients with KRAS mutations ($N = 670$)	All 2529 patients % (95% CI)	*P*
Sex								
Male	1,128	170	15.1 (13.1–17.3)	0.0001	959	248	25.9 (23.2–28.7)	0.58
Female	1,898	423	22.3 (20.5–24.2)		1,570	422	26.9 (24.7–29.1)	
Age, y								
Median (average)	66 (65)	66 (65)	NA		66 (65)	66 (66)	NA	
Range	15–96	24–90	NA		15–96	30–88	NA	
Ethnicity/race								
White	2,736	478	17.5	NA	2,285	641	28.1	NA
Asian/Pacific	136	75	55.1 (46.8–63.2)	0.0001	114	7	6.1 (2.8–12.4)	0.0001
Black	77	16	20.8 (13.1–31.2)	0.45	66	12	18.2 (10.6–29.3)	0.09
Asian/Indian	28	14	50 (32.6–67.4)	0.0001	23	0	0	0.0007
Other/Unknown	49	10	20.4 (11.3–33.8)		41	10	24.4 (13.7–39.5)	
Stage								
I	902	165	18.3 (15.9–21.0)		760	207	27.2 (24.2–30.5)	
II	188	35	18.6 (13.7–24.8)		158	43	27.2 (20.9–34.7)	
IIIA	260	40	15.4 (11.5–20.3)		210	54	25.7 (20.3–32.0)	
IIIB and IV	1,676	353	21.1 (19.2–23.7)		1,401	366	26.1 (23.9–28.5)	
Smoking history								
Never smokers	828	352	42.5 (39.2–45.9)	NA	669	43	6.4 (4.8–8.6)	NA
Former smokers	1,548	209	13.5 (11.9–15.3)	0.0001	1,297	419	32.3 (29.8–34.9)	0.0001
Current smokers	650	32	4.9 (3.5–7.7)	0.0001	563	208	36.9 (33.1–41.0)	0.0001

NOTE: *P* values compare the frequency of *EGFR* or *KRAS* mutations between men and women, between White patients and other ethnicities/races, and between never smokers and former and current smokers, respectively.

(41%). Patients with *EGFR* L858R tended to be older than exon 19- mutated (median age 68 vs. 64; $P = 8.1 \times 10^{-5}$), reflected by an exon 19 del to L858R ratio of 3.5 less than age 50 ($P = 0.002$), and of 1.0 in patients aged 70 and more ($P = 0.004$; Fig. 1). Men with *EGFR* mutations were more likely than women to present at late stage (i.e., IIIB/IV) of disease (118/170, 69% vs. 235/423, 56%; $P = 0.002$), whereas women predominated at stage I (31% vs. 19%, $P = 0.004$; Supplementary Fig. S1A). Tumors with *EGFR* L858R presented more often at stage I than tumors with exon 19 del (83/246, 34% vs. 82/347, 24%; $P = 0.009$; Supplementary Fig. S1B). Testing of 2,529 cases for *KRAS* mutations (codons 12 and 13) detected 670 (26%) mutations, including G12C (39%), G12V (21%), G12D (17%), G12A (11%), and other G12 and G13 mutations (12%). Although none of the *EGFR*-mutated tumors in the present clinical data set were tested for concomitant *KRAS* mutations, our more recent experience using multiplex genotyping by MALDI-TOF mass spectrometry (Sequenom) further confirm their mutually exclusive occurrence pattern (27). No significant differences in age or stage at presentation were noted between different subtypes of *KRAS* mutations (Supplementary Fig. S2).

The positive and negative associations of *KRAS* and *EGFR* mutations, respectively, with smoking are well known but had not previously been analyzed in detail in a single large dataset. Figure 2 illustrates the frequency of *EGFR* and *KRAS* mutations in relation to smoking history and smoking pack-years. *EGFR* mutations were found in 352 of 828 (43%) of never smokers and in 241 of 2,198 (11%) former and current smokers. There was no significant difference in frequency of *EGFR* exon 19 del versus *EGFR* L858R relative to smoking pack-years (data not shown). *KRAS* mutations were found in 627/1,860 (34%) of former and current smokers and in 43/669 (6%) of never smokers, the latter proportion being notably lower than in a smaller study from our center but within the confidence interval of the previously reported higher percentage (21). Although any smoking history significantly decreased the likelihood of *EGFR* mutations, no difference was noted among smokers with less than 10 pack-years smoking history. Furthermore, in smokers of more than 10 pack-years, *EGFR* mutations were 5-fold less likely to be found than in never smokers ($P = 0.0001$). In contrast, the proportion of *KRAS*-mutated lung cancers was significantly higher in smokers with any smoking history than in never smokers; among smokers, we



	EGFR exon 19 del (*N* = 347)	*EGFR* L858R (*N* = 246)	*P* value
Age, y			
Median (average)	64 (64)	68 (67)	8.1×10^{-5}
Range			
<40	7 (2%)	1 (0.4%)	0.002^
40–49	35 (10.1%)	11 (4.5%)	
50–69	191 (55%)	124 (50.4%)	
70+	114 (32.9%)	110 (44.7%)	0.004*

Figure 1. Age distribution of *EGFR* exon 19 del and *EGFR* L858R. Patients with *EGFR* L858R mutant tumors presented at older age than those harboring *EGFR* exon 19 del (median age 68 vs. 64; $P = 8.1 \times 10^{-5}$). Fisher exact test, *P* value < 0.01 is considered significant.

found 15 pack-years as a cut-point above which the likelihood of a lung cancer harboring *KRAS* mutations was 6-fold higher than in never smokers ($P = 0.0001$). Notably, even in patients with up to 10 pack-years of smoking, tumors with *EGFR* mutations were still more common than those with *KRAS* mutations.

The effect of smoking and smoking-free period on the likelihood of *EGFR* mutation has been previously reported in Asian patients with lung adenocarcinoma (28, 29), but the impact of these 2 smoking variables on the proportions of lung adenocarcinomas with either *EGFR* or *KRAS* mutations has not been previously investigated in a predominantly white patient population. Because smoking-free years and pack-years of smoking are partly dependent variables, we conducted a multivariate logistic regression analysis to examine the effect of these 2 parameters in current and former Caucasian smokers. Interestingly, this showed that, among patients with lung cancer, smoking-free years change the likelihood of *EGFR* mutation but not that of *KRAS* mutation (Supplementary Table S1).

Given the variety of possible nucleotide substitutions leading to missense mutations of *KRAS* G12 and G13, we examined their association with smoking in this large dataset. Among never smokers, the most common *KRAS* mutation was G12D (56%), and G12C was the most frequent mutation among former and current smokers (41%; Fig. 3A). Never smokers were significantly more likely than former and current smokers to have G > A transition mutations (as in G12D; 58% vs. 19% vs. 21%; $P = 0.0001$), whereas G > T transversion mutations (as in G12C), a typical change associated with tobacco carcinogens, was the most common nucleotide change in former and current smokers (67% and 71%, respectively; Fig. 3B). Compared with other *KRAS* mutations types, G12C was more frequent in women ($P = 0.007$; Fig. 3C), who were also younger than men with the same mutation (median age 65 vs. 69; $P = 0.0008$). Intriguingly, women with G > T transversions had smoked less (average 34 pack-years vs. 40 pack-years; $P = 0.001$; Supplementary Table S2) and were younger than men with the same nucleotide change

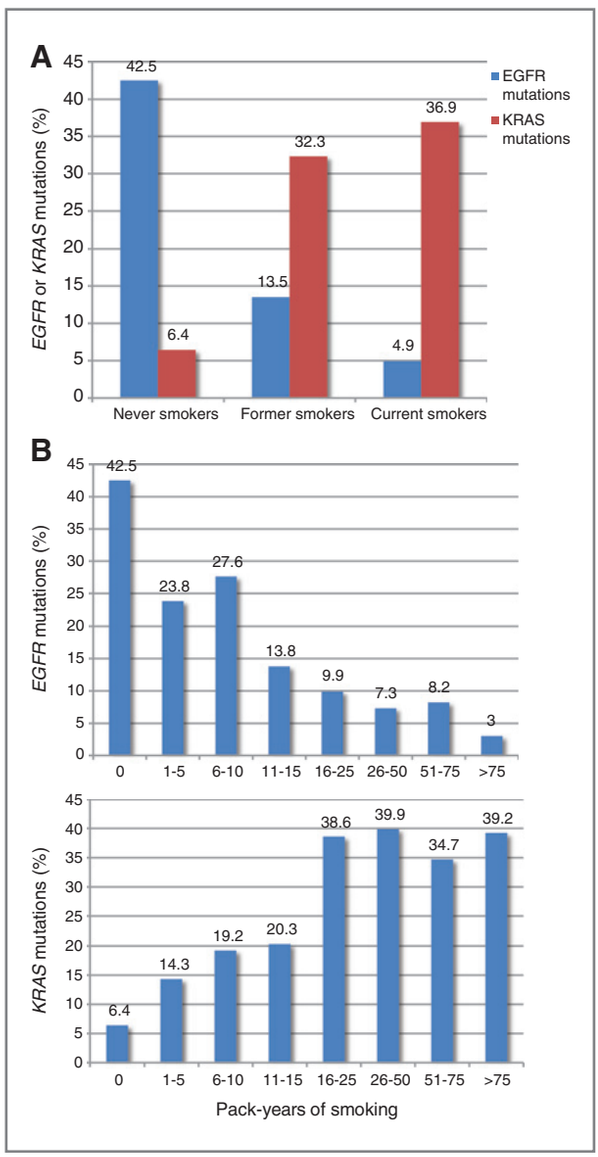
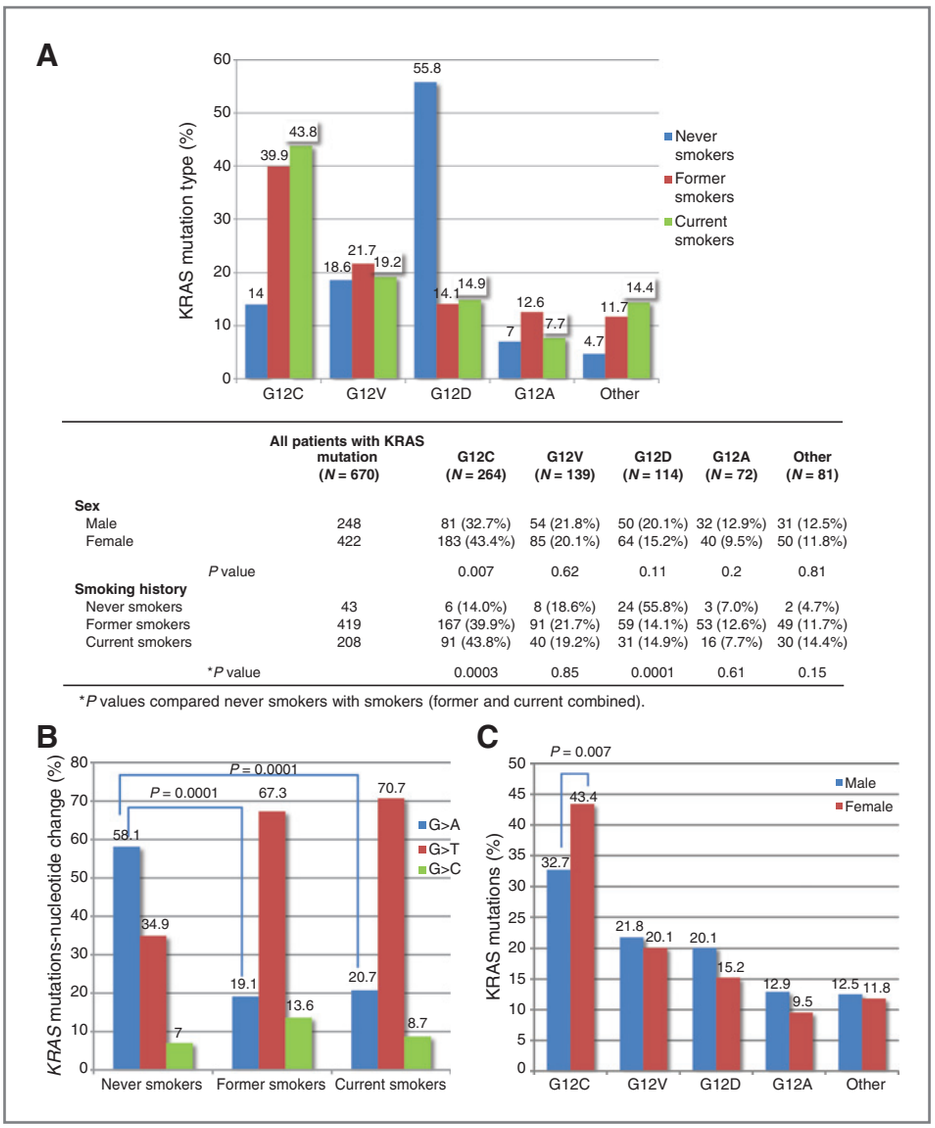


Figure 2. A, frequency of *EGFR* and *KRAS* mutations by smoking history. B, frequency of *EGFR* and *KRAS* mutations by pack-years of smoking. In the range of up to 10 pack-years, tumors with *EGFR* mutations are still more common than *KRAS* mutations.



	All patients with KRAS mutation (*N* = 670)	G12C (*N* = 264)	G12V (*N* = 139)	G12D (*N* = 114)	G12A (*N* = 72)	Other (*N* = 81)
Sex						
Male	248	81 (32.7%)	54 (21.8%)	50 (20.1%)	32 (12.9%)	31 (12.5%)
Female	422	183 (43.4%)	85 (20.1%)	64 (15.2%)	40 (9.5%)	50 (11.8%)
P value		0.007	0.62	0.11	0.2	0.81
Smoking history						
Never smokers	43	6 (14.0%)	8 (18.6%)	24 (55.8%)	3 (7.0%)	2 (4.7%)
Former smokers	419	167 (39.9%)	91 (21.7%)	59 (14.1%)	53 (12.6%)	49 (11.7%)
Current smokers	208	91 (43.8%)	40 (19.2%)	31 (14.9%)	16 (7.7%)	30 (14.4%)
**P* value		0.0003	0.85	0.0001	0.61	0.15

**P* values compared never smokers with smokers (former and current combined).

Figure 3. *KRAS* mutation type as a function of smoking history. A, *KRAS* G12D is the most common mutation in never smokers and *KRAS* G12C is the most frequent mutation among former and current smokers. B, never smokers are significantly more likely to have G > A transition mutation ($P < 0.0001$). G > T transversion is the most common nucleotide change in former and current smokers ($P < 0.0001$). C, *KRAS* G12C was relatively more frequent in women than in men ($P = 0.007$). Fisher exact test, *P* value <0.01 is considered significant.

(median age 64 vs. 67; $P = 0.006$). As discussed below, this pattern of findings suggests an increased susceptibility to tobacco carcinogenesis in women.

There is continuing interest in using clinical variables to prioritize *EGFR* mutation testing. Certain patient subsets, such as Asians and never-smokers are routinely tested, whereas other subsets, such as male Caucasian smokers, are considered of lower priority for testing. However, it is also becoming clear that these patient characteristics should not be used individually to exclude patients from testing, as shown in a recent analysis of a subset of the present data (30). Given the significant associations of *EGFR* mutation with sex ($P = 0.01$), pack-years of smoking ($P < 0.0001$), and smoking-free years ($P = 0.002$), we used these variables along with age to generate a nomogram to predict the *EGFR* status specifically in Caucasian smokers (current and former). We excluded Asians and never smokers from the nomogram dataset because it is generally agreed that patients in these groups should be tested regardless. The area under the receiver operating characteristic (ROC) curve was 0.70 (Fig. 4). To validate the performance of our nomogram in an independent dataset, we used the Caucasian smokers from the Boston cohorts used in the study by Girard and colleagues (25). In this independent set of patients, our nomogram generated an area under the ROC curve for predicting *EGFR* status of 0.71 (Supplementary Fig. S3). In the MSKCC training dataset ($n = 2078$), 16 had a predicted probability of *EGFR* mutation of 1% or less, and none were *EGFR*-mutated, and 421 had a predicted probability of 0.05 or lower, of which 14 (3%) had *EGFR* mutations. In the Boston dataset ($n = 375$) used for validation, 10 patients had probability below 1%, one of which was *EGFR* mutated and 145 had a probability below 5%, including 10 (7%) *EGFR*-mutated cases. As discussed below, we view these results as indicating that, even in the context of a rigorously developed nomogram, clinical

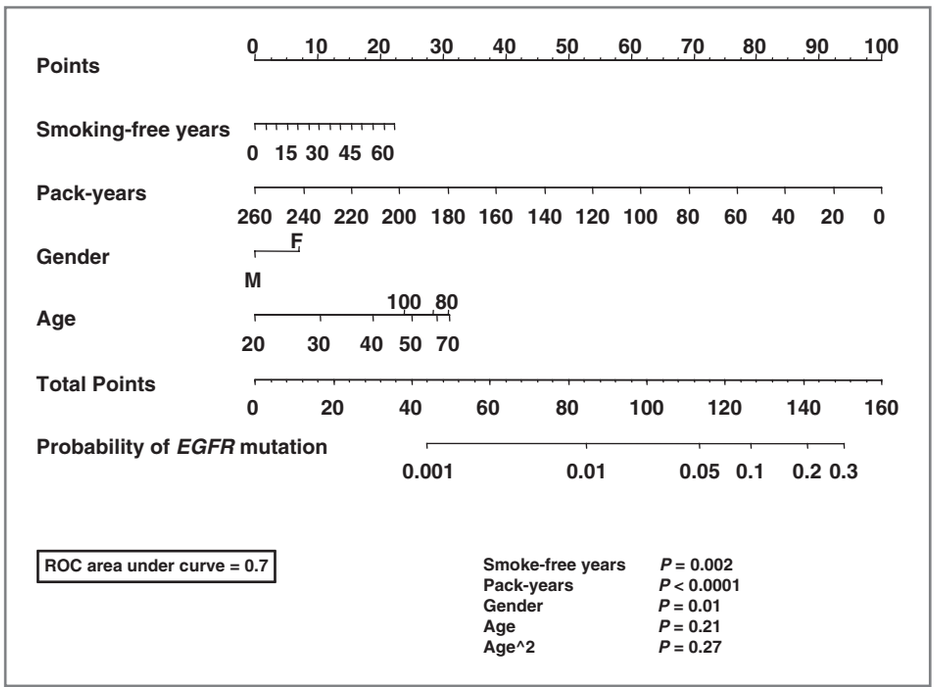


Figure 4. Development of a nomogram including clinical variables and smoking history data for prediction of *EGFR* mutant status among Caucasian smokers (current or former). Mark the smoking-free years on the axis and draw vertical line up to the points axis to determine the number of points. Repeat the same for pack-years, gender, and age, and sum the total points for all 4 variables. Plot the given number on the total points axis and draw a vertical line down to the probability of *EGFR* mutation.

variables cannot be used to robustly identify patients with a negligible chance of harboring a *EGFR*-mutated lung cancer.

Discussion

To accurately and reliably determine the frequency of the major mutations in *EGFR* and *KRAS* in lung adenocarcinoma in relation to patient characteristics and different levels of smoking, a sufficiently large number of case subjects is necessary to provide statistical power for more detailed analyses. Here, we carried out a retrospective analysis of our large clinical database of lung adenocarcinomas with established *EGFR*/*KRAS* mutation status. (i) We found distinct differences in sex, age, and stage distribution of 2 most common types of *EGFR* mutations; (ii) we determined the likelihood of *EGFR* and *KRAS* mutations by intensity and duration of smoking; (iii) evaluated the effects of smoking-free period on the proportions of *EGFR* and *KRAS* mutations in lung cancers arising in former smokers; (iv) we designed a nomogram to predict presence of *EGFR* mutation in Caucasian smokers; (v) we noted a distinct distribution of types of *KRAS* mutations in smokers versus never smokers; and (vi) we observed significant sex and age differences in the frequency of G12C as the most common smoking-related *KRAS* mutation.

EGFR exon 19 del was relatively more common than L858R mutation in younger patients. Notably, of 8 patients below the age of 40 years with *EGFR*-mutated lung adenocarcinoma, 7 were *EGFR* exon 19 del. In contrast, L858R occurred in a relatively older age distribution and the patients more often presented with stage I disease. These findings may suggest a potentially more aggressive natural history of adenocarcinomas with *EGFR* exon 19 del compared with tumors with the L858R mutation. Differences between *EGFR* exon 19 del and L858R-mutated tumors have been reported in patients treated with TKI or chemotherapy. *EGFR* exon 19 deletions have been associated with better response to TKI and with a longer time-to-progression (TTP) and overall survival (OS) in patients with advanced adenocarcinoma (31–34). However, the better clinical outcome of patients with *EGFR* exon 19 del compared with patients harboring *EGFR* L858R mutations remains controversial; 2 prospective randomized phase III trial studies did not confirm these observations (35, 36). A distinct age and stage distribution as well as differences in response to molecular targeted therapy may suggest subtle differences in biology and/or etiology for *EGFR* exon 19 del and the L858R mutation.

Although typically seen in the absence of smoking history, a significant minority (11%) of former and current smokers harbors *EGFR*-mutated tumors, arguing against excluding smokers from *EGFR* testing. Moreover, among smokers with less than 10 pack-years, *EGFR* mutations were more common than *KRAS* mutations. In a study of 265 lung adenocarcinomas, some of which are included in the current dataset, Pham and colleagues found significantly fewer *EGFR* mutations in people who smoked for more than 15 pack-years or stopped smoking less than 25 years ago compared with individuals who never smoked (7). Our extended dataset allowed for more accurate risk stratification by pack-years categories and showed, that any smoking history at or above one pack-year significantly decreased the likelihood of *EGFR*-mutated tumors with no notable difference up to 10 pack-years. Although our patient population was primarily Caucasian, the results seem generalizable, as a similar relationship of *EGFR* mutations to pack-

years and smoke-free years has also been reported in Asian patients with lung cancer (28, 29, 37).

As expected, most of the *KRAS* mutations were found among current and former smokers, and consistent with other studies (38), we identified 6% of never smokers with *KRAS*-mutated tumors. In our earlier study that included 102 *KRAS*-mutated tumors (21), we failed to show predictive value of pack-years for the presence of *KRAS* mutations likely due to small number of cases. Here, we have shown that any smoking history significantly increases the likelihood of a *KRAS* mutation being found in the lung cancer. Smoking-free years provided additional value in predicting the likelihood of *EGFR* mutations but not that of *KRAS* mutations, independent of pack-years of smoking. These multivariate results suggest a model in which *KRAS* mutations occur at the time of smoking and may lead to cancer eventually, explaining the lack of impact of smoke-free years. This is also supported by the observation that former and current smokers have similar proportions of *KRAS*-mutated lung cancers (Fig. 2A). Overall, this further supports the notion that permanent DNA damage by tobacco carcinogens acquired at the time of smoking is the major source of most *KRAS*-mutated lung adenocarcinomas. Thus, the likelihood that a patient with lung cancer has a *KRAS* mutation is determined by pack-years of smoking and does not decrease significantly over time upon smoking cessation; in contrast, because overall lung cancer incidence decreases with increasing smoke-free years, the relative proportion of nonsmoking-associated cancers (represented by *EGFR*-mutated tumors) increases. Importantly, these data should not be misinterpreted as supporting a "protective" effect of smoking on the risk of *EGFR*-mutated lung adenocarcinoma.

On the basis of the need for efficient medical resource utilization and concerns regarding health care costs and possible treatment delays due to testing, there is continuing controversy regarding routine *EGFR* mutation testing in certain patient subsets perceived as having a low chance of *EGFR* mutation in their lung cancer, such as male Caucasian smokers. Using the readily available clinical parameters of age, sex, pack-years, and smoking-free years, we developed a nomogram to predict the likelihood of *EGFR* mutation in Caucasian current or former smokers with lung adenocarcinoma. We should note that a similar, recently published nomogram (25) differs in 2 important ways from the one we have developed. First, it includes never smokers, a group in which the value of *EGFR* testing is no longer in question. Second, it includes the histologic subtype of adenocarcinoma, which usually can only be properly analyzed in resection specimens, whereas decisions regarding *EGFR* testing often have to be made in advanced stage patients in whom the available small biopsies are sometimes suboptimal for histologic subtyping. The accuracy of our nomogram was 70% on the source dataset and 71% in an independent validation dataset. On the basis of clinical considerations (for instance, the fact that testing of *ALK* fusions present in only 3–5% of lung adenocarcinomas is now indicated to select patients for crizotinib; ref. 39), we deemed that only a probability of harboring an *EGFR* mutation of less than 1% was clinically negligible and therefore actionable in terms of bypassing *EGFR* testing. However, only a very small proportion of patients fall in this category, 0.8% in the source dataset and 2.7% in the validation dataset and the latter included one incorrect prediction (10% error rate). Overall, the 70% to 71% accuracy of nomogram prediction, along with the very low proportion of predictions below 1%, suggests that clinical variables cannot be used to robustly identify patients with a negligible chance of harboring an *EGFR*-mutated lung cancer. Nonetheless, our nomogram may still be helpful in situations where mutation analysis for *EGFR* is simply not possible and the clinical parameters and smoking history are used to direct the treatment decision.

In a previous smaller study, we showed that never smokers were significantly more likely than former or current smokers to have a *KRAS* transition mutation (G > A) rather than the transversion mutations known to be smoking-related (G > T or G > C; ref. 21). The much larger number of cases in the present series allowed us to robustly confirm these earlier findings as well as to detect sex and age differences in the frequency of the most common smoking-related G > T transversion mutation, *KRAS* G12C. These findings support the notion that most *KRAS*-mutant lung adenocarcinomas in never smokers are not likely to be caused by environmental (second-hand) tobacco smoke, a potentially important observation in assessing the level of risk posed by such exposure.

Sex differences in sensitivity to tobacco smoke have been well documented (40). Zang and Wynder have reported that the odds ratios for major lung cancer types are consistently higher in women than in men at every level of exposure to cigarette smoke and that these differences cannot be explained by differences in baseline exposure, smoking history, or body size, but are likely because of a higher susceptibility to tobacco carcinogens in women (41). Computed tomographic screening data suggest that female smokers are almost twice as likely as male smokers to have a lung cancer detected in spite of lesser smoking histories (42). Consistent with our findings in *KRAS*, studies of the mutational spectrum of *TP53* in relation to smoking and sex showed that cancers arising in female smokers had significantly more tobacco-related mutations (G > T transversions) than in male smokers (43, 44). Therefore, taken together, the relatively higher percentage of the female patients with tumors containing *KRAS* G12C (because of G > T transversion), their younger age at diagnosis, and the fewer pack-years of smoking in women with this *KRAS* mutation, compared with men with the same *KRAS* mutation, provide yet another type of data supporting the hypothesis that women are more susceptible to tobacco carcinogens.

The apparent increased susceptibility of women to tobacco carcinogenesis may reflect constitutive differences in genes encoding tobacco carcinogen–metabolizing enzymes. For example, the cytochrome P450 phase I detoxifying enzyme *CYP1A1* shows higher expression in the normal

lung tissue of female smokers than male smokers (45). The most common polymorphism found in cytochrome P450 phase II detoxification enzymes is the *GSTM1*-null genotype, which is present in 40% to 50% of the general population due to homozygosity for a deletion polymorphism and the impact of this *GSTM1* genotype may be enhanced in female smokers (46).

In summary, several observations emerge from this large analysis of the molecular epidemiology of *EGFR* and *KRAS* mutations in lung adenocarcinoma. Pack-years of smoking have a significant predictive value for the presence of *EGFR* and *KRAS* mutations and smoking-free years have additional predictive value for presence of *EGFR* mutations but not that of *KRAS* mutations. However, even in the context of a rigorously developed nomogram incorporating these clinical variables, it remains difficult to reliably identify a significant subset of smokers who would have an *EGFR* mutation likelihood of less than 1%, and therefore our data do not support excluding any subset of patients with lung adenocarcinoma from *EGFR* testing. Our results suggest a different etiology of *KRAS* mutations in smokers versus never smokers and firmly support earlier observations of increased susceptibility to tobacco carcinogenesis in women. More broadly, our observations strengthen the notion that careful consideration of histologic subtypes (focusing on adenocarcinoma instead of mixing all lung cancer types) and molecular subtypes defined by distinct, nonoverlapping driver mutations (*EGFR, KRAS*) can help to clarify epidemiologic associations that may otherwise remain elusive (47, 48). This approach, which recognizes the possible etiologic diversity represented by different histologic and molecular subtypes, has recently been termed molecular pathologic epidemiology (49).

Disclosure of Potential Conflicts of Interest

M.L. Johnson is employed (other than primary affiliation; e.g., consulting) in Astellas and as a consultant in Federal Government Affairs. M.L. Johnson also has a commercial research grant from Novartis and is a consultant/advisory board member of Genentech, Boehringer-Ingelheim, Chugai, Ariad, Daiichi, Novartis, Abbott Molecular, Foundation Medicine, and Celgene. M.G. Kris has a commercial research grant from Pfizer Inc. and Boehringer Ingelheim and is a consultant/advisory board member of Pfizer Inc., Boehringer Ingelheim, Roche/Genentech, Clovis, and Millenium Pharmaceuticals. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: S. Dogan, M. G. Kris, M. Ladanyi
Development of methodology: S. Dogan, M. G. Kris
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Dogan, D. C. Ang, M. L. Johnson, S. P. D'Angelo, P. K. Paik, G. J. Riely, M. G. Kris
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S. Dogan, R. Shen, P. K. Paik, G. J. Riely, M. G. Kris, M. F. Zakowski
Writing, review, and/or revision of the manuscript: S. Dogan, R. Shen, M. L. Johnson, P. K. Paik, G. J. Riely, M. G. Kris, M. F. Zakowski, M. Ladanyi
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. Dogan, E. B. Brzostowski, M. G. Kris
Study supervision: S. Dogan, M. G. Kris, M. Ladanyi

Acknowledgments

The authors thank Justyna Sadowska, Jacklyn Casanova, and Lin Dong for excellent technical support. The authors also thank Dr. Cameron Brennan for helpful discussions.

Grant Support

This work is supported by grants from NIH P01 CA129243 (to M. Ladanyi and M. G. Kris).

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

Received January 3, 2012; revised August 31, 2012; accepted September 14, 2012; published OnlineFirst September 26, 2012.

References

1. Pao W, Iafrate AJ, Su Z. Genetically informed lung cancer medicine. J Pathol 2011;223:231–41.
2. 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.
3. Paez JG, Jänne 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.
4. 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.
5. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 2005;97:339–46.
6. Gazdar AF. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene 2009;28 Suppl 1:S24–31.
7. Pham D, Kris MG, Riely GJ, Sarkaria IS, McDonough T, Chuai S, et al. Use of cigarette smoking history to estimate the likelihood of mutations in epidermal growth factor receptor gene exons 19 and21 in lung adenocarcinomas. J Clin Oncol 2006;24:1700–4.
8. Lee YJ, Shim HS, Kang YA, Hong SJ, Kim HK, Kim H, et al. Dose effect of cigarette smoking on frequency and spectrum of epidermal growth factor receptor gene mutations in Korean patients with non-small cell lung cancer. J Cancer Res Clin Oncol 2010;136:1937–44.
9. Santos E, Martin-Zanca D, Reddy E, Pierotti MA, Della Porta G, Barbacid M. Malignant activation of a K-RAS oncogene in lung carcinoma but not in normal tissue of the same patient. Science 1984;223:661–4.
10. Rodenhuis S, van de Wetering ML, Mooi WJ, Evers SG, van Zandwijk N, Bos JL. Mutational activation of the K-ras oncogene: a possible pathogenetic factor in adenocarcinoma of the lung. N Engl J Med 1987;317:929–35.
11. Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, et al. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. PLoS Med 2005;2:e17.
12. Eberhard DA, Johnson BE, Amler LC, Goddard AD, Heldens SL, Herbst RS, et al. Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. J Clin Oncol 2005;23:5900–9.
13. Graziano SL, Gamble GP, Newman NB, Abbott LZ, Rooney M, Mookherjee S, et al. Prognostic significance of K-ras codon 12 mutations in patients with resected stage I and II non-small-cell lung cancer. J Clin Oncol 1999;17:668–75.

14. Slebos RJ, Kibbelaar RE, Dalesio O, Kooistra A, Stam J, Meijer CJ, et al. K-RAS oncogene activation as a prognostic marker in adenocarcinoma of the lung. N Engl J Med 1990;323:561–5.
15. Rosell R, Molina F, Moreno I, Martínez E, Pifarré A, Font A, et al. Mutated K–ras gene analysis in a randomized trial of preoperative chemotherapy plus surgery versus surgery in stage IIIA non-small cell lung cancer. Lung Cancer 1995;12 Suppl 1: S59–70.
16. Johnson ML, Sima CS, Chaft J, Paik PK, Pao W, Kris MG, et al. Association of KRAS and EGFR mutations with survival in patients with advanced lung adenocarcinoma. J Clin Oncol 28:7s, 2010 (suppl; abstr 7541)
17. Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, et al. Cosmic 2005. Br J Cancer 2006;94:318–22.
18. Buttitta F, Barassi F, Fresu G, Felicioni L, Chella A, Paolizzi D, et al. Mutational analysis of the HER2 gene in lung tumors from Caucasian patients: mutations are mainly present in adenocarcinomas with bronchioloalveolar features. Int J Cancer 2006;119:2586–91.
19. Suzuki M, Shigematsu H, Iizasa T, Hiroshima K, Nakatani Y, Minna JD, et al. Exclusive mutation in epidermal growth factor receptor gene, HER-2, and KRAS, and synchronous methylation of nonsmall cell lung cancer. Cancer 2006;106:2200–7.
20. Ahrendt SA, Decker PA, Alawi EA, Zhu Yr 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.
21. Riely GJ, Kris MG, Rosenbaum D, Marks J, Li A, Chitale DA, et al. Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. Clin Cancer Res 2008;14: 5731–4.
22. Girard N, Deshpande C, Azzoli CG, Rusch VW, Travis WD, Ladanyi M, et al. Use of EGFR/KRAS mutation testing to define clonal relationships among multiple lung adenocarcinomas: comparison with clinical guidelines. Chest 2010;137:46–52.
23. Pan Q, Pao W, Ladanyi M. Rapid polymerase chain reaction-based detection of epidermal growth factor receptor gene mutations in lung adenocarcinomas. J Mol Diagn 2005;7:396–403.
24. Arcila M, Lau C, Nafa K, Ladanyi M. Detection of *KRAS* and *BRAF* mutations in colorectal carcinoma roles for high-sensitivity locked nucleic acid-PCR sequencing and broad-spectrum mass spectrometry genotyping. J Mol Diagn 2011;13:64–73.
25. Girard N, Sima CS, Jackman DM, Sequist LV, Chen H, Yang JC, et al. Nomogram to predict the presence of *EGFR* activating mutation in lung adenocarcinoma. Eur Respir J 2012;39:366–72.
26. D'Angelo SP, Park B, Azzoli CG, Kris MG, Rusch V, Ladanyi M, et al. Reflex testing of resected stage I through III lung adenocarcinomas for EGFR and KRAS mutation: report on initial experience and clinical utility at a single center. J Thor Cardiovasc Surg 2011;141:476–80.
27. Arcila ME, Chaft JE, Nafa K, Roy-Chowdhuri S, Lau C, Zaidinski M, et al. Prevalence, clinicopathologic associations and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. Clin Cancer Res 2012;18:4910–8.
28. Sugio K, Uramoto H, Ono K, Oyama T, Hanagiri T, Sugaya M, et al. Mutations within the tyrosine kinase domain of EGFR gene specifically occur in lung adenocarcinoma patients with a low exposure of tobacco smoking. Br J Cancer 2006;94:896–903.
29. Matsuo K, Ito H, Yatabe Y, Hiraki A, Hirose K, Wakai K, et al. Risk factors differ for non-small-cell lung cancers with and without EGFR mutation: assessment of smoking and sex by a case-control study in Japanese. Cancer Sci 2007;98:96–101.
30. D'Angelo SP, Pietanza MC, Johnson ML, Riely GJ, Miller VA, Sima CS, et al. Incidence of EGFR exon 19 deletions and L858R mutations in tumor specimens from men and cigarette smokers with lung adenocarcinomas. J Clin Oncol 2011;29:2066–70.
31. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 2009;361:958–67.
32. Jackman DM, Miller VA, Cioffredi LA, Yeap BY, Jänne PA, Riely GJ, et al. Impact of epidermal growth factor receptor and KRAS mutations on clinical outcomes in previously untreated non-small cell lung cancer patients: results of an online tumor registry of clinical trials. Clin Cancer Res 2009;15:5267–73.
33. Won YW, Han JY, Lee GK, Park SY, Lim KY, Yoon KA, et al. Comparison of clinical outcome of patients with non-small-cell lung cancer harbouring epidermal growth factor receptor exon 19 or exon 21 mutations. J Clin Pathol 2011;64:947–52.
34. Kim DW, Lee SH, Lee JS, Lee MA, Kang JH, Kim SY, et al. A multicenter phase II study to evaluate the efficacy and safety of gefitinib as first-line treatment for Korean patients with advanced pulmonary adenocarcinoma harboring EGFR mutations. Lung Cancer 2011;71:65–9.
35. Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, et al. 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;11:121–8.
36. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med 2010;362:2380–8.
37. Huang YS, Yang JJ, Zhang XC, Yang XN, Huang YJ, Xu CR, et al. Impact of smoking status and pathologic type on epidermal growth factor receptor mutations in lung cancer. Chin Med J (Engl) 2011;124:2457–60.
38. Mao C, Qiu LX, Liao RY, Du FB, Ding H, Yang WC, et al. KRAS mutations and resistance to EGFR-TKIs treatment in patients with non-small cell lung cancer: A meta-analysis of 22 studies. Lung Cancer 2010;69:272–8.
39. Riely GJ, Chaft JE, Ladanyi M, Kris MG. Incorporation of crizotinib into the NCCN Guidelines. J Natl Compr Canc Netw 2011;9:1328–30.
40. Kiyohara C, Ohno Y. Sex differences in lung cancer susceptibility: a review. Gend Med 2010;7:381–401.
41. Zang EA, Wynder EL. Differences in lung cancer risk between men and women: examination of the evidence. J Natl Cancer Inst 1996;88: 183–92.
42. International Early Lung Cancer Action Program Investigators. Women's susceptibility to tobacco carcinogens and survival after diagnosis of lung cancer. JAMA 2006;296:180–4.
43. Bennett WP, Hussain SP, Vahakangas KH, Khan MA, Shields PG, Harris CC, et al. Molecular epidemiology of human cancer risk: gene-environment interactions and p53 mutation spectrum in human lung cancer. J Pathol 1999;187:8–18.
44. Toyooka S, Tsuda T, Gazdar AF. The TP53 gene, tobacco exposure, and lung cancer. Hum Mutat 2003;21:229–39.
45. Uppstad H, Osnes GH, Cole KJ, Phillips DH, Haugen A, Mollerup S. Sex differences in susceptibility to PAHs is an intrinsic property of human lung adenocarcinoma cells. Lung Cancer 2011;71:264–70.
46. Tang DL, Rundle A, Warburton D, Santella RM, Tsai WY, Chiamprasert S, et al. Associations between both genetic and environmental biomarkers and lung cancer: evidence of a greater risk of lung cancer in women smokers. Carcinogenesis 1998;19:1949–53.
47. Wakelee HA, Gomez SL, Chang ET. Sex differences in lung-cancer susceptibility: a smoke screen? Lancet Oncol 2008;9:609–10.
48. Gazdar AF, Thun MJ. Lung cancer, smoke exposure, and sex. J Clin Oncol 2007;25:469–71.
49. Ogino S, Chan AT, Fuchs CS, Giovannucci E. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. Gut 2011;60:397–411.