Molecular epidemiology of lung cancer and geographic variations with special reference to *EGFR* mutations

Tetsuya Mitsudomi

Division of Thoracic Surgery, Department of Surgery, Kinki University Faculty of Medicine, Osaka-Sayama, Japan

Correspondence to: Tetsuya Mitsudomi. Division of Thoracic Surgery, Department of Surgery, Kinki University Faculty of Medicine, Osaka-Sayama 589-8511, Japan. Email: mitsudom@surg.med.kindai.ac.jp.

Abstract: Lung cancer is a leading cause of cancer-related mortality in many countries. Although recent advances in targeted therapy against driver oncogenes have significantly improved patient outcome, cure of this disease is still exceptional. Although tobacco is a known cause of lung cancer, not all smokers develop lung cancer, and conversely many patients, especially Asian female patients with lung cancer, are lifetime never-smokers. Therefore, efforts to understand the basis for different susceptibilities to lung cancer among individuals with different genetic, biologic, ethnic, and social backgrounds are important to help develop effective preventive measures. Lung cancer in never-smokers has many different characteristics to lung cancer in smokers, such as adenocarcinoma predominance and high frequency of epidermal growth factor receptor (*EGFR*) mutation yet low number of genetic changes. Epidemiologic studies suggest that East Asians are more susceptible to smoking-unrelated lung cancer but less susceptible to smoking-related lung cancer compared with Caucasians. Mutations in the *EGFR* gene are more common in Asian females and never-smokers. Our case-control study suggests that *EGFR* mutation occurs independent of smoking, and that the apparent low frequency of *EGFR* mutations in smokers may be the result of dilution by smoking-related lung cancer. The frequencies of three *EGFR* gene polymorphisms associated with increased protein expression are significantly different between East Asians and Caucasians, favoring lower protein expression in East Asians. Although these may be associated with preferred expression of the *EGFR* mutant allele, it is difficult to explain the frequent EGFR mutation in Asian patients. Genome wide association studies (GWAS) revealed several loci related to lung cancer susceptibility. In the future, GWAS may identify loci that are specifically related to EGFR-targeted carcinogenesis, leading to identification of carcinogens that induce *EGFR* mutations and effective prevention measures.

Keywords: Susceptibility; tobacco smoke; never smokers; ethnic difference; adenocarcinoma; genome wide association studies (GWAS)

Submitted Aug 01, 2014. Accepted for publication Aug 07, 2014.
doi: 10.3978/j.issn.2218-6751.2014.08.04
View this article at: http://dx.doi: 10.3978/j.issn.2218-6751.2014.08.04

Introduction

Recent progress in the molecular understanding of lung cancer revealed that lung cancer is very heterogeneous in terms of driver oncogenes and hence the response to different drugs. The incidence of molecular alteration of a given driver oncogene is dependent on various epidemiologic factors, such as ethnicity, sex, and age of the patient. For example, mutation of the epidermal growth factor receptor (*EGFR*) gene is frequently observed in patients who are female, non-smoking and with East Asian ethnicity (1). However, translocation of the anaplastic lymphoma kinase (*ALK*) gene is common in young patients with non-smoking history, with no apparent ethnic differences (2). In this review, the molecular epidemiology of lung cancer is discussed with special reference to ethnic differences of *EGFR* mutations in lung cancer.

Table 1 Comparison of percentages of population smoking with lung cancer mortality among selected countries (3,4)

Countries	Population smoking (%) (A)	Lung cancer mortality 2009 (age standardized rate per 100,000 population) (%) (B)	B/A
United States	14.8	82.6	5.6
Iceland	14.3	73.4	5.1
United Kingdom	19.6	71.7	3.7
Hungary	26.5	95.5	3.6
Netherlands	20.8	74.5	3.6
Australia	15.1	53.0	3.5
Germany	21.9	51.8	2.4
Italy	22.5	51.6	2.3
Korea	23.2	52.3	2.3
France	23.3	51.5	2.2
Japan	20.1	43.8	2.2
Chile	29.8	30.3	1.0

Lung cancer mortality and smoking rate ranking are not correlated

Although tobacco is well known as a major cause of lung cancer, the international ranking of the percentage of population aged 15 years or older who smoke daily by country (3) does not correlate with ranking of lung cancer mortality (4), according to the Organization for Economic Cooperation and Development (OECD) data as shown in *Table 1*. For example, the United States has the 6th lowest smoking rate (at 14.8%), yet the age-standardized rate of lung cancer mortality per 100,000 people is 82.6, which is the 4th highest. Conversely, in Chile, the smoking rate is high (29.8%), but lung cancer mortality is low (30.3) compared with other countries. One can speculate that a great deal of heterogeneity exists related to ethnic/racial differences in terms of lung cancer susceptibility related to smoking, e.g., in the United States, Iceland, and United Kingdom, this susceptibility appears to be high, while it is low in Chile, Japan, France, and Korea.

Difference in smoking rate between Asia and other countries according to sex

Figure 1A illustrates the smoking rate in various countries according to sex. In general, the female smoking rate is low and male smoking rate is high in Asian countries, resulting in a large difference in smoking rates between men and women. There is also a difference in the smoking rate in lung cancer patients by sex and by country (*Figure 1B*). In our cohort of Japanese patients with lung cancer, as high as 83% of female patients were never-smokers, compared with only 10% of male patients (5). This trend is also observed in Singapore, where never-smokers account for 31.5% and 68.5% of males and females, respectively (6). In contrast, data from the Western world show that the majority of patients with lung cancer are smokers irrespective of sex of the patients (7,8). For example, only 15% of 706 females and 6% of 1,347 males with lung cancer are never-smokers (8). Thus, it is noteworthy that the smoking rate in female patients with lung cancer is very close to that in females without lung cancer in Japan, but this is not the case with US females.

Higher susceptibility to lung cancer in Caucasians than in East Asians in smokers, but opposite relationship in never-smokers

In an analysis of 13 cohorts and 22 cancer registry studies, the lung cancer mortality rates of current smokers in US males, Asian (Korean) males, US females and Asian females were 302.7, 120.8, 142.7 and 36.3, respectively (*Table 2*) (9). Thus, males are more susceptible to tobacco-related lung cancer, as are US people compared with East Asians. However, gender difference of susceptibility to lung cancer risk has been controversial. Several studies indicated higher odds ratios in women than in men. For example, Risch *et al*. reported that the odds ratio of lung cancer for individuals with a 40 pack/year smoking rate relative to never smokers was 27.9 in women compared with 9.6 in men (10). A

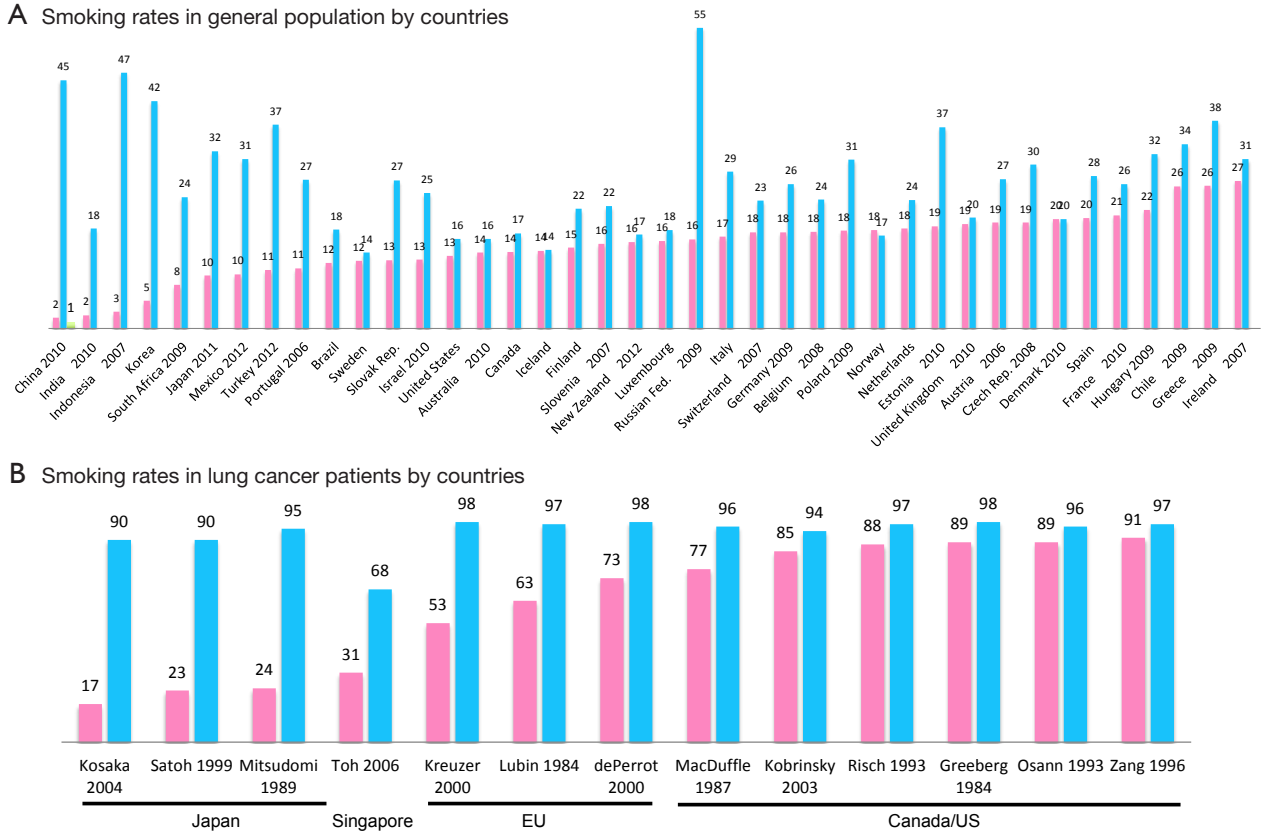

Figure 1 (A) Smoking rates in general population (A) (3) and in lung cancer patients (B) (5-8) by countries and sex.

Table 2 Lung cancer mortality rates (per 100,000) among current smokers and never smokers by sex [compiled from supplementary tables from (7)]

	Male					Female					
	CPS-1	CPS-II	HPFS	KCPS	Three prefectures	CPS-1	CPS-II	WHS	NHS	KCPS	Three prefectures
Country		US		Korea	Japan		US			Korea	Japan
Current smokers		302.7		120.8			143			36.3	
Never smokers	15.3	13.4	12.6	27.5	34.9	9.3	10.6	4	10.3	16.1	18.1

CPS-I, Cancer Prevention Study I; CPS-II, Cancer Prevention Study II; HPFS, Health Professional's Follow-up Study; KCPS, Korean Cancer Prevention Study; WHS, Women's Health Study; NHS, Nurses' Health Study; Three prefectures, three prefectures study.

similar higher susceptibility among women was reported by other groups (11,12). The reason for this discrepancy may be attributed to several factors. Exposure to environmental tobacco smoke may be different between men and women. The fact that women with lung cancer tend to have better prognosis would cause differences in the incidence and mortality data. In contrast, lung cancer mortality in Caucasian never-smokers is lower than that in the Asian population, both in males and females.

Smoking and histologic types of lung cancer

The contribution of smoking status to the development of lung cancer is dependent on histologic types. For small cell or squamous cell carcinomas, the odds ratio (OR) for smoking in male/female patients are 21.4/12.1 and 18.1/9/7,

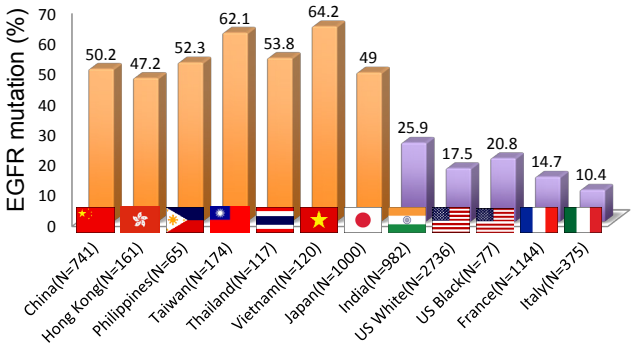

Figure 2 Frequency of *EGFR* mutations in lung cancer according to ethnic or geographic backgrounds (16-20). *EGFR*, epidermal growth factor receptor.

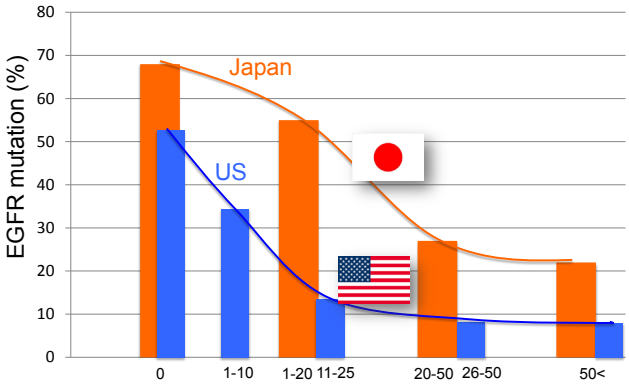

Figure 3 Incidence of *EGFR* mutations according to smoking dose in Japan (5) with that in US (21). *EGFR*, epidermal growth factor receptor.

respectively (13). In line with these data, it is rare to see small cell or squamous cell carcinomas in never-smokers. In contrast, the OR for adenocarcinomas is 1.9 for male and 1.3 for female (13). In other words, the contribution of smoking (at least direct smoking) to adenocarcinoma of the lung is lower, especially in females.

Genome wide association studies (GWAS) to identify lung cancer risk

Many studies have identified genetic susceptibility to lung cancer using GWAS. According to a recent comprehensive review article, 20 GWAS in lung cancer compared with controls identified 5p15 (*TERT-CLPTMIL*) and 15q25 (*CHRNA3*) as lung cancer susceptibility loci 11 and 7 times, respectively, in mainly Caucasian populations (14). Because characteristics of lung cancer, including *EGFR* mutation incidence or smoking habit in conduction with sex, are markedly different between Caucasians and East Asians as discussed earlier, it is of interest to evaluate whether these two loci are also associated with lung cancer susceptibility in the Asian population. Using 21 case-control studies for 11,645 lung cancer patients and 14,954 control subjects, 85% of which were Caucasians and 15% were Asians, similar associations between Caucasians and Asians were identified for the chromosome 5p15 region (15). ORs for rs2736100 (*TERT*) were 1.15 and 1.23 in Caucasians and Asians, respectively, and those for rs407210 (*CLPTM1L*) were 1.14 and 1.15 in Caucasians and Asians, respectively (15). Conversely, there were inconsistent results at 15q25 between Caucasians and Asians. ORs for rs16969968 (*CHRNA3*) were 1.26 per allele in Caucasians ($P=2\times10^{-26}$). However, the OR was 0.94 in Asians without statistical significance (P=0.67) (15). When Ito *et al.* considered the combined effect of 15q25 loci and smoking, they found that the OR for heavy smoking was 4.03 and 8.09 in patients without and with the risk allele of rs931794, respectively. Therefore, it appears that the 15q25 region modifies the effect of smoking on the risk of lung cancer in the Asian population, and its contribution to smoking-independent carcinogenesis is not significant.

EGFR mutations

EGFR is a receptor tyrosine kinase that is frequently activated in adenocarcinoma of the lung mainly by small in-frame deletions in exon 19 or a point mutation occurring at codon 858 that changes leucine to arginine (L858R) (1). *EGFR* mutation is more frequent in adenocarcinoma of the lung in patients with East Asian ethnicity without smoking history (1). *Figure 2* illustrates the frequency of *EGFR* mutations in lung cancer according to ethnic or geographic backgrounds. *EGFR* mutations are clearly more frequent in East Asian countries, including Japan, Korea, Philippines, Taiwan, Thailand, Vietnam, and Hong Kong. These differences may be partly explained by marked differences in smoking rate by countries and by sex. However, when the incidence of *EGFR* mutations according to smoking dose in Japan is compared with the United States (*Figure 3*), there is still a big difference in the incidence of *EGFR* mutation. For example, the *EGFR* mutation rate is less than 10% in patients with a 50 pack/year history or higher in the United States and it is more than 20% in Japanese patients (1). This pattern of *EGFR* mutations contrasts with that of

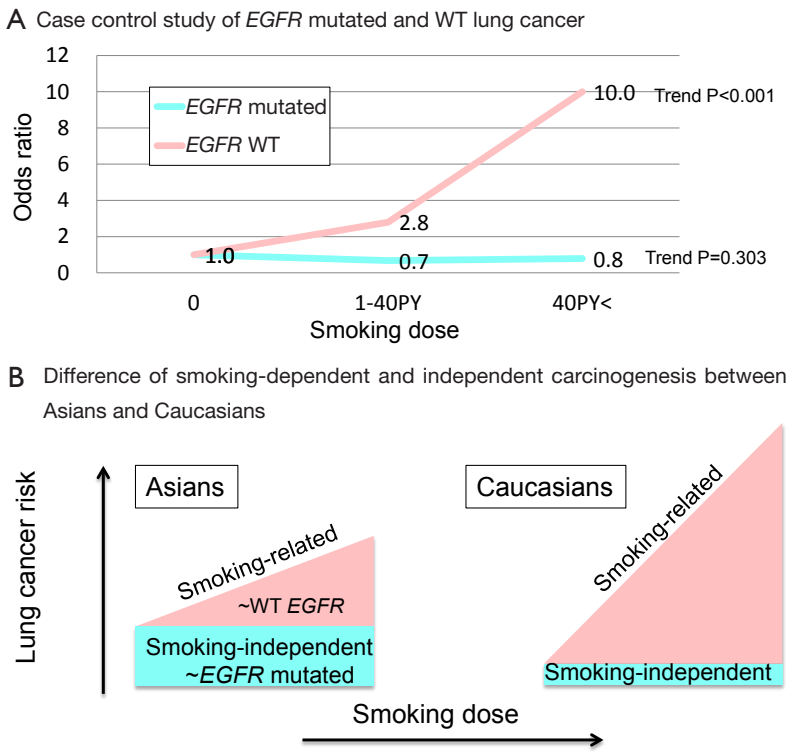

Figure 4 (A) Odds ratio of EGFR-mutated and WT lung cancer according to smoking dose; (B) proposed model to explain smoking-dependent and—independent lung carcinogenesis with reference to ethnic difference. EGFR, epidermal growth factor receptor; WT, wild-type.

KRAS and *TP53* mutations, in which heavier smokers have a higher incidence of gene mutation (5).

The question remains as to why *EGFR* mutations are more frequent among non-smokers. To address this issue, we performed a case-control study to assess the effect of smoking on *EGFR* mutations using 435 lung cancer patients and 2,175 matched controls. Among the 435 patients, 152 exhibited EGFR mutations and 283 contained wild-type (WT) *EGFR*. Comprehensive life-style data were obtained from the hospital-based epidemiologic research program at Aichi cancer center (HERPACC) (22). When the OR was calculated for *EGFR* mutated and wild-type lung cancers by tiers of smoking doses, there was a significant contribution of smoking dose to *EGFR* WT lung cancer with an OR for ever-smokers of 4.05 and 95% confidence interval (CI) of 2.79-5.88 (trend P<0.001) (22). However, there was no such relationship in *EGFR*-mutated lung cancer (OR, 0.73; CI, 0.46-1.14) (*Figure 4A*) (22). These results strongly indicate that *EGFR* mutation occurs independent of smoking. Because of increased risk of EGFR WT lung cancer by smoking dose, the ratio of *EGFR* mutated to *EGFR* WT lung cancer becomes lower in heavy smokers, resulting in low *EGFR* mutation rate among heavy smokers with lung cancer.

To explain the lower incidence of *EGFR* mutations and higher susceptibility of smoking-related lung cancer in Caucasian populations compared with East Asians, I propose the model depicted in *Figure 4B*. The risk of lung cancer that is independent of smoking as exemplified by those with *EGFR* mutation are not affected by smoking dose (light blue box), and East Asians have a higher risk for this type of lung cancer than Caucasians. In contrast, smoking-related lung cancer increases with smoking dose. Since Caucasians are more susceptible to this type of lung cancer by epidemiologic data as discussed previously, the slope of pink triangles is steeper in Caucasians.

Search for factors associated with susceptibility to *EGFR* mutation

If *EGFR* mutation is not affected by smoking, then what determines the occurrence of the *EGFR* mutation? In our case-control study described above, sex was the sole risk factor for *EGFR* mutated lung cancer (OR for women relative to men, 2.19; CI, 1.41-3.39) and total fertile

years showed a significant positive association with *EGFR* mutated NSCLC but not with *EGFR* WT lung cancer (22).

Other approaches to this issue include the exploration of the association between genetic polymorphisms and lung cancer with *EGFR* mutation. There are at least three polymorphisms within the *EGFR* gene. CA-SSR1 (CA simple sequence repeat 1) is located in intron 1 and the number of CA repeats ranges from 14 to 22 (23), with 16 being the most common number of repeats. East Asians tend to have longer repeats (23). Shorter repeats are associated with increased transcription and protein expression (24,25). Two single nucleotide polymorphisms (SNPs) in the promoter region with effects on increased transcription and expression of *EGFR* mRNA have also been identified. However, both of these SNPs (-216G/T, -191C/A) are less common in East Asians (26).

Nomura *et al.* examined the relationship between these three polymorphisms and *EGFR* mutation in samples including 556 resected non-small cell lung cancers, of which 336 are East Asians, and several normal tissue samples from different ethnicities (27). In line with the previous studies, the authors confirmed that these three polymorphisms associated with increased EGFR protein production were rare in East Asians (27). They also found that in East Asians, *EGFR* mutations occur on the shorter CA-SSR1 allele (suggestive of greater protein production) then allele specific amplification of the mutant allele follows. It is interesting to note three events leading to enhanced mutant protein production, such as increased transcription, mutation and amplification, target the same allele (27). However, it appears difficult to explain the higher frequency of *EGFR* mutations in East Asians by these findings.

Conclusions

In this review, I have discussed ethnic differences in lung cancer. Several factors appear to significantly modify the ethnic differences, i.e., sex, smoking habit, genetic polymorphism, driver gene mutations, and histologic subtypes. Our understanding of these relationships remains far from complete. In the future, GWAS may identify loci that are specifically related to EGFR-targeted carcinogenesis, leading to identification of carcinogens for *EGFR* mutations and effective prevention measures.

Acknowledgements

The author is grateful to Prof. Keitaro Matsuo of Kyushu University for helpful discussion and comments.

Disclosure: The author declares no conflict of interest.

References

1. Mitsudomi T, Yatabe Y. Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer. Cancer Sci 2007;98:1817-24.
2. Sasaki T, Rodig SJ, Chirieac LR, et al. The biology and treatment of EML4-ALK non-small cell lung cancer. Eur J Cancer 2010;46:1773-80.
3. OECD. Health at a Glance 2013: OECD Indicators. Paris: OECD Publishing, 2013.
4. OECD. Cancer Care: Assuring Quality to Improve Survival, Paris: OECD Publishing, 2013.
5. Kosaka T, Yatabe Y, Endoh H, et al. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res 2004;64:8919-23.
6. Toh CK, Gao F, Lim WT, et al. Never-smokers with lung cancer: epidemiologic evidence of a distinct disease entity. J Clin Oncol 2006;24:2245-51.
7. Siegfried JM. Women and lung cancer: does oestrogen play a role? Lancet Oncol 2001;2:506-13.
8. Kobrinsky NL, Klug MG, Hokanson PJ, et al. Impact of smoking on cancer stage at diagnosis. J Clin Oncol 2003;21:907-13.
9. Thun MJ, Hannan LM, Adams-Campbell LL, et al. Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. PLoS Med 2008;5:e185.
10. Risch HA, Howe GR, Jain M, et al. Are female smokers at higher risk for lung cancer than male smokers? A case-control analysis by histologic type. Am J Epidemiol 1993;138:281-93.
11. 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.
12. Osann KE, Anton-Culver H, Kurosaki T, et al. Sex differences in lung-cancer risk associated with cigarette smoking. Int J Cancer 1993;54:44-8.
13. Sobue T, Suzuki T, Fujimoto I, et al. Case-control study for lung cancer and cigarette smoking in Osaka, Japan: comparison with the results from Western Europe. Jpn J Cancer Res 1994;85:464-73.
14. Yang IA, Holloway JW, Fong KM. Genetic susceptibility to lung cancer and co-morbidities. J Thorac Dis

2013;5:S454-62.
15. Truong T, Hung RJ, Amos CI, et al. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. J Natl Cancer Inst 2010;102:959-71.
16. Shi Y, Au JS, Thongprasert S, et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). J Thorac Oncol 2014;9:154-62.
17. Chougule A, Prabhash K, Noronha V, et al. Frequency of EGFR mutations in 907 lung adenocarcioma patients of Indian ethnicity. PLoS One 2013;8:e76164.
18. Dogan S, Shen R, Ang DC, et al. Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related KRAS-mutant cancers. Clin Cancer Res 2012;18:6169-77.
19. Vallee A, Sagan C, Le Loupp AG, et al. Detection of EGFR gene mutations in non-small cell lung cancer: lessons from a single-institution routine analysis of 1,403 tumor samples. Int J Oncol 2013;43:1045-51.
20. Marchetti A, Martella C, Felicioni L, et al. EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol 2005;23:857-65.
21. D'Angelo SP1, Pietanza MC, Johnson ML, et al. Incidence of EGFR exon 19 deletions and L858R in tumor specimens from men and cigarette smokers with lung adenocarcinomas. J Clin Oncol 2011;29:2066-70.
22. Matsuo K, Ito H, Yatabe Y, 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.
23. Liu W, Innocenti F, Chen P, et al. Interethnic difference in the allelic distribution of human epidermal growth factor receptor intron 1 polymorphism. Clin Cancer Res 2003;9:1009-12.
24. Gebhardt F, Zänker KS, Brandt B. Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J Biol Chem 1999;274:13176-80.
25. Buerger H, Gebhardt F, Schmidt H, et al. Length and loss of heterozygosity of an intron 1 polymorphic sequence of egfr is related to cytogenetic alterations and epithelial growth factor receptor expression. Cancer Res 2000;60:854-7.
26. Liu W, Innocenti F, Wu MH, et al. A functional common polymorphism in a Sp1 recognition site of the epidermal growth factor receptor gene promoter. Cancer Res 2005;65:46-53.
27. Nomura M, Shigematsu H, Li L, et al. Polymorphisms, mutations, and amplification of the EGFR gene in non-small cell lung cancers. PLoS Med 2007;4:e125.

Cite this article as: Mitsudomi T. Molecular epidemiology of lung cancer and geographic variations with special reference to EGFR mutations. Transl Lung Cancer Res 2014;3(4):205-211. doi: 10.3978/j.issn.2218-6751.2014.08.04

