

Priority Report

Functional *EGFR* Germline Polymorphisms May Confer Risk for *EGFR* Somatic Mutations in Non–Small Cell Lung Cancer, with a Predominant Effect on Exon 19 Microdeletions

Wanqing Liu¹, Lijun He¹, Jacqueline Ramirez¹, Soundararajan Krishnaswamy¹, Rajani Kanteti¹, Yi-Ching Wang⁴, Ravi Salgia^{1,2}, and Mark J. Ratain^{1,2,3}

Abstract

Somatic mutations in the *EGFR* tyrosine kinase domain play a critical role in the development and treatment of non–small cell lung cancer (NSCLC). Strong genetic influence on susceptibility to these mutations has been suggested. To identify the genetic factors conferring risk for the *EGFR* tyrosine kinase mutations in NSCLC, a case–control study was conducted in 141 Taiwanese NSCLC patients by focusing on three functional polymorphisms in the *EGFR* gene [–216G/T, intron 1 (CA)_n, and R497K]. Allelic imbalance of the *EGFR* –216G/T polymorphism was also tested in the heterozygous patients and in the NCI-60 cancer cell lines to further verify its function. We found that the frequencies of the alleles –216T and CA-19 are significantly higher in the patients with any mutation ($P = 0.032$ and 0.01 , respectively), in particular in those with exon 19 microdeletions ($P = 0.006$ and 0.033 , respectively), but not in the patients with L858R mutation. The –216T allele is favored to be amplified in both tumor DNA of lung cancer patients and cancer cell lines. We conclude that the local haplotype structures across the *EGFR* gene may favor the development of cellular malignancies and thus significantly confer risk to the occurrence of *EGFR* mutations in NSCLC, particularly the exon 19 microdeletions. *Cancer Res*; 71(7); 2423–7. ©2011 AACR.

Introduction

The discovery of somatic mutations in the *EGFR* tyrosine kinase domain constitutes one of the most important findings in lung cancer in recent years. The missense point mutation L858R in exon 21 and microdeletions in exon 19 represent approximately 85% to 90% of all *EGFR* mutations and are significantly associated with clinical response to EGFR inhibitors (1–3). Why and how somatic *EGFR* mutations develop during carcinogenesis, however, remains largely unknown. The observation that the mutations are found more often in individuals who have never smoked excludes the involvement of carcinogens in tobacco smoke. The mutations are more prevalent in East Asian populations (30%–50%) but are relatively rare in individuals of European and African descent (<20%; refs. 1–3). Moreover, the prevalence of these mutations

in East Asians who have migrated to other countries remains high, suggesting that the origin of these mutations is related more to ethnicity than environment (3). These observations imply a germline susceptibility to this mutagenesis event. We therefore hypothesize that lung cancer carrying *EGFR* mutations may be attributable to susceptibility alleles with significant ethnic differences in their frequency. Moreover, these alleles may be distinct from those associated with lung cancer in general, and the molecular epidemiology of each somatic mutation may also be different.

We and others have previously shown that *EGFR* is highly polymorphic and its expression and activity are significantly regulated by polymorphisms (4–8). A germline (CA)_n polymorphism (rs45559542) in the intron 1 of *EGFR* has been associated with *EGFR* gene expression, with the shorter CA-repeats being associated with higher transcription level of *EGFR* (7). We also discovered a promoter single-nucleotide polymorphism (SNP; –216G>T/rs712829) that increases *EGFR* expression (4). A nonsynonymous common *EGFR* polymorphism R497K (G>A) (rs2227983) SNP was also identified, and the K (A) allele seems to decrease the activity of EGFR (8). We have shown that the allele frequency of these polymorphisms differs significantly among ethnic groups (4–6). Whether these polymorphisms confer risk for development of *EGFR* mutations, and the ethnic distribution of these polymorphisms further leads to a higher mutation rate in Asian patients have not yet been systematically evaluated. Here, we performed an association study between *EGFR* mutations and the 3 aforementioned germline polymorphisms.

Authors' Affiliations: ¹Department of Medicine; ²Comprehensive Cancer Center, and ³Committee on Clinical Pharmacology and Pharmacogenomics, The University of Chicago, Chicago, Illinois; ⁴Department of Pharmacology, National Cheng Kung University, Tainan, Taiwan, Republic of China

Note: Supplementary material for this article is available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corresponding Author: Mark J. Ratain, Section of Hematology/Oncology, Department of Medicine, 5841 S. Maryland Ave., MC 2115, The University of Chicago, Chicago, IL 60637. Phone: 773-702-4400; Fax: 773-702-3969. E-mail: mratain@medicine.bsd.uchicago.edu

doi: 10.1158/0008-5472.CAN-10-2689

©2011 American Association for Cancer Research.

Materials and Methods

EGFR mutation detection and genotyping in NSCLC patient samples

EGFR mutation data of exons 18 to 21 in a total of 141 NSCLC patients collected from Taiwan have been previously published (9). The 3 functional polymorphisms were genotyped in the germline DNA extracted from peripheral blood samples of these patients. Candidate polymorphisms were genotyped according to protocols published previously (4–6). Genotyping assays were repeated in 10% of patient samples (randomly chosen), and 100% concordance between the replicates and original data was observed. The –216G/T and intron 1 (CA)_n were also genotyped in the HapMap HCB (Han Chinese in Beijing, China; *n* = 45) and CEU (Utah residents with Northern and Western European Ancestry; *n* = 60) samples to compare their allele frequencies with our patient population (Supplementary Materials).

Analysis of allelic imbalance of the EGFR –216G/T polymorphism

The peak heights of the G and T alleles in NSCLC tumor DNAs of the heterozygous individuals were assessed by SNaPshot (Applied Biosystems). Allelic imbalance (AI) was defined by taking the ratio of the peak height of each allele and normalizing to that of the germline DNA. For NCI-60 cancer cell lines, we selected heterozygous samples based on our previous study (6) and used the mean value of the peak height ratio from 10 randomly selected normal germline DNA samples as controls for data normalization. A final ratio of less than 0.60 or more than 1.67 was set as the cutoff for AI.

Statistics

χ^2 Test or Fisher's exact test was used to test for Hardy-Weinberg equilibrium and the allelic association between polymorphisms and EGFR mutations. We regarded the patients bearing EGFR mutations as "cases" and those bearing EGFR wild type as "controls." Because of the biological difference between the EGFR exon 19 microdeletions and the exon 21 L858R point mutation (1), we also tested for association between each polymorphism and each of these 2 mutations. In addition, we conducted an exploratory analysis of the combined effect of the "risk" alleles on the presence of EGFR mutations. Odds ratio (OR), 95% confidence interval (CI), and *P* values (2-sided) were calculated for each allelic association. *P* = 0.05 was set as the cutoff for statistical significance, without adjustment for multiple testing.

Results

With the exception of the EGFR intron 1 (CA)_n polymorphism that seemed ambiguous in 3 samples, all polymorphisms were successfully genotyped in all samples. No significant deviation from Hardy-Weinberg equilibrium was observed (χ^2 test, *P* > 0.05 for all tests; data not shown).

Both the –216T and 19 CA-repeat (CA-19) alleles of the intron 1 (CA)_n polymorphism were significantly associated with the exon 19 mutations, but not with the L858R mutations. No statistically significant differences were found between the R497K polymorphism and EGFR mutations (Table 1).

The association between the number of "risk" alleles and EGFR mutations was further tested by combining –216T and intron 1 CA-19 alleles with and without the 497G allele, using the patients with no risk alleles as a reference. As a result, an

Table 1. Allelic association between EGFR polymorphisms and mutations

Polymorphism	Allele	All	All	OR	<i>P</i>	Ex19	Ex19	OR	<i>P</i>	858	858	OR	<i>P</i>
		MT	WT	(95% CI)		del+	del–	(95% CI)		L>R+	L>R–	(95% CI)	
–216G/T	T	8	6	3.13 (1.05–9.33)	0.03	6	8	4.28 (1.41–12.98)	0.006	2	12	1.07 (0.23–5.00)	1.00
	G	80	188			40	228			36	232		
R497K	G	42	89	1.08 (0.65–1.79)	0.77	26	105	1.62 (0.86–3.07)	0.14	14	117	0.63 (0.31–1.28)	0.20
	A	46	105			20	131			24	127		
Intron 1 (CA) _n	CA-19	12	9	3.43 (1.27–7.76)	0.01	7	14	2.77 (1.05–7.30)	0.03	4	17	1.53 (0.49–4.82)	0.51
	others	76	179			39	216			34	221		
	CA-15	3	8			2	9			1	10		
	CA-16	14	35			7	42			6	43		
	CA-17	3	5			2	6			0	8		
	CA-18	4	6	–	0.18	2	8	–	0.60	2	8	–	0.38
	CA-19	12	9			7	14			4	17		
	CA-20	44	109			23	130			20	133		
	CA-21	4	13			2	15			2	15		
CA-22	4	3			1	6			3	4			

Abbreviations: Ex19 del, exon 19 deletion; MT, mutant; WT, wild type.

Table 2. Association between the EGFR mutations and the number of risk alleles combining the genotypes

Combination of polymorphisms	Number of risk alleles	All WT MT	All OR (95% CI)	P	Ex19 del- del-	Ex19 del+ del+	OR (95% CI)	P	858 L>R- L>R+	OR (95% CI)	P	
-216G/T + intron 1 (CA)n	0	83	1.00 (referent)	-	98	14	1.00 (referent)	-	98	14	1.00 (referent)	-
	1	13	2.20 (0.87-5.56)	0.13	18	5	1.94 (0.62-6.07)	0.32	19	4	1.47 (0.44-4.97)	0.51
	2	1	14.30 (1.60-127.10)	0.008	2	4	14.00 (2.34-83.67)	0.005	5	1	1.40 (0.15-12.88)	0.57
-216G/T+Intron 1 (CA)n+R497K	0	25	1.00 (referent)	-	31	3	1.00 (referent)	-	29	5	1.00 (referent)	-
	1	46	0.97 (0.37-2.50)	1	54	8	1.53 (0.38-6.20)	0.74	54	8	0.86 (0.26-2.87)	1.00
	2	20	1.81 (0.64-5.08)	0.31	26	7	2.78 (0.65-11.86)	0.19	27	6	1.29 (0.35-4.72)	0.75
	3	6	1.85 (0.42-8.11)	0.45	7	3	4.43 (0.73-26.76)	0.12	10	0	2.26 (0.01-5.03)	0.57
	4	0	13.42 (0.59-306.10)	0.09	0	2	45.00 (1.78-1,140.00)	0.02	2	0	1.07 (0.05-25.56)	1.00

NOTE: The risk alleles refer to -216T, Intron 1 CA-19, and 497R (G).

enhanced association between the combined alleles and EGFR mutations, in particular the exon 19 microdeletions, was shown. Although 497G/A alone is not associated with EGFR mutations, combination of the 497G allele with -216T and CA-19 alleles showed a significant increase of the OR compared with the patients with no risk alleles (Table 2).

To further evaluate the function of the -216T allele, AI analysis of -216G/T alleles was successfully performed in 9 of 12 heterozygous NSCLC patients. Significant AI was observed in 4 tumors (44%) and showed a relative gain of the T allele or loss of the G allele. This preference for retention of the T allele was also observed in the NCI-60 cancer cell lines ($n = 58$) whereas AI was observed in 12 (55%) out of 22 heterozygous cell lines, 10 of which (83%) contained a relative gain or retention of the T allele (Fig. 1).

Discussion

The ethnic differences in the incidence of EGFR mutations in NSCLC remain incompletely understood. We chose a population of Taiwanese patients to perform a case-control-based association study aiming at elucidating the relationship between functional EGFR polymorphisms and somatic mutations. The Taiwanese population consisted of more than 98% Han Chinese (10), which is genetically close to other major ethnicities in East Asia (11). We confirmed this by showing that the allele frequencies of all 3 tested polymorphisms are similar to previous reports in Asian populations (Supplementary Table S1). We also found a high incidence rate of EGFR mutations (31.2%) in our NSCLC patients, consistent with data reported for other East Asian populations (2).

Our results suggest that local functional polymorphisms at the EGFR locus together play a major role in the development

of EGFR mutations. Given the significant difference in the allele frequencies of these polymorphisms between Asian and Caucasian populations (6, 7), our data may further explain the ethnic differences in the EGFR mutation rate. Previous studies showed that exon 19 microdeletions and the L858R point mutation have different biological properties, for example, a differential response to EGFR inhibitors (1). A recent study further suggested that EGFR amplification is specifically associated with exon 19 deletions (12). Our study, however, suggests that these two types of mutations may also have a different genetic basis. This is consistent with the previous observation of an association between exon 19 deletions and the shorter alleles of intron 1 (CA)_n polymorphism (13). These lines of evidence collectively show that EGFR exon 19 deletions have a pathogenic process distinct from other mutations. Our findings have significant clinical implications and may shed light on the pathogenesis of lung cancer as well.

Increasing evidence has strongly suggested that germline polymorphisms can confer susceptibility to the retention of somatic mutations during cancer development. Known examples include the R72P polymorphism in the TP53 gene (14); melanocortin-1 receptor gene polymorphisms and BRAF gene mutations in melanoma (15); JAK2 SNP (rs10974944) and JAK2V617F mutation in myeloproliferative neoplasms (16); and a 5'-distal SNP and FGFR3 mutations in urinary bladder cancer (17). Our findings consistently highlight the crucial role of the combined effects of functional germline polymorphisms in the development of EGFR mutations. Given the proto-oncogenic nature of EGFR, it is possible that certain haplotypes of the EGFR gene might have a selective advantage because of increased EGFR activity, and thus are more likely to contribute to neoplasia. Previous studies have shown that the chromosome/haplotype bearing the EGFR mutations, in particular the exon 19 mutations, tends to be selectively amplified in NSCLC tumors (12, 18), suggesting that once certain haplotypes are mutated, they may create an "EGFR-addicted" environment and hence tend to be positively selected in the tumorigenesis process. The polymorphisms tested in this study are all functional, and have been associated with higher EGFR activity (8) and increased EGFR expression (4, 6). We further observed a preferential gain of the T allele or loss of the G allele. Because amplification of the EGFR gene is commonly observed in multiple human cancers as we previously reported in NCI60 (6), we infer that our observed AI is most likely because of selective amplification of the -216T-containing allele. Because NCI60 consists of various cancer types, the AI of -216G/T in these cells may suggest that specific EGFR haplotypes (e.g., the ones carrying these functional alleles) may benefit the cell transformation in general, whereas specifically in lung cancer, this intrinsic capacity confers higher capability for retaining EGFR mutations (especially exon 19 deletions).

The EGFR intron 1 (CA)_n polymorphism has also been associated with EGFR gene expression (7). It was shown that selective amplification of the shorter alleles occurred frequently in tumors harboring EGFR mutations, especially in patients of East Asian ethnicity (19). Another study also found a genetic association between exon 19 mutations and the shorter alleles (<17 repeats in the shorter allele; ref. 13).

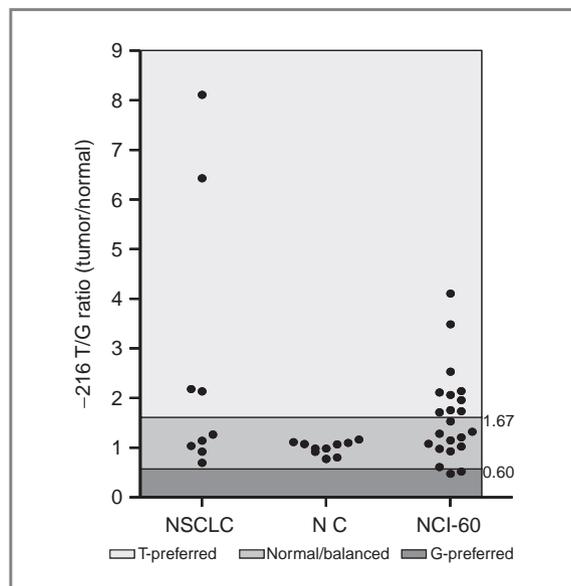


Figure 1. AI of the EGFR -216G/T polymorphism in NSCLC patients and the NCI-60 cancer cell lines. The relative T/G ratio in each sample and the threshold of AI are shown. The imbalance of each allele is shown by gradations of shading. NC, normal control.

However, this was not confirmed in our study (data not shown). The intron 1 (CA)_n polymorphism has more than 10 alleles (5), and it is challenging to set an appropriate cutoff for shorter and longer alleles due to the lack of clear biological rationale, or even evidence of a monotonic relationship with any phenotype. For instance, the study mentioned earlier used the median repeat length (17 repeats) in shorter alleles as a cutoff (13). In our population, however, the median is 19 repeats. Moreover, our previous studies suggested that -216G/T and the (CA)_n polymorphism are in linkage disequilibrium, and the shorter CA alleles tend to cosegregate with the -216T allele (4, 6). This may confound the association between the (CA)_n polymorphism and *EGFR* mutations, and the previous observation (13) between "shorter" CA alleles and *EGFR* mutations may be because of their linkage with the -216T allele. Following this reasoning, instead of testing for an association between "shorter" alleles and mutations, we tested for an association between each of the CA-repeat alleles (versus all other alleles) and *EGFR* mutations. As a result, we observed significant association between the allele CA-19 and *EGFR* mutations. However, no clear biological function has been specifically associated with CA-19 allele thus far. There is also no linkage disequilibrium between the -216T and CA-19 alleles (data not shown). Thus, it seems that CA-19 (or an unknown functional polymorphism in linkage disequilibrium with CA-19), alone or in combination with the -216T allele, may predispose to the *EGFR* mutagenesis.

We recognize certain limitations of our study. Because the *EGFR* gene is large (>188 kb) and a large number of SNPs

(>600) have been identified in the region, a comprehensive study in a much larger sample set will be necessary to completely elucidate the somatic and germline genetics at the *EGFR* locus. In addition, the absence of a replication dataset does not exclude the possibility that the associations we report are false discovery, or are limited to a subpopulation due to the potential heterogeneity among East Asian populations. Independent validation is warranted, ideally in the context of a prospective study, because germline DNA has been infrequently collected in conjunction with tumor DNA samples, and thus the potential value of large-scale sample collection of matching blood and tumor samples is emphasized. Nevertheless, the data collected in this study consistently support the hypothesis that certain haplotypes consisting of *cis*-acting functional polymorphisms may play a critical role in the accumulation of the *EGFR* exon 19 deletions during lung cancer development, which provides a strong rationale for further investigation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This work was supported by the NIH/NIGMS Grant U01GM61393 (M.J. Ratain) and Grant NIH R01CA125541-03 (R. Salgia).

Received July 22, 2010; revised January 13, 2011; accepted January 17, 2011; published OnlineFirst February 3, 2011.

References

- 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.
- 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.
- Tsao AS, Tang XM, Sabloff B, Xiao L, Shigematsu H, Roth J, et al. Clinicopathologic characteristics of the EGFR gene mutation in non-small cell lung cancer. *J Thorac Oncol* 2006;1:231-9.
- Liu W, Innocenti F, Wu MH, Desai AA, Dolan ME, Cook EH Jr, 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.
- Liu W, Innocenti F, Chen P, Das S, Cook EH Jr, Ratain MJ. Interethnic difference in the allelic distribution of human epidermal growth factor receptor intron 1 polymorphism. *Clin Cancer Res* 2003;9:1009-12.
- Liu W, Wu X, Zhang W, Montenegro RC, Fackenthal DL, Spitz JA, et al. EGFR mutations, expression, amplification, polymorphisms and their interrelationship with sensitivity/resistance to EGFR inhibitors in the NCI60 cell lines. *Clin Cancer Res* 2007;13:6788-95.
- Brandt B, Meyer-Staeckling S, Schmidt H, Agelopoulos K, Buerger H. Mechanisms of egfr gene transcription modulation: relationship to cancer risk and therapy response. *Clin Cancer Res* 2006;12:7252-60.
- Moriai T, Kobrin MS, Hope C, Speck L, Korc M. A variant epidermal growth factor receptor exhibits altered type α transforming growth factor binding and transmembrane signaling. *Proc Natl Acad Sci U S A* 1994; 91:10217-21.
- Krishnaswamy S, Kanteti R, Duke-Cohan JS, Loganathan S, Liu W, Ma PC, et al. Ethnic differences and functional analysis of MET mutations in lung cancer. *Clin Cancer Res* 2009;15:5714-23.
- Spain D. Taiwan: country profile. *Int Demogr* 1984;3:4-8.
- Tian C, Kosoy R, Lee A, Ransom M, Belmont JW, Gregersen PK, et al. Analysis of East Asia genetic substructure using genome-wide SNP arrays. *PLoS One* 2008;3:e3862.
- Sholl LM, Yeap BY, Iafrate AJ, Holmes-Tisch AJ, Chou YP, Wu MT, et al. Lung adenocarcinoma with EGFR amplification has distinct clinicopathologic and molecular features in never-smokers. *Cancer Res* 2009;69:8341-8.
- Sueoka-Aragane N, Imai K, Komiya K, Sato A, Tomimasu R, Hisatomi T, et al. Exon 19 of EGFR mutation in relation to the CA-repeat polymorphism in intron 1. *Cancer Sci* 2008;99:1180-7.
- Marin MC, Jost CA, Brooks LA, Irwin MS, O'Nions J, Tidy JA, et al. A common polymorphism acts as an intragenic modifier of mutant p53 behaviour. *Nat Genet* 2000;25:47-54.
- Landi MT, Bauer J, Pfeiffer RM, Elder DE, Hulley B, Minghetti P, et al. MC1R germline variants confer risk for BRAF-mutant melanoma. *Science* 2006;313:521-2.
- Campbell PJ. Somatic and germline genetics at the JAK2 locus. *Nat Genet* 2009;41:385-6.
- Kiemeny LA, Sulem P, Besenbacher S, Vermeulen SH, Sigurdsson A, Thorleifsson G, et al. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nat Genet* 2010;42: 415-9.
- Takano T, Ohe Y, Sakamoto H, Tsuta K, Matsuno Y, Tateishi U, et al. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. *J Clin Oncol* 2005;23:6829-37.
- Nomura M, Shigematsu H, Li L, Suzuki M, Takahashi T, Estess P, et al. Polymorphisms, mutations, and amplification of the EGFR gene in non-small cell lung cancers. *PLoS Med* 2007;4: e125.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Functional *EGFR* Germline Polymorphisms May Confer Risk for *EGFR* Somatic Mutations in Non–Small Cell Lung Cancer, with a Predominant Effect on Exon 19 Microdeletions

Wanqing Liu, Lijun He, Jacqueline Ramírez, et al.

Cancer Res 2011;71:2423-2427. Published OnlineFirst February 3, 2011.

Updated version Access the most recent version of this article at:
doi:[10.1158/0008-5472.CAN-10-2689](https://doi.org/10.1158/0008-5472.CAN-10-2689)

Supplementary Material Access the most recent supplemental material at:
<http://cancerres.aacrjournals.org/content/suppl/2011/02/03/0008-5472.CAN-10-2689.DC1>

Cited articles This article cites 19 articles, 9 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/71/7/2423.full#ref-list-1>

Citing articles This article has been cited by 4 HighWire-hosted articles. Access the articles at:
<http://cancerres.aacrjournals.org/content/71/7/2423.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cancerres.aacrjournals.org/content/71/7/2423>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.