classifiers have been validated with a sufficient number of independent cases.[12]

In this study, we report successful identification of a relapse-related molecular signature in adenocarcinomas through analysis of genome-wide expression profiles using a training set of 60 patients with lung adenocarcinomas. General applicability of the resultant classifier was successfully validated in a blind test set of 27 cases with stage I to III disease as well as with another independent cohort of 30 stage I patients. Moreover, additional validation using two data sets on a different platform further confirmed the predictive power of the genes comprising the relapse-related molecular signature.

## PATIENTS AND METHODS

### Patient Samples

Eighty seven lung adenocarcinoma samples from patients who underwent potential curative resection between December 1995 and August 1999 were collected at Aichi Cancer Center, Nagoya, Japan (herein referred to as data set I; online-only Appendix Table A1). An additional independent cohort of 30 adenocarcinoma samples from patients with pathologic stage (pStage) I disease were also collected at Aichi Cancer Center between February 2002 and December 2004 (herein referred to as data set II; Appendix Table A1). None of the 117 patients received adjuvant chemotherapy. General schedule of follow-up examinations was chest x-ray (every month for the first 3 months, and 3 months interval thereafter) and chest and abdominal computed tomography (CT; every year) until 5 years after surgery. Additional examinations, such as CT, bone scan, and brain magnetic resonance imaging, were also considered, if any signs of possible relapse were suspected. The median follow-up periods for patients alive at the last follow-up examination in data set I and data set II were 90 months (range, 64 to 108 months) and 64 months (range, 55 to 75 months), respectively. All tumor specimens were collected under approval from the institutional review boards of Aichi Cancer Center and Nagoya University with written informed consent from each patient.

### Acquisition of Expression Profiles and Analysis of EGFR, p53, and K-ras Mutations

Double-stranded cDNA was synthesized from 500 ng of total RNA using Moloney murine leukemia virus-reverse transcriptase (Agilent Technologies, Palo Alto, CA) and poly dT primer incorporating the T7 promoter. Cy5-sample cRNA and Cy3-common reference cRNA were generated and hybridized to a Whole Human Genome oligo DNA microarray kit (G4112F, Agilent Technologies) with 41,000 distinct probes, which was scanned using an Agilent DNA microarray scanner (G2505B, Agilent Technologies), basically as described previously.[13] The mutation status of *EGFR*, *p53*, and *K-ras* was previously reported in the same set of patients.[13] All the microarray data and the pathologic and clinical data used for this study are available at Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/; accession number GSE13213). Cross-platform validation was carried out using the Duke[11] and Director's Challenge Consortium[12] data sets as detailed in the online-onlyAppendix.

### Biostatistical and Bioinformatic Analyses

To identify a relapse-related signature using signals that were expressed above the background in at least 90% of samples, we used a weighted voting algorithm, in which each weight value was calculated as the signal-to-noise ratio, basically according to the detailed method that we described previously.[14] Kaplan-Meier survival curves and Cox proportional hazards model analyses (Stata, version 7.0; Stata Corp, College Station, TX) were used to analyze the relationships of the resultant relapse-related signature with overall and relapse-free survival. All statistical tests were two sided. The CLUSTER[15] program was used for average linkage hierarchical clustering of both genes and cases, and the TREEVIEW[15] program was used for display (http://rana.lbl.gov/EisenSoftware.htm).

## RESULTS

### Identification of Relapse-Related Signature

A schematic diagram of our strategy for constructing and validating a relapse-related signature in surgically treated lung adenocarcinoma patients is shown in Figure 1, which was formed with the intention of blocking any information leakage between the training and validation data sets. First, we divided expression profile data obtained from 87 patients into 60 training and 27 validation data sets, the latter of which was completely set aside during training. In order to identify a generic signature with clear associations with relapse in the training set of patients with lung adenocarcinomas, we selected 28 favorable samples (alive > 5 years after surgery without any evidence of relapse) and 21 fatal samples (dead in 5 years after initial surgery with evidence of relapse). The remaining 11 patients in the training set were excluded from analysis of a possible relapse-related signature, because of ambiguity related to the aggressiveness of their tumors,

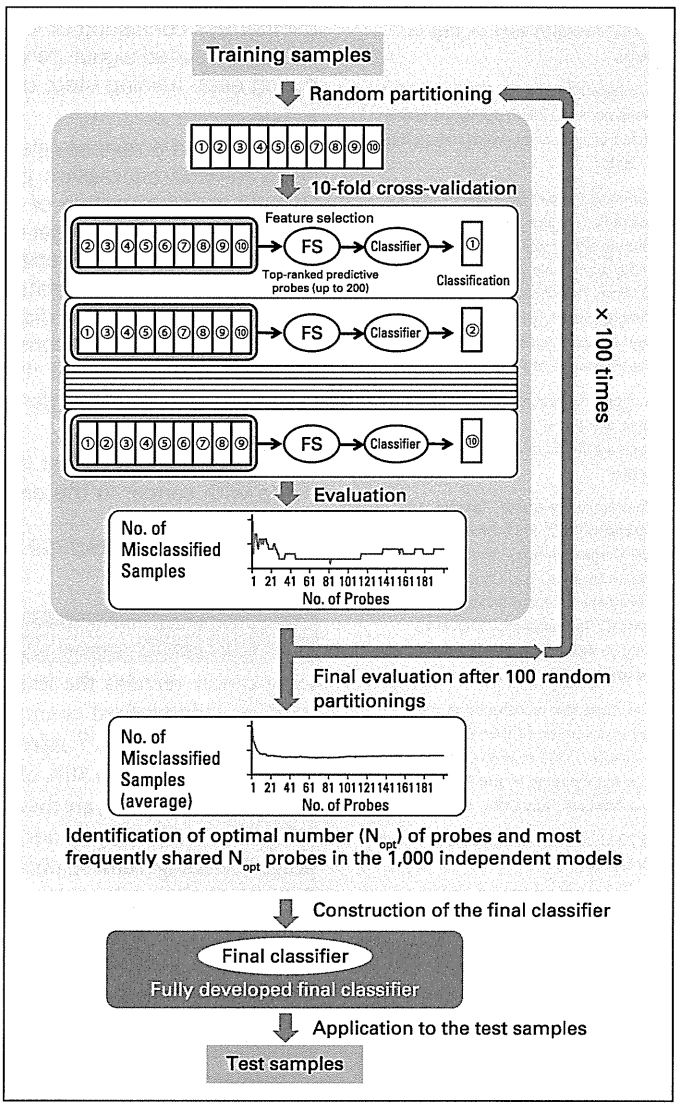


**Fig 1.** Schematic diagram of our training-validation stratergy for identifying replapse-related signature using 10-fold cross-validation procedures with 100 random partitions of the training data set.

  

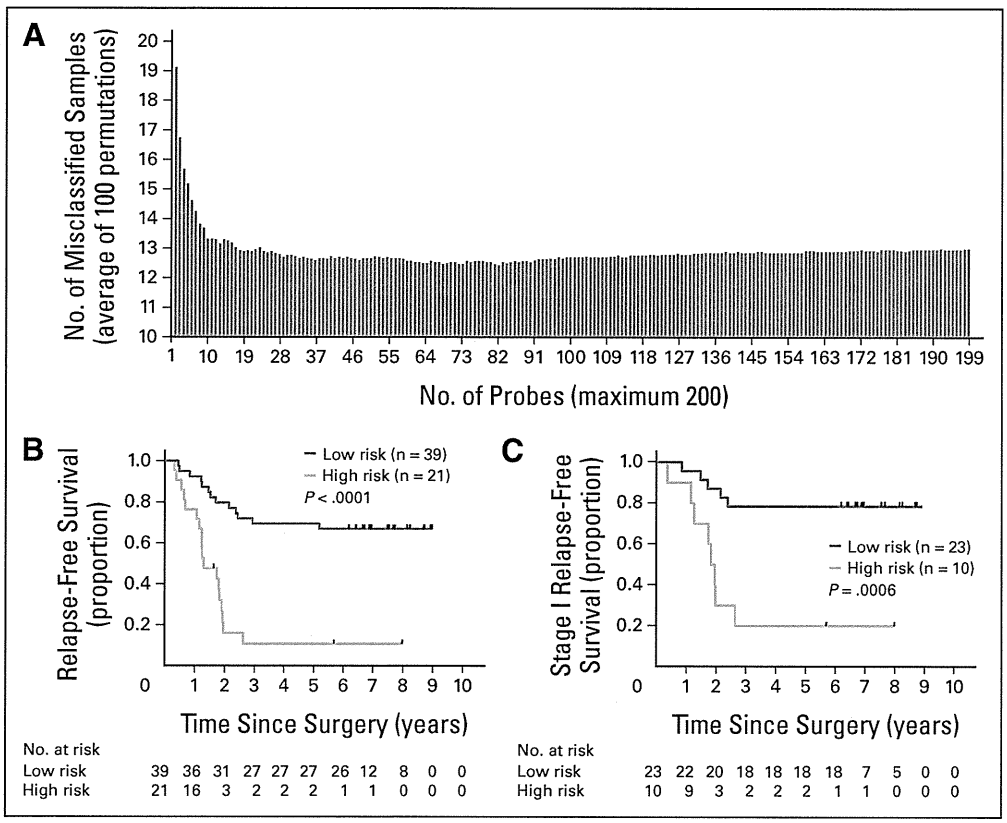


**Fig 2.** Results of the training procedure for identifying a relapse-related signature. (A) Results of our search for the optimum number of probes for defining a relapse-related signature. Kaplan-Meier survival curves were used to estimate survival in the training cohort. Relapse-free survival curves for patients in (B) all stages and (C) stage I.

which were five who survived for more than 5 years with some signs of relapse during follow-up, five who died of cancer after surviving for more than 5 years, and one who died within 5 years without evidence of relapse.

Of the 41,000 probes in the entire genome microarray, 23,828 passed the initial filtering criteria for selecting informative probes, and were then ranked according to a signal-to-noise metric and used to identify a relapse-related signature that could best distinguish patients who died with relapse from those cured by surgery. The learning errors for each model, to which increasing numbers of the predictive probes were applied, were calculated using 10-fold cross-validation and repeated with new randomly partitioned data sets 100 times. Thus, 1,000 independent sets consisting of up to 200 predictive probes each were selected for constructing a relapse-related signature-based classifier. As a result, 82 predictive probes were found to yield the fewest numbers of learning errors (Fig 2A), and the group of 82 probes most frequently shared among each of the 1,000 independent sets of 82 predictive probes was identified as a relapse-related signature (hereafter referred to as RRS-82; online-only Appendix Table A2). RRS-82 was able to distinguish patients with a very poor prognosis when all stages or only stage I were considered (Figs 2B and 2C for relapse-free survival and online-only Appendix Fig. A1 for overall survival). There were no associations of RRS-82 with the presence of *EGFR*, *K-ras*, or *p53* gene mutations, none of which showed any prognostic significance (Appendix Fig. A2).

### *Validation of RRS-82 in the Test Cohort of Data Set I*

To evaluate the robustness of RRS-82, we analyzed its discriminatory power using a completely blinded data set of 27 adenocarcinomas. Results with the validation data set indicated that RRS-82 could distinguish between patients with high and low risks of recurrence and death. Relapse-free survival was significantly different between the two groups ($P = .0003$; Fig 3A), and the proportions of relapse-free patients in the high- and low-risk groups were 38% and 78%, respectively, after 2 years. In the high-risk group, the overall survival rate after surgical resection was also significantly lower than that in the low-risk group ($P = .026$; Fig 3B). It was of note that all stage I patients, who were predicted as high-risk based on RRS-82, experienced relapse within 5 years, and died during the follow-up period (Figure 3C for relapse-free survival; $P = .0008$; Fig 3D for overall survival; $P = .043$; both by log-rank test). Interestingly, Kaplan-Meier curves for both relapse-free and overall survival showed tendencies to have modest associations with pathologic disease stage ($P = .15$ for relapse-free survival and $P = .18$ for overall survival) among patients in the low-risk group but not in patients with high-risk RRS-82 (online-only Appendix Fig. A3). The presence of a high risk signature of RRS-82 was not associated with site of relapse (online-only Appendix Table A3).

### *Further Validation of RRS-82 With an Additional Independent Cohort of pStage I Patients*

Further validation of the predictive power of RRS-82 in early-stage patients was conducted using another completely independent cohort of 30 stage I adenocarcinomas in patients who underwent surgery during a different period of time (data set II). RRS-82 was again shown capable of predicting which stage I patients were at extreme high risk (Figs 4A and 4B). In the combined validation cohort

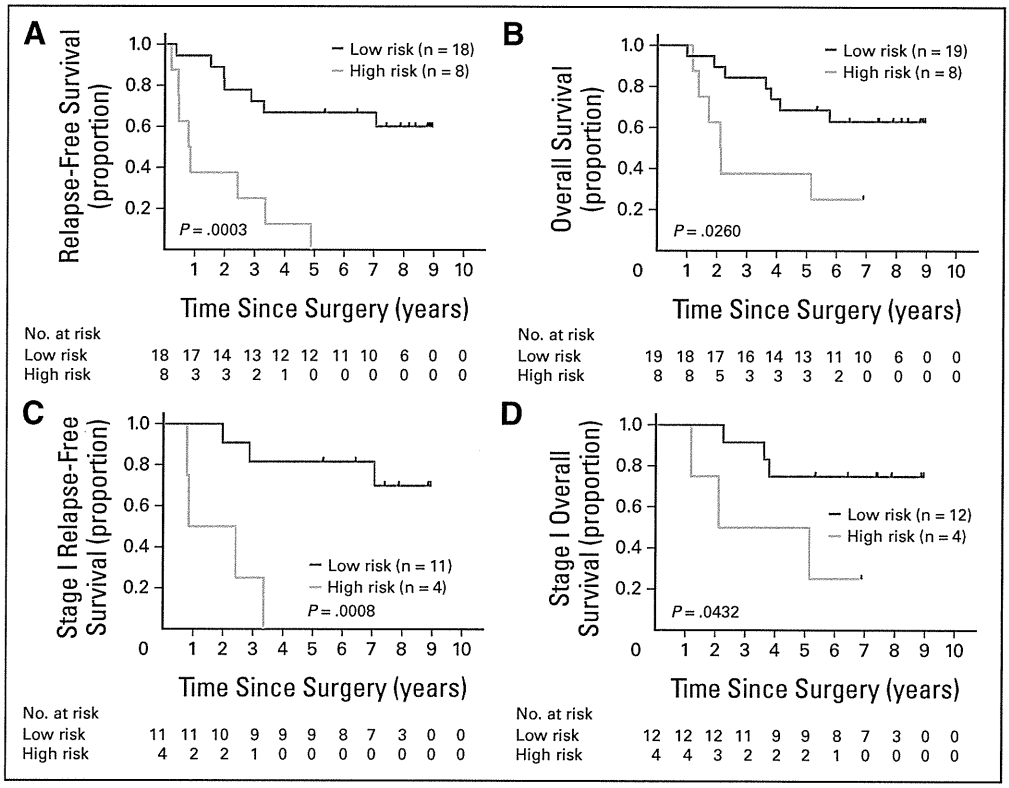


**Fig 3.** Validation of the RRS-82 signature with the use of completely blinded data set of 27 patients. Relapse-free survival curves for (A) all stages, (C) stage I. Overall survival curves for (B) all stages and (D) stage I.

consisting of 46 stage I cases (16 and 30 from datasets I and II, respectively), Kaplan-Meier survival curves based on RRS-82–based predictions were markedly different, showing relapse-free survival in 74% and 10% of patients with low- and high-risk signatures, respectively ($P < .0001$; Fig 4C). Overall survival was also significantly worse in the high-risk group as compared with the low-risk group ($P = .002$; Fig 4D). Data for patients in all stages are shown in online-only Appendix Figure A4. Multivariate Cox regression analysis of the combined validation data sets, in which the results of RRS-82–based predictions were considered as one of the variables, revealed that RRS-82 was highly predictive and independent of disease stage for both relapse-free survival ($P < .001$) and overall survival ($P = .005$; Table 1).

### *Confirmation of Predictive Capability of RRS-82 Using Two Additional Data Sets With a Different Platform*

The robustness of RRS-82 for predicting survival of patients with lung adenocarcinomas was further validated using a completely independent Duke University data set of 39 lung adenocarcinomas. We conducted an unsupervised hierarchical clustering based on the expression profiles of the 46 genes, which corresponded to those constituting RRS-82 (Appendix Table A4). Thirty-nine adenocarcinomas were clearly clustered into two distinct subsets (Fig 5A), with significantly different postoperative survival results shown ($P = .028$; Fig 5B). The vast majority of genes corresponding to those related to relapse in RRS-82 showed a higher expression in patients in cluster 2, who had a poor prognosis, supporting the general applicability of RRS-82 for lung adenocarcinomas.

We further confirmed the predictive capability of the gene set constituting RRS-82 with a different approach by utilizing recently reported large training-testing, multisite data sets (Fig 5C). Using the University of Michigan data set consisting of 75 alive and 102 dead patients, we calculated each weighted value for 31 genes, which corresponded to the gene set constituting RRS-82, as the signal-to-noise ratio and then applied it to the 104 Memorial Sloan-Kettering samples, all of which had valuable information regarding relapse. The resultant RRS-82-based classifier built on the University of Michigan data set was able to predict patients at high risk in the Memorial Sloan-Kettering validation data set (Fig 5D). Taken together, these results demonstrated the predictive power of the gene set constituting RRS-82 for identifying patients at high risk for disease recurrence. Since the 31 genes in the set were selected based only on the presence of corresponding genes between the two distinct platforms, our findings suggest that potential future development of an optimally downsized classifier with sufficient predictive power based on RRS-82 is possible.

## DISCUSSION

In this study, we identified a molecular signature, termed RRS-82, which was significantly associated with relapse and death in patients with adenocarcinomas of the lung. Based on the RRS-82 signature, we were able to construct a prognosis prediction classifier, which may ultimately aid in patient-tailored selection for therapeutic strategies. The robustness of the RRS-82 signature was successfully validated through application in four attempts with two independent Nagoya data sets as well as with the Duke and Director's Challenge Consortium data sets. Notably, the RRS-82–based classifier clearly distinguished patients with very poor prognosis from those with favorable outcome, including the duration of relapse-free survival, even in stage

  

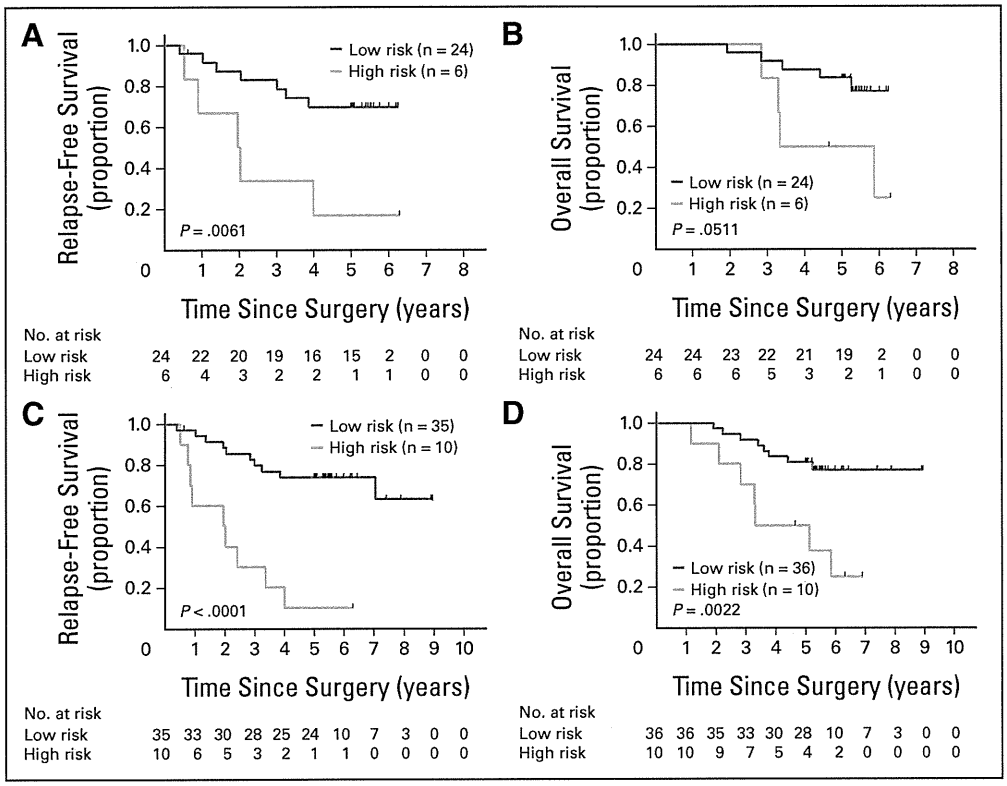


**Fig 4.** Independent validation of the RRS-82 signature using an additional independent cohort of 30 patients with stage I disease. (A) Relapse-free survival curves and (B) overall survival curves. Kaplan-Meier survival curves were used to estimate (C) relapse-free survival and (D) overall survival in the 46 stage I patients from data sets I and II.

I cases. These findings suggest that patients with the high-risk RRS-82 signature, who are overlooked using current diagnostic procedures for staging because of the inability of detection, are likely to have minimal residual disease. We previously reported that a 25-peak proteomic signature could also identify patients with very unfavorable outcome after surgery with curative intent at the protein level,[14] similarly to the present RRS-82 signature. Taken together, these findings support the notion that patients with very poor prognosis are certainly predictable even in stage I cases and that inclusion of molecular signature-based prognosis predictions, which take molecular and biologic characteristics manifested as signatures into consideration, may improve our capabilities for evaluating each patient with the ultimate aim of better therapeutic options.

Several studies have presented evidence supporting a model in which the propensity to metastasize reflects the predominant genetic/epigenetic state of a primary tumor, rather than the emergence of rare cells with a metastatic phenotype.[16-18] In this regard, it is interesting that disease stage at surgery appeared to have a modest tendency to affect patient outcome only in patients with a low-risk RRS-82 signature and not in those with a high-risk signature. A similar tendency was consistently observed in our previous proteomic analysis using matrix-assisted laser desorption/ionisation time of flight mass spectrometry, in which a 25-peak-based prediction model was constructed.[14] These findings therefore suggest a potential difference in biologic aggressiveness between the groups with high- and low-risk RRS-82 signatures.

**Table 1.** Univariate and Multivariate Cox Regression Analysis for the Combined Test Cohort (n = 57)

| | | Univariate | | | Multivariate | | |
|---|---|---|---|---|---|---|---|
| Variable | Unfavorable/Favorable | Hazard Ratio | 95% CI | *P* | Hazard Ratio | 95% CI | *P* |
| Relapse-free survival (n = 56)* | | | | | | | |
| Age | > 61/≤ 61 | 0.68 | 0.32 to 1.47 | .331 | 0.91 | 0.41 to 2.02 | .817 |
| Sex | Male/female | 1.46 | 0.68 to 3.10 | .329 | 1.19 | 0.54 to 2.60 | .668 |
| Stage | II-III/I | 2.41 | 1.05 to 5.54 | .038 | 2.00 | 0.84 to 4.72 | .115 |
| RRS-82 | High risk/low risk | 5.48 | 2.50 to 12.0 | < .001 | 4.92 | 2.17 to 11.2 | < .001 |
| Overall survival (n = 57) | | | | | | | |
| Age | > 61/≤ 61 | 1.00 | 0.44 to 2.32 | .991 | 1.21 | 0.50 to 2.91 | .668 |
| Sex | Male/female | 1.61 | 0.70 to 3.74 | .265 | 1.33 | 0.55 to 3.19 | .526 |
| Stage | II-III/I | 2.56 | 1.04 to 6.32 | .041 | 2.15 | 0.84 to 5.47 | .106 |
| RRS-82 | High risk/low risk | 3.68 | 1.58 to 8.56 | .003 | 3.60 | 1.48 to 8.77 | .005 |

*Information of relapse was not available in a single case.

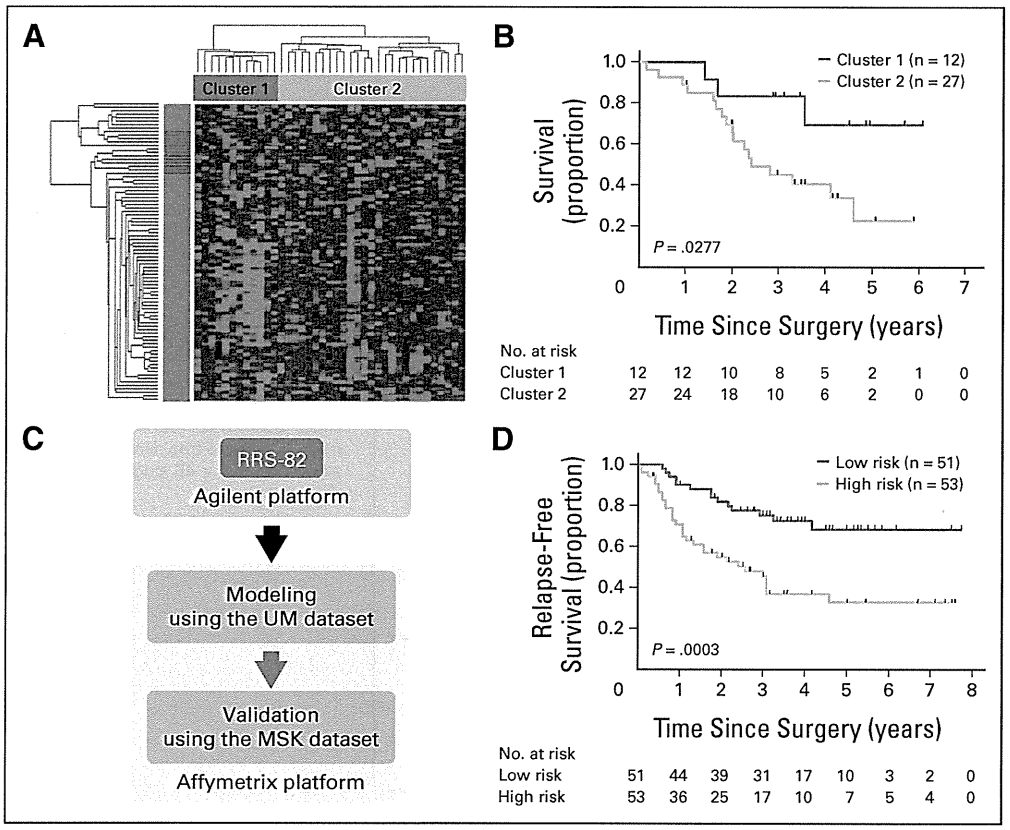


**Fig 5.** Results of (A) unsupervised hierarchical clustering analysis and (B) Kaplan-Meier survival curves for clusters I and II of the Duke data set. Schematic diagram showing further verification of RRS-82 constituents using another completely (C) independent training-validation data sets of 177 University of Michigan (UM) patients and 104 Memorial Sloan Kettering (MSK) patients, and (D) relapse-free survival curves for MSK patients.

The highly predictive nature of our RRS-82 signature, especially in terms of risk of relapse, may have been accomplished by our strategy used in the identification process, which paid special attention to relapse-free duration in a training cohort with high quality follow-up data. In fact, relapse within 5 years after surgery was observed in 80% and 90% of the patients with a high-risk RRS-82 signature in the training and combined validation cohorts with stage I disease, respectively. Although relapse-free survival data were not available for the 50 gene–based prediction classifier presented by the Michigan group[8] or the "A method" model by the Director's Challenge Consortium,[12] fatal outcome within 5 years after surgery was observed in approximately 55% and 60%, respectively, of those patients. In addition, the high-risk Duke metagene signature composed of nine metagene groups corresponding to 133 probes[11] was reported to correctly predicted relapse in 69% and 79% in their American College of Surgeons Oncology Group (stages I and II) and Cancer and Leukemia Group B (stages I to III) validation cohorts, respectively. Interestingly, the constituents of the RRS-82 signature do not have a significant overlap with other predictive signatures thus far reported by us and others.[8,9,11,12,19-23] Such variability among studies is commonly observed in molecular signatures for class prediction, and we suspect that it may reflect the use of different platforms for expression profiling and/or existence of distinct genes with similar predictive information, because of the presence of similarly coregulated genes that do not necessarily have similar biologic and/or biochemical properties.[24] For example, *PSMD12*, *FIP1L1*, and *UBE2V2*, included in RRS-82, are a part of the cluster six-gene set reported by the Director's Challenge Consortium, while *SMARCE1* in RRS-82 is included in the cluster 10-gene set. Additional analyses using the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) and Gene Ontology (http://www.geneontology.org/) databases identified only a few common pathways and networks containing predictive gene sets in such studies (examples shown in the Data Supplement). However, those results may not be surprising, since all of these studies including our own were not aimed at identifying functionally relevant gene sets or pathways associated with differences in clinical behavior such as relapse after surgery.

A number of negative results have been reported in regard to the benefits of adjuvant chemotherapy in patients with early-stage lung cancers,[25-28] although we believe that those do not preclude the potential clinical importance of molecular signature–based classification. Instead, such classification will likely add additional important information for patient-tailored evaluation of the nature of those diseases, considering that the current staging procedures, which rely on the measurement of disease spread by imaging techniques with insufficient power for detecting minute residual tumors, may be causing stage-migration of actual advanced cases into false early stages.

In conclusion, we succeeded in identifying a relapse-related molecular signature for use with patients diagnosed with adenocarcinomas, which was able to select those at extremely high risk for relapse, even in early-stage patients. In the field of breast cancer, a molecular signature–based prediction of surgically treated patients has been approved by the US Food and Drug Administration, and development of a similar useful means is urgent for lung cancer, which claims the highest number of lives each year. A future confirmatory study and clinical trial for patient-tailored adjuvant therapy with stratification according to the RRS-82 molecular signature are therefore warranted

  

to evaluate whether such selection may ultimately improve patient prognosis after surgery for this deadly cancer.

## AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

*Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.*

**Employment or Leadership Position:** Toshiyuki Takeuchi, Oncomics Co, Ltd (C) **Consultant or Advisory Role:** None **Stock Ownership:** Shuta Tomida, Oncomics Co, Ltd; Takashi Takahashi, Oncomics Co, Ltd, **Honoraria:** Tetsuya Mitsudomi, AstraZeneca Japan, Chugai Pharmaceutical, Astellas, Daiichi-Sanyo, Taiho, Eli-Lilly Japan, Kyowa hakko **Research Funding:** None **Expert Testimony:** Tetsuya Mitsudomi, AstraZeneca (U) **Other Remuneration:** None

## AUTHOR CONTRIBUTIONS

**Conception and design:** Shuta Tomida, Takashi Takahashi
**Financial support:** Shuta Tomida, Takashi Takahashi
**Administrative support:** Takashi Takahashi
**Provision of study materials or patients:** Tetsuya Mitsudomi, Yasushi Yatabe
**Collection and assembly of data:** Shuta Tomida, Toshiyuki Takeuchi, Yukako Shimada
**Data analysis and interpretation:** Shuta Tomida, Chinatsu Arima, Keitaro Matsuo, Takashi Takahashi
**Manuscript writing:** Shuta Tomida, Takashi Takahashi
**Final approval of manuscript:** Shuta Tomida, Takashi Takahashi

## REFERENCES


1. Jemal A, Siegel R, Ward E, et al: Cancer statistics, 2007. CA Cancer J Clin 57:43-66, 2007
2. McCracken M, Olsen M, Chen MS Jr, et al: Cancer incidence, mortality, and associated risk factors among Asian Americans of Chinese, Filipino, Vietnamese, Korean, and Japanese ethnicities. CA Cancer J Clin 57:190-205, 2007
3. Sun S, Schiller JH, Gazdar AF: Lung cancer in never smokers–a different disease. Nat Rev Cancer 7:778-790, 2007
4. Hoffman PC, Mauer AM, Vokes EE: Lung cancer. Lancet 355:479-485, 2000
5. Sotiriou C, Piccart MJ: Taking gene-expression profiling to the clinic: When will molecular signatures become relevant to patient care? Nat Rev Cancer 7:545-553, 2007
6. Garber ME, Troyanskaya OG, Schluens K, et al: Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A 98:13784-13789, 2001
7. Bhattacharjee A, Richards WG, Staunton J, et al: Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 98:13790-13795, 2001
8. Beer DG, Kardia SL, Huang CC, et al: Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 8:816-824, 2002
9. Tomida S, Koshikawa K, Yatabe Y, et al: Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. Oncogene 23:5360-5370, 2004
10. Jones MH, Virtanen C, Honjoh D, et al: Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. Lancet 363:775-781, 2004
11. Potti A, Mukherjee S, Petersen R, et al: A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 355:570-580, 2006
12. Shedden K, Taylor JM, Enkemann SA, et al: Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. Nat Med 14:822-827, 2008
13. Takeuchi T, Tomida S, Yatabe Y, et al: Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 24:1679-1688, 2006
14. Yanagisawa K, Tomida S, Shimada Y, et al: A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer. J Natl Cancer Inst 99:858-867, 2007
15. Eisen MB, Spellman PT, Brown PO, et al: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95:14863-14868, 1998
16. Ramaswamy S, Ross KN, Lander ES, et al: A molecular signature of metastasis in primary solid tumors. Nat Genet 33:49-54, 2003
17. Minn AJ, Gupta GP, Siegel PM, et al: Genes that mediate breast cancer metastasis to lung. Nature 436:518-524, 2005
18. Glinsky GV, Berezovska O, Glinskii AB: Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. J Clin Invest 115:1503-1521, 2005
19. Lau SK, Boutros PC, Pintilie M, et al: Three-gene prognostic classifier for early-stage non small-cell lung cancer. J Clin Oncol 25:5562-5569, 2007
20. Chen HY, Yu SL, Chen CH, et al: A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 356:11-20, 2007
21. Lu Y, Lemon W, Liu PY, et al: A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. PLoS Med 3:e467, 2006
22. Bianchi F, Nuciforo P, Vecchi M, et al: Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. J Clin Invest 117:3436-3444, 2007
23. Larsen JE, Pavey SJ, Passmore LH, et al: Gene expression signature predicts recurrence in lung adenocarcinoma. Clin Cancer Res 13:2946-2954, 2007
24. Fan C, Oh DS, Wessels L, et al: Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355:560-569, 2006
25. Winton T, Livingston R, Johnson D, et al: Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. N Engl J Med 352:2589-2597, 2005
26. Douillard JY, Rosell R, De Lena M, et al: Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): A randomised controlled trial. Lancet Oncol 7:719-727, 2006
27. Strauss GM, Herndon JE, Maddaus MA, et al: Adjuvant chemotherapy in stage IB non-small cell lung cancer (NSCLC): Update of Cancer and Leukemia Group B (CALGB) protocol 9633. J Clin Oncol 24:365s, 2006 (suppl; abstr 7007)
28. Wakelee H, Dubey S, Gandara D: Optimal adjuvant therapy for non-small cell lung cancer–how to handle stage I disease. Oncologist 12:331-337, 2007



## *Acknowledgment*

We thank Kiyoshi Yanagisawa at Nagoya University and Hirotaka Osada at Aichi Cancer Center Research Institute for their valuable discussion and critical reading of the manuscript.

# EGFR and HER2 Genomic Gain in Recurrent Non-small Cell Lung Cancer After Surgery

## *Impact on Outcome to Treatment with Gefitinib and Association with EGFR and KRAS Mutations in a Japanese Cohort*

*Marileila Varella-Garcia, PhD,* Tetsuya Mitsudomi, MD,† Yashushi Yatabe, MD,‡ Takayuki Kosaka, MD,† Eiji Nakajima, MD,* Ana Carolina Xavier, MD,* Margaret Skokan, BS,* Chan Zeng, PhD,* Wilbur A. Franklin, MD,* Paul A. Bunn, Jr., MD,* and Fred R. Hirsch, MD, PhD**

**Background:** Sensitivity to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) and frequency of activation mutations in EGFR is lower in Caucasian than Asian non small-cell lung cancer (NSCLC) patients. Increased EGFR gene copy numbers evaluated by fluorescence in situ hybridization (FISH) has been reported as predictor of clinical benefit from EGFR-TKIs in Caucasian NSCLC patients. This study was carried out to verify whether EGFR FISH had similar performance in Japanese patients.

**Methods:** A cohort of 44 Japanese patients with recurrent NSCLC after surgery was treated with gefitinib 250 mg daily. The cohort included 48% females and 52% never-smokers; 73% had prior chemotherapy and 57% had stage III-IV at the time of surgery. Adenocarcinoma was the most common histology (86%). FISH was performed using the EGFR/Chromosome Enumeration Probe 7 and PathVysion DNA probes (Abbott Molecular). Specimens were classified as FISH positive when showing gene amplification or high polysomy (≥4 copies of the gene in ≥40% of tumor cells). Tumor response to gefitinib was assessed by RECIST for 33 patients with measurable diseases.

**Results:** Twenty-nine tumors (66%) were EGFR FISH+ and 23 (53%) were HER2 FISH+. Overall response rate was 52%, representing 65% of EGFR FISH+ patients and 29% of EGFR FISH− patients ($p = 0.0777$). Survival was not impacted by the EGFR FISH ($p = 0.9395$) or the HER2 FISH ($p = 0.0671$) status. EGFR FISH+ was significantly associated with HER2 FISH+ ($p = 0.015$) and presence of EGFR mutation ($p = 0.0060$). EGFR mutation significantly correlated with response ($p < 0.0001$) and survival after gefitinib ($p = 0.0204$). EGFR and HER2 FISH status were not associated with KRAS mutation.

**Conclusion:** Frequency of EGFR FISH+ status was higher and its predictive power for TKI sensitivity was lower in this Japanese cohort than in Western NSCLC cohorts. These findings support differences in the mechanisms of EGFR pathway activation in NSCLC between Asian and Caucasian populations. Confirmation of these results in larger cohorts is warranted.

**Key Words:** FISH, EGFR, HER2, KRAS, Biomarkers, NSCLC, Tyrosine inhibitors.

(*J Thorac Oncol.* 2009;4: 318–325)

Tumor dependence on specific molecular pathways may identify the best target for therapy exploration. Activation of the epidermal growth factor receptor (EGFR)-related signaling pathways drives numerous cancer-promoting processes, such as cell proliferation, apoptosis inhibition, angiogenesis, cell adhesion, and motility and invasion, and also controls development of drug resistance.[1] Therefore, anti-EGFR approaches (antibodies directed against the extracellular domain and small inhibitors of the tyrosine kinase activity) have been one of the most successful examples of molecular target therapy in human solid neoplasias, mainly in

*University of Colorado Cancer Center, Aurora, Colorado; †Departments of Thoracic Surgery, Pathology, and Molecular Diagnostics; and ‡Aichi Cancer Center Hospital, Nagoya, Japan.

Disclosure: Dr. Hirsch has served on advisory boards for AstraZeneca, Pfizer, Merck Serono, BMS/Imclone, Syndax, Boehringer Ingelheim, Roche, and Lilly. He has received research grants from Astra Zeneca, OSI, Genentech, Syndax, and Merck. He is also the co-inventor of a University of Colorado owned patent: EGFR FISH As a Predictive Marker for EGFR Inhibitors (patent licensed to Abbot Diagnostics). Dr. Varella-Garcia received honorarium from Abbott Molecular to speak at the Association for Molecular Pathology annual meeting in 2008 about the EGFR FISH assay. She is also a co-inventor of a patent for use of the EGFR FISH assay to select NSCLC patients for therapy. Dr. Mitsudomi has received honorarium for speaking to a professional group from AstraZeneca and Chugai Pharmaceutical. He also provided testimony in court for gefitinib. Dr. Bunn was paid an honorarium and travel expenses by GlaxoSmithKline to attend an advisory board on the MAGG A3 vaccine. A clinical trial using the vaccine is being conducted at the University of Colorado Cancer Center. Dr. Bunn is not the PI and received no funding for this trial.

Address for correspondence: Dr. Fred R. Hirsch, University of Colorado Cancer Center, Department of Medicine and Pathology, PO Box 6511, Mails stop 8117, Aurora, CO 80045. E-mail: fred.hirsch@uccdenver.edu



ISSN: 1556-0864/09/0403-0318

non small-cell lung cancer (NSCLC), head and neck, pancreatic and colorectal carcinomas.[2]

Targeted therapies are expected to be effective when the targeted molecule is a major player in the tumor cellular processes, which usually does not occur in all patients with any specific solid tumor. Strategies for patient selection for targeted therapy are almost universally considered to be necessary but are not fully implemented, even for anti-EGFR therapies. In NSCLC, causally associated with EGFR activation are mutations in the adenosine triphosphate-binding site of the tyrosine kinase domain that sustain abnormal response to the ligand,[3,4] activate multiple signaling transduction pathways[5,6] and selectively activate AKT and signal transducers and activators of transcription signaling.[6,7] Increased gene copy numbers is also a well known mechanism for activation of EGFR-related pathways in gliomas,[8] breast,[9] colon,[10] head and neck cancers,[11]and NSCLC.[12]

In NSCLC, at least three molecular markers have been consistently associated with sensitivity or resistance to EGFR-TKIs (tyrosine kinase inhibitors): mutations and amplification/overrepresentation of the EGFR gene[3–5,12–14] and mutation in the KRAS genes.[15–18] The impact on survival has been extensively investigated for activating EGFR mutations,[19] and less for the EGFR gene copy numbers[12,14,20,21] or for the KRAS mutations[16,22] and results among studies have not been totally concordant. Distinct technologies have been used to identify mutations and genomic gain and part of the discrepancies among results from different studies may due to technical factors. However, other factors such as smoking status, gender, and ethnicity have been demonstrated to impact sensitivity to EGFR-TKIs. Patients of Eastern Asian origin have significantly better clinical outcome to EGFR-TKIs than western populations[23,24] but reasons for these differences are not completely understood. The most important factor so far accounting for this finding is that the Asian NSCLC patients including Japan, have high incidence of activating EGFR mutations.[4,25]

This study aimed to verify the role of EGFR genomic gain as a marker for sensitivity to gefinitib in a Japanese cohort using fluorescence in situ hybridization (FISH), a technology proved to be successful for prediction of outcome to EGFR TKIs in some Caucasian NSCLC populations. In addition, the study aimed to compare EGFR genomic gain with two other gefitinib-related markers, activating mutations in EGFR and resistant mutations in KRAS, which were previously investigated in this cohort.[13]

## PATIENTS AND METHODS

### Description of Patient Population and Definition of Effectiveness of Gefitinib Treatment

From a population of NSCLC patients who underwent surgery between 1999 and 2003 in the Aichi Cancer Center Hospital in Nagoya, Japan, 75 had recurrent disease and were treated with 250 mg/daily of gefitinib for recurrent disease. From those, response to treatment could not be evaluated in 6 cases, tumor material was not available in 24 cases, and FISH analyses failed in 4 cases. Thus, the current study reports on 44 patients, all of whom provided consent for the study.

Tumor materials obtained at initial tumor resection for these 44 NSCLC cases had been previously investigated for EGFR and KRAS mutations.[13,16] Tumor response to gefitinib treatment was evaluated for 33 patients eliminating 11 patients who did not have measurable diseases. Tumor response was judged according to the RECIST, without requirement of confirmation of tumor response no less than 4 weeks apart. The length of gefitinib therapy was used as a surrogate for disease free survival and overall survival was calculated form the start of gefitinib administration to death from any cause or the most recent date on which the patient was known to be alive.

### EGFR and HER2 Fluorescence In Situ Hybridization Assays

Formalin-fixed, paraffin-embedded tumor blocks were sectioned at 4 $\mu$m and submitted to dual-color FISH assays using the Locus Specific Indicator EGFR SpectrumOrange/CEP 7 SpectrumGreen and the PathVysion DNA Probe Kit (HER2 SpectrumOrange/CEP 17 SpectrumGreen Vysis/Abbott Molecular) following protocol previously described.[12] Briefly, slides were deparaffinized in CitriSolv (Fisher Scientific) and washed in 100% ethanol for 5 minutes. The slides were then sequentially incubated in 2× SSC (saline sodium citrate) at 75°C for 13 to 18 minutes, digested in 0.25 mg/ml Proteinase K/2× SSC at 45°C for 14 to 18 minutes, washed in 2× SSC for 5 minutes and dehydrated in ethanol series. Probes were applied according to the manufacturer instructions to the selected hybridization areas, which were covered and sealed. DNA denaturation was performed in dry oven for 15 minutes at 80°C and hybridization was allowed to occur for 20 hours at 37°C in a humidified chamber. Posthybridization washes were performed consecutively in 2× SSC/0.3% NP-40 at 72°C and 2× SSC for 2 minutes each. Following dehydration in ethanol, chromatin was counterstained with 4' = 6-diamidino-2-phenylindole (DAPI) (0.3 $\mu$g/ml in antifade Vectashield mounting medium, Vector Laboratories). Analysis was performed on epifluorescence microscopes using single interference filters sets for green, red (Texas red) and blue (DAPI) as well as dual (red/green) and triple (blue, red, green) band pass filters. For documentation, images were acquired using charged-coupled device camera with Z-stacking and merged using dedicated software (CytoVision, Applied Imaging).

At least 50 tumor nuclei were analyzed in tumor areas selected using the correspondent HE stained slide as a guide. Scoring system followed previous publications.[12] According to the frequency of tumor cells with specific number of copies of the gene and the CEP targets, the tumors were initially classified into six FISH categories (disomy, low and high trisomy, low and polysomy, and gene amplification) and finally grouped into two strata: (a) FISH negative including disomy to low polysomy tumors, which basically have ≥4 copies of the gene in <40% of cells; and (b) FISH positive including tumors with high polysomy (≥4 copies in ≥40% of cells) and gene amplification (defined by a ratio gene/chromosome per cell ≥2, presence of small or nonenumerable clusters of the gene signal or ≥15 copies of the gene signal in ≥10% of the analyzed cells).

**TABLE 1.** Population Characteristics

| Variable | Categories | Statistics |
|---|---|---|
| Age (years) | Median | 60.9 × 10.3 |
| | Range | 38–79 |
| Gender | Male | 23 (52.3%) |
| | Female | 21 (47.7%) |
| Smoking | Never | 23 (52.3%) |
| | Ever | 21 (47.7%) |
| Histology | Adenocarcinoma | 38 (86.4%) |
| | Nonadenocarcinoma (SqC, LC)[a] | 6 (13.6%) |
| Differentiation | Poor | 10 (26.3%) |
| | Moderate | 22 (57.9%) |
| | Well | 6 (15.8%) |
| | Not determined | 6 |
| Stage | Early (I–II) | 19 (43.2%) |
| | Advanced (III–IV) | 25 (56.8%) |
| Prior chemotherapy | Yes | 12 (27.3%) |
| | No | 32 (72.7%) |
| Survival after surgery (days) | Median | 2081 |
| | Range | 250–2655 |
| Tumor response (RECIST) | Yes | 17 (52%) |
| | No | 16 (48%) |
| Disease free interval (days) | Median | 375 |
| | Range | 99–1818 |
| Survival after gefitinib (days) | Median | 562 |
| | Range | 69–724 |
| Death | Dead | 15 (34.1%) |
| | Alive | 29 (65.9%) |

[a] SqC, Squamous cell carcinoma; LC, Large cell carcinoma.

## Statistical Analysis

For comparisons of proportions, the Pearson's $\chi^2$ test or the Fisher's exact test was used. Nonparametric Wilcoxon rank sum test or Kruskal-Wallis test was used to compare the difference in continuous variables. The Kaplan-Meier method was used to estimate the probability of survival as a function of time, and survival differences between groups were analyzed by the log-rank test. The two-sided significance level was set at $p < 0.05$. All analyses were performed using SAS version 9.1 (SAS Institute Inc, Cary, NC) software.

## RESULTS

Clinical and demographical characteristics are summarized in Table 1. The patients were evenly split between males and females, never or ever smokers and with early or advanced stage disease. Adenocarcinoma histology and poorly or moderately differentiated histologic grade were prevalent. Most patients had not received prior chemotherapy. Median disease free interval after surgery was 375 days, median survival after gefitinib treatment was 562 days, and 66% of patients were alive at the time of last follow up.

EGFR FISH and mutation status in relation to demographics are summarized in Table 2. While EGFR mutation was associated with female gender, never-smoking status, and adenocarcinoma histology, none of these was related with EGFR-FISH status.

Distribution of patients through the FISH categories is illustrated in Figure 1*A* for the EGFR gene and Figure 1*B* for the HER2 gene. The majority of tumors (29 cases [66%]) were EGFR FISH positive, predominantly due to a large representation of tumors with high polysomy (23 cases, 52%, Figure 2*A*) rather than gene amplification (6 cases, 14%, Figure 2*B*). Also, a high number of tumors (23 cases, 53%) were positive for HER2 FISH, of which 21 cases (48%) were represented by high polysomy and only 2 cases (5%) by gene amplification (illustrated in Figure 2*C*). EGFR and HER2 patterns were significantly associated ($p = 0.015$): 19 cases (43%) of tumors were positive and 11cases (25%) were negative for both genes, while 14 cases (32%) had discordant patterns; EGFR FISH positives were more likely to be HER2 FISH positives (19/29 = 66%) than EGFR FISH negatives (4/15 = 27%).

Overall, the specimens with amplification of the EGFR or HER2 genes exhibited clusters of loosely associated signals (Figures 2*B*, *C*) indicating that the amplification occurred as homogenously staining regions. However, one specimen displayed EGFR gene amplification as numerous, diffuse signals mimicking the extrachromosomal double minutes (Figure 2*D*). Heterogeneity for both EGFR and HER FISH

**TABLE 2.** EGFR FISH and Mutation Status According to Demographics

| | | EGFR FISH | | | EGFR Mutation | | |
|---|---|---|---|---|---|---|---|
| Variable | Categories | Positive | Negative | *p* | Positive | Negative | *p* |
| Age (years) | Median | 61.0 | 62.0 | | 61.0 | 61.0 | |
| Gender | Male | 15 (65%) | 8 | $p = 0.9193$ | 11 (48%) | 12 | $p = 0.0536$ |
| | Female | 14 (67%) | 7 | | 16 (76%) | 5 | |
| Smoking | Never | 15 (67%) | 8 | $p = 0.9193$ | 18 (78%) | 5 | $p = 0.016$ |
| | Ever | 14 (65%) | 7 | | 9 (43%) | 12 | |
| Histology | Adenocarcinoma | 25 (66%) | 13 | $p = 0.9664$ | 26 (68%) | 12 | $p = 0.0151$ |
| | Nonadenocarcinoma (SqC, LC)[a] | 4 (67%) | 2 | | 1 (17%) | 5 | |

FISH, fluorescence in situ hybridization; EGFR, epidermal growth factor receptor.
[a] SqC, Squamous cell carcinoma; LC, Large cell carcinoma.

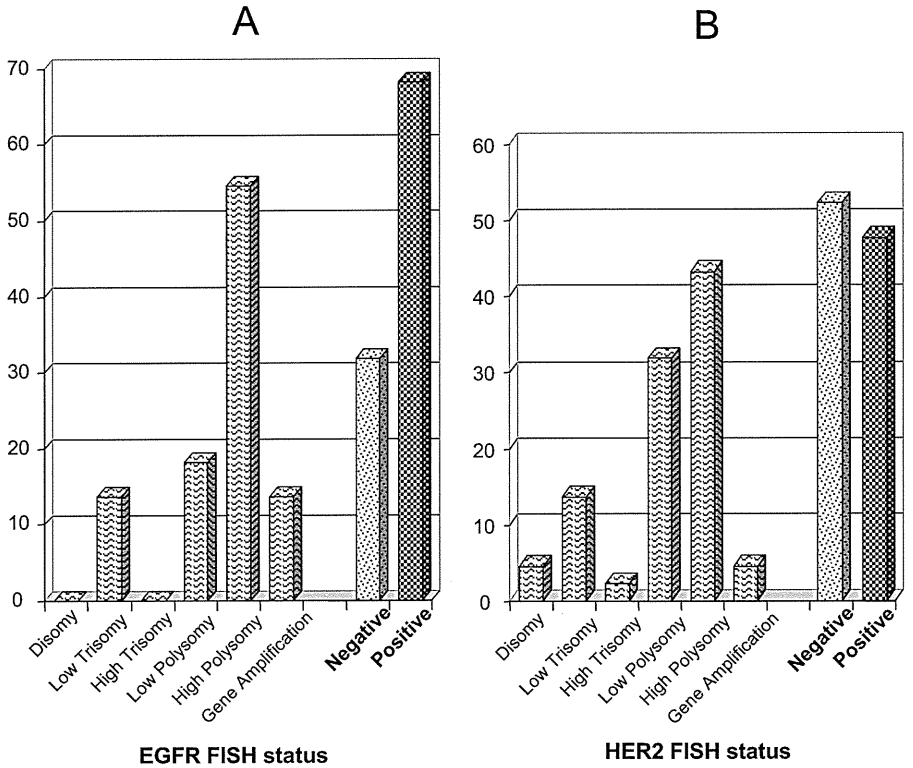


**FIGURE 1.** Frequencies of tumors with distinct categories for the epidermal growth factor receptor-fluorescence in situ hybridization (EGFR-FISH) (*A*) and the HER2 FISH (*B*) assays. Negative category includes disomy to low polysomy. Positive category includes high polysomy and gene amplification.

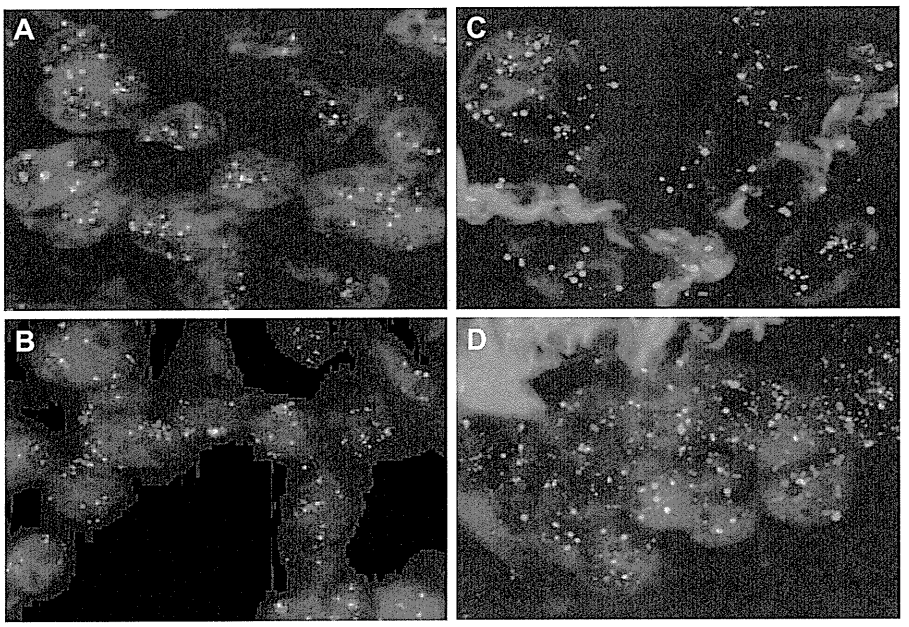


**FIGURE 2.** Hybridization of the non small-cell lung cancer (NSCLC) sections with the epidermal growth factor receptor EGFR/CEP7 (*A, B, D*) and the PathVysion probe set (*C*) showing EGFR high polysomy (*A*), EGFR clustered gene amplification (*B*), HER2 gene amplification (*C*) and EGFR amplification as double minutes (*D*).

patterns was common, with tumor foci showing nuclei with high copy numbers (including gene amplification) interspaced with nuclei with low copy numbers.

The association between FISH patterns and response to the gefitinib treatment for 33 patients with measurable diseases is shown in Table 3. Response to gefitinib was marginally higher in EGFR FISH positive (65%) than negative (29%) patients ($p = 0.0777$). Patients with EGFR gene amplification had a trend towards better benefit (response in 4 of 4 = 100%) than patients with high polysomy (response in 9 of 16 = 56%). HER2 FISH positive pattern trended no impact, including 47% of responders ($p =$

**TABLE 3.** Tumor Response in Relation to EGFR FISH, HER2 FISH, EGFR Mutation and KRAS Mutation Status

| | | Patients | | Tumor response | | | |
|---|---|---|---|---|---|---|---|
| **Molecular marker** | **Categories** | ***n*** | **%** | **PR (%)** | **SD** | **PD** | ***p*** |
| EGFR | Positive (+) | 20 | 61 | 13 (65%) | 1 | 6 | $p = 0.0777$ |
| | Negative (−) | 13 | 39 | 4 (29%) | 0 | 9 | |
| HER2 | Positive (+) | 17 | 52 | 8 (47%) | 0 | 9 | $p = 0.4426$ |
| | Negative (−) | 16 | 48 | 9 (56%) | 1 | 6 | |
| EGFR and HER2 | +/+ | 13 | 39 | 8 (62%) | 0 | 5 | $p = 0.2451$ |
| | +/− | 7 | 21 | 5 (71%) | 1 | 1 | |
| | −/+ | 4 | 12 | 0 (0%) | 0 | 4 | $p^a$ |
| | −/− | 9 | 27 | 4 (44%) | 0 | 5 | |
| EGFR mutation | Positive (+) | 20 | 61 | 17 (85%) | 1 | 2 | $p < 0.0001$ |
| | Negative (−) | 13 | 39 | 0 (0%) | 0 | 13 | |
| EGFR FISH and EGFR mutation | +/+ | 16 | 48 | 13 (81%) | 1 | 2 | $p = 0.0029$ |
| | +/− | 4 | 12 | 0 (0%) | 0 | 4 | |
| | −/+ | 4 | 12 | 4 (100%) | 0 | 0 | $p^a$ |
| | −/− | 9 | 27 | 0 (0%) | 0 | 9 | |
| KRAS mutation | Positive (+) | 4 | 13 | 0 (0%) | 0 | 4 | $p = 0.0995$ |
| | Negative (−) | 26 | 87 | 14 (54%) | 1 | 11 | |
| EGFR FISH and KRAS mutation | +/+ | 0 | 0 | 0 (0%) | 0 | 0 | $p^a$ |
| | +/− | 17 | 57 | 10 (59%) | 1 | 6 | |
| | −/+ | 4 | 13 | 0 (0%) | 0 | 4 | $p^a$ |
| | −/− | 9 | 30 | 4 (44%) | 0 | 5 | |

FISH, fluorescence in situ hybridization; EGFR, epidermal growth factor receptor; PR, partial response; PD, progressive disease.
[a] *p* value could not be calculated because of blank cells.

**TABLE 4.** Time to Treatment Failure According to EGFRFISH, HER2 FISH, EGFR Mutation and KRAS Mutation Status

| | | Patients | | TTF after Gefitinib (Days) | |
|---|---|---|---|---|---|
| **Molecular marker** | **Categories** | ***n*** | **%** | **Median** | ***p*** |
| EGFR | Positive (+) | 29 | 66 | 169 | 0.722 |
| | Negative (−) | 15 | 34 | 97 | |
| HER2 | Positive (+) | 23 | 53 | 121 | 0.1815 |
| | Negative (−) | 21 | 47 | 144 | |
| EGFR and HER2 | +/+ | 19 | 43 | 169 | 0.0179 |
| | +/− | 10 | 23 | 118 | |
| | −/+ | 4 | 9 | 56 | |
| | −/− | 11 | 25 | 144 | |
| EGFR mutation | Positive (+) | 27 | 61 | 311 | <0.0001 |
| | Negative (−) | 17 | 39 | 83 | |
| EGFR FISH and EGFR mutation | +/+ | 22 | 50 | 182 | <0.0001 |
| | +/− | 7 | 16 | 67 | |
| | −/+ | 5 | 11 | 916 | |
| | −/− | 10 | 23 | 83 | |
| KRAS mutation | Positive (+) | 5 | 12 | 87 | 0.0248 |
| | Negative (−) | 36 | 88 | 146 | |
| EGFR FISH and KRAS mutation | +/+ | 1 | 2 | 113 | 0.0767 |
| | +/− | 25 | 61 | 169 | |
| | −/+ | 4 | 10 | 57 | |
| | −/− | 11 | 25 | 144 | |

FISH, fluorescence in situ hybridization; EGFR, epidermal growth factor receptor; TTF, time to treatment failure.

0.4426). Response rate was 62% of patients with EGFR and HER2 FISH positive tumors, in 45% of patients with EGFR or HER2 FISH positive tumors, and in 44% of patients EGFR and HER2 FISH negative tumors. Time to treatment failure (TTF) was not significantly associated with EGFR or HER2 FISH positivity (Table 4). Overall survival was not associated with patterns of EGFR FISH ($p = 0.93$) or HER2 FISH ($p = 0.69$), as shown in Figure 3*A*, *B*. EGFR FISH+ patients with high polysomy (score 5) and true gene amplification (score 6) did not differ regarding survival ($p = 0.6607$; Figure 3*C*).

Among these 44 NSCLC patients, 27 (61%) had activating mutations in the tyrosine kinase domain of the EGFR gene and, among 41 who were tested for KRAS mutations, 5

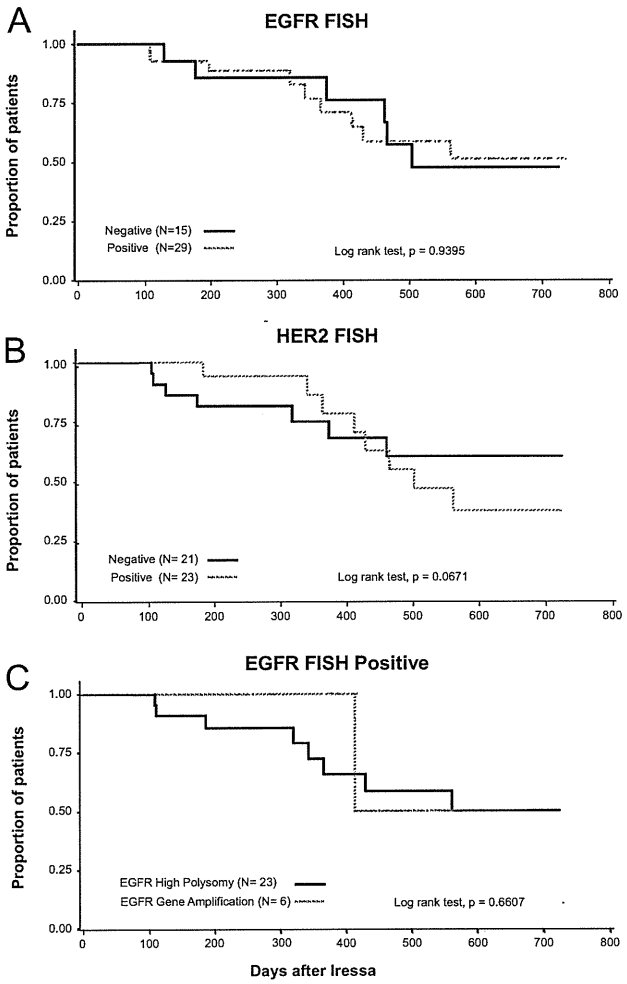


**FIGURE 3.** Effect on survival from the day of initiating gefitinib treatment in recurrent non small-cell lung cancer (NSCLC) after surgery by epidermal growth factor receptor fluorescence in situ hybridization (EGFR FISH) status (*A*), HER2 FISH status (*B*), and EGFR high polysomy and gene amplification (*C*).

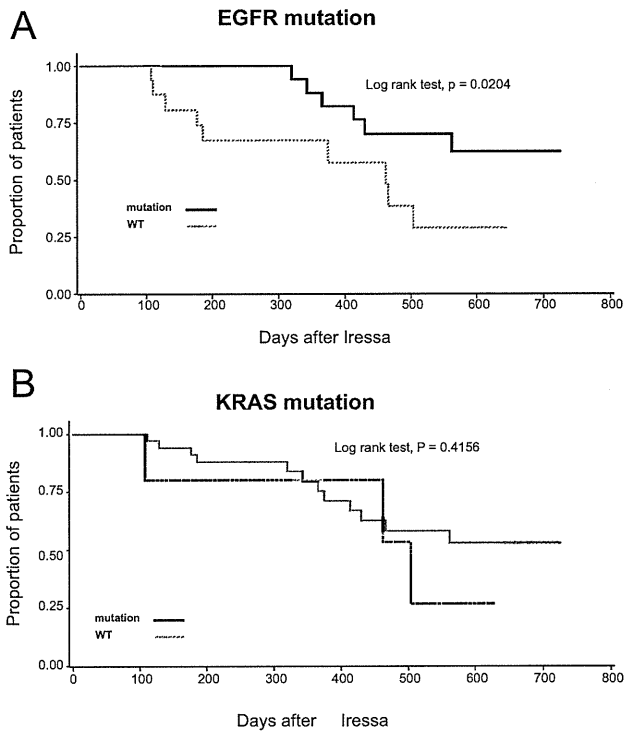


**FIGURE 4.** Effect on survival from the day of initiating gefitinib treatment in recurrent non small-cell lung cancer (NSCLC) after surgery by epidermal growth factor receptor (EGFR) activating mutation (*A*) and KRAS mutation (*B*) status.

(12%) had point mutations in codons 12 or 13. Table 3 also shows tumor response according to presence or absence of EGFR and KRAS mutations, both individually and in combination with EGFR FISH. EGFR mutation was significantly associated with tumor response ($p < 0.0001$) and prolonged TTF ($p < 0.0001$) or survival ($p = 0.02$; Figure 4*A* and Table 4). EGFR FISH positivity was significantly associated with presence of EGFR mutation ($p = 0.0060$). Patients with EGFR mutation were more likely to be EGFR FISH positive (22/27 = 81%) than patients with wild type EGFR (7/17 = 41%). EGFR mutations were present in all 6 tumors with EGFR gene amplification and in 16 out of 23 tumors with EGFR high polysomy (70%). Response rate was 81% of 16 cases positive for both EGFR FISH and mutation and all 4 EGFR FISH negative/EGFR mutation positive cases responded to gefitinib (Table 3).

Conversely, none of the 4 patients with KRAS mutation (none of whom were EGFR FISH positive) or of the 13 patients with EGFR wild type (4 of whom were EGFR FISH positive) benefited from gefitinib treatment. Presence of KRAS mutation was significantly associated with TTF ($p = 0.0248$) but not with lack of response ($p = 0.0995$) or overall survival ($p = 0.4156$, Figure 4*B*).

## DISCUSSION

The EGFR FISH positive status had a borderline association to response of gefitinib treatment, but no impact on

survival in this cohort of Japanese NSCLC patients. These results do not support a predictive role of the established EGFR FISH assay to gefitinib sensitivity in Japanese NSCLC patients. This observation contrasts with previous findings in Caucasian NSCLC populations obtained by our group[12,20,21] and others,[14] that had identified EGFR genomic gain by FISH as a significant predictor of outcome to EGFR-TKIs. In the current study, EGFR mutation was highly predictive of both response and survival to gefitinib. Lack of predictive value of EGFR FISH or EGFR gene copy numbers as assessed by quantitative polymerase chain reaction have also been reported by Korean[17] and Japanese[26] groups. Therefore, there seems to be ethnic differences as to whether EGFR genomic gain is predictive for response or survival after geftinib treatment.

The clinical and demographical characteristics of this Japanese cohort were distinctive, including high proportion of female, never smokers, early stage disease, no prior chemotherapy, and adenocarcinomas. Unselected cohorts of Asian origin usually have higher frequency of females (40%[27]) and never smokers (40%[27]) than Caucasians (34% for females, 9% for never smokers according to Kobrinsky et al.[28]). In addition, this cohort had one of the highest reported frequencies of EGFR FISH+ tumors (68%) and EGFR mutations (61%). Taken only studies that evaluated gene copy numbers by FISH with identical or similar scoring criteria, the frequency of EGFR FISH+ tumors ranged from 44 to 48% in Asian patients[17,26,29] and from 32 to 45% in Caucasian NSCLCs.[14,21] EGFR activating mutations are well known to be more prevalent in Asian (40–50% of adenocarcinomas[27,30]) than Caucasian NSCLCs (10% of adenocarcinomas[25]). Altogether, these findings substantiate the interesting hypothesis that there are ethnicity-associated molecular peculiarities in NSCLC.

The two EGFR gene markers, activating mutation and genomic gain, were significantly correlated in this cohort. Association between EGFR gene amplification and activating mutations has been reported in NSCLC cell lines[31] and clinical specimens of Caucasian[12] and Asian origins.[17,32] Furthermore, the selective amplification of the mutant allele was verified in the cell lines H3255, H827, PC-9, KT-2, KT-4 and Ma-1,[31] as well as in Asian patients.[32] These findings support the hypothesis that there is a selection of cells carrying the amplification of the mutant allele in lung tumorigenesis. Interestingly, high EGFR copy numbers due to chromosomal aneusomy or structural rearrangements (high polysomy) were also associated with mutations in this cohort and in Caucasian NSCLC.[33]

Status of the HER2 gene in NSCLC has been poorly explored and discrepant results have been reported in association with outcome to EGFR-TKIs.[34] In this cohort, HER2 genomic gain showed up as a negative impact factor for survival after gefinitib treatment, in contrast to our previous results in an Italian cohort.[34] Conversely, none of the five KRAS mutant tumors showed treatment efficacy in this study, in agreement with previously findings that KRAS mutations are primary resistance factors to EGFR-TKIs.[18,35]

In summary, the study showed that the EGFR FISH scoring criteria proposed for stratification of NSCLC for therapy with EGFR-TKIs was not effective in Japanese patients as in Caucasian patients. Confirmation of these results in larger cohorts is warranted and investigation of factors that may underlie distinct molecular mechanisms of activation of the EGFR pathway in these populations should be investigated.

## REFERENCES


1. Hynes NE, Lane HA. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 2005;5:341–354.
2. Ciardiello F, Tortora G. EGFR antagonists in cancer treatment. *N Engl J Med* 2008;358:1160–1174.
3. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–2139.
4. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304: 1497–1500.
5. Pao W, Miller V, Zakowski M, 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 U S A* 2004;101:13306–13311.
6. Sordella R, Bell DW, Haber DA, et al. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* 2004;305:1163–1167.
7. Tracy S, Mukohara T, Hansen M, et al. Gefitinib induces apoptosis in the EGFRL858R non-small-cell lung cancer cell line H3255. *Cancer Res* 2004;64:7241–7244.
8. Ohgaki H, Dessen P, Jourde B, et al. Genetic pathways to glioblastoma: a population-based study. *Cancer Res* 2004;64:6892–6899.
9. Slamon DJ, Clark GM, Wong SG, et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235:177–182.
10. Moroni M, Veronese S, Benvenuti S, et al. Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. *Lancet Oncol* 2005;6:279–286.
11. Chung CH, Ely K, McGavran L, et al. Increased epidermal growth factor receptor gene copy number is associated with poor prognosis in head and neck squamous cell carcinomas. *J Clin Oncol* 2006;24: 4170–4176.
12. Cappuzzo F, Hirsch FR, Rossi E, et al. Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small cell lung cancer. *J Natl Cancer Inst* 2005;97:643–655.
13. Mitsudomi T, Kosaka T, Endoh H, et al. Mutations of the epidermal growth factor receptor gene predict prolonged survival after gefitinib treatment in patients with non-small-cell lung cancer with postoperative recurrence. *J Clin Oncol* 2005;23:2513–2520.
14. Tsao MS, Sakurada A, Cutz JC, et al. Erlotinib in lung cancer - molecular and clinical predictors of outcome. *N Engl J Med* 2005;353: 133–144.
15. Eberhard DA, Johnson BE, Amler LC, 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;23:5900–5909.
16. Endoh H, Yatabe Y, Kosaka T, et al. PTEN and PIK3A expression is associated with prolonged survival after gefitinib treatment in EGFR-mutated lung cancer patients. *J Thorac Oncol* 2006;1:629–634.
17. Han SW, Kim TY, Jeon YK, et al. Optimization of patient selection for gefitinib in non-small cell lung cancer by combined analysis of epidermal growth factor receptor mutation, K-ras mutation, and Akt phosphorylation. *Clin Cancer Res* 2006;12:2538–2544.
18. Pao W, Wang TY, Riely GJ, et al. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med* 2005;2:e17.
19. Uramoto H, Mitsudomi T. Which biomarker predicts benefit from EGFR-TKI treatment for patients with lung cancer? *Br J Cancer* 2007; 96:857–863.
20. Hirsch FR, Varella-Garcia M, Bunn PA Jr, et al. Molecular predictors

of outcome with gefitinib in a phase III placebo-controlled study in advanced non-small-cell lung cancer. *J Clin Oncol* 2006;24:5034–5042.
21. Hirsch FR, Varella-Garcia M, McCoy J, et al. Increased epidermal growth factor receptor gene copy number detected by fluorescence in situ hybridization associates with increased sensitivity to gefitinib in patients with bronchioloalveolar carcinoma subtypes: a Southwest Oncology Group Study. *J Clin Oncol* 2005;23:6838–6845.
22. 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;23:857–865.
23. Calvo E, Baselga J. Ethnic differences in response to epidermal growth factor receptor tyrosine kinase inhibitors. *J Clin Oncol* 2006;24:2158–2163.
24. Shigematsu H, Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer* 2006;118:257–262.
25. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 2005;97:339–346.
26. Sone T, Kasahara K, Kimura H, et al. Comparative analysis of epidermal growth factor receptor mutations and gene amplification as predictors of gefitinib efficacy in Japanese patients with nonsmall cell lung cancer. *Cancer* 2007;109:1836–1844.
27. 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;64:8919–8923.
28. Kobrinsky NL, Klug MG, Hokanson PJ, et al. Impact of smoking on cancer stage at diagnosis. *J Clin Oncol* 2003;21:907–913.
29. Sasaki H, Endo K, Okuda K, et al. Epidermal growth factor receptor gene amplification and gefitinib sensitivity in patients with recurrent lung cancer. *J Cancer Res Clin Oncol* 2008;134:569–577.
30. Tokumo M, Toyooka S, Kiura K, et al. The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin Cancer Res* 2005;11:1167–1173.
31. Okabe T, Okamoto I, Tamura K, et al. Differential constitutive activation of the epidermal growth factor receptor in non-small cell lung cancer cells bearing EGFR gene mutation and amplification. *Cancer Res* 2007;67:2046–2053.
32. Takano T, Ohe Y, Sakamoto H, 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;23:6829–6837.
33. Daniele L, Macri L, Schena M, et al. Predicting gefitinib responsiveness in lung cancer by fluorescence in situ hybridization/chromogenic in situ hybridization analysis of EGFR and HER2 in biopsy and cytology specimens. *Mol Cancer Ther* 2007;6:1223–1229.
34. Cappuzzo F, Varella-Garcia M, Shigematsu H, et al. Increased HER2 gene copy number is associated with response to gefitinib therapy in epidermal growth factor receptor-positive non-small-cell lung cancer patients. *J Clin Oncol* 2005;23:5007–5018.
35. Massarelli E, Varella-Garcia M, Tang X, et al. KRAS mutation is an important predictor of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. *Clin Cancer Res* 2007;13:2890–2896.

 

the FEBS Journal

MINIREVIEW

# Epidermal growth factor receptor in relation to tumor development: *EGFR* gene and cancer

Tetsuya Mitsudomi and Yasushi Yatabe

Department of Thoracic Surgery, Pathology and Molecular Diagnostics, Aichi Cancer Center Hospital, Nagoya, Japan

**Keywords**
cancer; epidermal growth factor receptor (EGFR); gefitinib; non-small cell lung carcinoma (NSCLC); tyrosine kinase inhibitor (TKI)

**Correspondence**
T. Mitsudomi, Department of Thoracic Surgery, Aichi Cancer Center Hospital, 1-1 Kanokoden, Chikusa-ku, Nagoya 464-8681, Japan
Fax: +81 52 764 2963
Tel: +81 52 762 6111
E-mail: mitsudom@aichi-cc.jp

(Received 17 July 2009, accepted 13 September 2009)

doi:10.1111/j.1742-4658.2009.07448.x

Epidermal growth factor receptor (EGFR) and its three related proteins (the ERBB family) are receptor tyrosine kinases that play essential roles in both normal physiological conditions and cancerous conditions. Upon binding its ligands, dynamic conformational changes occur in both extracellular and intracellular domains of the receptor tyrosine kinases, resulting in the transphosphorylation of tyrosine residues in the C-terminal regulatory domain. These provide docking sites for downstream molecules and lead to the evasion of apoptosis, to proliferation, to invasion and to metastases, all of which are important for the cancer phenotype. Mutation in the tyrosine kinase domain of the *EGFR* gene was found in a subset of lung cancers in 2002. Lung cancers with an *EGFR* mutation are highly sensitive to EGFR tyrosine kinase inhibitors, such as gefitinib and erlotinib. Here, we review the discovery of EGFR, the EGFR signal transduction pathway and mutations of the *EGFR* gene in lung cancers and glioblastomas. The biological significance of such mutations and their relationship with other activated genes in lung cancers are also discussed.

## Identification of epidermal growth factor, epidermal growth factor receptor and ERBB family proteins

Epidermal growth factor (EGF) was originally isolated by Stanley Cohen in 1962 as a protein extracted from the mouse submaxillary gland that accelerated incisor eruption and eyelid opening in the newborn animal [1]. Therefore, it was originally termed 'tooth-lid factor', but was later renamed EGF because it stimulated the proliferation of epithelial cells [1]. In 1972, the amino acid sequence of the EGF was determined. The presence of a specific binding site for EGF, the EGF receptor (EGFR), was confirmed in 1975 by showing that $^{125}$I-labeled EGF binds specifically to the surface of fibroblasts [1].

In 1978, EGFR was identified as a 170kDa protein that showed increased phosphorylation when bound to EGF in the A431 squamous cell carcinoma cell line that had an amplified *EGFR* gene. The discovery (in 1980) that the transforming protein of Rous sarcoma virus, v-src, has tyrosine-phosphorylation activity led to the discovery that EGFR is a tyrosine kinase activated by binding EGF [1]. In 1984, the cDNA of human *EGFR* was isolated and characterized. A high degree of similarity was found between the amino acid sequence of *EGFR* and that of v-erbB, an oncogene of the avian erythroblastosis virus [1].

**Abbreviations**
ALK, anaplastic lymphoma kinase; BAC, bronchioloalveolar cell carcinoma; EGF, epidermal growth factor; EGFR, epidermal growth factor receptor; EML4, echinoderm microtubule-associated protein-like 4; NRG, neuregulin; STAT, signal transducer and activator of transcription; TKI, tyrosine kinase inhibitor; TRU, terminal respiratory unit.

Screening of cDNA libraries using an EGFR probe identified a family of proteins closely related to EGFR. This family consists of EGFR (also known as ERBB1/HER1), ERBB2/HER2/NEU, ERBB3/HER3 and ERBB4/HER4. ERBB2, ERBB3 and ERBB4 show extracellular homologies, relative to the EGFR, of 44, 36 and 48%, respectively, while those for the tyrosine kinase domain are 82, 59 and 79%, respectively. The degrees of homology in the C-terminal regulatory domain are relatively low, being 33, 24 and 28%, respectively.

## Structure of the ERBB proteins and diversity of their ligands

The EGFR gene is located on chromosome 7p12-13 and codes for a 170kDa receptor tyrosine kinase. All ERBB proteins have four functional domains: an extracellular ligand-binding domain; a transmembrane domain; an intracellular tyrosine kinase domain; and a C-terminal regulatory domain [2]. The extracellular domain is subdivided further into four domains. The tyrosine kinase domain consists of an N-lobe and a C-lobe, and ATP binds to the cleft formed between these two lobes. The C-terminal regulatory domain has several tyrosine residues that are phosphorylated specifically upon ligand binding, as described below (Fig. 1A).

Eleven ligands are known to bind to the ERBB family of receptors [3]. These can be classified into three groups (a) ligands that specifically bind to EGFR (including EGF, transforming growth factor-a, amphiregulin and epigen); (b) those that bind to EGFR and ERBB4 (including betacellulin, heparin-binding EGF and epiregulin); and (c) neuregulin (NRG) (also known as heregulin) that binds to ERBB3 and ERBB4. NRG1 and NRG2 bind to both ERBB3 and ERBB4, whereas NRG3 and NRG4 only bind to ERBB4 [3]. Although these ligands show redundancy, heparin-binding-EGF is the only ligand whose absence in knockout mice results in postnatal lethality as a result of heart and lung problems, while mice lacking other EGF ligands, or even triple null mice deficient for amphiregulin, EGF and transforming growth factor-a are viable [4]. These ligands are synthesized as transmembrane proteins, and soluble ligands (growth factors) are released into the extracellular environment via proteolytic processing. This shedding is mediated by ADAM (a disintegrin and metalloprotease) proteins that are membrane-anchored metalloproteases [4].

## Signal transduction by ERBB proteins

Binding of a family of specific ligands to the extracellular domain of ERBB (except for ERBB2, see below) leads to the formation of homodimers and heterodimers. This process is mediated by rotation of domains I and II, leading to promotion from a tethered configuration to an extended configuration (Fig. 1B) [2]. This exposes the dimerization domain. ERBB2 does not have corresponding ligands but is expressed constitutively in the extended configuration. ERBB2 is a preferred dimerization partner, and heterodimers containing ERBB2 mediate stronger signals

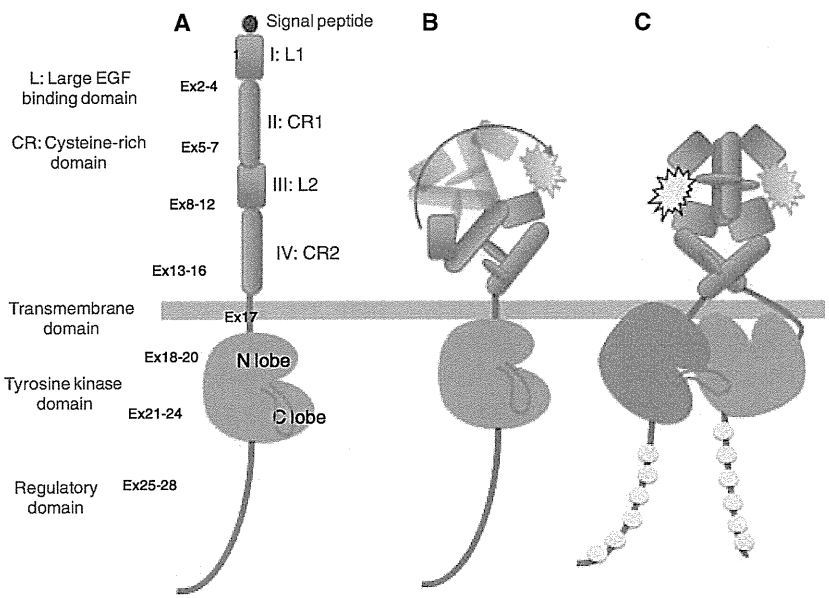


**Fig. 1.** Structure of the EGFR protein (A), activation (B) and dimerization by ligand binding (C).

than other dimers. In the cytoplasm, the kinase domain dimerizes asymmetrically in a tail-to-head orientation (Fig 1C) [5]. In this manner, tyrosine kinase becomes activated, as in the case of activation of cyclin-dependent kinases by cylclins. Dimerization consequently stimulates intrinsic tyrosine kinase activity of the receptors and triggers autophosphorylation of specific tyrosine residues within the cytoplasmic regulatory domain.

These phosphorylated tyrosines serve as specific binding sites for several adaptor proteins, such as phospholipase Cg, CBL, GRB2, SHC and p85. For example, tyrosine-X-X-methionine (where X is any amino acid) is a motif for the p85 binding site. Several signal transducers then bind to these adaptors to initiate multiple signalling pathways, including mitogen-activated protein kinase, phosphatidylinositol 3-kinase/AKT and the signal transducer and activator of transcription (STAT)3 and STAT5 pathways (Fig. 2) [3]. These eventually result in cell proliferation, migration and metastasis, evasion from apoptosis, or in angiogenesis, all of which are associated with cancer phenotypes. ERBB3 lacks tyrosine kinase activity because of substitutions in crucial residues in the tyrosine kinase domain. However, it has many binding sites for p85, a regulatory subunit of phosphatidylinositol 3-kinase, and thus is a preferred dimerization partner.

## EGFR overexpression and cancer

EGFR is expressed in a variety of human tumors, including those in the lung, head and neck, colon, pancreas, breast, ovary, bladder and kidney, and in gliomas. EGFR expression and cancer prognosis have been investigated in many human cancers. Although there some discrepancies have been reported, patients with tumors that show high expression of EGFR tend to have a poorer prognosis in general. However, it was not possible to predict super-responder of gefitinib degree of EGFR expression, as determined by immunohistochemistry or immunoblotting.

## Mutations of the extracellular domain are frequent in glioblastomas

Three different types of deletion mutations (categorized according to the extent of deletion, and termed *EGFR vI, EGFR vII* and *EGFR vIII*) have been reported in the extracellular domain of the *EGFR* gene [6]. In the *EGFR vI* mutation, the extracellular domain has been totally deleted and resembles the v-erbB oncoprotein. In the *EGFR vII* mutation, 83 amino acids in domain IV of the extracellular domain have been deleted; however, this mutation does not appear to contribute to a malignant phenotype. The most

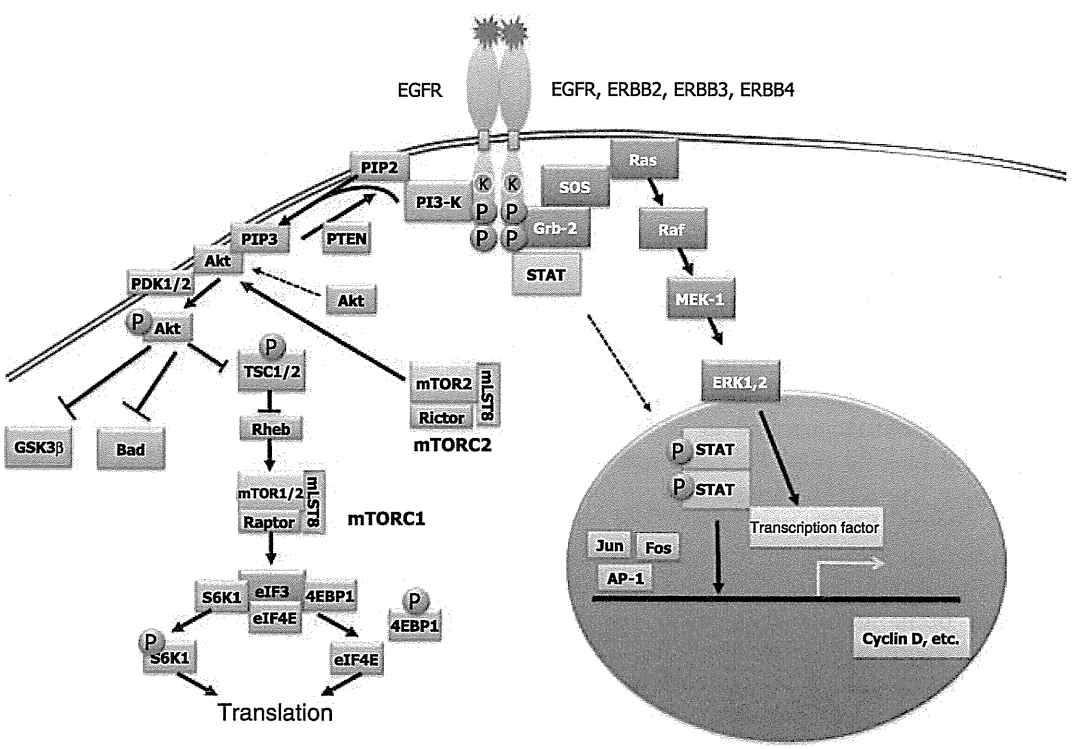


**Fig. 2.** EGFR and ERBB proteins and their downstream pathways.

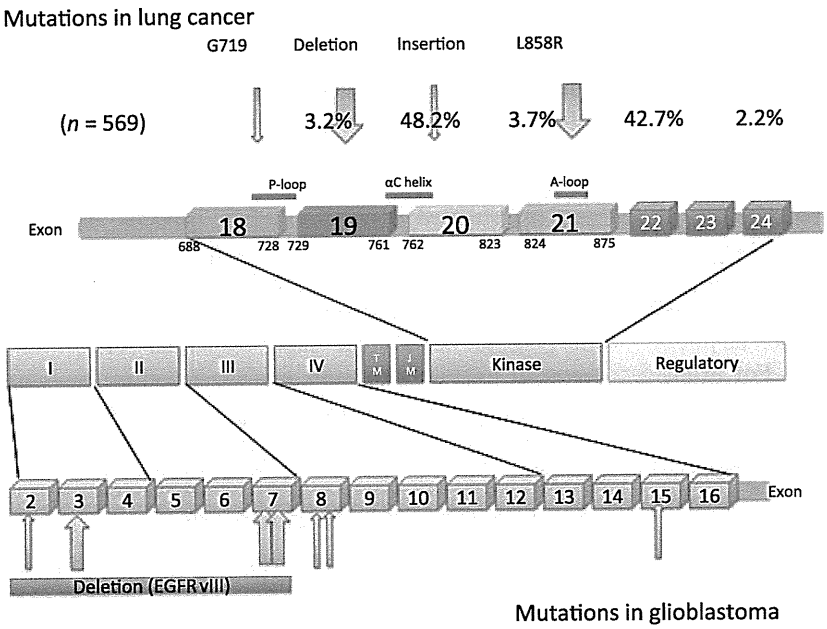


**Fig. 3.** Distribution and frequency of *EGFR* mutations occurring in the kinase domain in lung cancer (upper part of the figure) [12] and in the extracellular domain in glioblastoma (lower part of the figure) [8].

common of the three types of deletion mutations is *EGFR vIII*. This mutation often accompanies gene amplification, resulting in the overexpression of EGFR lacking amino acids 30–297, corresponding to domains I and II. In this case, the EGFR tyrosine kinase is activated constitutively without ligand binding, as in the case of *EGFR vI*. *EGFR vIII* is reported to occur in 30–50% of glioblastomas [6]. In lung cancers, *EGFR vIII* is found in 5% of squamous cell carcinomas, while none of 123 adenocarcinomas were found to harbor this mutation [7]. It is also known that tissue-specific expression of *EGFR vIII* leads to the development of lung cancer [7]. There is also a suggestion that lung tumors with *EGFR vIII* are sensitive to the irreversible EGFR tyrosine kinase inhibitor (TKI), HKI272, despite the fact these tumors are relatively resistant to the reversible inhibitors, gefitinib and erlotinib [7].

Recently, novel missense mutations in the extracellular domain of the *EGFR* gene have been identified in 13.6% (18/132) of glioblastomas and in 12.5% (1/8) of glioblastoma cell lines [8] (Fig. 3). There appear to be several hot spots: five R108K mutations were found in domain I, three T263P mutations and five A289V/D/T mutations were found in domain II, and two G598V mutations were found in domain IV. These *EGFR* mutations occur independently of *EGFR vIII* and provide an alternative mechanism for *EGFR* activation in glioblastomas [8]. Furthermore, these mutations are associated with increased *EGFR* gene dosage and confer anchorage-independent growth and tumorigenicity to NIH-3T3 cells. Cells transformed by expression of these *EGFR* mutants are sensitive to small-molecule EGFR kinase inhibitors [8]. In contrast, none of 119 primary lung tumors was found to harbor these ectodomain mutations [8].

## *EGFR* mutations in the tyrosine kinase domain

In April 2004, two groups of researchers in Boston [9,10], and subsequently a group in New York [11], reported that activating mutations of the *EGFR* gene are present in a subset of non-small cell lung cancer and that tumors with *EGFR* mutations are highly sensitive to EGFR-TKIs. This discovery solved the enigma of why female, nonsmoking, adenocarcinoma patients of East Asian origin with lung cancers had a higher response to EGFR-TKIs, because patients with these characteristics have a higher incidence of *EGFR* mutations. Figure 4 shows the incidence of *EGFR* mutations found in 559 mutations in 2880 lung cancer patients in the literature [12]. It is also intriguing that *EGFR* mutations in the tyrosine kinase domain are almost exclusively seen in lung cancers and not in other types of tumor.

It is of particular interest that *EGFR* mutations are the first molecular aberrations found in lung cancer that are more frequent among patients without a smoking history than among those with one. Furthermore, the *EGFR* mutation frequency is inversely associated with the total amount of tobacco smoked [13]. However, it should be noted that *EGFR* mutations

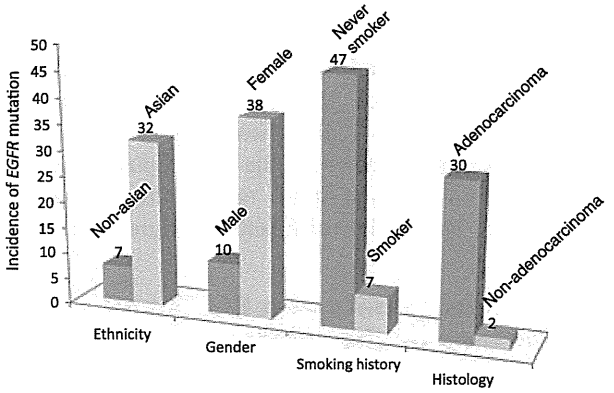


**Fig. 4.** Incidences of *EGFR* mutations in lung cancer in various different clinical backgrounds [12]. Hx, history; adeno, adenocarcinoma.

have been detected in more than 20% of patients with a history of heavy smoking [13]. These findings do not necessarily mean that smoking has a preventive effect on *EGFR* mutations. Rather, they suggest that *EGFR* mutations are caused by carcinogen(s) other than those contained in tobacco smoke, and indicate that the apparent negative correlation with smoking dose occurs as a result of diluting the number of tumors containing *EGFR* mutations with an increased number of tumors containing wild-type *EGFR* as the smoking dose increases. Indeed, this was shown in our case–control study [14].

## Pathology of lung cancers with *EGFR* gene mutations

Bronchioloalveolar cell carcinoma (BAC) is defined as a carcinoma *in situ* without stromal, vascular or pleural invasion, showing growth of neoplastic cells along pre-existing alveolar structures (lepidic growth). Although it is relatively rare to present with pure BAC, invasive adenocarcinomas with areas exhibiting lepidic growth are frequently seen. This type of adenocarcinoma is sometimes referred to as an adenocarcinoma with BAC features. Such tumors respond more to gefitinib than do other types of adenocarcinoma [15] and thus have a higher incidence of *EGFR* mutations. As expected, adenocarcinomas with BAC features are more common in adenocarcinomas of never-smoking patients (13%) than in smokers (5%).

We proposed a terminal respiratory unit (TRU)-type of adenocarcinoma [16]. This type of cancer is characterized by distinct cellular features (expression of thyroid transcription factor 1 and surfactant proteins, and lepidic growth in the periphery), and it resembles adenocarcinomas with nonmucinous BAC features. Although, according to the World Health Organization classification, mucinous BACs form a subset of BACs, this type of BAC does not express thyroid transcription factor 1 or surfactant apoprotein, and is thus not a TRU-type adenocarcinoma. It is also known that *KRAS* mutations are more frequent in mucinous BAC than in nonmucinous BAC.

In our series of 195 adenocarcinomas, 149 were of the TRU type and 46 were of other types [17]. TRU-type adenocarcinomas are associated with a significantly higher incidence of female patients, never-smokers and *EGFR* mutations, but with fewer *KRAS* and *TP53* mutations than other types of adenocarcinoma [17]. An *EGFR* mutation was detected in 97/195 adenocarcinomas, in 91/149 TRU-type adenocarcinomas and in 6/46 tumors of other types. Conversely, 91/97 *EGFR*-mutated adenocarcinomas were categorized as TRU-type adenocarcinomas [17]. In addition, *EGFR* mutations were detected in some cases of atypical adenomatous hyperplasias known to be precursor lesions for BAC [17]. These findings further confirm that the TRU-type adenocarcinoma is a distinct adenocarcinoma subset involving a particular molecular pathway. It is of note that *EGFR* mutations can also occur in poorly differentiated adenocarcinomas, as long as the tumor belongs to the TRU cellular lineage.

## Types of *EGFR* mutations

*EGFR* mutations are mainly present in the first four exons of the gene encoding the tyrosine kinase domain (Fig. 3) [12]. About 90% of the *EGFR* mutations are either small deletions encompassing five amino acids from codons 746–750 (ELREA) or missense mutations resulting in a substitution of leucine with arginine at codon 858 (L858R). There are more than 20 variant types of deletion, including larger deletions, deletions plus point mutations and deletions plus insertions. About 3% of the mutations occur at codon 719, resulting in the substitution of glycine with cysteine, alanine or serine (G719X). In addition, about 3% are in-frame insertion mutations in exon 20. These four types of mutations seldom occur simultaneously. There are many rare point mutations, some of which occur together with L858R [12].

Exon 19 deletional mutation and L858R result in increased and sustained phosphorylation of EGFR and other ERBB family proteins without ligand stimulation. It has been shown that mutant EGFR selectively activates the AKT and STAT signaling pathways that promote cell survival, but has no effect on the mitogen-activated protein kinase pathway that induces cell proliferation [18]. EGFR mutants in the

kinase domain are oncogenic [19]. The mutant EGFR protein can transform both fibroblasts and lung epithelial cells in the absence of exogenous EGFR, as evidenced by anchorage-independent growth, focus formation and tumor formation in immunocompromised mice [19]. Transformation is associated with constitutive autophosphorylation of EGFR, SHC phosphorylation and STAT pathway activation [19]. Whereas transformation by most EGFR mutants confers cell sensitivity to erlotinib and gefitinib, transformation by an exon 20 insertion (D770insNPG) makes cells resistant to these inhibitors but more sensitive to the irreversible inhibitor CL-387,785 [19]. In that study, the G719S mutation of exon 18 showed intermediate sensitivity *in vitro* [19]. However, the authors did not observe any difference between the exon 19 deletion and L858R in their cell-based assay. However, biochemical analysis of the kinetics of purified wild-type and mutant kinases revealed that mutant kinases have a higher $K_m$ for ATP (wild-type, 5 $\mu mol \cdot L^{-1}$; L858R, 10.9 $\mu mol \cdot L^{-1}$; deletion, 129.0 $\mu mol \cdot L^{-1}$) and a lower $K_i$ for erlotinib (wild-type, 17.5 $\mu mol \cdot L^{-1}$; L858R, 6.25 $\mu mol \cdot L^{-1}$; deletion, 3.3 $\mu mol \cdot L^{-1}$;) [20]. Mulloy *et al.* [21] showed that the Del747–753 kinase had a higher autophosphorylation rate and higher sensitivity to erlotinib than L858R kinase. These data reflect differences in the clinical response rate between the exon 19 deletion and L858R.

Oncogenic activity of *EGFR* mutants has also been shown *in vivo*. Two groups of researchers have developed transgenic mice that express either the exon 19 deletion mutant or the L858R mutant in type II pneumocytes under the control of doxycyclin [22,23]. Expression of either *EGFR* mutant led to the development of adenocarcinomas similar to human BACs, and the withdrawal of doxycycline to reduce expression of the transgene, or erlotinib treatment, resulted in tumor regression. These experiments show that persistent EGFR signaling is required for tumor maintenance in human lung adenocarcinomas expressing *EGFR* mutants.

## *EGFR* gene copy numbers

*EGFR* amplification is detectable in 40% of human gliomas and is often associated with deletion mutations, as discussed below. When the topographical distribution of *EGFR* amplification in lung cancers with confirmed mutations was examined, gene amplification was found in 11 of 48 specimens [24]. Nine of the cancers showed heterogeneous distribution, and amplification was associated with higher histological tumor grades or invasive growth [24]. However, the amplification status of the metastatic lymph node was not always associated with gene amplification of the primary tumors [24]. Only one of 21 carcinomas *in situ*, and none of 17 precursor lesions, harbored gene amplifications [24]. These results suggest that mutations occur early in the development of lung adenocarcinomas and that amplification might be acquired in association with tumor progression.

## Relationship between EGFR and mutations of the related genes

The activating mutation of the *KRAS* gene was one of the earliest discoveries of genetic alterations in lung cancer, and has been known as a poor prognostic indicator since 1990 [25]. We were the first group to report that the occurrence of *EGFR* and *KRAS* mutations are strictly mutually exclusive [13]. One explanation is that the KRAS–mitogen-activated protein kinase pathway is one of the downstream signaling pathways of EGFR. Interestingly, *KRAS* mutations predominantly occur in White people with a history of smoking. Mutations of the *ERBB2* gene are present in a very small fraction ($\sim$ 3%) of adenocarcinomas and they appear to target the same population targeted by *EGFR* mutations: never-smokers and female patients [26]. Most of the *ERBB2* mutations are insertion mutations in exon 20 [26]. As anticipated, tumors with *ERBB2* mutations are resistant to treatment with EGFR-TKIs [27] because constitutively activated ERBB2 kinase will phosphorylate other ERBB family proteins, resulting in the activation of downstream molecules even when the EGFR tyrosine kinase is blocked. Mutation of the *BRAF* gene occurs in about 1–3% of lung adenocarcinomas.

By retrieving transforming genes from mouse 3T3 fibroblasts transfected with a cDNA expression library constructed from a lung adenocarcinoma arising in a male smoker, Soda *et al.* [28] identified the gene resulting from the fusion of that for transforming echinoderm microtubule-associated protein-like 4 (*EML4*) and the gene for anaplastic lymphoma kinase (*ALK*). This *EML4–ALK* fusion gene resulted from a small inversion within chromosome 2p. The *EML4–ALK* fusion transcript is detected in about 5% of non-small cell lung cancers. *ALK* translocation was associated with patients being never-smokers of a younger age and acinar-type adenocarcinomas, in a larger study [29]. It is also noteworthy that *EGFR*, *ERBB2*, *BRAF*, *KRAS* and *ALK* mutations almost never occur simultaneously in individual patients, suggesting a complementary role of these mutations in lung carcinogenesis.