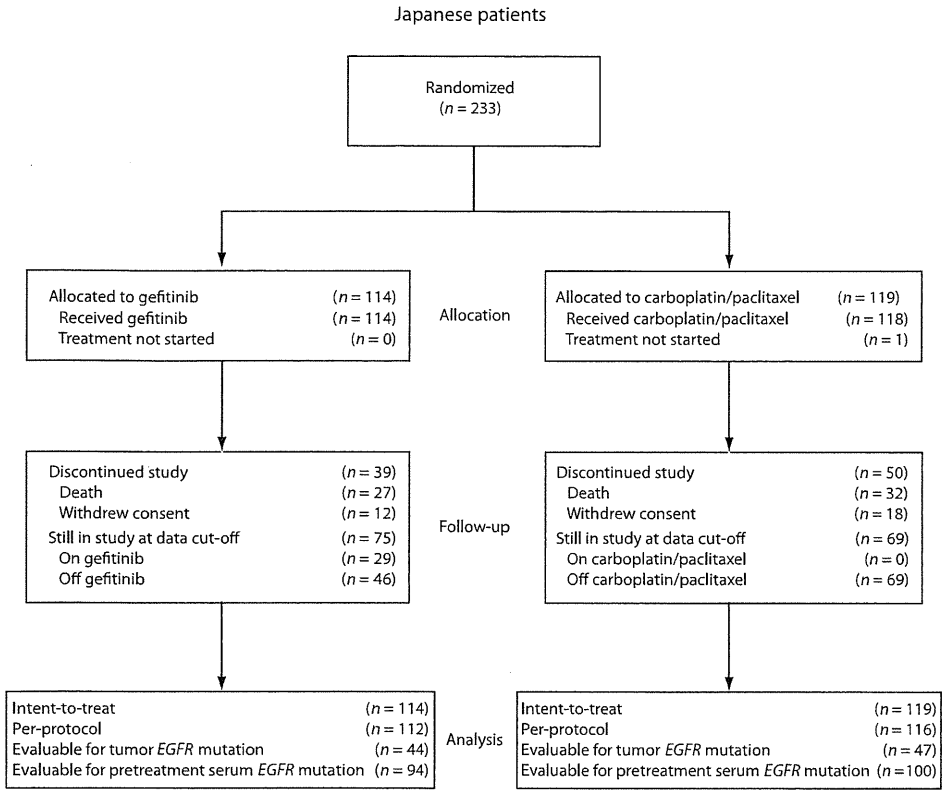

FIGURE 1. CONSORT diagram representing patient disposition (including number of patients with tumor tissue or serum evaluable for *EGFR* mutation status). *EGFR*, epidermal growth factor receptor.

and negative predictive values and their exact 95% CIs, and the kappa coefficient and 95% CI, for *EGFR* mutation status in serum samples, were evaluated assuming that the *EGFR* mutation status in tumor tissue was a true reflection of tumor biology. The proportion of concordance between cfDNA and tumor was calculated on a similar basis by excluding patients judged as unknown using either cfDNA or tumor samples.

RESULTS

Patients

In total, 233 patients from Japan were randomized to study treatment (19.1% of the overall IPASS population). Preplanned evaluations of efficacy, quality of life, and safety for the overall Japanese study population have been previously presented[15,16] and are summarized in Supplemental Digital Content 2 (Results http://links.lww.com/JTO/A153) and 3 (Figure http://links.lww.com/JTO/A154). The patient disposition for the Japanese subset of IPASS is shown in Figure 1.

EGFR Mutation Status

An evaluable DNA sample for *EGFR* mutation status derived from tumor tissue was available for 91 patients; of these, 56 (61.5%) patients were *EGFR* M+, with a lower proportion of *EGFR* M+ patients in the gefitinib group compared with the carboplatin/paclitaxel group (52.3% [23/44] versus 70.2% [33/47]) (Figure 2). A total of 194 patients provided a pretreatment serum sample for mutation analysis; all were evaluable. Of these, 46 (23.7%) patients were cfDNA *EGFR* M+ (25.5% [24/94] and 22.0% [22/100] in the gefitinib and carboplatin/paclitaxel groups, respectively) (Figure 2). Data from pretreatment pleural effusion (9 patients) and postprogression serum analyses (144 patients) are presented in Supplemental Digital Content 2 (Results http://links.lww.com/JTO/A153) and 4 (Table http://links.lww.com/JTO/A155).

Demographic and Baseline Characteristics of Patients with Known *EGFR* Mutation Status

Key demographic and baseline characteristics for patients with known (i.e., evaluable) cfDNA or tumor *EGFR* mutation status were generally consistent with the overall Japanese study population (Table 1).

Pretreatment cfDNA *EGFR* Mutation Status and Clinical Outcomes

The subset of patients with known cfDNA *EGFR* mutation status could be assumed to be representative of the overall Japanese study population (and therefore the overall study population) as shown by similar PFS and ORR results (Table 1).

A significant interaction between cfDNA *EGFR* mutation status and treatment was evident for PFS (interaction test $p = 0.045$). PFS was significantly longer with gefitinib than carboplatin/paclitaxel in the cfDNA *EGFR* M+ subgroup (HR, 0.29; 95% CI, 0.14–0.60; $p < 0.001$) (Figure 3*A*). In the cfDNA *EGFR* M− subgroup, there were no significant differences for PFS with gefitinib compared with carboplatin/paclitaxel (HR, 0.88; 95% CI, 0.61–1.28; $p = 0.50$) (Figure 3*B*). However, the HR was not constant over time. We

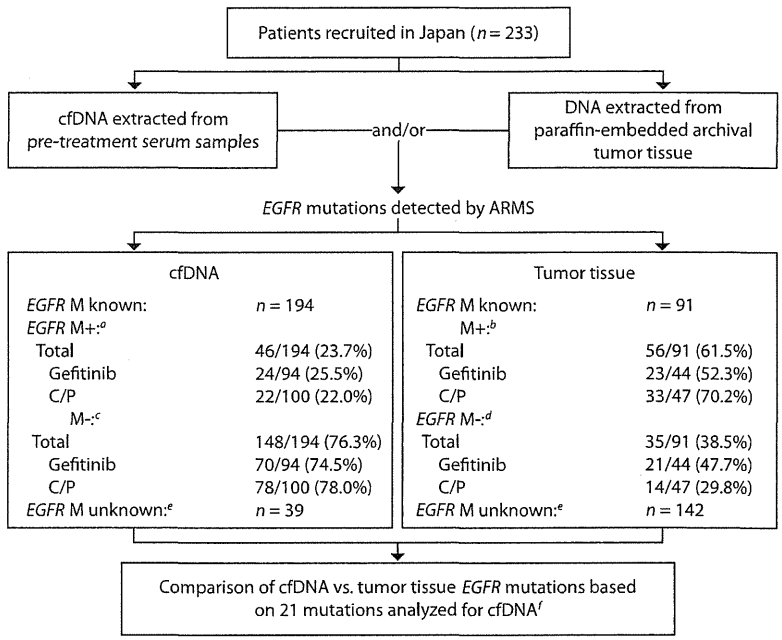

FIGURE 2. Flow and results of *EGFR* mutation analysis. [a]Sample positive for ≥1 of 21 mutations tested; detected 19 deletions in exon 19, L858R, and T790M. [b]Sample positive for ≥1 of 29 mutations tested; detected 19 deletions in exon 19, L858R, T790M, L861Q, G719S, G719A, G719C, S768I; 3 insertions in exon 20. [c]Sample negative for all 21 mutations tested. [d]Sample negative for all 29 mutations tested. [e]Unknown *EGFR* mutations: no sample available or failed analysis. [f]86 patients had known mutation status by both tumor tissue and cfDNA. C/P, carboplatin/paclitaxel; *EGFR*, epidermal growth factor receptor; M, mutation; M+, mutation-positive; M−, mutation-negative.

TABLE 1. Patient Demographics, Baseline Characteristics, and Efficacy (PFS and ORR) for Patients with Samples (cfDNA or Tumor) Evaluable for *EGFR* Mutation Status Compared with the Overall Japanese[a] Study Population (Japanese ITT Population)

	Evaluable for *EGFR* Mutation Status (cfDNA) (n = 194)[b]	Evaluable for *EGFR* Mutation Status (Tumor) (n = 91)[b]	Overall Japanese Study Population (n = 233)
Demography, n (%)			
Female	172 (88.7)	84 (92.3)	204 (87.6)
WHO PS 0/1	185 (95.4)	89 (97.8)	223 (95.7)
Never-smoker	177 (91.2)	83 (91.2)	212 (91.0)
Stage IIIB	66 (34.0)	27 (29.7)	73 (31.3)
Age <65 yr	97 (50.0)	45 (49.5)	121 (51.9)
Efficacy			
PFS HR[c] (95% CI)	0.68 (0.49–0.95)	1.08 (0.68–1.72)	0.69 (0.51–0.94)
ORR OR[d] (95% CI)	1.45 (0.80–2.61)	0.99 (0.41–2.40)[e]	1.34 (0.78–2.30)

[a] Refers to the country of recruitment and not necessarily to racial origin.
[b] Includes both mutation-positive and mutation-negative samples.
[c] HR <1 indicates a difference in favor of gefitinib.
[d] OR >1 indicates a greater chance of response on gefitinib.
[e] These results should be interpreted with caution as the logistic regression model did not converge.
cfDNA, circulating free DNA; CI, confidence interval; *EGFR*, epidermal growth factor receptor; HR, hazard ratio; ITT, intent-to-treat; OR, odds ratio; ORR, objective response rate; PFS, progression-free survival; PS, performance status; WHO, World Health Organization.

believe that this result was due to the high rate of false negative results as described later (i.e., this group included both tumor *EGFR* M+ and M− patients).

In the cfDNA M+ subgroup, ORR was not significantly different in the gefitinib group compared with carboplatin/paclitaxel treatment (75.0% [18/24] and 63.6% [14/22], respectively; odds ratio [OR], 1.71; 95% CI, 0.48–6.09; p = 0.40). In the cfDNA M− subgroup, there were no significant differences in ORR with gefitinib compared with carboplatin/paclitaxel (27.1% [19/70] and 21.8% [17/78], respectively; OR, 1.34; 95% CI, 0.63–2.84; p = 0.45) (Figure 4). Again, this subgroup included both tumor *EGFR* M+ and M− patients as described later.

The results for clinical outcome by *EGFR* mutation status (M+, M−) for the Japanese subset of patients with known tumor *EGFR* mutation status (n = 91) are included in Supplemental Digital Content 2 (Results http://links.lww.com/JTO/A153).

Comparison of *EGFR* Mutation Status in Pretreatment cfDNA and Tumor Tissue

A total of 108 patients had a known mutation result by cfDNA but not by tumor; 5 patients had a known mutation

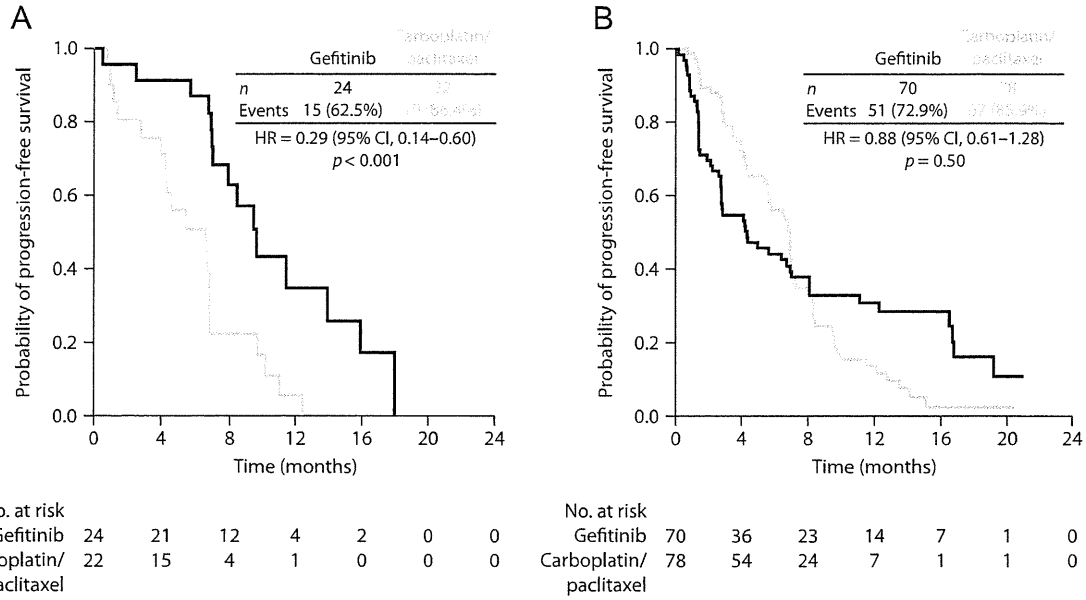

FIGURE 3. Kaplan-Meier curves of progression-free survival in cfDNA *EGFR* mutation-positive (*A*) and cfDNA *EGFR* mutation-negative (*B*) patients in the Japanese subset of IPASS. HR <1 indicates a difference in favor of gefitinib. CI, confidence interval; cfDNA, circulating free DNA; *EGFR*, epidermal growth factor receptor; HR, hazard ratio.

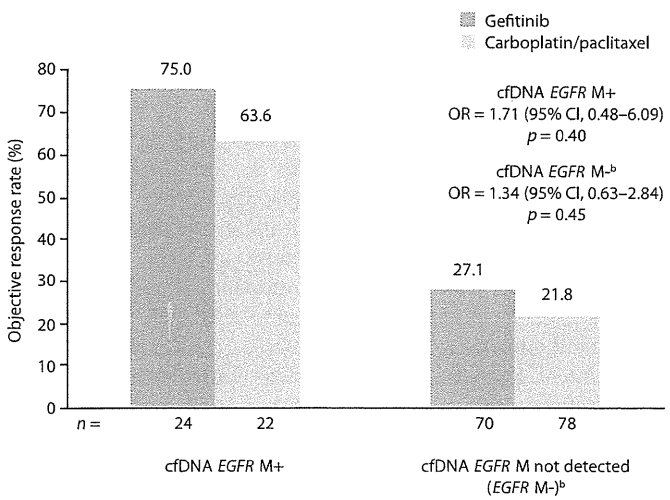

FIGURE 4. Objective response rates by treatment and by cfDNA (serum) *EGFR* mutation status (Japanese ITT population[a]). [a]Refers to the country of recruitment and not necessarily to racial origin. [b]There was a high rate of false-negative results, i.e., this group included both tumor *EGFR* M+ and M− patients. OR >1 implies a greater chance of response on gefitinib. OR, CI, and *p* values from logistic regression. cfDNA, circulating free DNA; CI, confidence interval; *EGFR*, epidermal growth factor receptor; ITT, intent-to-treat; M+, mutation-positive; M−, mutation-negative; OR, odds ratio.

result by tumor but not cfDNA (no serum sample provided); and 86 patients had a known mutation status by both tumor and cfDNA.

Of the 86 patients who had a known tumor and cfDNA mutation status, no false positives were identified (i.e., no samples were tumor M− but cfDNA M+). All 22 patients identified as cfDNA *EGFR* M+ were tumor *EGFR* M+, i.e., the positive predictive value was 100% (all samples that were cfDNA M+ were tumor M+) and the specificity was 100% (all samples that were tumor M− were cfDNA M−) (Table 2). However, the rate of false negatives was high: 29/51 (56.9%) of patients identified as tumor *EGFR* M+ were cfDNA *EGFR* M− (Table 2).

TABLE 2. Comparison of *EGFR* Mutation Status in cfDNA and Tumor Samples in 86 Patients with a Known *EGFR* Mutation Status Using Both Methods (Japanese[a] ITT Population)

	Mutation Status (Tumor Tissue), *n*		
	M+	M−	Total
Mutation status (cfDNA), *n*			
M+	22	0	22
M−	29	35	64
Total	51	35	86

Sensitivity = 43.1% (22 cfDNA M+ out of 51 tumor M+).[b]
Specificity = 100% (all 35 tumor M- were cfDNA M−).[b]
Positive predictive value = 100% (all 22 cfDNA M+ were tumor M+).[b]
Negative predictive value = 54.7% (35 tumor M- out of 64 cfDNA M−).[b]
Concordance = 66.3% (cfDNA and tumor results agreed in 57 of 86 cases).[b,c]

[a] Refers to the country of recruitment and not necessarily to racial origin.
[b] Those with a known *EGFR* mutation status using both methods.
[c] Kappa coefficient 0.38 (95% CI, 0.24–0.53).
cfDNA, circulating free DNA; CI, confidence interval; *EGFR*, epidermal growth factor receptor; ITT, intent-to-treat; M+, mutation positive; M−, mutation negative.

EGFR Mutation Types in Pretreatment cfDNA and Tumor Tissue

Of the patients classified as *EGFR* M+ at pretreatment by both tumor and cfDNA, all had the same mutation type in

TABLE 3. *EGFR* Mutations in Pretreatment cfDNA vs. Tumor Samples (Japanese[a] ITT Population)

	Tumor *EGFR* Mutation[b]						
cfDNA *EGFR* Mutation	**Exon 19 Deletions Only**	**Exon 20 T790M Only**	**Exon 21 L858R Only**	**Exon 20 T790M and Exon 21 L858R**	**Negative**	**Unknown**	**Total**
Exon 19 deletions only	11	0	0	0	0	15	26
Exon 20 T790M only	0	0	0	1	0	1	2
Exon 21 L858R only	0	0	10	0	0	8	18
Exon 20 T790M and exon 21 L858R	0	0	0	0	0	0	0
Negative	18	0	11	0	35	84	148
Unknown	2	1	0	0	2	34	39
Total	31	1	21	1	37	142	233

The categories are mutually exclusive. The categories "Exon 19 deletions and exon 20 T790M" and "Exon 19 deletions and exon 21 L858R" were 0 for both tumor and cfDNA and have been omitted from the table.

[a] Refers to the country of recruitment and not necessarily to racial origin.

[b] Mutations that were tested in tumor tissue samples but not serum included: exon 20 insertion, exon 21 L861Q, exon 18 G719X, and exon 20 S768I. Two patients with tumor samples had these mutations (1 with exon 20 insertion and 1 with exon 21 L861Q). These patients were excluded from the comparative analysis of mutation detection by sample type.

cfDNA, circulating free DNA; *EGFR*, epidermal growth factor receptor; ITT, intent-to-treat.

tumor and cfDNA except one patient who had exon 20 T790M and exon 21 L858R by tumor but exon 20 T790M only by cfDNA (Table 3).

DISCUSSION

The feasibility of using cfDNA to detect *EGFR* mutations was assessed in the Japanese subset of patients from the IPASS study. The proportion of patients identified as *EGFR* M+ was lower when assessed in cfDNA (23.7%) compared with tumor tissue (61.5%). Although cfDNA results identified no false positives, a high rate of false negatives (56.9%) was observed, with more than half of the tumor M+ patients not detected by cfDNA testing (of patients with evaluable mutation status from both cfDNA and tumor). Further research into appropriate methods and analysis needs to be performed before it could be accepted as an option in the diagnostic or screening setting. If larger patient series confirmed the absence of false-positive results and demonstrated an improvement or lowering of false-negative results, serum testing may prove useful for patients for whom tumor samples are not available.

Testing of biopsied tumor tissue remains the current recommended method for *EGFR* mutation analysis.[8] However, tumor tissue is often difficult to obtain, particularly from patients with advanced non-small cell lung cancer (NSCLC), and a lack of tumor cells in a given sample and subsequently failure on pathological examination can make *EGFR* mutation analysis very difficult. The increased recognition of the relevance of mutation testing to treatment selection may stimulate efforts to better obtain tissue for *EGFR* mutation testing in the future. In the meantime, detection of *EGFR* mutation status in cfDNA derived from serum/plasma may allow patients without diagnostic tumor material the opportunity to benefit from personalized treatment and also has a use in the clinical trial setting where tumor material is not always available.

Although minimally invasive, the use of serum as a nontumor surrogate sample may be limited by the amount of cfDNA available in the sample, meaning that some positive samples are not detected. In addition, some patients may not have cf tumor DNA as their tumors may not be releasing this material into the bloodstream, giving rise to false-negative results. Because of the limited yields of cfDNA obtained from serum, two changes (in addition to duplicate tests) were made to the *EGFR* mutation ARMS kit used to detect *EGFR* mutations in this study: an increase in the number of PCR cycles and an alteration of the cutoffs used to define M+ samples (dCt values). Further analysis is underway to investigate whether these conditions are the most appropriate and whether less stringent settings could result in more true positives (fewer false negatives) while retaining no false positives.

There have been several reports on the detection of cfDNA *EGFR* mutation status using different methods. A significant correlation between cfDNA *EGFR* mutation status and clinical response to gefitinib was found in two previous small studies that assessed cfDNA *EGFR* mutation status using the ARMS method of detection, a highly sensitive (1% sensitive) targeted technique to detect specific known *EGFR* mutations.[9,11] Other screening techniques detect all *EGFR* mutations, known and novel variants, by PCR amplification followed by sequencing, pyrosequencing, or melt analysis (10–30% sensitivity).[8] However, although these methods are widely used for *EGFR* mutation analysis of DNA derived from tumor tissue, not all of these methods have demonstrated utility for *EGFR* mutation analysis of cfDNA. In a small study that used DNA sequencing to detect *EGFR* mutations in serum, mutations were more frequently observed in patients experiencing partial response or stable disease compared with those whose disease progressed, although the difference did not reach statistical significance.[10] No statistically significant association between cfDNA *EGFR* mutation status and PFS by multivariate analysis (HR, 1.48; 95% CI, 0.93–2.36; $p = 0.09$) was found in the study by Rosell et al.[12] which assessed *EGFR* mutations by PCR-based methods in the presence of a protein nucleic acid (PNA) clamp in the cfDNA extracted from serum of 164 patients

treated with erlotinib. In another study that used denaturing high-performance liquid chromatography to analyze for mutations in exons 19 and 21 from matched plasma and tumor samples, patients with plasma *EGFR* mutations had significantly higher ORR and prolonged PFS.[7] The present study using ARMS demonstrated that the treatment effect for the Japanese cfDNA *EGFR* M+ subgroup followed the same pattern as the tumor *EGFR* M+ subgroup of the overall IPASS population (i.e., PFS HR significantly in favor of gefitinib and higher ORR with gefitinib versus carboplatin/paclitaxel).[6] There was a significant interaction between cfDNA *EGFR* mutation status and treatment for PFS.

Any variance in concordance rates for mutation results between pretreatment serum versus tumor tissue (66.3% in our study and between 58 and 93% in previously reported studies)[7,9–11] may be attributed to different methods of extraction, detection, run conditions, the size and yield of the DNA fragments, and the fact that cfDNA may not be present in the circulation of all patients with NSCLC. For example, targeted sequences amplified by ARMS are short, at 100–150 bp, leading to decreased assay failure rates (particularly from formalin-fixed paraffin-embedded material or fragments of cfDNA) compared with sequencing methods, which tend to involve the amplification of longer target sequences of 150–250 bp or above.[8,13,14,17,18]

In patients who were cfDNA *EGFR* M− in this study, no significant difference for PFS was seen with gefitinib compared with carboplatin/paclitaxel; however, the HR was not constant over time (as was observed for the overall Japanese study population). These results should be interpreted with caution as there was a high rate of false negatives, and this subgroup is likely to include tumor *EGFR* M+ and M− patients.

In conclusion, these results merit further investigation to determine whether alternative samples, including serum or plasma, may be considered for determining *EGFR* mutation status in future, particularly in cases where diagnostic tumor material is not available. Currently, analysis of tumor material is the recommended method for determining *EGFR* mutation status.

ACKNOWLEDGMENTS

Supported by AstraZeneca.

The authors thank the patients and investigators for their participation in this study and Annette Smith, PhD, from Complete Medical Communications, who provided medical writing support funded by AstraZeneca.

REFERENCES

1. Modjtahedi H, Essapen S. Epidermal growth factor receptor inhibitors in cancer treatment: advances, challenges and opportunities. *Anticancer Drugs* 2009;20:851–855.
2. 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.
3. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500.
4. 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 USA* 2004;101:13306–13311.
5. Mitsudomi T, Kosaka T, Yatabe Y. Biological and clinical implications of EGFR mutations in lung cancer. *Int J Clin Oncol* 2006;11:190–198.
6. Mok TS, Wu Y-L, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;361:947–957.
7. Bai H, Mao L, Wang HS, et al. Epidermal growth factor receptor mutations in plasma DNA samples predict tumor response in Chinese patients with stages IIIB to IV non-small-cell lung cancer. *J Clin Oncol* 2009;27:2653–2659.
8. Eberhard DA, Giaccone G, Johnson BE. Biomarkers of response to epidermal growth factor receptor inhibitors in Non-Small-Cell Lung Cancer Working Group: standardization for use in the clinical trial setting. *J Clin Oncol* 2008;26:983–994.
9. Kimura H, Kasahara K, Kawaishi M, et al. Detection of epidermal growth factor receptor mutations in serum as a predictor of the response to gefitinib in patients with non-small-cell lung cancer. *Clin Cancer Res* 2006;12:3915–3921.
10. Kimura H, Kasahara K, Shibata K, et al. EGFR mutation of tumor and serum in gefitinib-treated patients with chemotherapy-naive non-small cell lung cancer. *J Thorac Oncol* 2006;1:260–267.
11. Kimura H, Suminoe M, Kasahara K, et al. Evaluation of epidermal growth factor receptor mutation status in serum DNA as a predictor of response to gefitinib (IRESSA). *Br J Cancer* 2007;97:778–784.
12. Rosell R, Moran T, Queralt C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* 2009;361:958–967.
13. Newton CR, Graham A, Heptinstall LE, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 1989;17:2503–2516.
14. Whitcombe D, Theaker J, Guy SP, et al. Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol* 1999;17:804–807.
15. Ohe Y, Ichinose Y, Nishiwaki Y, et al. Phase III, randomized, open-label, first-line study of gefitinib vs carboplatin/paclitaxel in selected patients with advanced non-small cell lung cancer (IPASS): evaluation of recruits in Japan. Poster 8044 presented at the ASCO Annual Meeting, Orlando, FL, USA, May 29 to June 2, 2009.
16. Ichinose Y, Nishiwaki Y, Ohe Y, et al. Analyses of Japanese patients recruited in IPASS, a phase III, randomized, open-label, first-line study of gefitinib vs carboplatin/paclitaxel in selected patients with advanced non-small cell lung cancer. Poster PD3.1.3 presented at the 13th World Conference on Lung Cancer, International Association for the Study of Lung Cancer, San Francisco, CA, USA, July 31 to August 4, 2009.
17. Board RE, Thelwell NJ, Ravetto PF, et al. Multiplexed assays for detection of mutations in PIK3CA. *Clin Chem* 2008;54:757–760.
18. Holland PM, Abramson RD, Watson R, et al. Detection of specific polymerase chain reaction product by utilizing the 5′—3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci USA* 1991;88:7276–7280.

ROS1-Rearranged Lung Cancer

A Clinicopathologic and Molecular Study of 15 Surgical Cases

Akihiko Yoshida, MD, PhD, Takashi Kohno, PhD,† Koji Tsuta, MD, PhD,* Susumu Wakai, MT,* Yasuhito Arai, PhD,‡ Yoko Shimada, MSc,† Hisao Asamura, MD, PhD,§ Koh Furuta, MD, PhD,* Tatsuhiro Shibata, MD, PhD,‡ and Hitoshi Tsuda, MD, PhD**

Abstract: Recent discovery of *ROS1* gene fusion in a subset of lung cancers has raised clinical interest, because *ROS1* fusion–positive cancers are reportedly sensitive to kinase inhibitors. To better understand these tumors, we examined 799 surgically resected non–small cell lung cancers by reverse transcriptase polymerase chain reaction and identified 15 tumors harboring *ROS1* fusion transcripts (2.5% of adenocarcinomas). The most frequent fusion partner was *CD74* followed by *EZR*. The affected patients were often younger nonsmoking female individuals, and they had overall survival rates similar to those of the *ROS1* fusion–negative cancer patients. All the *ROS1* fusion–positive tumors were adenocarcinomas except 1, which was an adenosquamous carcinoma. Histologic examination identified an at least focal presence of either solid growth with signet-ring cells or cribriform architecture with abundant extracellular mucus in 53% of the cases. These 2 patterns are reportedly also characteristic of *anaplastic lymphoma kinase* (*ALK*)-rearranged lung cancers, and our data suggest a phenotypic resemblance between the *ROS1*-rearranged and *ALK*-rearranged tumors. All tumors except 1 were immunoreactive to thyroid transcription factor-1. Fluorescence in situ hybridization using *ROS1* break-apart probes revealed positive rearrangement signals in 23% to 93% of the tumor cells in *ROS1* fusion–positive cancers, which were readily distinguished using a 15% cutoff value from 50 *ROS1* fusion–negative tumors tested, which showed 0% to 6% rearrangement signals. However, this perfect test performance was achieved only when isolated 3′ signals were included along with classic split signals in the definition of rearrangement positivity. Fluorescence in situ hybridization signal patterns were unrelated to 5′ fusion partner genes. All *ROS1* fusion–positive tumors lacked alteration of *EGFR*, *KRAS*, *HER2*, *ALK*, and *RET* genes.

Key Words: ROS1, lung, carcinoma, diagnosis

(*Am J Surg Pathol* 2013;00:000–000)

Lung cancer is the leading cause of cancer-related deaths worldwide, and it is responsible for 1.38 million mortalities per year.[1] The most common type, adenocarcinoma, has recently been undergoing extensive molecular subclassification with each subtype being characterized by distinct epidemiological, histologic, and immunohistochemical profiles.[2–6] Molecular data can have a significant clinical impact, because genetic changes dictate clinical activity of targeted therapy. For example, cancers with a mutation in epidermal growth factor receptor (*EGFR*) respond to EGFR tyrosine kinase inhibitors (eg, erlotinib and gefitinib),[7] and those with anaplastic lymphoma kinase (*ALK*) gene rearrangement are excellent candidates for ALK inhibitor (eg, crizotinib) therapy.[8]

In 2007, Rikova et al[9] identified *ROS1* gene rearrangement in a cell line and a clinical sample of lung adenocarcinoma, and this observation was confirmed by subsequent investigators who identified a larger number of such cases.[10–14] ROS1 is an orphan receptor tyrosine kinase that is phylogenetically related, albeit remotely, to ALK.[15,16] ROS1 is normally expressed in the lung and other organs, although its physiological function remains unclear.[15] *ROS1* rearrangement has also been reported in glioblastoma cell lines,[17] cholangiocarcinoma,[18] and ovarian serous tumor of low malignant potential.[19] In lung cancer, several *ROS1* fusion partners have been discovered including *CD74*, solute carrier family 34 (sodium phosphate), member 2 (*SLC34A2*), ezrin (*EZR*), leucine-rich repeats and immunoglobulin-like domains 3 (*LRIG3*), syndecan 4 (*SDC4*), tropomyosin 3 (*TPM3*), and fused in glioblastoma (*FIG*).[9,10,12–14] In all these fusions, the *ROS1* gene is disintegrated near the sequence coding the transmembrane domain, and the 3′ kinase

From the *Division of Pathology and Clinical Laboratories; §Division of Thoracic Surgery, National Cancer Center Hospital; †Division of Genome Biology; and ‡Division of Cancer Genomics, Center for Medical Genomics, National Cancer Center Research Institute, Tokyo, Japan.

Conflicts of Interest and Source of Funding: Supported in part by the Program for Promotion of Fundamental Studies in Health Sciences from the National Institute of Biomedical Innovation (NIBIO), Grants-in-Aid from the Ministry of Health, Labour and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control, and National Cancer Center Research and Development Fund. The National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund, Japan.

Correspondence: Akihiko Yoshida, MD, PhD, Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan (e-mail: akyoshid@ncc.go.jp).

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website (www.ajsp.com).

domain–coding component is fused to the 5′ part of the respective partner gene. The fusion results in a chimeric protein, which demonstrates constitutive phosphorylation of the kinase that activates intracellular signal cascades and exerts an oncogenic effect. All these *ROS1* fusion transcripts were proven to have transforming capabilities.[13,18] Significantly, in vitro data suggested that *ROS1*-rearranged cancers respond to ALK inhibitors,[10,12,14] and a recent clinical trial[20] revealed a marked crizotinib activity in this molecular subclass. These data highlight the importance of identifying *ROS1*-rearranged cancers to optimally tailor molecular-based treatments.

As *ROS1*-rearranged lung cancers are rare, they still remain poorly characterized aside from demographic features and fusion types. In particular, the histology of this type of cancers has not been widely investigated, and it is unclear whether this subset is associated with any characteristic morphologic appearance. In addition, although fluorescence in situ hybridization (FISH) has been used in the diagnosis of these cancers,[10,13,14] the actual FISH signal patterns have rarely been detailed, and the accuracy of the modality has been only incompletely validated against reverse transcriptase polymerase chain reaction (RT-PCR). Therefore, we conducted a clinicopathologic study of 15 lung cancers that carried RT-PCR-proven *ROS1* fusion transcripts, with a particular view to clarifying their histologic features using well-sampled surgical materials. A detailed FISH analysis was also performed, and the data were correlated with the RT-PCR results to better understand the biology of these tumors and to establish accurate FISH diagnostic criteria.

MATERIALS AND METHODS

Case Selection

This study was approved by the institutional review board of the National Cancer Center, Tokyo, Japan. Samples of 799 non–small cell lung carcinomas that were frozen immediately after surgical excision were provided by the National Cancer Center Biobank, Japan. The series comprised 569 adenocarcinomas and 230 squamous cell carcinomas according to the World Health Organization (WHO) classification system.[21] All the samples were investigated by RT-PCR for the known *ROS1* fusion genes (*CD74-ROS1*, *EZR-ROS1*, *SLC34A2-ROS1*, *FIG-ROS1*, *LRIG3-ROS1*, *SDC4-ROS1*, and *TPM3-ROS1*). In brief, total RNA (500 ng) was reverse transcribed to cDNA using Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA). cDNA (corresponding to 10 ng total RNA) or 10 ng genomic DNA was subjected to PCR amplification using KAPA Taq DNA Polymerase (KAPA Biosystems, Woburn, MA). The reactions were carried out in a thermal cycler under the following conditions: 40 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 2 minutes, with a final extension for 10 minutes at 72°C. The gene encoding glyceraldehyde-3-phosphate dehydrogenase was amplified to estimate the efficiency of cDNA synthesis. PCR products were subjected to agarose gel electrophoresis. cDNA of the HCC78 lung cancer cell line with an *SLC34A2-ROS1* fusion[9] was used as a positive control. When visible bands were detected, the PCR products were subjected to Sanger sequencing. The PCR primers used are listed in Supplementary Table 1 (Supplemental Digital Content, http://links.lww.com/PAS/A160).

The foregoing examination identified 14 adenocarcinomas that harbored *ROS1* fusion genes (2.5% of 569 adenocarcinomas). Nine cases had *CD74-ROS1* (C6;R34), 4 had *EZR-ROS1* (E10;R34), and 1 had *SLC34A2-ROS1* (S13del2046;R34). In addition, 1 adenosquamous carcinoma positive for *CD74-ROS1* (C6;R34) was identified by examining a separate small set of clinical samples and was also included in the analysis. No tumors carried the other *ROS1* fusion types tested, and none of the squamous cell carcinomas tested showed *ROS1* fusion. Among 555 *ROS1* fusion–negative adenocarcinomas, 50 tumors were randomly selected to serve as controls to test diagnostic performance of the FISH assay.

All *ROS1* fusion–positive and *ROS1* fusion–negative tumors were assembled into tissue microarrays (TMA) using duplicate 2.0-mm cores sampled from 2 different representative areas of each tumor (Azumaya, Tokyo, Japan). Immunohistochemistry (IHC) and FISH assays were performed on the TMA, except for 1 fusion-positive tumor that was evaluated using the whole section because of limited tissue availability.

Clinicopathologic Analysis

Clinical data including age, sex, smoking history, stage,[22] and overall survival were retrieved from the patients' medical records. Hematoxylin and eosin-stained slides were available for all of the 15 cases with *ROS1* gene fusion. The number of tumor slides available was between 2 and 11 (median, 5). All tumors measuring 3 cm or less in diameter were submitted in their entirety, and larger tumors were sampled extensively. One or more Elastica van Gieson stains were available for all of the cases to facilitate the detection of vascular or pleural invasion. All the glass slides were reviewed to evaluate the following parameters: tumor size, predominant growth pattern based on the WHO description[21] with a recent modification (lepidic, papillary, micropapillary, acinar, or solid),[2,23] any existing growth pattern, degree of differentiation, nuclear features, lymphovascular invasion, pleural invasion, intrapulmonary metastasis, necrosis, and any coexisting tumor histotype other than adenocarcinoma. In addition, to determine whether *ROS1*-rearranged lung cancer has any histologic similarity to *ALK*-rearranged lung cancers, as alluded in a few reported cases,[12,13] we paid particular attention to the presence or absence of a solid growth pattern containing signet-ring cells (solid signet-ring cell pattern) and a cribriform structure associated with abundant extracellular mucus (mucinous cribriform pattern), 2 unique features characteristic of *ALK*-rearranged lung cancer.[5,13,24–26] Background uninvolved lung parenchyma and metastatic tumor tissues concurrently or subsequently obtained from the lymph nodes and/or pleura were also reviewed when available.

IHC for Thyroid Transcription Factor-1

Four-micrometer-thick sections were deparaffinized. Heat-induced epitope retrieval was performed with targeted retrieval solution (pH 9) (Dako, Carpinteria, CA). The slides were treated with 3% hydrogen peroxide for 20 minutes to block endogenous peroxidase activity. The slides were then incubated with a primary antibody against TTF-1 (1:100, 8G7G3/1; Dako) for 1 hour. Immunoreactions were detected using the Envision-Plus system (Dako). The reactions were visualized with 3,3′-diaminobenzidine, followed by counterstaining with hematoxylin. Appropriate positive and negative controls were used, and nuclear staining of $\geq 10\%$ of tumor cells was considered positive.

Fluorescence In Situ Hybridization

FISH assays were performed using a *ROS1* break-apart probe set (Chromosome Science Labo Inc., Sapporo, Japan), which hybridizes with the neighboring 5′ centromeric (RP11-835I21, labeled with Spectrum Green) and 3′ telomeric (RP11-1036C2, labeled with Spectrum Red) sequences of the *ROS1* gene. FISH images were captured using the Metafer Slide Scanning Platform (MetaSystems, Altlussheim, Germany) to facilitate analysis. One-hundred nonoverlapping tumor cells with at least 1 each of 5′ and 3′ signals, whether fused or separated, were examined for each core, and a detailed signal pattern was recorded. A fused 5′/3′ signal may appear yellow because of colocalization of green (5′) and red (3′) signals. A split signal was defined by 5′ and 3′ probes observed at a distance >1 time the signal size, and signals separated by less than that distance were regarded as fused signals. The rearrangement-positive cells were defined as having any split signal or any isolated red (3′) signal. The rate of rearrangement-positive cells was calculated for each case and compared between RT-PCR-proven *ROS1* fusion–positive and *ROS1* fusion–negative tumors. For *ROS1* fusion–positive cases, the rate of rearrangement-positive cells was internally compared between 2 TMA cores to address the intratumoral distribution of gene rearrangement.

Status of Other Driver Genes

ROS1 fusion–positive lung cancers were also analyzed by PCR for the mutation status of the *EGFR*, *KRAS*, and *human epidermal growth factor receptor type 2* (*HER2*) genes, and the rearrangement status of *ALK* and *RET*, according to methods previously described elsewhere.[6]

Statistical Analyses

All data analyses were performed using SPSS version 20.0 (IBM Corporation, Somers, NY). The Fisher exact test was used for categorical data, and the Mann-Whitney *U* test was used for continuous data. Overall survival, measured from the date of surgery, was determined using the Kaplan-Meier method, and survival difference was compared using the log rank test. All *P* values were 2-tailed, and $P < 0.05$ was considered significant.

RESULTS

Clinical Characteristics

The clinical data of 15 *ROS1* fusion–positive lung cancers are summarized in Table 1 and detailed in Table 2. The patients included 3 men and 12 women with a median age of 59 years (range, 37 to 68 y). Twelve patients were never smokers, 1 was a light smoker (< 20 pack-years), and 2 were heavy smokers ($\geq$ 20 pack-years). Five tumors were found to be at stage I, 2 were at stage II, and 8 were at stage III. Thirteen tumors were treated by standard lobectomy and lymph node dissection, and 1 each of the remaining cases was treated by wedge resection and total pneumonectomy. None of the patients received crizotinib. Follow-up data were available for all 15 patients for a median of 2034 days (range, 236 to 4043 d). At the last follow-up, 5 patients were alive with no evidence of disease, 6 patients were alive with recurrent tumors, and the remaining 4 patients died of the disease. The estimated 5-year overall survival rate was 69.2%. When compared with 555 surgically resected lung adenocarcinomas that lacked a *ROS1* fusion gene (Table 1), *ROS1* fusion–positive lung cancers were significantly more common in never smokers ($P = 0.024$) and women ($P = 0.008$) and tended to occur in younger patients ($P = 0.078$). The tumor stage was not different ($P = 0.54$) between *ROS1* fusion–positive and *ROS1* fusion–negative patients. Follow-up data were available for all 555 *ROS1* fusion–negative patients for a median of 1630 days (range, 10 to 5438 d), and their overall survival was not different from that of *ROS1* fusion–positive patients ($P = 0.99$, Fig. 1).

Histologic Characteristics

Histologic findings are summarized in Table 2. The median tumor size was 2.5 cm (range, 1.6 to 6.5 cm). All of the 15 *ROS1*-rearranged tumors harbored an adenocarcinoma component. The predominant growth

TABLE 1. Comparison of Clinical Data between Patients with *ROS1* Fusion–Positive and *ROS1* Fusion–Negative Lung Cancers

Clinical Characteristics	*ROS1* Fusion Positive (N = 15)	*ROS1* Fusion Negative (N = 555)	*P*
Age (y)			0.078
Median	59	61	
Range	37-68	28-88	
Sex			0.008
Male	3	305	
Female	12	250	
Smoking (pack-years)			0.024
0	12	246	
< 20	1	69	
≥ 20	2	237	
Stage			0.54
I	5	260	
II	2	77	
III	8	216	
IV	0	2	

Smoking history was unavailable for 3 *ROS1* fusion–negative patients.

pattern was lepidic in 3 cases, papillary in 5 cases, micropapillary in 1 case, acinar in 2 cases, and solid in 4 cases. At least focal lepidic, papillary, micropapillary, acinar, and solid patterns were identified in 8, 10, 11, 12, and 11 cases, respectively. Lepidic growth accounted for < 5% of the total tumor volume in 4 of 8 cases. Two cases were well differentiated, 9 were moderately differentiated, and 4 were poorly differentiated. Tumor cell nuclei were low grade in 7 cases, intermediate grade in 7 cases, and high grade in 1 case. Twelve cases exhibited a relatively monotonous nuclear appearance, whereas 3 showed at least focal nuclear pleomorphism. Pleural invasion, lymphovascular invasion, intrapulmonary metastasis, and necrosis were seen in 7, 9, 2, and 7 cases, respectively. A solid growth pattern containing signet-ring cells (solid signet-ring cell pattern) was at least focally present in 5 cases (Figs. 2A, B). The amount of solid signet-ring cell patterns ranged from < 5% to > 95% of the tumor volume. A cribriform structure associated with abundant extracellular mucus (mucinous cribriform pattern) was identified at least focally in 5 cases (Figs. 2C, D) and accounted for < 5% to 80% of the tumor volume. Two cases showed both patterns, and 8 cases (53%) showed either or both of them constituting < 5% to 100% of the tumor volume (mean, 33%; median, 10%). These patterns were present in 5 of 10 *CD74-ROS1*-positive tumors and 3 of 4 *EZR-ROS1*-positive tumors and were also seen in 2 of 10 metastatic tumor tissues available for review.

Overall, histologic findings did not predict *ROS1* fusion partners, although *EZR-ROS1*-positive tumors seemed less well differentiated than those with 2 other fusion types. One case (case 15) focally exhibited unequivocal squamous differentiation, immunopositive for cytokeratin 5/6 and p40, comprising 10% of the total tumor volume (Fig. 2E), thus meeting the WHO criteria of adenosquamous carcinoma.[21] Other notable findings included non–signet-ring mucous (goblet) cells in 1 case associated with a mucinous cribriform pattern (Fig. 2D) and psammomatous calcifications in 5 cases. Background uninvolved lung parenchyma was well sampled in all cases with a median of 6 slides per specimen, and none of them harbored atypical adenomatous hyperplasia. One patient (case 14) underwent a simultaneous resection of a 2.2-cm well-differentiated invasive adenocarcinoma in the contralateral lung. That lepidic-predominant tumor lacked *ROS1* rearrangement as determined by FISH and was considered an independent primary.

Immunohistochemistry

TTF-1 was immunoreactive in 14 of 15 tumors, with most (93%) cases showing diffuse (> 50%) and at least moderate reactivity (Fig. 2F). One *EZR-ROS1*-positive carcinoma (case 8) was TTF-1 negative (0% positive), and the negativity was confirmed by staining the whole section.

FISH Analysis of *ROS1* Fusion–Positive Cancer

The results are summarized in Table 2. Although the FISH probe did not hybridize in 2 cases likely because of the old age of the blocks, the assay was successful in the remaining 13 cases. *ROS1* rearrangement signals (ie, splits of 5′ and 3′ signals, or isolated 3′ signals) ranged from 23% to 93% (median, 65%). All the cases showed > 50% rearrangement-positive cells, except for 2 cases

TABLE 2. Clinicopathologic and Molecular Details of 15 Patients with *ROS1* Fusion–Positive Lung Cancers

No.	Sex	Age	Smoking (pack-years)	Stage	Follow-up (mo)	Size (cm)	Diagnosis	Predominant Growth	Pattern S*	Pattern M†	Fusion Gene	FISH+ Cell Rate‡ (%)	FISH Pattern§
1	Female	55	0	IA	NED (66)	2.3	Ad	Lepidic	—	—	*CD74-ROS1*	63	FGR
2	Female	40	0	IA	AWD (118)	2.5	Ad	Papillary	—	—	*CD74-ROS1*	75	GR
3	Male	47	0	IIIB	AWD (8)	6.5	Ad	Papillary	—	—	*CD74-ROS1*	62	FR
4	Female	64	0	IIIA	DOD (29)	2.8	Ad	Solid	10%	—	*CD74-ROS1*	68	FGR
5	Female	59	0	IA	DOD (37)	2.0	Ad	Acinar	—	80%	*CD74-ROS1*	78	FR
6	Female	68	0	IA	AWD (75)	2.2	Ad	Lepidic	—	—	*CD74-ROS1*	N/A	N/A
7	Female	61	0	IIIA	AWD (95)	3.4	Ad	Papillary	—	10%	*CD74-ROS1*	93	GR
8	Female	54	0	IIB	NED (7)	3.4	Ad	Solid	> 95%	< 5%	*EZR-ROS1*	38	FGR‖
9	Female	60	30	IIIA	DOD (45)	1.6	Ad	Papillary	—	—	*EZR-ROS1*	N/A	N/A
10	Female	66	0	IA	NED (63)	2.5	Ad	Lepidic	—	5%	*CD74-ROS1*	65	FGR
11	Female	56	0	IIIA	DOD (31)	4.5	Ad	Solid	5%	—	*EZR-ROS1*	79	GR
12	Female	63	0	IIIB	NED (80)	3.0	Ad	Solid	50%	5%	*EZR-ROS1*	70	FR
13	Male	52	35	IIIB	NED (78)	3.4	Ad	Micropapillary	—	—	*SLC34A2-ROS1*	61	FR
14	Male	59	11	IIIA	AWD (67)	2.4	Ad	Papillary	—	—	*CD74-ROS1*	23	FGR‖
15	Female	37	0	IIA	AWD (132)	2.3	AdSq	Acinar	< 5%	—	*CD74-ROS1*	53	FGR

*The percentage of "solid signet-ring-cell pattern."
†The percentage of "mucinous cribriform pattern."
‡The percentage of tumor cells with rearrangement-positive FISH signals.
§Predominant FISH signal pattern in rearrangement-positive cells.
‖FGR pattern with narrow splits.
Ad indicates adenocarcinoma; AdSq, adenosquamous carcinoma; AWD, alive with disease; DOD, dead of disease; FGR, fused, green, and red signals; FR, fused and red signals without green signals; GR, green and red signals without fused signals; NA, data not available; NED, no evidence of disease.

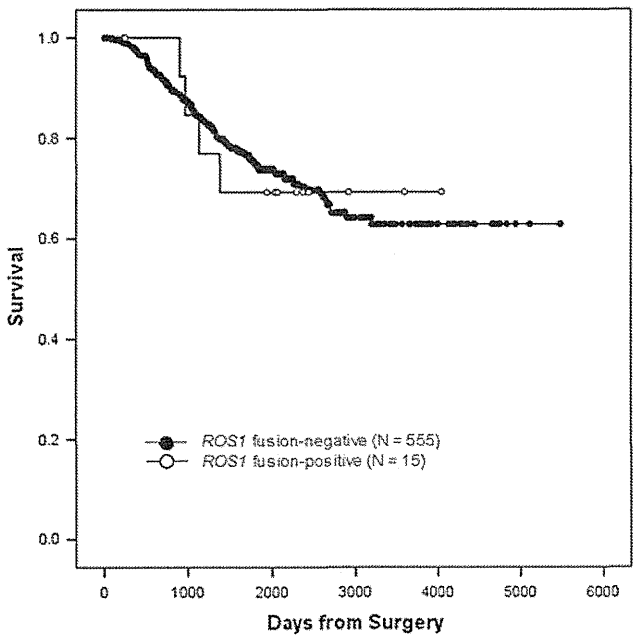

FIGURE 1. Kaplan-Meier analysis of overall survival of 15 *ROS1* fusion–positive carcinomas and 555 *ROS1* fusion–negative adenocarcinomas that were surgically resected. No significant difference in survival was observed between the 2 groups as determined by the log rank test ($P = 0.99$).

that showed a smaller number of positive cells (38% and 23%). The most common rearrangement signal pattern was a classic combination of fused, green, and red signals (FGR pattern) observed in 6 cases (Fig. 3A). A split was wide and easily discernible from a fused signal in all the cases, except for 2 in which it was narrow in most occasions and difficult to differentiate from a fused signal. Notably, these 2 cases corresponded to those that showed < 50% rearrangement signals (Fig. 3B). In 3 cases, fused 5′/3′ signals were lost in > 85% of the tumor cells that showed rearrangement signals, and cells with only split green and red signals (GR pattern) predominated (Fig. 3C). In 4 cases, isolated 5′ green signals were lost in > 85% of the tumor cells that showed rearrangement signals, and cells with fused and isolated 3′ red signals (FR pattern) predominated (Fig. 3D).

The *ROS1* gene fusion partner and rearrangement signal profile were not significantly associated. In 9 *CD74-ROS1*-positive cases, 5 showed an FGR pattern, 2 showed a GR pattern, and 2 showed an FR pattern. In 3 *EZR-ROS1*-positive cases, 1 case each showed an FGR, a GR, and an FR pattern. The only *SLC34A2-ROS1*-positive case showed an FR pattern. A low level of signal number gain (such as > 2 red signals in a cell) was uncommonly found in all cases tested, but a significant copy number increase up to 9 copies was seen in only 1 case. In 12 *ROS1*-positive cases, the FISH signal profiles were successfully compared between 2 TMA cores sampled from separate foci of a tumor. The rate of rearrangement-positive cells and the predominant signal pattern (ie, FGR, GR, or FR pattern) were comparable between the 2 cores.

Validation of FISH Diagnostic Criteria

Analysis of 50 *ROS1* fusion–negative tumors showed rearrangement signals in 0% to 6% of the cells (median, 2%). Therefore, when the cutoff value was set at 15%, *ROS1* fusion–positive and *ROS1* fusion–negative tumors were readily distinguishable with 100% sensitivity and specificity. To determine whether evaluating fewer cells influences test performance, a separate evaluation was made using 50 instead of 100 tumor cells for each case. Using this method, the rearrangement-positive cell rate in *ROS1* fusion–positive cancers (range, 26% to 92%; median, 66%) and *ROS1* fusion–negative cancers (range, 0% to 8%; median, 2%) were still patently different and could be readily distinguished with 100% sensitivity and specificity using the same cutoff value. Importantly, when the definition of rearrangement positivity were limited to split signals only, as adopted by some studies,[10,12] 4 *ROS1* fusion–positive tumors that predominantly showed an FR pattern were misclassified as *ROS1* fusion–negative, and the sensitivity decreased to 69%.

Status of Other Driver Genes

All of the 15 *ROS1* fusion–positive lung cancers lacked mutations in the *EGFR*, *KRAS*, and *HER2* genes and rearrangements in the *ALK* and *RET* genes.

DISCUSSIONS

Consistent with an earlier report,[10] our study showed that *ROS1*-rearranged lung cancer typically occurs in a younger age group with less smoking history. In addition, female patients were overrepresented in our cohort, and the skewed distribution was much more pronounced than in prior studies.[10,11,13] The tendency of *ROS1*-rearranged lung cancer to affect younger never smokers is reminiscent of *EGFR*-mutated[3] or *ALK*-rearranged lung cancers.[5,13,25] Similarly, *EGFR*-mutated tumors are known to affect women more frequently,[3] and large series of *ALK*-rearranged lung cancers also documented a slight predilection toward women.[5,13] *ROS1*-rearranged lung cancer thus seems to share demographic characteristics with 2 other molecular subtypes of lung cancer targetable by kinase inhibitors. No difference in tumor stage and overall survival was observed between *ROS1*-rearranged and wild-type cases in this exclusively surgical cohort, in agreement with a study that mainly analyzed nonsurgical patients.[10]

Although most *ROS1*-rearranged lung cancers reported were adenocarcinomas,[9–14] their histology has not been well characterized. Meanwhile, a small number of cases have been reported that demonstrate a unique histologic appearance for this type of cancer. Specifically, 3 *CD74-ROS1*-positive lung cancers in 1 series[13] were described as having a "mucinous cribriform pattern," and 1 *ROS1*-rearranged case in another series[12] was designated as a signet-ring cell type. The present study using well-sampled surgical materials confirmed these prior observations and found that 53% of *ROS1*-rearranged lung cancers at least focally exhibited a "solid signet-ring cell

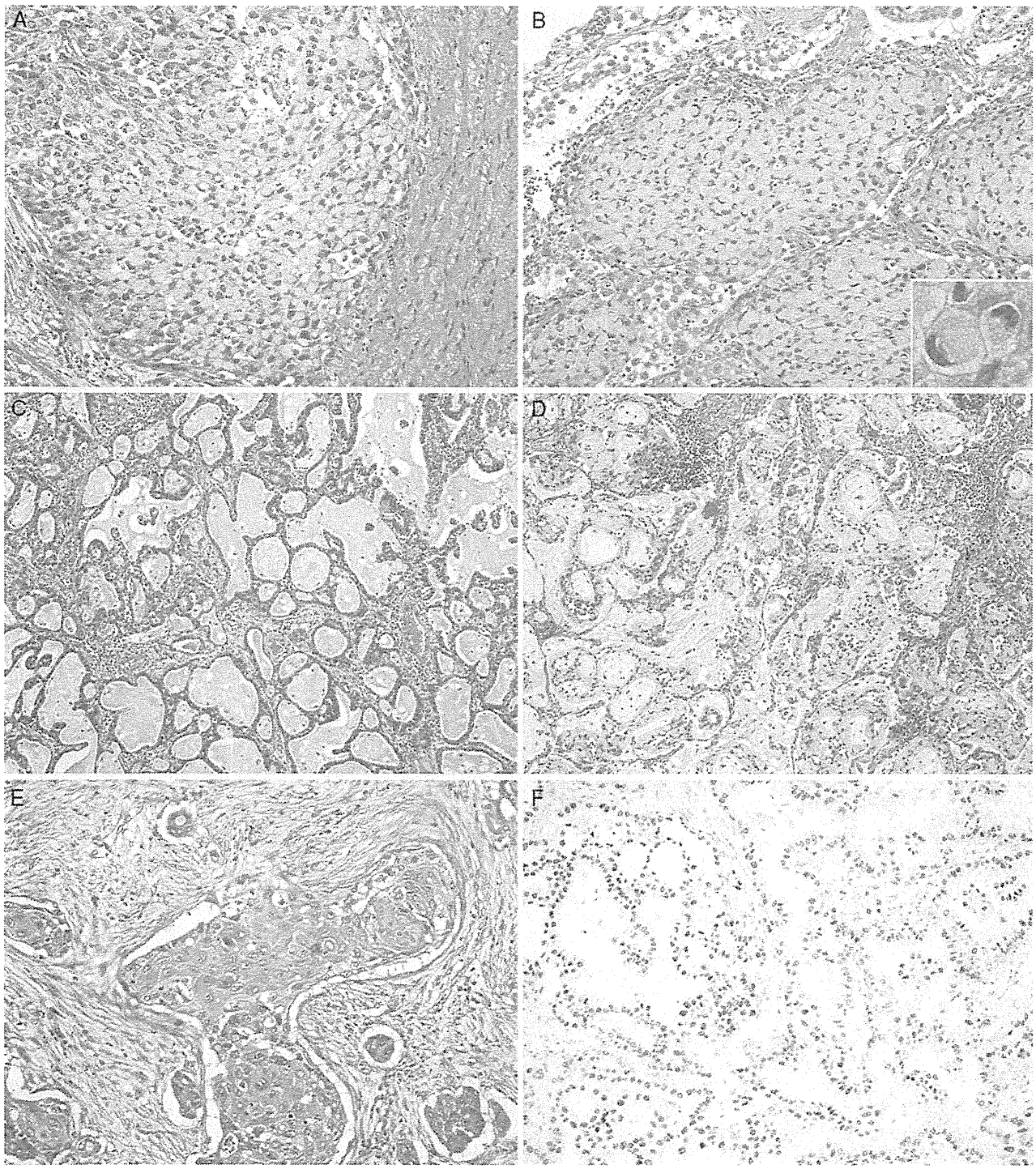

FIGURE 2. Histologic features of *ROS1* fusion–positive lung cancers. A and B, Five tumors showed solid growth containing signet-ring cells (solid signet-ring cell pattern) (A, case 4; B, case 8; H&E; inset shows a magnified view of signet-ring cells). C and D, Five tumors showed a cribriform structure associated with abundant extracellular mucus (mucinous cribriform pattern) (C, case 5; D, case 12; H&E). These 2 patterns were collectively seen in 8 cases (53%), whereas the remaining 7 cases exhibited nondescript histology without characteristic patterns (not shown). E, One tumor (case 15) showed focal squamous differentiation (H&E). F, Most tumors were diffusely positive for TTF-1 (case 13, TTF-1 IHC).

pattern" or "mucinous cribriform pattern." Notably, these 2 patterns were previously shown to be characteristic of *ALK*-rearranged lung cancers.[5,13,25,26] In our recent study,[5] these patterns were seen in 76% of 54 *ALK*-rearranged tumors but in only 1% of 100 consecutive *ALK* wild-type tumors, making them the most powerful

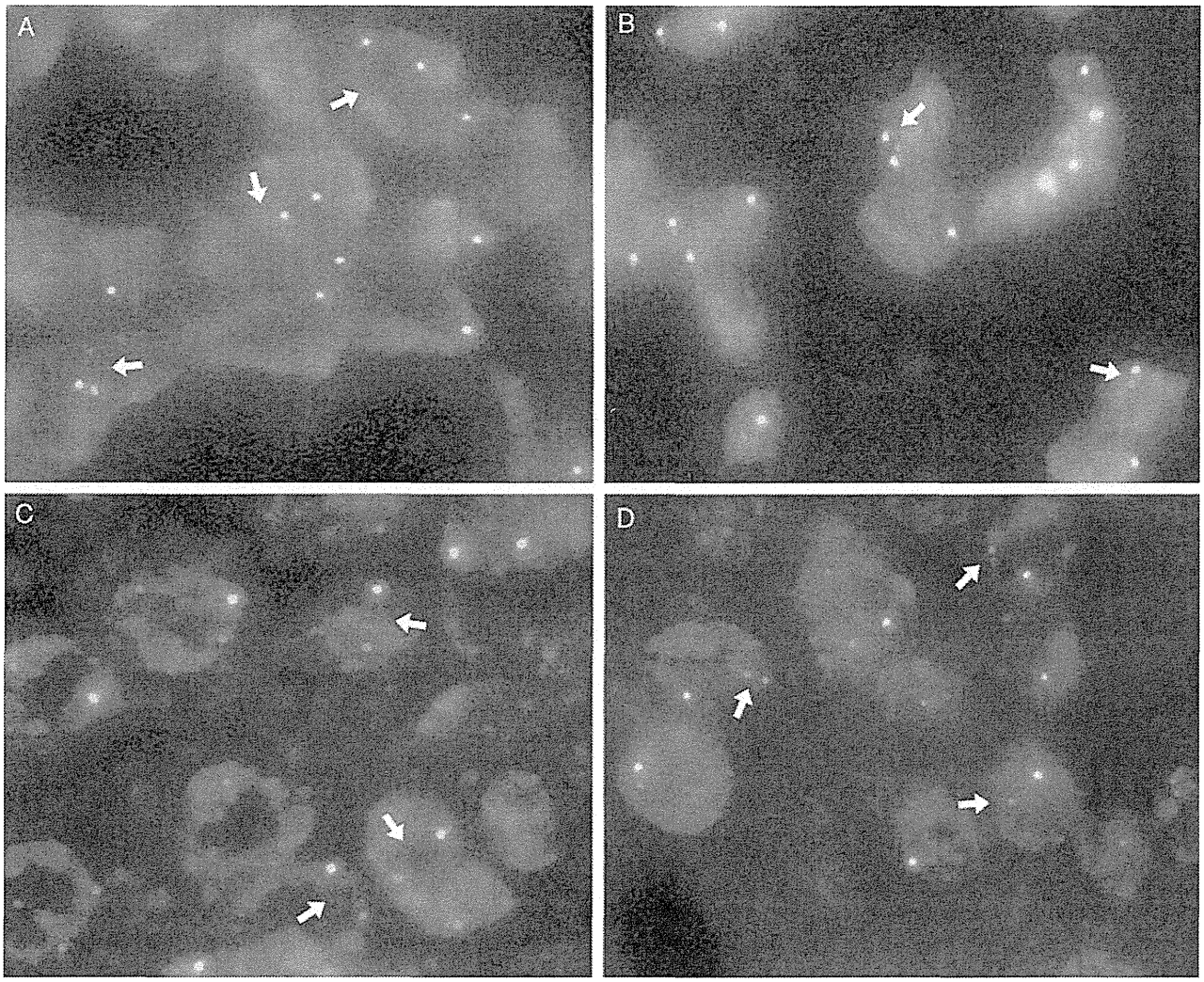

FIGURE 3. FISH pattern of *ROS1*-rearranged lung cancers. In 6 cases, the predominant pattern was "FGR" consisting of 5′/3′ fused signals and a split of 5′ green and 3′ red signals (A, case 15; arrows indicate splits). In 2 of the cases, the distance between the 5′ and 3′ signals was narrow, and rearrangement-positive cells were relatively infrequent (23% and 38%) (B, case 8; arrows indicate narrow splits). In 3 cases, most tumor cells showed a "GR" pattern comprising a split of 5′ green and 3′ red signals without accompanying 5′/3′ fused signals (C, case 11; arrows indicate splits). In 4 cases, most tumor cells showed an "FR" pattern in which 5′/3′ fused signals and isolated 3′ red signals were seen without isolated 5′ green signals (D, case 12; arrows indicate isolated 3′ signals).

histologic indicators of *ALK* rearrangement by multivariate analysis. The present results, therefore, suggest a close phenotypic link between *ROS1*-rearranged and *ALK*-rearranged lung cancers at least in a subset of cases and lead us to suspect that these patterns may be histologic signatures of fusion gene–associated lung cancers. Postulating a similar hypothesis, Takeuchi et al[13] recently described a lung cancer with a *CCDC6-RET* fusion that exhibited a "mucinous cribriform pattern." We have also encountered 2 cases of *RET*-translocated lung cancer that showed a prominent "solid signet-ring cell pattern" (A. Yoshida and T. Shibata, 2012, unpublished data). Future research focusing on these histologic patterns may discover as yet unidentified oncogenic fusion genes in lung adenocarcinoma. In contrast to our results, Bergethon et al[10] did not reveal a clear correlation between *ROS1* status and a histologic subtype. The reason for this discrepancy is unclear, but may be related to the differences in the ethnic background and/or specimen type (eg, biopsy vs. resection).

Detailed FISH analysis revealed several important findings. First and foremost, we successfully validated the existing FISH diagnostic criteria using a cutoff value of 15%.[10,12] However, we noted that the criteria did require modification, because 4 of 13 RT-PCR-proven fusion-positive cases showed almost exclusive FR signal patterns and were misclassified as fusion negative if only split signals were interpreted as rearrangement. Hence, we propose that isolated 3′ signals should be considered as positive for rearrangement to attain an optimal test performance, similar to the strategy implemented for diagnosing *ALK*-rearranged lung cancer.[5,8] This proposal is

biologically reasonable, because the ROS1 kinase domain is encoded at the 3′ part of the fusion gene, and therefore, the presence of an oncogenic fusion is represented by an unpaired 3′ FISH signal irrespective of the presence of accompanying 5′ signals. A few prior reports also considered isolated *ROS1* 3′ signals as gene rearrangement.[13,14] Evaluating 50 tumor cells per case also provided perfect concordance between RT-PCR and FISH, and this finding should be of value when only a limited number of cells are available for testing.

Another important finding from the FISH analysis was that rearrangement signals were present in >50% tumor cells in most *ROS1* fusion–positive cases. In addition, the percentage of rearrangement-positive cells and actual signal patterns were comparable between 2 separate TMA cores representing distant foci in all the fusion-positive tumors successfully tested. These data suggest that *ROS1* rearrangement is diffusely distributed in a homogenous manner, a property typically associated with a "driver" oncogenic event rather than a secondary alteration. Further support for this idea comes from the finding that all the *ROS1* rearrangements occurred mutually exclusive of other oncogenic changes involving *EGFR*, *KRAS*, *HER2*, *ALK*, and *RET* in our cohort. Two fusion-positive tumors (cases 8 and 14) showed only a modest number of rearrangement-positive cells, and one may suspect that gene rearrangement was an acquired event in these cases. However, in both cases, the splits between 5′ and 3′ signals were narrow with a distance approaching the diameter of 1 to 2 hybridization signals. Some of these narrow splits may not have been distinguishable from fused signals, and this technical difficulty may have lowered the apparent number of rearrangement-positive cells. From a practical standpoint, the occurrence of narrow splits in a break-apart FISH assay warrants cautious interpretation for the accurate diagnosis of *ROS1*-rearranged lung cancer.

FISH analysis showed no clear association between fusion partners and signal patterns. Because the *CD74* (5q32) and *SLC34A2* (4p15.2) genes are located in different chromosomes from *ROS1* (6q22.1), a simple reciprocal translocation would suffice to create a fusion that would manifest as a classic FGR FISH pattern. Similarly, because the *EZR* gene (6q25.3) is located 41 Mbp away from *ROS1* (6q22.1) in the same transcriptional orientation on the same arm of chromosome 6, the most simple mechanism conceivable to generate fusion would be the interstitial deletion that would result in an FR FISH pattern. However, these expected patterns were observed in less than half of the cases, and 5 of 10 *CD74-ROS1*-positive or *SLC34A2-ROS1*-positive tumors showed a GR or FR pattern, and 2 of 3 *EZR-ROS1*-positive tumors showed an FGR or GR pattern. Likewise, the narrow split observed in the 2 aforementioned cases cannot be explained by the expected genomic mechanisms to generate *EZR-ROS1* (case 8) and *CD74-ROS1* (case 14). Signal number gain involving fused, 5′, or 3′ signals, albeit at a low level, was also noted. Collectively, these data suggest that the fusion genes develop through complex genomic events that may involve a combination of deletion, insertion, and copy number increase. Although paying attention to the FR pattern reportedly contributed to the discovery of *FIG* (6q22.1) as a fusion partner of *ROS1* in lung cancer,[12] we conclude that it is generally impossible to infer the fusion partner on the basis of the *ROS1* break-apart FISH results.

Because *ROS1*-rearranged lung cancer is rare, comprising only 1% to 3% of lung adenocarcinomas,[9–14] and molecular assays require significant cost and labor, an algorithm is needed with which a subpopulation of the patients can be efficiently preselected who are more likely to yield positive molecular results. Although *ROS1*-rearranged cancer tends to occur in young female nonsmokers, clinical parameters are not sensitive enough for successful triage. The characteristic histologic features described above should be helpful for prompting pathologists for genetic testings; however, these features were present in only a subset of fusion-positive cases and accounted for a small part of total tumor volume in the majority of such cases. Moreover, these patterns were even less frequent in metastatic sites. Hence, we suspect that tumor morphology likely plays a limited role in identifying *ROS1*-rearranged tumors presenting at inoperable stages for which only small samples are available. TTF-1 immunoreactivity was observed in most *ROS1*-rearranged tumors, but it is far from specific, because many other types of adenocarcinoma, such as *EGFR* mutant,[2] label for this marker. As highly sensitive ALK IHC is making a significant contribution to the diagnosis of *ALK*-rearranged cancer,[5,27,28] ROS1 IHC may, in theory, also be expected to simplify the screening process. However, unlike *ALK*, *ROS1* mRNA is known to be overexpressed in 20% to 30% of lung adenocarcinomas[15] irrespective of gene rearrangement status,[11] which casts a shadow over the potential utility of ROS1 IHC. A recently developed novel anti-ROS1 rabbit monoclonal antibody may be promising,[12] but it still needs to be validated when the reagent becomes commercially available. Finally, because driver mutations are typically mutually exclusive of one another, it is eminently reasonable to initiate testing from more frequent genetic events such as *EGFR* and *KRAS* mutations, before exploring the rarer abnormalities like *ALK* or *ROS1* rearrangement.

In summary, our RT-PCR study of a large number of surgical cases identified 15 *ROS1* fusion–positive lung cancers. The most common fusion partner was *CD74* followed by *EZR*. The tumors most commonly occurred in younger never-smoking women, and the affected patients showed similar overall survival to the *ROS1* fusion–negative population. All 15 cases contained an adenocarcinoma component, and half of them were characterized by the "solid signet-ring cell pattern" and/or the "mucinous cribriform pattern," suggesting a close phenotypic kinship between *ALK*-rearranged and *ROS1*-rearranged cancers. We successfully validated *ROS1* break-apart FISH as a reliable method to diagnose *ROS1* rearrangement and verified that the proposed 15% cutoff value is reasonable. In addition, we proposed that isolated 3′ signals should be considered positive evidence of rearrangement to improve the test performance. FISH

signal patterns were unrelated to 5′ fusion partner genes. All the *ROS1* fusion–positive tumors lacked alteration of *EGFR*, *KRAS*, *HER2*, *ALK*, and *RET* genes. Our study has clarified clinicopathologic and molecular characteristics of *ROS1*-rearranged lung cancer, and the results presented here should aid in our understanding of this rare but important molecular subset that is likely responsive to emerging targeted therapy.

ACKNOWLEDGMENTS

The authors thank Sachiko Miura, Chizu Kina, and Ryosuke Yamaga for technical assistance. They also appreciate Dr. Fumiaki Takahashi for helpful inputs.

REFERENCES

1. Ferlay J, Shin HR, Bray F, et al. *GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]*. Lyon, France: International Agency for Research on Cancer. Available at: http://globocan.iarc.fr. Accessed June 14, 2012; 2010.
2. Motoi N, Szoke J, Riely GJ, et al. Lung adenocarcinoma: modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. *Am J Surg Pathol.* 2008;32:810–827.
3. 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.
4. Yatabe Y, Koga T, Mitsudomi T, et al. CK20 expression, CDX2 expression, K-ras mutation, and goblet cell morphology in a subset of lung adenocarcinomas. *J Pathol.* 2004;203:645–652.
5. Yoshida A, Tsuta K, Nakamura H, et al. Comprehensive histologic analysis of ALK-rearranged lung carcinomas. *Am J Surg Pathol.* 2011;35:1226–1234.
6. Kohno T, Ichikawa H, Totoki Y, et al. KIF5B-RET fusions in lung adenocarcinoma. *Nat Med.* 2012;18:375–377.
7. 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.
8. Kwak EL, Bang YJ, Camidge DR, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med.* 2010;363:1693–1703.
9. Rikova K, Guo A, Zeng Q, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell.* 2007;131:1190–1203.
10. Bergethon K, Shaw AT, Ou SH, et al. ROS1 rearrangements define a unique molecular class of lung cancers. *J Clin Oncol.* 2012;30:863–870.
11. Li C, Fang R, Sun Y, et al. Spectrum of oncogenic driver mutations in lung adenocarcinomas from East Asian never smokers. *PLoS One.* 2011;6:e28204.
12. Rimkunas VM, Crosby K, Kelly M, et al. Analysis of receptor tyrosine kinase ROS1 positive tumors in non-small cell lung cancer: identification of a FIG-ROS1 fusion. *Clin Cancer Res.* 2012;18:4449–4457.
13. Takeuchi K, Soda M, Togashi Y, et al. RET, ROS1 and ALK fusions in lung cancer. *Nat Med.* 2012;18:378–381.
14. Davies KD, Le AT, Theodoro MF, et al. Identifying and targeting ROS1 gene fusions in non-small cell lung cancer. *Clin Cancer Res.* 2012;18:4570–4579.
15. Acquaviva J, Wong R, Charest A. The multifaceted roles of the receptor tyrosine kinase ROS in development and cancer. *Biochim Biophys Acta.* 2009;1795:37–52.
16. Robinson DR, Wu YM, Lin SF. The protein tyrosine kinase family of the human genome. *Oncogene.* 2000;19:5548–5557.
17. Charest A, Lane K, McMahon K, et al. Fusion of FIG to the receptor tyrosine kinase ROS in a glioblastoma with an interstitial del(6)(q21q21). *Genes Chromosomes Cancer.* 2003;37:58–71.
18. Gu TL, Deng X, Huang F, et al. Survey of tyrosine kinase signaling reveals ROS kinase fusions in human cholangiocarcinoma. *PLoS One.* 2011;6:e15640.
19. Birch AH, Arcand SL, Oros KK, et al. Chromosome 3 anomalies investigated by genome wide SNP analysis of benign, low malignant potential and low grade ovarian serous tumours. *PLoS One.* 2011;6:e28250.
20. Shaw A, Camidge D, Engelman J, et al. Clinical activity of crizotinib in advanced non-small cell lung cancer (NSCLC) harboring ROS1 gene rearrangement. *J Clin Oncol.* 2012;30(suppl):abstr 7508.
21. Travis WD, Brambilla E, Muller-Hermelink HK, et al. *Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart.* Lyon, France: IARC Press; 2004.
22. Sobin L, Wittekind C. *International Union Against Cancer (UICC) TNM Classification of Malignant Tumours.* 6th ed. New York: Wiley-Liss; 2002.
23. Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol.* 2011;6:244–285.
24. Inamura K, Takeuchi K, Togashi Y, et al. EML4-ALK lung cancers are characterized by rare other mutations, a TTF-1 cell lineage, an acinar histology, and young onset. *Mod Pathol.* 2009;22:508–515.
25. Rodig SJ, Mino-Kenudson M, Dacic S, et al. Unique clinicopathologic features characterize ALK-rearranged lung adenocarcinoma in the western population. *Clin Cancer Res.* 2009;15:5216–5223.
26. Jokoji R, Yamasaki T, Minami S, et al. Combination of morphological feature analysis and immunohistochemistry is useful for screening of EML4-ALK-positive lung adenocarcinoma. *J Clin Pathol.* 2010;63:1066–1070.
27. Mino-Kenudson M, Chirieac LR, Law K, et al. A novel, highly sensitive antibody allows for the routine detection of ALK-rearranged lung adenocarcinomas by standard immunohistochemistry. *Clin Cancer Res.* 2010;16:1561–1571.
28. Takeuchi K, Choi YL, Togashi Y, et al. KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer. *Clin Cancer Res.* 2009;15:3143–3149.

nature medicine

KIF5B-RET fusions in lung adenocarcinoma

Takashi Kohno[1,15], Hitoshi Ichikawa[2,15], Yasushi Totoki[3], Kazuki Yasuda[4], Masaki Hiramoto[4], Takao Nammo[4], Hiromi Sakamoto[2], Koji Tsuta[5], Koh Furuta[5], Yoko Shimada[1], Reika Iwakawa[6], Hideaki Ogiwara[1], Takahiro Oike[6], Masato Enari[7], Aaron J Schetter[8], Hirokazu Okayama[6,8], Aage Haugen[9], Vidar Skaug[9], Suenori Chiku[10], Itaru Yamanaka[11], Yasuhito Arai[3], Shun-ichi Watanabe[12], Ikuo Sekine[13], Seishi Ogawa[14], Curtis C Harris[8], Hitoshi Tsuda[5], Teruhiko Yoshida[2], Jun Yokota[6] & Tatsuhiro Shibata[3]

We identified in-frame fusion transcripts of *KIF5B* (the kinesin family 5B gene) and the *RET* oncogene, which are present in 1–2% of lung adenocarcinomas (LADCs) from people from Japan and the United States, using whole-transcriptome sequencing. The *KIF5B-RET* fusion leads to aberrant activation of RET kinase and is considered to be a new driver mutation of LADC because it segregates from mutations or fusions in *EGFR*, *KRAS*, *HER2* and *ALK*, and a RET tyrosine kinase inhibitor, vandetanib, suppresses the fusion-induced anchorage-independent growth activity of NIH3T3 cells.

A considerable proportion of LADCs, the most common histological type of lung cancer that comprises ~40% of the total cases, develops through activation of oncogenes, for example, somatic mutations in *EGFR* (10–50% of cases) or *KRAS* (10–30% of cases) or fusion of *ALK* (5% of cases), in a mutually exclusive manner[1–4]. Tyrosine kinase inhibitors (TKIs) targeting the EGFR and ALK proteins are effective in the treatment of LADCs that carry *EGFR* mutations and *ALK* fusions[1–3], respectively.

We performed whole-transcriptome sequencing (RNA sequencing)[5] of 30 LADC specimens from Japanese individuals to identify new chimeric fusion transcripts that could be targets for therapy[3,5,6]. These LADCs were 2 carcinomas with *EML4-ALK* fusions, 4 with *EGFR* or *KRAS* mutations and 24 without these fusions or mutations (**Supplementary Table 1**). Identifying candidate fusions represented by >20 paired-end reads and validation by Sanger sequencing of the RT-PCR products (**Supplementary Methods**) led to the identification of seven fusion transcripts, including *EML4-ALK* (**Supplementary Table 1**). We detected one of these fusions between *KIF5B* on chromosome 10p11.2 and *RET* on chromosome 10q11.2 in subject BR0020 (**Fig. 1** and **Supplementary Fig. 1a**). We then further investigated this fusion, as fusions between *RET* and genes other than *KIF5B* have previously been shown to drive papillary thyroid tumor formation[6,7].

RT-PCR and a Sanger sequencing analysis of 319 LADC specimens from Japanese individuals (**Supplementary Table 2**), including 30 that had been subjected to whole-transcriptome sequencing, revealed that 1.9% (6 out of 319) expressed *KIF5B-RET* fusion transcripts (**Fig. 1b** and **Supplementary Fig. 1b**). We identified four variants in these six tumors, and all of these variants were in frame (**Fig. 1a**).

A genomic PCR analysis of the six tumors that were positive for RET fusions revealed somatic fusions of the *KIF5B* introns 15, 16, 23 or 24 at chromosome 10p11.2 with the *RET* introns 7 or 11 at 10q11.2 (**Supplementary Fig. 1c,d**), indicating that a chromosomal inversion had occurred between the long and short arms in the centromeric region of chromosome 10 (**Supplementary Figs. 1e** and **2**). We verified this chromosomal inversion using fluorescence *in situ* hybridization, which revealed a split in the signals for the probes that flank the *RET* translocation sites in tumors positive for the *KIF5B-RET* fusion (**Supplementary Fig. 2**).

The tumors positive for the *KIF5B-RET* fusion were all well or moderately differentiated (**Table 1** and **Supplementary Fig. 3**). None of the subjects with these tumors had a history of thyroid cancer, and none showed abnormal findings in their thyroid tissues as determined by computed tomography or positron emission tomography before surgery for LADC. All five examined tumors with the *KIF5B-RET* fusion were positive for thyroid transcription factor 1 (TTF-1) and napsin A aspartic proteinase (Napsin A)[8] but were negative for thyroglobulin[9], indicating that they were of pulmonary origin (**Table 1** and **Supplementary Fig. 3**). The LADCs that were positive for the *KIF5B-RET* fusion showed twofold to 30-fold higher *RET* expression than non-cancerous lung tissues (**Fig. 1b** and **Supplementary Figs. 4** and **5**). An immunohistochemical analysis using an antibody against the C-terminal region of the RET protein detected positive cytoplasmic staining in the tumor cells of the fusion-positive LADCs (**Table 1** and **Supplementary Fig. 3b**) but did not detect this staining in any of the non-cancerous lung cells. A western blot analysis confirmed the expression of the fusion proteins in the LADCs (**Supplementary Fig. 6**).

To address the prevalence of *KIF5B-RET* fusions in LADCs from individuals of non-Asian ancestry, we examined LADCs in cohorts from the United States and Norway (**Supplementary Table 2**). We detected a fusion transcript in 1 of the 80 (1.3%) subjects from the

[1]Division of Genome Biology, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [2]Division of Genetics, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [3]Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [4]Department of Metabolic Disorder, Diabetes Research Center, Research Institute, National Center for Global Health and Medicine, Shinjuku-ku, Tokyo, Japan. [5]Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan. [6]Division of Multistep Carcinogenesis, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [7]Division of Refractory Cancer Research, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [8]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, US National Institutes of Health, Bethesda, Maryland, USA. [9]Section of Toxicology, Department of Chemical and Biological Working Environment, National Institute of Occupational Health, Oslo, Norway. [10]Science Solutions Division, Mizuho Information and Research Institute, Chiyoda-ku, Tokyo, Japan. [11]Statistical Genetics Analysis Division, StaGen, Taito-ku, Tokyo, Japan. [12]Division of Thoracic Surgery, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan. [13]Division of Thoracic Oncology, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan. [14]Cancer Genomics Project, University of Tokyo, Bunkyo-ku, Tokyo, Japan. [15]These authors equally contributed to this work. Correspondence should be addressed to T.K. (tkkohno@ncc.go.jp).

Received 23 August 2011; accepted 16 December 2011; published online 12 February 2012; doi:10.1038/nm.2644

Cancer Research

Identification of Genes Upregulated in *ALK*-Positive and *EGFR/KRAS/ALK*-Negative Lung Adenocarcinomas

Hirokazu Okayama[1,2,10], Takashi Kohno[2], Yuko Ishii[1,2], Yoko Shimada[2], Kouya Shiraishi[2], Reika Iwakawa[1], Koh Furuta[5], Koji Tsuta[5], Tatsuhiro Shibata[3], Seiichiro Yamamoto[7], Shun-ichi Watanabe[6], Hiromi Sakamoto[4], Kensuke Kumamoto[10], Seiichi Takenoshita[10], Noriko Gotoh[8], Hideaki Mizuno[11,12], Akinori Sarai[11], Shuichi Kawano[9], Rui Yamaguchi[9], Satoru Miyano[9], and Jun Yokota[1]

Abstract

Activation of the *EGFR*, *KRAS*, and *ALK* oncogenes defines 3 different pathways of molecular pathogenesis in lung adenocarcinoma. However, many tumors lack activation of any pathway (triple-negative lung adenocarcinomas) posing a challenge for prognosis and treatment. Here, we report an extensive genome-wide expression profiling of 226 primary human stage I–II lung adenocarcinomas that elucidates molecular characteristics of tumors that harbor *ALK* mutations or that lack *EGFR*, *KRAS*, and *ALK* mutations, that is, triple-negative adenocarcinomas. One hundred and seventy-four genes were selected as being upregulated specifically in 79 lung adenocarcinomas without *EGFR* and *KRAS* mutations. Unsupervised clustering using a 174-gene signature, including *ALK* itself, classified these 2 groups of tumors into *ALK*-positive cases and 2 distinct groups of triple-negative cases (groups A and B). Notably, group A triple-negative cases had a worse prognosis for relapse and death, compared with cases with *EGFR*, *KRAS*, or *ALK* mutations or group B triple-negative cases. In *ALK*-positive tumors, 30 genes, including *ALK* and *GRIN2A*, were commonly overexpressed, whereas in group A triple-negative cases, 9 genes were commonly overexpressed, including a candidate diagnostic/therapeutic target *DEPDC1*, that were determined to be critical for predicting a worse prognosis. Our findings are important because they provide a molecular basis of *ALK*-positive lung adenocarcinomas and triple-negative lung adenocarcinomas and further stratify more or less aggressive subgroups of triple-negative lung ADC, possibly helping identify patients who may gain the most benefit from adjuvant chemotherapy after surgical resection. *Cancer Res; 72(1); 100–11. ©2011 AACR.*

Introduction

Lung cancer is the leading cause of cancer death worldwide (1, 2). Adenocarcinoma, which accounts for more than 50% of non-small-cell lung cancers (NSCLC), is the most frequent type and is increasing. Lung adenocarcinoma has a heterogeneous nature in various aspects, including clinicopathologic features (3). Recent molecular studies have revealed at least 3 major molecular pathways for the development of lung adenocarcinoma (4–8). A considerable fraction (30%–60%) of lung adenocarcinomas develops through acquisition of mutations either in the *EGFR*, *KRAS*, or *ALK* genes in a mutually exclusive manner, and the remaining lung adenocarcinomas, that is, those without *EGFR*, *KRAS*, and *ALK* mutations (herein designated "triple-negative adenocarcinomas"), develop with mutations of several other genes. *HER2*, *BRAF*, etc. are known to be mutated also mutually exclusively with the *EGFR*, *KRAS*, and *ALK* genes; however, frequencies of their mutations are very low (<5%; refs. 4–7). Therefore, genes responsible for the development of triple-negative adenocarcinomas are largely unknown.

Mutations in the *EGFR* gene are prevalent in females and never-smokers, and the frequencies are considerably higher in Asians (40%–60%) than in Europeans/Americans (~10%; refs. 5–7, 9). *EGFR* mutations make tumor cells dependent on epidermal growth factor receptor (EGFR) signaling and define patients who respond to EGFR tyrosine kinase inhibitors (TKI), such as gefitinib (10, 11). On the other hand, mutations in the *KRAS* gene occur predominantly in males and ever-smokers, and their frequencies are higher in Europeans/Americans (>15%) than in Asians (10%; ref. 9). Specific inhibitors against KRAS activity are being developed (12). Therefore, clinicopathologic features of lung adenocarcinomas with *EGFR* mutations (herein designated "*EGFR*-positive adenocarcinomas") and

Authors' Affiliations: Divisions of [1]Multistep Carcinogenesis, [2]Genome Biology, [3]Cancer Genomics and [4]Genetics, National Cancer Center Research Institute; [5]Division of Pathology and Clinical Laboratories and [6]Thoracic Surgery Division, National Cancer Center Hospital; [7]Cancer Information Services and Surveillance Division, Center for Cancer Control and Information Services, National Cancer Center; [8]Division of Systems Biomedical Technology and [9]Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo; [10]Department of Organ Regulatory Surgery, Fukushima Medical University School of Medicine, Fukushima; [11]Department of Biosciences and Bioinformatics, Kyushu Institute of Technology, Fukuoka; and [12]Discovery Science & Technology Department, Chugai Pharmaceutical Co., Ltd., Kanagawa, Japan

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

Corresponding Author: Jun Yokota, Divison of Multistep Carcinogenesis, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104-0045, Japan. Phone: 81-3-3547-5272; Fax: 81-3-3542-0807; E-mail: jyokota@ncc.go.jp

doi: 10.1158/0008-5472.CAN-11-1403

those with *KRAS* mutations (herein designated "*KRAS*-positive adenocarcinomas") are considerably different from each other. Recently, a small subset of *EGFR*- and *KRAS*-negative lung adenocarcinomas (~5%) was shown to have rearrangements of the *ALK* gene generating gene fusion transcripts (13), and patients with *ALK* rearrangements tend to be younger and have little or no smoking histories (4, 6–8). Because lung adenocarcinoma cells with *ALK* rearrangements (herein designated "*ALK*-positive adenocarcinomas") are specifically sensitive to ALK TKIs, *ALK*-positive adenocarcinomas have been recently considered to be another subset of adenocarcinomas by considering the differences in therapeutic targets (4, 6–8). In contrast, clinicopathologic features of triple-negative lung adenocarcinomas have not been precisely characterized because of the lack of sufficient genetic information in these adenocarcinomas.

There have been several studies which attempted to characterize gene expression profiles in particular types of lung adenocarcinoma, including *EGFR*-positive and *KRAS*-positive adenocarcinomas (14–17). However, such information is limited for *ALK*-positive adenocarcinomas and triple-negative adenocarcinomas. Therefore, in this study, we aimed to elucidate clinicopathologic features and gene expression profiles of *ALK*-positive adenocarcinomas and triple-negative adenocarcinomas in comparison with those of *EGFR*-positive adenocarcinomas and *KRAS*-positive adenocarcinomas. We conducted a genome-wide gene expression profiling of 226 lung adenocarcinomas, consisting of 127 *EGFR*-positive adenocarcinomas, 20 *KRAS*-positive adenocarcinomas, 11 *ALK*-positive adenocarcinomas, and 68 triple-negative adenocarcinomas. To identify genes useful for molecular diagnosis and applicable to targeted therapy of *ALK*-positive adenocarcinomas and triple-negative adenocarcinomas, we focused on genes that were upregulated in these adenocarcinomas by selecting genes with low expression in *EGFR*-positive and *KRAS*-positive adenocarcinomas. Several genes were identified as being specifically and significantly upregulated in *ALK*-positive adenocarcinomas. In particular, the *ALK* gene itself was highly expressed exclusively in *ALK*-positive adenocarcinomas. More importantly, a distinct group of triple-negative adenocarcinomas with unfavorable outcome was identified. This group of triple-negative adenocarcinomas showed much worse prognosis than the other group of triple-negative adenocarcinomas, *EGFR*-positive adenocarcinomas, *KRAS*-positive adenocarcinomas, and *ALK*-positive adenocarcinomas. Several genes were identified as being upregulated and critical for predicting prognosis of patients in this group of adenocarcinomas.

Materials and Methods

Patients

The tumors were pathologically classified according to the TNM classification of malignant tumors (18). A total of 226 lung adenocarcinoma cases subjected to expression profiling were selected from 393 stage I–II cases who underwent potential curative resection between 1998 and 2008 at the National Cancer Center Hospital as follows (ref. 19; Supplementary Fig. S1). Among the 393 cases, 363 cases, consisting of 305 stage I and 58 stage II cases, were eligible by the criteria of cases who did not receive any neoadjuvant therapies before surgery and had not been diagnosed with cancer in the 5 years before lung adenocarcinoma diagnosis. All 58 stage II cases were subjected to expression profiling. The 305 stage I cases included 37 cases with relapse and 268 cases without relapse. To improve statistical efficiency, all the 37 relapsed cases and 131 matched unrelapsed cases selected by the incidence density sampling method (20, 21) were subjected to expression profiling. In total, 226 cases, consisting of 168 stage I and 58 stage II cases, were subjected to the expression profiling. Among the 226 cases, 204 who received complete resection (i.e., free resection margins and no involvement of mediastinal lymph nodes examined by mediastinal dissection) and did not receive postoperative chemotherapy and/or radiotherapy, unless relapsed, were subjected to survival analyses. This study was approved by the Institutional Review Boards of the National Cancer Center.

Microarray experiments and data processing

Total RNA was extracted using TRIzol reagent (Invitrogen), purified by an RNeasy kit (Qiagen), and qualified with a model 2100 Bioanalyzer (Agilent). All samples showed RNA Integrity Numbers more than 6.0 and were subjected to microarray experiments. Two micrograms of total RNA were labeled using a 5X MEGAscript T7 Kit (Ambion) and analyzed by Affymetrix U133Plus2.0 arrays. The data were processed by the MAS5 algorithm, and the mean expression level of a total of 54,675 probes was adjusted to 1,000 for each sample. Microarray data are available at National Center for Biotechnology Information Gene Expression Omnibus (GSE31210).

Probe selection for unsupervised clustering

One hundred and seventy-four genes (190 probes), preferentially expressed in *ALK*-positive and triple-negative adenocarcinomas, were selected by the following criteria; probes whose expression levels were less than 1,000 in any adenocarcinomas with *EGFR* or *KRAS* mutations, and probes whose averaged expression levels in *ALK*-positive and triple-negative adenocarcinomas were more than 1.5-fold higher than those in *EGFR*-positive and *KRAS*-positive adenocarcinomas with P values less than 0.05 by t test. Expression levels for these 190 probes were log-transformed and median-centered, both for probes and samples, and were subjected to an unsupervised hierarchical clustering. The clustering was done by the centroid linkage method using the Cluster 3.0 program, and the results were visualized using the Java Treeview program (22).

Mutation analyses

Genomic DNAs from all 226 lung adenocarcinomas were analyzed for *EGFR* and *KRAS* mutations by the high-resolution melting method as described (23, 24). Total RNAs from the 226 adenocarcinomas were examined for expression of fusion transcripts between *ALK* and *EML4* or *KIF5* using a multiplex reverse transcription PCR (RT-PCR) method (25).

Statistics

Cumulative survival was estimated by the Kaplan–Meier method, and differences in the survivals between 2 groups were

analyzed by log-rank test. Influences of variables on relapse-free survival (RFS) and overall survival (OS) were evaluated by uni- and multivariate analyses of the Cox proportional hazard model. For all analyses, smoking status was polarized as never-smokers (0 pack years) and ever-smokers (>0 pack years). Pathologic TNM staging was categorized as stage I versus stage II. For multivariate analysis, all variables were included that were moderately associated ($P < 0.1$) with RFS or OS in any of the analyses.

Bioinformatics

Associations of gene expression levels with prognosis of NSCLC patients in 7 other expression profile studies were obtained from the PrognoScan database (26). In the PrognoScan database, association of gene expression with survival of patients was evaluated by the minimum P value approach. Briefly, patients were first arranged by expression levels of a given gene. They were then divided into high- and low-expression groups at all possible cutoff points, and the risk differences of any 2 groups were estimated by the log-rank test. Finally, the cutoff point that gave the most pronounced P value was selected.

Results

EGFR/KRAS/ALK mutations and clinicopathologic characteristics of lung adenocarcinomas subjected to gene expression profiling

Among 226 stages I and II lung adenocarcinomas, *EGFR* and *KRAS* mutations were mutually exclusively detected in 127 (56%) and 20 (9%) cases, respectively, and an *EML4–ALK* fusion gene was expressed in 11 (4.9%) cases (Table 1). *EGFR* or *KRAS* mutations were not detected in any of the 11 cases with *EML4–ALK* fusion expression; thus, the occurrence of *ALK* rearrangements in a mutually exclusive manner with *EGFR* and *KRAS* mutations in lung adenocarcinoma was confirmed. The incidence and the fraction of *EGFR*-, *KRAS*-, and *ALK*-positive cases in this study were consistent with those in previous studies (5–7, 9, 13). Accordingly, the remaining 68 (30%) cases were defined as "triple-negative adenocarcinomas" because of the absence of *EGFR*, *KRAS*, and *ALK* mutations. Clinicopathologic features of *EGFR*-positive adenocarcinomas and *KRAS*-positive adenocarcinomas in this study are well consistent with those in previous studies of Japanese populations (27, 28). Patients with *ALK*-positive adenocarcinomas were younger and more likely to be never-smokers, as previously indicated (4, 6–8). Triple-negative adenocarcinomas showed similar clinicopathologic features to those of *KRAS*-positive adenocarcinomas, that is, a predominance of males, ever-smokers, and advanced stages.

Expression profile unique to *ALK*-positive lung adenocarcinomas

All 226 cases were subjected to genome-wide expression profiling using Affymetrix U133Plus2.0 arrays. One hundred and seventy-four genes, evaluated with 190 probes (Supplementary Table S1), were selected as those preferentially expressed in either *ALK*-positive adenocarcinomas or triple-negative adenocarcinomas under the criteria described in Materials and Methods. In particular, 10 genes evaluated with 11 probes were markedly upregulated according to the criteria of fold-differences more than 2.0 with P values less than 0.05 (Supplementary Table S2). It was noted that 2 probes for the *ALK* gene were present among them, and 1 of them (probe ID = 208212_s_at) showed the highest fold-difference of 8.7 between *ALK*-positive/triple-negative adenocarcinomas and *EGFR*-positive/*KRAS*-positive adenocarcinomas among the 190 probes. This result indicated that there is a subset of adenocarcinomas in which *ALK* was overexpressed. Therefore, an unsupervised

Table 1. Clinicopathologic characteristics of 226 lung adenocarcinomas subjected to expression profile analysis

		Mutation				Expression profile	
Variable	All	EGFR (+)	KRAS (+)	ALK (+)	Triple (−)	Group A	Group B
No. of cases	226	127	20	11	68	36	32
Age							
Mean	60	60	60	54	61	61	60
Range	30–76	35–72	46–75	30–68	46–76	46–71	47–76
Sex							
Male	105	50	12	2	41	25	16
Female	121	77	8	9	27	11	16
Smoking habit							
Never-smoker	115	67	10	7	31	10	21
Ever-smoker	111	60	10	4	37	26	11
pStage							
IA	114	77	6	3	28	10	18
IB	54	26	8	0	20	12	8
II	58	24	6	8	20	14	6

Table 2. Genes upregulated in *ALK*-positive lung adenocarcinomas

Gene symbol[a]	Gene name	Probe ID	Fold difference
ALK	Anaplastic lymphoma receptor tyrosine kinase	208212_s_at	55.2
EST	Transcribed locus	242964_at	26.8
ALK	Anaplastic lymphoma receptor tyrosine kinase	208211_s_at	17.2
GRIN2A	Glutamate receptor, ionotropic, *N*-methyl D-aspartate 2A	242286_at	13.0
GRIN2A	Glutamate receptor, ionotropic, *N*-methyl D-aspartate 2A	231384_at	12.4
MUC5AC /// MUC5B	Mucin 5AC, oligomeric mucus/gel-forming /// mucin 5B, oligomeric mucus/gel-forming	222268_x_at	9.2
EST	Transcribed locus	1570291_at	8.1
LOC100292909	Hypothetical protein LOC100292909	241535_at	7.7
BLID	BH3-like motif containing, cell death inducer	1555675_at	7.4
LOC100130894	Hypothetical LOC100130894	1564158_a_at	6.1
CLDN10	Claudin 10	1556687_a_at	6.0
KRT16	Keratin 16	209800_at	5.9
PROM2	Prominin 2	1562378_s_at	5.6
GJB5	Gap junction protein, beta 5, 31.1 kDa	206156_at	5.0
KIAA1644	KIAA1644	221901_at	4.8
EPHB1	EPH receptor B1	210753_s_at	4.5
LRRC4	Leucine rich repeat containing 4	223552_at	4.2
EST	Transcribed locus	235373_at	3.4
tcag7.1188	Hypothetical LOC340340	1561254_at	3.3
SBNO2	Strawberry notch homolog 2 (*Drosophila*)	204166_at	3.3
EST	Transcribed locus	241083_at	3.1
SLC25A37	Solute carrier family 25, member 37	222528_s_at	3.1
NDP	Norrie disease (pseudoglioma)	206022_at	3.1
EST	Transcribed locus	243478_at	3.0
EST	Transcribed locus	239136_at	2.9
RHOV	ras homolog gene family, member V	241990_at	2.9
YIF1B	Yip1 interacting factor homolog B (*S. cerevisiae*)	231211_s_at	2.9
RPRM	Reprimo, TP53 dependent G2 arrest mediator candidate	219370_at	2.5
SYT12	Synaptotagmin XII	228072_at	2.5
HES2	Hairy and enhancer of split 2 (*Drosophila*)	231928_at	2.4
CDH11	Cadherin 11, type 2, OB-cadherin (osteoblast)	239769_at	2.2
IRAK3	Interleukin-1 receptor-associated kinase 3	220034_at	2.1

[a]Genes with fold difference >2.0 and $P < 0.05$ between *ALK*-positive and *ALK*-negative adenocarcinomas are shown.

hierarchical clustering using these 190 probes was done on 11 *ALK*-positive adenocarcinomas and 68 triple-negative adenocarcinomas (Supplementary Figs. S1 and S2). There were 3 distinct sets of genes/probes, as indicated by red, yellow, and blue bars on the left of the heat map. Two probes for the *ALK* gene were present in the gene/probe set with a yellow bar, and 11 cases with extremely high levels of *ALK* expression comprised a small subcluster in the right side of cluster 1. All the 11 cases corresponded to the ones with *EML4–ALK* fusion gene expression.

The results strongly indicated that *ALK*-positive adenocarcinomas have distinct expression profiles in comparison with *ALK*-negative adenocarcinomas, including not only triple-negative adenocarcinomas but also *EGFR*-positive and *KRAS*-positive adenocarcinomas. Therefore, genes with fold-differences more than 2.0 and *P* values less than 0.05 in their expression between *ALK*-positive adenocarcinomas and *ALK*-negative adenocarcinomas were further selected from the 190 probes. Thirty genes with 32 probes were then selected (Table 2). The *ALK* gene showed the highest level of fold difference in *ALK*-positive adenocarcinomas. Therefore, as previously reported (29–31), *ALK*-positive adenocarcinomas express high levels of *ALK* gene products, supporting that upregulation of the *ALK* gene is a biological consequence of *ALK* rearrangements in lung adenocarcinoma cells. Expression profiling further revealed that various other genes are distinctly upregulated in *ALK*-positive adenocarcinomas. In particular, fold differences of *GRIN2A* (glutamate receptor, ionotropic, *N*-methyl D-aspartate 2A) expression were more than 10, as with *ALK* expression. Moreover, *GRIN2A* was branched most closely to *ALK* in the heat map (Supplementary Fig. S2). Therefore, high levels of *GRIN2A* expression can be a characteristic unique to *ALK*-positive adenocarcinomas, in addition to upregulation of the *ALK* gene itself. The levels of *GRIN2A* expression in *ALK*-

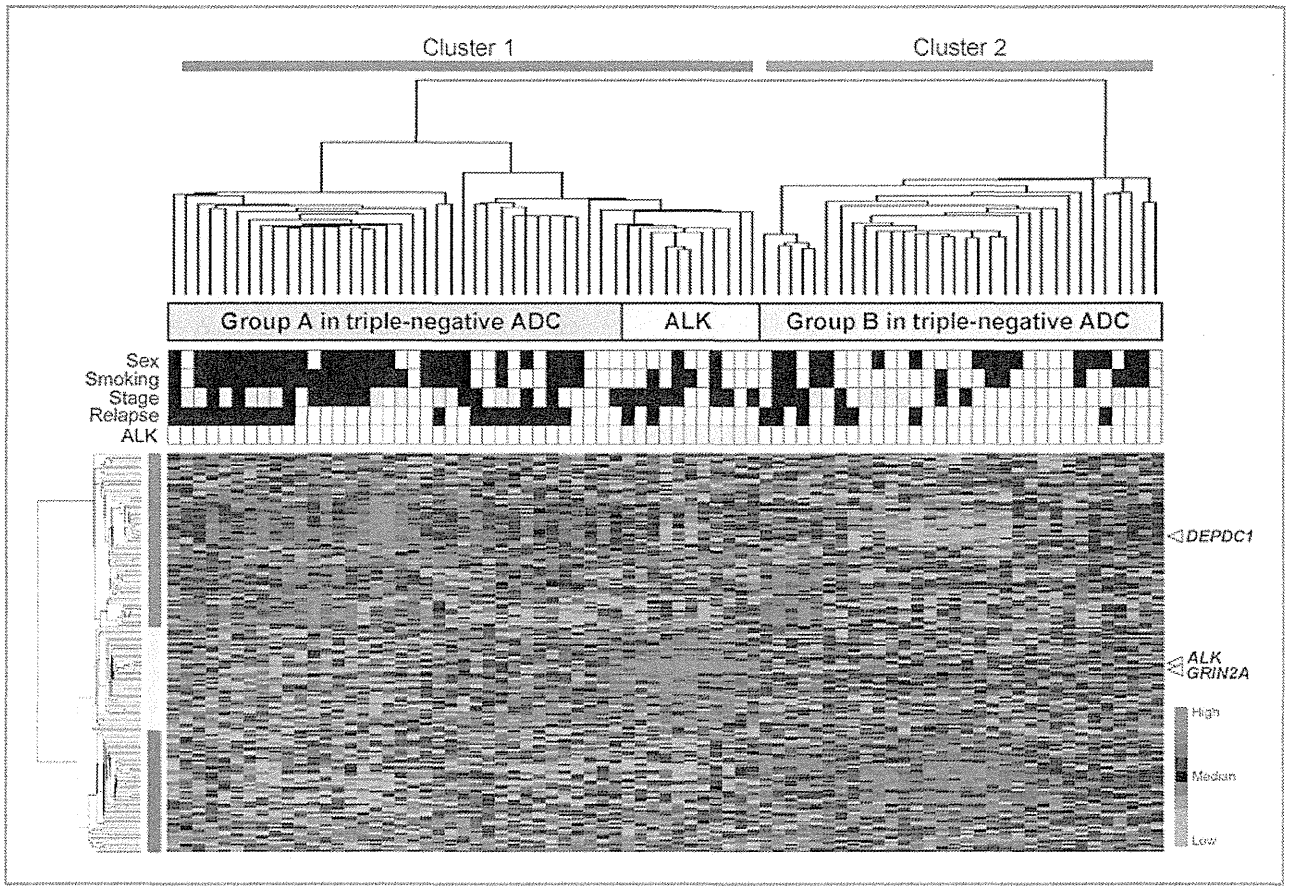

Figure 1. Unsupervised hierarchical clustering of 11 *ALK*-positive adenocarcinomas and 68 triple-negative adenocarcinomas. Triple-negative adenocarcinomas were separated into 36 group A cases and 32 group B cases, and group A cases construct cluster 1 with 11 *ALK*-positive adenocarcinoma cases. Clinical and genetic features are shown below the tree; sex (black, male; white, female); smoking status (black, ever-smoker; white, never-smoker); pathologic stage (black, stage II; gray, stage IB; white, stage IA); relapse (black, evidence of relapse; white, no evidence of relapse); *ALK* (yellow, *ALK*-fusion gene expression positive; white, negative). Three colored bars according to the main branches of probes/genes are shown on the left. Positions of probes for *ALK*, *GRIN2A*, and *DEPDC1* are shown on the right. ADC, adenocarcinoma.

positive adenocarcinomas were significantly higher than those in *ALK*-negative adenocarcinomas by quantitative RT-PCR analysis (Supplementary Fig. S3).

Triple-negative lung adenocarcinomas with poor prognosis identified by gene expression profiling

By the unsupervised hierarchical clustering, 68 triple-negative adenocarcinomas were separated into 2 major groups, one containing 36 cases and the other 32 cases, designated as groups A and B, respectively (Fig. 1). Group A comprised cluster 1 with 11 *ALK*-positive adenocarcinomas. Group A cases were dominant in males, ever-smokers, and advanced stages, whereas group B cases were dominant in never-smokers and early stages (Table 1), indicating that group A cases comprise an aggressive type in triple-negative adenocarcinomas. Therefore, we next compared RFS and OS among the 5 groups of patients; groups A and B, *EGFR*-positive cases, *KRAS*-positive cases, and *ALK*-positive cases (Fig. 2). Among the 226 cases, 204 cases that received complete resection and did not receive postoperative chemotherapy and/or radiotherapy were subjected to survival analysis. Group A cases ($n = 32$) showed the worst prognosis for both RFS and OS among the 5 groups (Fig. 2A and B). In particular, group A cases showed significantly worse prognosis ($P < 0.05$) for both RFS and OS than group B cases ($n = 30$) and *EGFR*-positive cases ($n = 116$) by the log-rank test. Such differences were marginally significant between group A cases and *KRAS*-positive cases ($n = 19$) and not significant between group A cases and *ALK*-positive cases ($n = 7$), probably because the numbers of *KRAS*-positive and *ALK*-positive cases were smaller than those of group B and *EGFR*-positive cases.

Similar results were obtained from the analysis of 162 patients with stage I adenocarcinomas (Fig. 2C and D), indicating the independency of these associations with staging. Therefore, we next carried out multivariate analyses on RFS and OS of these 5 groups (Table 3). In the analysis of 204 stages I and II patients, RFS and OS of group A cases were significantly worse than those of *EGFR*-positive and group B cases, and the differences were independent of staging. HRs of *ALK*-positive and *KRAS*-positive cases were also as high as *EGFR*-positive and group B cases, although only the difference in RFS was statistically significant between group A cases and *KRAS*-positive cases. This could be also due to the small numbers