The IRESSA Pan-Asia Study (IPASS) compared gefitinib with carboplatin/paclitaxel as first-line treatment in 1217 never-smokers/light ex-smokers with advanced adenocarcinoma of the lung in East Asia.[6] Subgroup analysis of patients with *EGFR* mutations ($n = 261$) detected in DNA derived from tumor tissue samples demonstrated significantly longer progression-free survival (PFS) with gefitinib versus carboplatin/paclitaxel (hazard ratio [HR], 0.48; 95% confidence interval [CI], 0.36–0.64; $p < 0.001$).[6] In the *EGFR* mutation-negative (M−) subgroup ($n = 176$), PFS was significantly longer with carboplatin/paclitaxel versus gefitinib (HR, 2.85; 95% CI, 2.05–3.98; $p < 0.001$). Objective response rates (ORR) were 71.2% versus 47.3% ($p < 0.001$) and 1.1% versus 23.5% ($p = 0.001$) with gefitinib versus carboplatin/paclitaxel in *EGFR* M+ and M− patients, respectively.

The difficulties of collecting sufficient tumor tissue for biomarker analyses have stimulated interest in analyses using surrogate samples, such as serum and plasma samples, which frequently contain circulating free (cf) DNA derived from tumor tissues. Previous studies in relatively few patients had detected *EGFR* mutations in cfDNA in serum or plasma samples and suggested that using such methodology to predict response to gefitinib was worthy of further evaluation.[7–12] However, most of these studies were retrospective.

Herein, we report the evaluation of *EGFR* mutations in cfDNA from serum samples of patients in the IPASS study recruited in Japan. This preplanned, exploratory analysis was conducted to increase the understanding of the use of surrogate samples, such as serum, versus tumor biopsy samples for determining *EGFR* mutation status.

MATERIALS AND METHODS

Study Design and Patients

Full details of the IPASS study design (ClinicalTrials.gov identifier NCT00322452) have been published previously.[6] Planned objectives of this substudy of IPASS were evaluations of efficacy between the gefitinib and carboplatin/paclitaxel treatment groups by cfDNA *EGFR* mutation status from pretreatment serum samples and evaluation of the concordance between *EGFR* mutation status in pretreatment cfDNA versus tumor. Comparison of *EGFR* mutation status in pretreatment versus postprogression serum samples was also performed; however, not all patients with a pretreatment sample had a postprogression sample, which limited the comparison. In addition, comparisons with postprogression serum and pretreatment pleural effusion samples are reported in Supplemental Digital Content 1 (Methods http://links.lww.com/JTO/A152). Preplanned analysis of the Japanese subset of the IPASS population was performed to meet Japanese regulatory requirements.

All patients provided written informed consent. Provision of samples for biomarker research was optional and involved separate consent procedures for tumor and serum sampling. An independent ethics committee at each participating institution approved the study protocol. The study was conducted in accordance with the Declaration of Helsinki, the International Conference on Harmonisation Guidelines for Good Clinical Practice, applicable regulatory requirements, and AstraZeneca's policy on bioethics.

Biomarker Analyses

Sample collection and DNA extraction are described in Supplemental Digital Content 1 (Methods http://links.lww.com/JTO/A152). *EGFR* mutations were detected using the DxS *EGFR* Mutation Test Kit for Research Use Only (DxS, Manchester, UK), which combines Amplification Refractory Mutation System (ARMS) (allele-specific polymerase chain reaction [PCR]) with the Scorpions real-time PCR technology.[13,14] Modified run conditions and cutoffs (delta Ct values [dCt]) used to define M+ samples for cfDNA derived from serum and pleural effusion samples were as follows: 50 cycles of PCR were carried out and the dCt for exon 19 deletions was 12, L858R was 14, and T790M was 8 (for tumor DNA, 40 cycles of PCR were carried out and the dCt cutoffs were 9, 11, and 8, respectively). In analyses of tumor DNA, all 29 mutations detected by the kit were assayed (19 deletions in exon 19, L858R, T790M, L861Q, G719X [S, A, or C], S768I, and 3 insertions in exon 20); whereas for serum and pleural effusion samples, the 21 most common mutations (19 deletions in exon 19, L858R, and T790M) were assayed (to make the best use of limited cfDNA yield). Samples were tested in duplicate, and only if both replicates were positive for at least one of the mutations was the sample defined as M+. Patients without a tumor sample evaluable for mutation analysis and samples which were not successfully analyzed were classified as *EGFR* mutation unknown. Biomarker samples were assayed blinded to clinical outcome and randomized treatment.

Statistical Analyses

Serum samples were collected for patients recruited in Japan and who consented to this optional analysis. Analyses of efficacy end points comparing treatment groups in the Japanese subset (intent-to-treat [ITT] population) were assessed as described previously for the overall IPASS population.[6] However, for the analyses in the cfDNA M+ and M− subgroups, the prespecified covariates of World Health Organization (WHO) performance status (PS), smoking history, and sex could not be included as covariates because of the small number of patients who had a WHO PS 2, were ex-smokers, or were males; therefore, models without covariates were used. Because of the lack of power to detect treatment differences, the result of the Japanese subset should be interpreted with caution, taking into account the associated variability and overlap in plausible range of effects (CIs). Analyses comparing treatment groups were performed for PFS (by Cox proportional hazards model) and ORR (by logistic regression model) in subgroups defined by cfDNA *EGFR* mutation status. A test for interaction between cfDNA *EGFR* mutation status (M+ or M−) and treatment was used to assess whether the PFS treatment effect was statistically different between subgroups.

Comparison of pretreatment cfDNA versus tumor *EGFR* mutations was based on the 21 mutations analyzed for cfDNA using patients with known mutation status (M+ or M−) in both samples. The sensitivity, specificity, positive

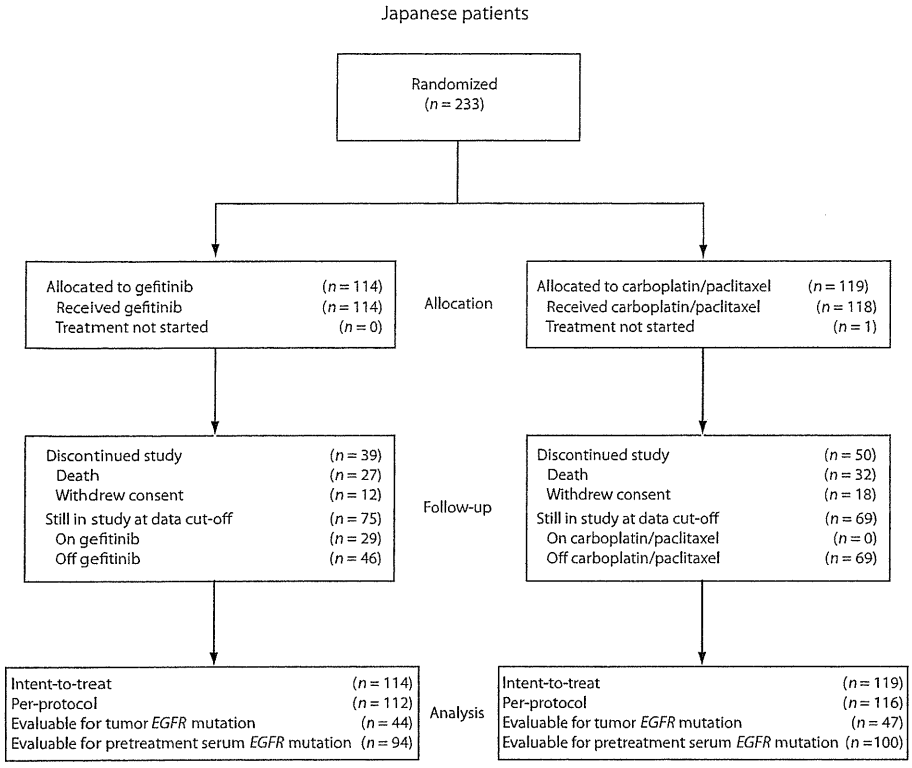

FIGURE 1. CONSORT diagram representing patient disposition (including number of patients with tumor tissue or serum evaluable for *EGFR* mutation status). *EGFR*, epidermal growth factor receptor.

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

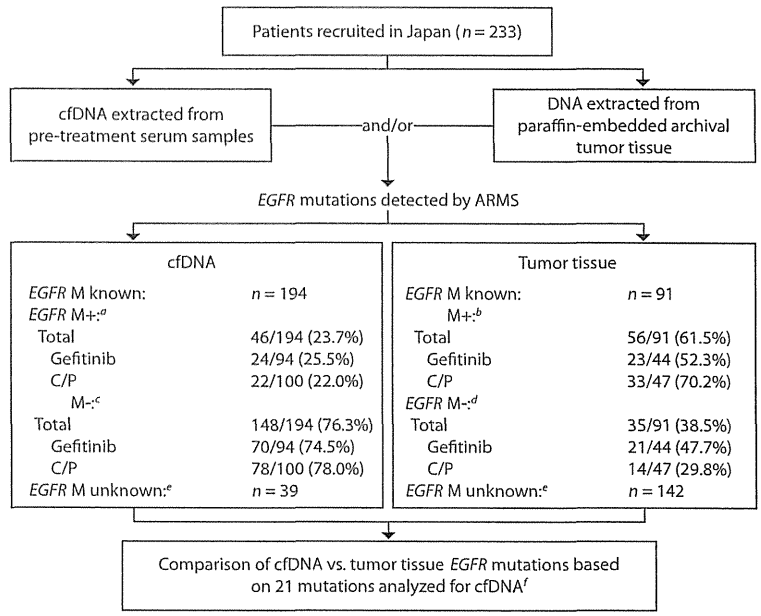

FIGURE 2. Flow and results of *EGFR* mutation analysis. [a]Sample positive for ≥1 of 21 mutations tested; detected 19 deletions in exon 19, L858R, and T790M. [b]Sample positive for ≥1 of 29 mutations tested; detected 19 deletions in exon 19, L858R, T790M, L861Q, G719S, G719A, G719C, S768I; 3 insertions 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

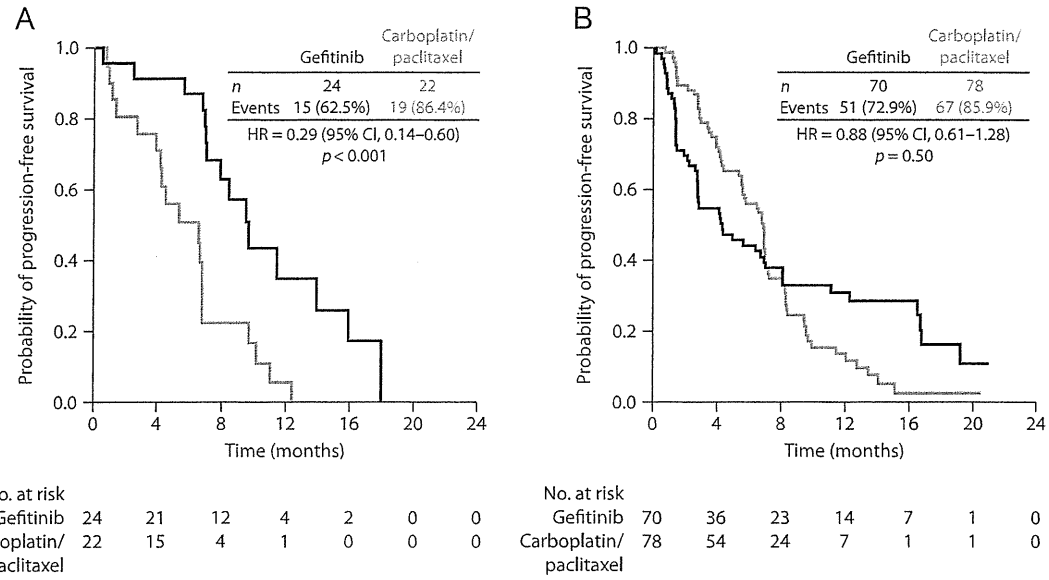

FIGURE 3. Kaplan-Meier curves of progression-free survival in cfDNA *EGFR* mutation-positive (*A*) and cfDNA *EGFR* mutation-negative (*B*) patients in the Japanese subset of IPASS. HR <1 indicates a difference in favor of gefitinib. CI, confidence interval; cfDNA, circulating free DNA; *EGFR*, epidermal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

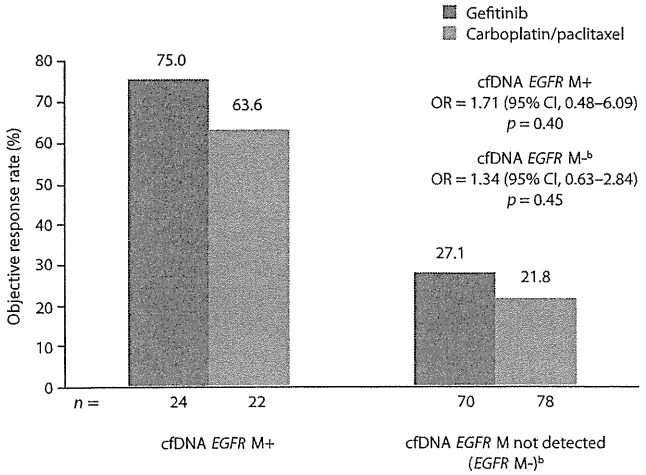

FIGURE 4. Objective response rates by treatment and by cfDNA (serum) *EGFR* mutation status (Japanese ITT population[a]). [a]Refers to the country of recruitment and not necessarily to racial origin. [b]There was a high rate of false-negative results, i.e., this group included both tumor *EGFR* M+ and 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.

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

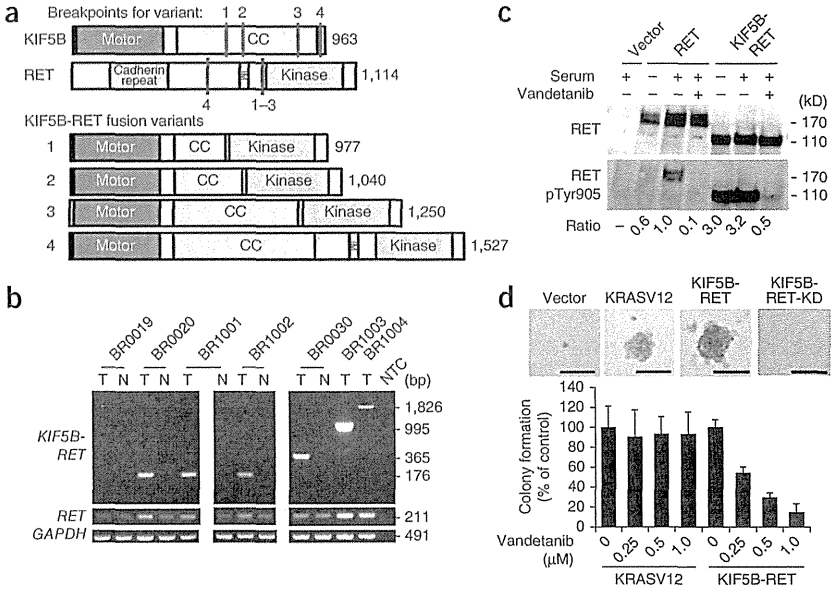

Figure 1 *KIF5B-RET* fusions in LADC. (**a**) Schematic representations of the wild-type KIF5B and RET proteins as well as the four fusion variants identified in this study. The breakpoints for each variant are indicated with red lines. CC, coiled coil; TM, transmembrane. (**b**) Detection of *KIF5B-RET* fusions by RT-PCR. RT-PCR products for the *RET* kinase domain (exons 12 and 13) and *GAPDH* are shown below. Six LADCs positive for *KIF5B-RET* fusions (T) are shown, with four corresponding non-cancerous lung tissues (N), a no-template control (NTC) and one LADC that was negative for the fusion (BR0019). (**c**) Activation of RET kinase activity in the KIF5B-RET protein and the suppression of this activity by vandetanib. H1299 lung cancer cells were transfected with an empty vector, wild-type RET (RET) or KIF5B-RET expression plasmids and treated either with DMSO (serum) or vandetanib, as indicated. The ratios of phosphorylated Tyr905 (pTyr905) RET to total RET signals with respect to wild-type RET after the serum treatment are listed below the gels. (**d**) Anchorage-independent growth of NIH3T3 cells expressing KIF5B-RET protein and the suppression of this growth by vandetanib. Representative pictures of colonies without vandetanib treatment (top). Scale bars, 50 μm. Bar graph showing the percentage (± s.d.) of colonies formed after treatment with the indicated amounts of vandetanib (average results of three independent experiments) with respect to those formed by DMSO-treated cells. The study was approved by the institutional review boards of institutions participating in this study.

United States (an individual of European ancestry) (**Supplementary Fig. 7**), but we detected no fusion transcripts in the 34 subjects from Norway (**Supplementary Table 3**); *KIF5B-RET* fusions occurred in 1–2% of LADCs in both Asians and non-Asians. The individual from the United States with the *RET* fusion was classified as an 'ever smoker', whereas the six individuals from Japan with the *RET* fusion were 'never smokers' (**Table 1**). Therefore, prevalence of LADC with regard to smoking status is unclear. We did not detect the *KIF5B-RET* fusion in other major subtypes of lung cancer, including 234 squamous-cell, 17 large-cell and 20 small-cell lung carcinomas (**Supplementary Table 3**). The fusion was also not present in other types of adenocarcinomas, including those of the ovary ($n = 100$) and colon ($n = 200$) (data not shown), suggesting that it is specific to LADC.

All seven subjects with LADC harboring the *KIF5B-RET* fusion were negative for *EGFR*, *KRAS* and *ALK* mutations or fusions and were negative for mutations in *HER2*, which is an additional driver mutation in LADC[10] (**Table 1** and **Supplementary Table 4**). The mutually exclusive nature of the *RET* fusions and other oncogenic alterations[1,2,11] suggests that the *KIF5B-RET* fusion is a driver mutation. All proteins encoded by the four *KIF5B-RET* fusion variants contained the KIF5B coiled-coil domain, which functions in protein dimerization[12], and retained the full RET kinase domain, similar to other types of oncogenic *RET* fusions observed in thyroid tumors (**Fig. 1a**)[13]. The KIF5B-RET proteins are likely to form a homodimer through the coiled-coil domain of KIF5B, causing an aberrant activation of the kinase function of RET in a manner similar to the *PTC-RET* and *KIF5B-ALK* fusions[7,14]. In fact, the N-terminal portion of the KIF5B coiled-coil region, which is retained in all variants, has been predicted to have the ability to dimerize through two coiled-coil structures[15]. Consistently, when the *KIF5B-RET* variant 1 was exogenously expressed in H1299 human lung cancer cells without wild-type or fusion *RET* expression, Tyr905, which is located in the activation loop of the RET kinase site[15,16], was phosphorylated in the absence of serum stimulation, indicating an aberrant activation of *RET* kinase[16,17] by fusion with *KIF5B* (**Fig. 1c**). This phosphorylation was suppressed by vandetanib, a TKI against RET (as well as other tyrosine kinases, including EGFR and VEGFR)[18] (**Fig. 1c** and **Supplementary Fig. 8**).

Expression of exogenous *KIF5B-RET*, but not *KIF5B-RET*-KD (a kinase-dead mutant corresponding to S765P in wild-type RET[17]), induced morphological transformation (**Supplementary Fig. 9**) and anchorage-independent growth of NIH3T3 fibroblasts in a way that was analogous to the induction caused by mutant *KRAS* (KRASV12) (**Fig. 1d**). Consistently, phosphorylation of Tyr905 was higher in the KIF5B-RET

Table 1 Characteristics of lung adenocarcinomas with the *KIF5B-RET* fusion

Sample	Country	Sex	Age[a]	Smoking	*KIF5B-RET* fusion[b]	Pathological stage	Pathological findings	RET staining	TTF-1 staining	Napsin A staining	Thyroglobulin staining
BR0020	Japan	Male	57	Never	K15; R12 (variant 1)	IIB	Moderately differentiated ADC	+	+	+	–
BR1001	Japan	Female	65	Never	K15; R12 (variant 1)	IB	Well differentiated ADC	+	+	+	–
BR1002	Japan	Female	64	Never	K15; R12 (variant 1)	IB	Well differentiated ADC	+	+	+	–
BR0030	Japan	Male	57	Never	K16; R12 (variant 2)	IA	Well differentiated ADC	+	+	+	–
BR1003	Japan	Male	28	Never	K23; R12 (variant 3)	IA	Well differentiated ADC	+	+	+	–
BR1004	Japan	Female	71	Never	K24; R8 (variant 4)	IA	Moderately differentiated ADC	NT	NT	NT	NT
NCI1580	USA	Male	63	Ever[c]	K15; R12 (variant 1)	II	Moderately differentiated ADC	NT	NT	NT	NT

[a]Age in years. [b]Fused exon numbers of *KIF5B* (K) and *RET* (R); and variant types (in parentheses) are shown. None of the subjects had oncogenic *EGFR*, *KRAS*, *HER2* or *ALK* mutations or fusions. [c]The number of pack years smoked for this subject is not known. NT, not tested.

protein than in the KIF5B-RET-KD protein. The anchorage-independent growth induced by *KIF5B-RET* was suppressed by treatment with vandetanib (<1 μM), whereas the growth induced by mutant *KRAS* was not (**Fig. 1d**). These results are similar to those observed for *RET* fusions in thyroid cancer[19]. We also detected phosphorylation of the KIF5B-RET protein at Tyr905 in fusion-positive LADC specimens (**Supplementary Fig. 6**). These results suggest that the *RET* fusions are a previously unidentified LADC driver mutation and a potential target for existing TKIs, including vandetanib, which has been recently approved by the US Food and Drug Administration for the treatment of thyroid cancer[18]. Further studies are warranted to promote molecular subtype diagnoses and personalized therapy options for LADC. For this purpose, both the clinical and biological features of this fusion are being investigated. For further information, please see the **Supplementary Note** and **Supplementary Tables 5** and **6**.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

This work was supported in part by the program for promotion of Fundamental Studies in Health Sciences of 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, the National Cancer Center Research and Development Fund and the Norwegian Cancer Society. National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund, Japan. We thank T. Urushidate, S. Ohashi, S. Mitani, K. Yokozawa, S. Wakai, C. Otsubo and H. Isomura of the National Cancer Center and D. Suzuki and K. Nagase of the National Center for Global Health and Medicine for technical assistance. We also thank J.D. Minna and L. Girard of the University of Texas Southwestern Medical Center, K. Kumamoto of Saitama Medical University and A. Okamoto of Jikei University for *RET* fusion screening, N. Morii of the National Institute of Advanced Industrial Science and Technology (AIST) for thermodynamic characterization of the KIF5B protein and M. Maekawa of the GSP laboratory for rapid preparation of the FISH probes.

AUTHOR CONTRIBUTIONS

RNA sequencing: H.I., K.Y., M.H., T.N. and H.S. Sequence data processing: Y.T., S.C. and I.Y. Molecular biological analyses: T.K., Y.S., R.I., H. Ogiwara, T.O., M.E., A.J.S., H. Okayama, A.H., Y.A. and S.O. Clinical and pathological analyses: K.T., K.F., V.S., S.W., I.S. and H.T. Manuscript writing: T.K., H.I. and T.S. Study design: T.K., H.I., C.C.H., T.Y., J.Y. and T.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemedicine/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Janku, F., Stewart, D.J. & Kurzrock, R. *Nat. Rev. Clin. Oncol.* **7**, 401–414 (2010).
2. Gerber, D.E. & Minna, J.D. *Cancer Cell* **18**, 548–551 (2010).
3. Lovly, C.M. & Carbone, D.P. *Nat. Rev. Clin. Oncol.* **8**, 68–70 (2011).
4. Soda, M. *et al. Nature* **448**, 561–566 (2007).
5. Meyerson, M., Gabriel, S. & Getz, G. *Nat. Rev. Genet.* **11**, 685–696 (2010).
6. Mani, R.S. & Chinnaiyan, A.M. *Nat. Rev. Genet.* **11**, 819–829 (2010).
7. Wells, S.A. Jr. & Santoro, M. *Clin. Cancer Res.* **15**, 7119–7123 (2009).
8. Bishop, J.A., Sharma, R. & Illei, P.B. *Hum. Pathol.* **41**, 20–25 (2010).
9. DeLellis, R.A., Shin, S.J. & Treaba, D.O. *Immunohistology of Endocrine Tumors* (Saunders, Philadelphia, Pennsylvania, USA, 2010).
10. Shigematsu, H. *et al. Cancer Res.* **65**, 1642–1646 (2005).
11. Herbst, R.S., Heymach, J.V. & Lippman, S.M. *N. Engl. J. Med.* **359**, 1367–1380 (2008).
12. Hirokawa, N., Noda, Y., Tanaka, Y. & Niwa, S. *Nat. Rev. Mol. Cell Biol.* **10**, 682–696 (2009).
13. Jhiang, S.M. *Oncogene* **19**, 5590–5597 (2000).
14. Takeuchi, K. *et al. Clin. Cancer Res.* **15**, 3143–3149 (2009).
15. Morii, H., Takenawa, T., Arisaka, F. & Shimizu, T. *Biochemistry* **36**, 1933–1942 (1997).
16. Vitagliano, D. *et al. Endocr. Relat. Cancer* **18**, 1–11 (2011).
17. Croyle, M. *et al. Cancer Res.* **68**, 4183–4191 (2008).
18. Commander, H., Whiteside, G. & Perry, C. *Drugs* **71**, 1355–1365 (2011).
19. Carlomagno, F. *et al. Cancer Res.* **62**, 7284–7290 (2002).

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

Molecular and Cellular Pathobiology

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

Published OnlineFirst November 11, 2011; 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

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

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

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

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*-

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

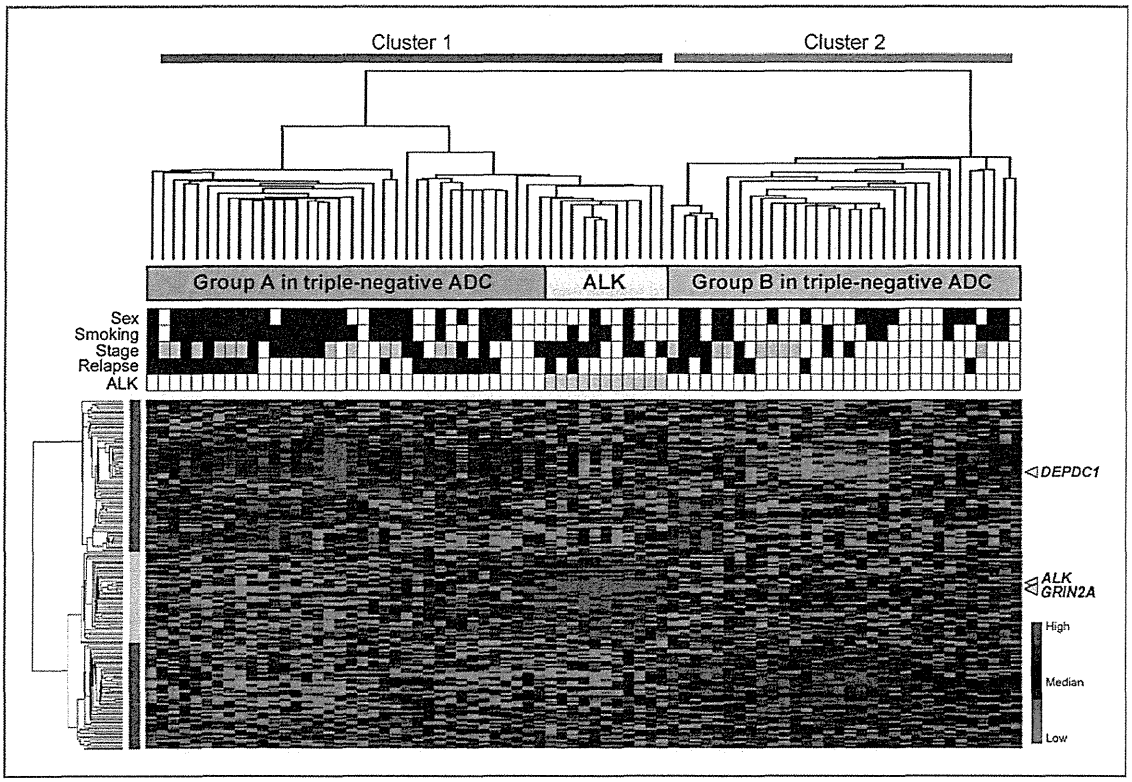

Figure 1. Unsupervised hierarchical clustering of 11 *ALK*-positive adenocarcinomas and 68 triple-negative adenocarcinomas. Triple-negative adenocarcinomas were separated into 36 group A cases and 32 group B cases, and group A cases construct cluster 1 with 11 *ALK*-positive adenocarcinoma cases. Clinical 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

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

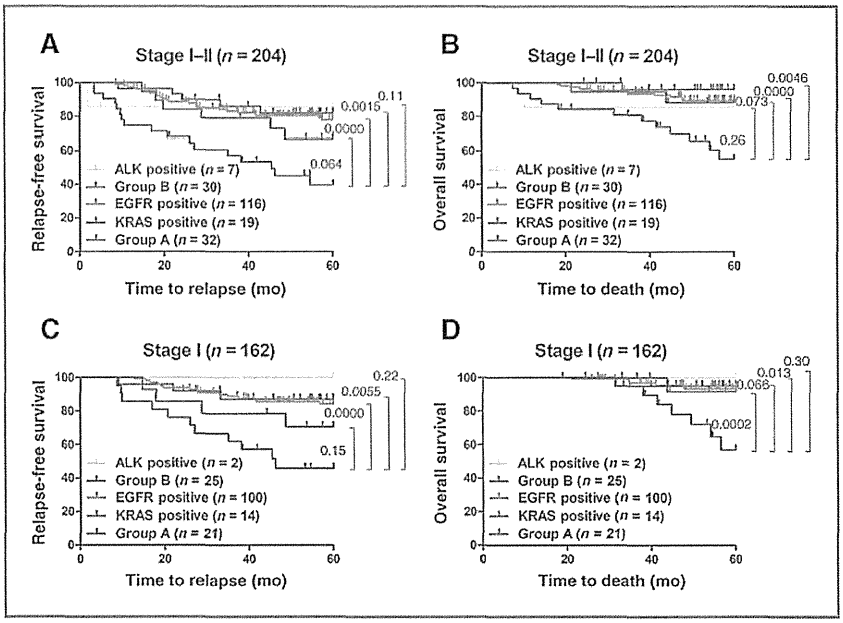

Figure 2. Kaplan–Meier survival curves for RFS and OS of 204 lung adenocarcinoma cases according to *EGFR*-positive, *KRAS*-positive, *ALK*-positive, group A, and group B. RFS and OS of stage I–II (A, B) and stage I (C, D) cases are shown.

of *KRAS*-positive and *ALK*-positive cases. Accordingly, multivariate analyses of 162 stage I patients further showed significant differences in RFS and OS between group A cases and *EGFR*-positive cases, and also between group A cases and group B cases. Because numbers of *KRAS*-positive cases and *ALK*-positive cases were small, we next compared RFS and OS between group A patients and patients in all 4 other groups combined ("Others" in Table 3). Differences in RFS as well as those in OS were highly significant and independent of staging. These results strongly indicated that group A patients comprise a distinct subclass of *EGFR/KRAS/ALK*-negative lung adenocarcinomas, and the prognoses of group A patients were the worst among the 5 groups of patients.

Clustering of lung adenocarcinomas with poor prognosis by gene expression profiling

We next carried out unsupervised hierarchical clustering of all the 226 adenocarcinoma cases, including 127 *EGFR*-positive cases and 20 *KRAS*-positive cases, to investigate whether expression profiling with a set of 174 genes with 190 probes could extract group A cases as a unique subset among all adenocarcinomas, and whether the profiling could be useful for prognosis prediction of patients with any genotypes of adenocarcinomas in general. As shown in Supplementary Fig. S4, clustering patterns of all the 226 patients were very similar to those of the 79 patients consisting of 11 *ALK*-positive cases and 68 triple-negative cases. In particular, the 11 *ALK*-positive cases comprised a small cluster in the right side of Cluster 1 (Cluster 1b), supporting that *ALK*-positive adenocarcinomas show unique expression profiles among all adenocarcinomas. Group A and group B cases also have a tendency to accumulate in Clusters 1a and Cluster 2, respectively. However, group A cases often comprise clusters with the *KRAS*-positive cases, whereas group B cases were distributed with the *EGFR*-positive cases. Therefore, group A and group B triple-negative adenocarcinomas were not exclusive with the *EGFR*-positive and *KRAS*-positive adenocarcinomas by expression profiling of these 174 genes. Therefore, expression profiling with a set of the 174 genes was concluded to be useful to distinguish *ALK*-positive adenocarcinomas among all lung adenocarcinomas.

However, RFS of 119 patients in Cluster 1 was significantly worse than RFS of 85 patients in Cluster 2 (HR = 3.73, P = 0.00016). When Cluster 1 was further divided into 2 subclasses 1a and 1b of the right and left sides, respectively, Cluster 1a containing most of group A patients showed the worst prognosis among the 3 subclasses (Supplementary Fig. S4). Therefore, the expression signature of these 174 genes was indicated to be useful for prognostic prediction of adenocarcinoma patients, in particular of triple-negative adenocarcinoma patients.

Minimum set of genes characterizing triple-negative lung adenocarcinomas with poor prognosis

The above results implied that triple-negative adenocarcinomas can be classified into 2 distinct subgroups by expression profiling and prognoses of these 2 groups are significantly different from each other. Accordingly, expression of several genes among the 174 genes was expected to be independently associated with prognosis of triple-negative adenocarcinoma patients. Therefore, we next selected genes whose expression was associated with prognosis from the 174 genes evaluated by the 190 probes. To evaluate the prognostic value of each probe and to make a comparative study for association of gene expression with prognosis in other cohorts possible, we took a minimum P value approach for grouping the patients for survival analysis because of the following reason. A database

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

Table 3. Hazard ratios for relapse-free and overall survivals in lung adenocarcinomas

			Univariate		Multivariate	
Survival	**Case (*n*)**	**Variable**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Relapse free	Stage I–II (204)	Age	1.03 (0.99–1.07)	0.12	1.04 (0.99–1.08)	0.092
		Sex (male/female)	1.39 (0.82–2.38)	0.22	1.00 (0.49–2.05)	0.99
		Smoking habit (ever/never)	1.43 (0.84–2.44)	0.19	1.10 (0.54–2.24)	0.80
		pStage (II/I)	1.86 (1.41–2.45)	1.3E-05	3.50 (1.93–6.34)	3.6E-05
		Subgroup				
		Group A/ALK (+)	4.78 (0.63–35.99)	0.13	6.01 (0.76–47.82)	0.09
		Group A/KRAS (+)	2.43 (0.96–6.17)	0.062	2.85 (1.10–7.35)	0.031
		Group A/EGFR (+)	3.58 (1.93–6.64)	5.3E-05	2.76 (1.44–5.29)	0.0022
		Group A/Group B	4.58 (1.69–12.42)	0.0028	4.10 (1.50–11.24)	0.0061
		Group A/Others	3.56 (2.00–6.34)	1.6E-05	3.04 (1.68–5.53)	2.5E-04
	Stage I (162)	Age	1.01 (0.96–1.06)	0.69	1.00 (0.95–1.05)	0.97
		Sex (male/female)	0.99 (0.50–1.96)	0.98	0.83 (0.33–2.07)	0.69
		Smoking habit (ever/never)	1.06 (0.54–2.08)	0.87	0.97 (0.39–2.45)	0.95
		Subgroup				
		Group A/ALK (+)	—	—	—	—
		Group A/KRAS (+)	2.31 (0.73–7.28)	0.15	2.36 (0.73–7.62)	0.15
		Group A/EGFR (+)	4.33 (2.00–9.35)	2.0E-04	4.51 (2.05–9.91)	1.7E-04
		Group A/Group B	5.36 (1.49–19.24)	0.010	5.52 (1.50–20.37)	0.010
		Group A/Others	4.18 (2.03–8.60)	1.0E-04	4.32 (2.06–9.09)	1.1E-04
Overall	Stage I–II (204)	Age	1.03 (0.98–1.08)	0.33	1.03 (0.98–1.09)	0.21
		Sex (male/female)	1.69 (0.82–3.48)	0.16	0.89 (0.33–2.41)	0.82
		Smoking habit (ever/never)	1.91 (0.92–3.97)	0.084	1.46 (0.54–3.92)	0.45
		pStage (II/I)	2.07 (1.45–2.97)	7.2E-05	3.93 (1.83–8.44)	4.6E-04
		Subgroup				
		Group A/ALK (+)	2.95 (0.38–22.78)	0.30	3.50 (0.41–29.85)	0.25
		Group A/KRAS (+)	3.12 (0.88–11.09)	0.079	3.31 (0.91–12.03)	0.069
		Group A/EGFR (+)	4.59 (2.06–10.23)	2.0E-04	3.35 (1.44–7.81)	0.005
		Group A/Group B	6.83 (1.53–30.54)	0.012	5.68 (1.24–25.95)	0.025
		Group A/Others	4.50 (2.17–9.36)	5.7E-05	3.61 (1.68–7.78)	0.0010
	Stage I (162)	Age	0.99 (0.93–1.06)	0.73	0.98 (0.91–1.04)	0.45
		Sex (male/female)	1.15 (0.43–3.08)	0.79	0.77 (0.20–3.00)	0.70
		Smoking habit (ever/never)	1.47 (0.55–3.91)	0.45	1.26 (0.32–4.89)	0.74
		Subgroup				
		Group A/ALK (+)	—	—	—	—
		Group A/KRAS (+)	5.79 (0.71–47.3)	0.10	5.61 (0.67–46.84)	0.11
		Group A/EGFR (+)	5.83 (2.04–16.71)	0.0010	6.06 (2.08–17.71)	9.8E-04
		Group A/Group B	9.13 (1.12–74.34)	0.039	9.32 (1.10–78.61)	0.040
		Group A/Others	6.30 (2.34–16.99)	2.8E-04	6.47 (2.33–17.98)	3.4E-04

named PrognoScan was recently developed by coauthors of this study (26). In the PrognoScan database, minimum *P* values for the association of gene expression with prognosis of all probes in a platform are available for a number of cohorts that have been published. Therefore, it was possible to validate the present findings using data from various other cohorts by the same criteria. According to the method described previously (26), corrected minimum *P* values were calculated for each probe to control the error rate for the evaluation of the association with RFS and OS. Expression of 11 genes evaluated with 12 probes (2 probes for the *DEPDC1* gene) showed significant associations with both RFS and OS in 62 triple-negative adenocarcinomas and also in 46 stage I triple-negative adenocarcinomas (Table 4). Among the 11 genes, expression of 10 genes was positively correlated with poor prognosis, whereas that of the remaining 1 gene, *KIF19*, expression was negatively correlated with poor prognosis.

We first selected 174 genes as being preferentially expressed in either *ALK*-positive adenocarcinomas or triple-negative adenocarcinomas by the criteria of "probes whose expression levels in any adenocarcinomas with *EGFR* or *KRAS* mutations were lower than the mean expression

Table 4. List of genes whose expression is associated with relapse free survival and overall survival of patients with lung adenocarcinoma

		NCC														CAN/DF		HLM		MSK		UM		Nagoya		Duke		Seoul	
		TN, Stage I–II				TN, Stage I				All Stage I–II				All Stage I		Stage I–III		Stage I–III		Stage I–III		Stage I–III		Stage I–III		Stage I–III		Stage I–III	
Dataset		RFS $n = 62$		OS $n = 62$		RFS $n = 46$		OS $n = 46$		RFS $n = 204$		OS $n = 204$		RFS $n = 162$		OS $n = 162$		OS $n = 82$		OS $n = 79$		OS $n = 104$		OS $n = 178$		OS $n = 117$		RFS $n = 111$		RFS $n = 138$	
Gene symbol	Probe ID (for NCC)	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR	*P*	HR
DEPDC1	222958_s_at	0.00	2.3	0.00	3.0	0.00	3.0	0.02	2.7	0.00	2.1	0.00	1.8	0.00	2.0	0.00	2.1	—	—	—	—	0.03	1.1	—	—	0.00	1.6	0.04	1.0	0.01	0.9
DEPDC1	235545_at	0.00	1.8	0.01	2.3	0.00	2.4	0.04	2.6	0.00	1.4	0.00	1.6	0.00	1.3	0.00	2.2														
FOSL2	218881_s_at	0.01	1.7	0.03	1.7	0.02	1.8	0.00	3.3	0.00	1.2	0.00	1.7	0.00	1.4	0.00	2.4	—	—	0.00	1.7	—	—	0.02	0.7	—	—	—	—	0.01	1.0
MCM4	222037_at	0.00	1.8	0.00	3.0	0.01	2.0	0.04	2.6	0.00	1.4	0.00	1.8	0.00	1.5	0.00	2.1	—	—	—	—	—	—	—	—	0.00	1.7	—	—	—	—
UBE2S	202779_s_at	0.00	3.0	0.02	16.0	0.01	2.8	0.02	16.6	0.00	1.6	0.02	1.4	0.00	1.6	0.00	2.1	—	—	—	—	0.05	1.0	—	—	—	—	—	—	—	—
CD300A	217078_s_at	0.01	1.7	0.00	2.1	0.04	1.7	0.00	2.8	0.00	1.1	0.00	1.5	0.01	1.2	0.01	1.7	—	—	—	—	0.00	1.5	—	—	—	—	—	—	—	—
SLITRK4	232636_at	0.02	1.7	0.03	1.7	0.00	2.9	0.00	2.5	0.01	1.1	0.00	1.4	0.04	1.1	0.00	2.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—
KRT16	209800_at	0.00	2.0	0.00	2.5	0.00	2.4	0.00	2.7	0.00	1.2	0.00	1.4	0.01	1.2	0.01	1.7	—	—	—	—	—	—	—	—	—	—	—	—	—	—
SIGLEC9	210569_s_at	0.00	1.9	0.01	2.0	0.00	2.1	0.04	2.1	0.00	1.6	0.00	1.4	0.00	1.7	0.00	2.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—
DIAPH3	232596_at	0.02	1.5	0.00	3.0	0.05	1.6	0.03	2.6	0.00	1.2	0.00	2.1	0.00	1.5	0.00	2.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—
LOC152225	1562048_at	0.01	1.5	0.00	2.3	0.02	1.7	0.00	2.7	0.00	1.3	0.00	1.9	—	—	0.00	1.9	—	—	—	—	—	—	—	—	—	—	—	—	—	—
KIF19	1553314_a_at	0.01	-1.5	0.05	-1.6	0.00	-3.0	0.00	−2.5	—	—	0.03	−1.4	—	—	—	—	—	—	—	—	—	—	—	—	0.00	−1.4	—	—	—	—

Abbreviations: NCC, National Cancer Center; TN, Triple-negative.

HRs (log2 ratio) with corrected *P* value < 0.05 are shown.

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

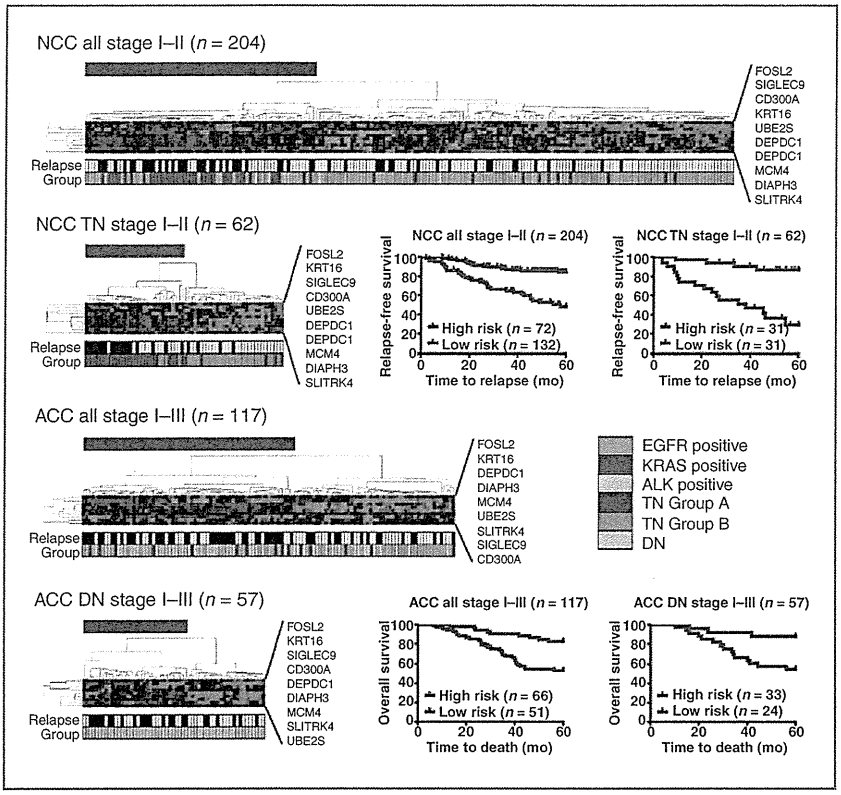

Figure 3. Unsupervised hierarchical clustering based on the expression of a set of 9 genes. All 204 stage I–II adenocarcinomas and 62 triple-negative (TN) stage I–II adenocarcinomas of the National Cancer Center (NCC) data set subjected to survival analysis were analyzed, and a cluster with higher expression of these genes than the other cluster was recognized as a high-risk group (red bar). Results of 117 adenocarcinomas, including 57 double-negative (DN) adenocarcinomas, of the Aichi Cancer Center (ACC) data set are shown below.

level of a total of 54,675 probes." Then, 11 of the 174 genes were further selected as being associated with prognosis of patients with triple-negative adenocarcinomas. Therefore, higher expression of several genes among the 11 genes was predicted to be associated with poorer prognosis, even when all adenocarcinoma cases, including *EGFR*-positive, *KRAS*-positive, and *ALK*-positive adenocarcinomas were analyzed together. Furthermore, triple-negative adenocarcinomas with poor prognosis would be separated into a high-risk group classified with this procedure. For this reason, we next analyzed all 204 adenocarcinoma cases. Among the 11 genes with 12 probes, 9 genes with 10 probes showed significant associations with both RFS and OS in all 204 adenocarcinoma cases and also in 162 stage I adenocarcinoma cases. *LOC152225* and *KIF19* were excluded because of no significant associations in stage I adenocarcinoma cases. As predicted, higher expression of the 9 genes was correlated with poorer prognosis in the analysis of RFS and OS among 204 stages I and II cases and also among 162 stage I cases.

The result strongly indicated that unsupervised hierarchical clustering using this 10 probe set (9 genes) would separate the patients into high-risk and low-risk groups for prognosis and that all group A triple-negative adenocarcinoma patients with poor prognosis would be classified into the high-risk group (Fig. 3 and Supplementary Table S3). As expected, expression profiling of these 9 genes successfully separated the 204 patients into high-risk and low-risk groups with significantly different RFS (HR = 3.79, 95% CI = 2.19–6.55, P = 1.9E-06) as well as OS (HR = 5.72, 95% CI = 2.53–12.87, P = 2.5E-05). Furthermore, if 62 triple-negative cases only were separated with these 9 genes, HRs for both RFS and OS were much higher than those with separation of all the 204 cases. All the relapsed cases in group A were separated into the high-risk group in the analyses of both cases (all the 204 cases and the 62 triple-negative cases only), supporting that triple-negative adenocarcinomas cases with poor prognosis can be selected as a high-risk group from all the adenocarcinoma cases by expression profiling of these 9 genes (Fig. 3). This profiling further separated 162 stage I cases as well as 46 stage I triple-negative adenocarcinoma cases into high-risk and low-risk groups with significantly different RFS as well as OS (Supplementary Fig. S5 and Supplementary Table S3). Again, HRs for both RFS and OS were much higher in triple-negative adenocarcinoma cases than in all adenocarcinoma cases. Accordingly, high levels of expression in these 9 genes were concluded to be distinct characteristics of triple-negative adenocarcinomas with poor prognosis.

Validation of associations using independent expression profiling data

To validate the present findings using the data of other cohorts, we searched for expression profiling data with

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

mutation data of the *EGFR*, *KRAS*, and *ALK* genes in various databases. However, there has been no cohort in which expression profiles specifically in triple-negative adenocarcinomas were analyzed. Therefore, unsupervised hierarchical clustering using these 9 genes was done on a cohort of 117 Japanese lung adenocarcinoma cases because expression profile data as well as *EGFR/KRAS* mutation data were available only in this cohort (32). This study included 57 adenocarcinoma cases without *EGFR* and *KRAS* mutations. Although a different array platform was used, the data for all the 9 genes were available for clustering. These cases were separated into 2 groups of 33 cases and 24 cases (Fig. 3). OS of the 33 cases was significantly shorter than that of the 24 cases (HR = 3.17, 95% CI = 1.17–8.63, $P = 2.4E-02$; Supplementary Table S3). As with our cohort, the high-risk group showed a significantly higher HR of 2.73, even when all the 117 cases were analyzed together. Although *ALK* mutation data were not available for this cohort, the results strongly supported that expression profiling of the 9 genes would be highly informative for prediction of prognosis of lung adenocarcinoma patients, in particular patients with *EGFR*- and *KRAS*-negative adenocarcinomas.

Associations of *DEPDC1* expression with prognosis of NSCLC patients

Associations of gene expression with prognosis in various cancers are available from the PrognoScan database (22). Therefore, associations of expression of these 9 genes with prognosis of NSCLC patients were examined in 7 other cohorts (Table 4). Notably, *DEPDC1* expression was positively associated with poor prognosis in 4 of the 7 cohorts; MSK, Nagoya, Duke, and Seoul. The results strongly indicated that *DEPDC1* expression can be a novel prognostic marker for patients with NSCLC. Representative data showing the association of *DEPDC1* expression with prognosis in 204 adenocarcinoma patients obtained from the minimum *P* value approach are shown in Supplementary Fig. S6. Associations of *DEPDC1* expression with RFS and OS were validated by quantitative RT-PCR analysis of 204 stages I and II cases and also of 162 stage I cases (Supplementary Fig. S3).

FOSL2 expression was associated with prognosis in 3 of the 7 cohorts, whereas *MCM4*, *CD300A*, and *UBE2S* expression was associated in 1 cohort, respectively (Table 4).

Discussion

In this study, we attempted to characterize *ALK*-positive adenocarcinomas and triple-negative adenocarcinomas by genome-wide expression profiling. For this purpose, we selected a set of genes that are not transcriptionally activated in any *EGFR*-positive and *KRAS*-positive adenocarcinomas, and obtained 2 pieces of unique evidence. One is that *ALK*-positive adenocarcinomas show unique expression profiles in comparison with any other types of adenocarcinomas. The other is that there is a group of patients with extremely poor prognosis among triple-negative adenocarcinomas. This group, herein designated as group A, of patients showed much worse prognoses than patients with *EGFR*, *KRAS*, or *ALK* mutations and also than the other group, group B, of patients with triple-negative adenocarcinomas.

ALK-positive adenocarcinomas are sensitive to ALK TKIs with an overall response rate of 55% (8). Therefore, for the clinical application of ALK-targeted therapy, it is indispensable to develop a simple and reliable method for detection of *ALK* rearrangements in lung adenocarcinomas. Here, we showed that *ALK* expression is exclusively high only in *ALK*-positive adenocarcinomas and that several other genes, including *GRIN2A*, are overexpressed together with *ALK* specifically in *ALK*-positive adenocarcinomas. Therefore, GRIN2A can be a biomarker for detection of *ALK*-positive adenocarcinomas. *GRIN2A* encodes an *N*-methyl-D-aspartate (NMDA) receptor, which is a neurotransmitter-gated ion channel involved in regulation of synaptic function in the central nervous system (33). It was noted that the *GRIN2A* gene was recently reported to be frequently mutated in melanoma (34). Therefore, although the biological significance of *GRIN2A* upregulation in *ALK*-positive adenocarcinomas remains unclear, *GRIN2A* expression may play some important role in the phenotype unique to *ALK*-positive adenocarcinomas. Expression profiles unique to *ALK*-positive adenocarcinomas, shown here, will be also informative to improve clinical detection of *ALK* rearrangements.

Group A cases were discriminated by expression profiling of 9 genes among stage I–II cases who received complete surgical resection of tumors. Therefore, this gene set will be applicable as biomarkers to select lung adenocarcinoma patients who will benefit from adjuvant therapy after surgery, in particular to select them among patients with triple-negative adenocarcinomas. For this reason, combined analyses of this expression profiling with mutational analyses of the *EGFR*, *KRAS*, and *ALK* genes will be appropriate to pick out triple-negative adenocarcinoma patients with poor prognosis from all the adenocarcinoma patients. Molecular targeting drugs against triple-negative adenocarcinomas are not available at present; therefore, genes upregulated in group A cases will also be applicable as targets for therapy. *DEPDC1* was previously identified as being upregulated in bladder cancer and breast cancer (35–37). Because *DEPDC1* expression was hardly detectable in any normal tissues except testis, it has been considered as a cancer/testis antigen and also as a promising target of therapeutic drugs (35, 36). This study showed that *DEPDC1* is preferentially expressed in triple-negative adenocarcinomas with poor prognosis. In the PrognoScan database, *DEPDC1* expression is shown to be positively associated with poor prognosis in bladder cancer, multiple myeloma, breast cancer, glioma, and melanoma. Therefore, *DEPDC1* could be a novel target for diagnosis as well as therapy in various cancers, including lung adenocarcinoma.

Identification of genetic alterations that occur specifically in group A cases will be also of great importance for the development of target therapy for stages I and II lung adenocarcinoma patients with poor outcomes. Group A cases include males and ever-smokers as a majority (Table 1); therefore, group A cases were likely to carry several genetic alterations induced by tobacco carcinogens leading to poor outcomes. Identification of genetic alterations in

Published OnlineFirst November 11, 2011; DOI: 10.1158/0008-5472.CAN-11-1403

group A adenocarcinomas will further facilitate the development of targeted therapies for lung adenocarcinomas with poor prognosis.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Dr. Teruhiko Yoshida and Ms. Sachiyo Mimaki for their efforts in expression profiling.

Grant Support

This work was supported in part by grants-in-aid from the Ministry of Health, Labor and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control and from the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO: 10–41). K. Shiraishi was an awardee of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan.

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

Received April 23, 2011; revised October 19, 2011; accepted October 21, 2011; published OnlineFirst November 11, 2011.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893–917.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108.
3. Colvy T, Noguchi M, Henschke C, Vazquez M, Geisinger K, Yokose T, et al. Adenocarcinoma. In:Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. World Health Organization classification of tumors: pathology and genetics; tumours of the lung, pleura, thymus and heart. IARC Press: Lyon, France; 2004. p. 35–44.
4. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. Lancet Oncol 2011;12:175–80.
5. Herbst RS, Heymach JV, Lippman SM: Lung cancer. N Engl J Med 2008;359:1367–80.
6. Janku F, Stewart DJ, Kurzrock R. Targeted therapy in non-small-cell lung cancer–is it becoming a reality? Nat Rev Clin Oncol 2010;7: 401–14.
7. Bronte G, Rizzo S, La Paglia L, Adamo V, Siragusa S, Ficorella C, et al. 2288; Driver mutations and differential sensitivity to targeted therapies: a new approach to the treatment of lung adenocarcinoma. Cancer Treat Rev 2010;36 Suppl 3:S21–9.
8. Gerber DE, Minna JD. ALK inhibition for non-small cell lung cancer: from discovery to therapy in record time. Cancer Cell 2010;18:548–51.
9. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers–a different disease. Nat Rev Cancer 2007;7:778–90.
10. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 2009;361:958–67.
11. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al.: Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009;361:947–57.
12. Konstantinopoulos PA, Karamouzis MV, Papavassiliou AG. Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets. Nat Rev Drug Discov 2007;6:541–55.
13. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 2007;448:561–6.
14. Takeuchi T, Tomida S, Yatabe Y, Kosaka T, Osada H, Yanagisawa K, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 2006;24:1679–88.
15. Angulo B, Suarez-Gauthier A, Lopez-Rios F, Medina PP, Conde E, Tang M, et al. Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK3CA overexpression by gene amplification. J Pathol 2008;214:347–56.
16. Motoi N, Szoke J, Riely GJ, Seshan VE, Kris MG, Rusch VW, 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;32:810–27.
17. Meyerson M, Carbone D. Genomic and proteomic profiling of lung cancers: lung cancer classification in the age of targeted therapy. J Clin Oncol 2005;23:3219–26.
18. Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. World Health Organization classification of tumors: pathology and genetics; tumours of the lung, pleura, thymus and heart. IARC Press: Lyon, France; 2004. p. 1–344.
19. Saito M, Schetter AJ, Mollerup S, Kohno T, Skaug V, Bowman ED, et al. The association of microRNA expression with prognosis and progression in early-stage, non-small cell lung adenocarcinoma: a retrospective analysis of three cohorts. Clin Cancer Res 2011;17: 1875–82.
20. Robins JM, Gail MH, Lubin JH. More on "Biased selection of controls for case-control analyses of cohort studies". Biometrics 1986;42: 293–9.
21. Richardson DB. An incidence density sampling program for nested case-control analyses. Occup Environ Med 2004;61:e59.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–8. Available from: http://rana.lbl.gov/eisen/.
23. Takano T, Ohe Y, Sakamoto H, Tsuta K, Matsuno Y, Tateishi U, et al. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. J Clin Oncol 2005;23: 6829–37.
24. Takano T, Ohe Y, Tsuta K, Fukui T, Sakamoto H, Yoshida T, et al. Epidermal growth factor receptor mutation detection using high-resolution melting analysis predicts outcomes in patients with advanced non small cell lung cancer treated with gefitinib. Clin Cancer Res 2007;13:5385–90.
25. Takeuchi K, Choi YL, Soda M, Inamura K, Togashi Y, Hatano S, et al. Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. Clin Cancer Res 2008;14:6618–24.
26. Mizuno H, Kitada K, Nakai K, Sarai A. 2288; PrognoScan: a new database for meta-analysis of the prognostic value of genes. 2288; BMC Med Genomics 2009;2:18. Available from: http://gibk21.bio.kyutech.ac.jp/PrognoScan/index.html.
27. Mitsudomi T. Advances in target therapy for lung cancer. Jpn J Clin Oncol 2010;40:101–6.
28. Suda K, Tomizawa K, Mitsudomi T. Biological and clinical significance of KRAS mutations in lung cancer: an oncogenic driver that contrasts with EGFR mutation. Cancer Metastasis Rev 2010;29: 49–60.
29. Zhang X, Zhang S, Yang X, Yang J, Zhou Q, Yin L, et al. Fusion of EML4 and ALK is associated with development of lung adenocarcinomas lacking EGFR and KRAS mutations and is correlated with ALK expression. Mol Cancer 2010;9:188.
30. Boland JM, Erdogan S, Vasmatzis G, Yang P, Tillmans LS, Johnson MR, et al. Anaplastic lymphoma kinase immunoreactivity correlates with ALK gene rearrangement and transcriptional up-regulation in non-small cell lung carcinomas. Hum Pathol 2009; 40:1152–8.
31. Mino-Kenudson M, Chirieac LR, Law K, Hornick JL, Lindeman N, Mark EJ, et al. A novel, highly sensitive antibody allows for the routine detection of ALK-rearranged lung adenocarcinomas by standard immunohistochemistry. Clin Cancer Res 2010;16:1561–71.