

Mass Spectrometry to Classify Non–Small-Cell Lung Cancer Patients for Clinical Outcome After Treatment With Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors: A Multicohort Cross-Institutional Study

Fumiko Taguchi, Benjamin Solomon, Vanesa Gregorc, Heinrich Roder, Robert Gray, Kazuo Kasahara, Makoto Nishio, Julie Brahmer, Anna Spreafico, Vienna Ludovini, Pierre P. Massion, Rafal Dziadziuszko, Joan Schiller, Julia Grigorieva, Maxim Tsybin, Stephen W. Hunsucker, Richard Caprioli, Mark W. Duncan, Fred R. Hirsch, Paul A. Bunn Jr, David P. Carbone

- Background** Some but not all patients with non–small-cell lung cancer (NSCLC) respond to treatment with epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs). We developed and tested the ability of a predictive algorithm based on matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) analysis of pretreatment serum to identify patients who are likely to benefit from treatment with EGFR TKIs.
- Methods** Serum collected from NSCLC patients before treatment with gefitinib or erlotinib were analyzed by MALDI MS. Spectra were acquired independently at two institutions. An algorithm to predict outcomes after treatment with EGFR TKIs was developed from a training set of 139 patients from three cohorts. The algorithm was then tested in two independent validation cohorts of 67 and 96 patients who were treated with gefitinib and erlotinib, respectively, and in three control cohorts of patients who were not treated with EGFR TKIs. The clinical outcomes of survival and time to progression were analyzed.
- Results** An algorithm based on eight distinct *m/z* features was developed based on outcomes after EGFR TKI therapy in training set patients. Classifications based on spectra acquired at the two institutions had a concordance of 97.1%. For both validation cohorts, the classifier identified patients who showed improved outcomes after EGFR TKI treatment. In one cohort, median survival of patients in the predicted “good” and “poor” groups was 207 and 92 days, respectively (hazard ratio [HR] of death in the good versus poor groups = 0.50, 95% confidence interval [CI] = 0.24 to 0.78). In the other cohort, median survivals were 306 versus 107 days (HR = 0.41, 95% CI = 0.17 to 0.63). The classifier did not predict outcomes in patients who did not receive EGFR TKI treatment.
- Conclusion** This MALDI MS algorithm was not merely prognostic but could classify NSCLC patients for good or poor outcomes after treatment with EGFR TKIs. This algorithm may thus assist in the pretreatment selection of appropriate subgroups of NSCLC patients for treatment with EGFR TKIs.

J Natl Cancer Inst 2007;99:838–46

The epidermal growth factor receptor (EGFR) was identified as a potential target for lung cancer therapy because of its frequent expression in lung cancer tissue and the importance of EGF and transforming growth factor α in supporting the growth of lung

Affiliations of authors: Departments of Medicine (FT, DPC), Medicine, Pulmonary Division (PPM), Biochemistry (RC), and Cancer Biology (DPC), Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN; Departments of Medical Oncology (BS, FRH, PAB), and Pediatrics (SWH, MWD), University of Colorado at Denver and Health Sciences Center, Aurora, CO; Department of Oncology, Scientific Institute University Hospital San Raffaele, Milan, Italy (VG, AS); Biodesix, Steamboat Springs, CO (HR, JG, MT); Department of Biostatistics and Computational Biology, ECOG Biostatistical Office, Boston, MA (RG); Department of Respiratory Medicine, Kanazawa University, Kanazawa, Japan (KK); Thoracic

Oncology Center, Japanese Foundation for Cancer Research, Tokyo, Japan (MN); Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD (JB); Department of Medical Oncology, Azienda Ospedaliera di Perugia, Perugia, Italy (VL); Department of Medical Oncology, Medical University of Gdansk, Poland (RD); Department of Medicine, University of Texas Southwestern Medical Center, Dallas, TX (JS).

Correspondence to: David P. Carbone, MD, PhD, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN 37232-6838 (e-mail: d.carbone@vanderbilt.edu).

See “Notes” following “References.”

DOI: 10.1093/jnci/djk195

© 2007 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

cancer cells (1). The tyrosine kinase domain of EGFR is necessary for its activity, and two highly selective EGFR tyrosine kinase inhibitors (TKIs), gefitinib and erlotinib, produce objective tumor responses and clinically important stable disease in some patients with advanced non-small-cell lung cancer (NSCLC) (2,3). However, many patients do not respond to EGFR TKIs, indicating that growth of only a subset of tumors is dependent on this signaling pathway. Nevertheless, randomized trials in unselected NSCLC patients comparing erlotinib or gefitinib with placebo (4,5) show a small survival advantage for erlotinib-treated patients and a trend toward improved survival for those treated with gefitinib. Thus, although a benefit for these drugs may exist for unselected patients, it is clearly desirable to identify—before initiation of therapy—patients who will benefit and those who will not.

Specific mutations in the tyrosine kinase domain of the EGFR gene (6–8), increased EGFR gene copy number as assessed by fluorescence in situ hybridization (FISH) (3,9–11), and high EGFR protein levels as assessed by immunohistochemistry (3,9,11) are associated with tumor sensitivity to gefitinib and erlotinib. Mutations of the KRAS gene have also been associated with resistance to EGFR TKIs (12). To date, EGFR copy number and EGFR protein levels are the only molecular features of tumors that have been shown to predict survival benefit in patients treated with gefitinib or erlotinib in randomized placebo-controlled trials (3,11). Some clinical parameters, especially a history of never smoking, are also associated with responsiveness to EGFR TKIs, but, importantly, a statistically significant survival benefit for erlotinib was observed in all of the clinical subsets evaluated (13). This finding indicates that clinical parameters alone are not sufficient to identify subsets of patients who will benefit from therapy and that this benefit cannot be explained by receptor mutation or gene amplification (13). Thus, better tools are needed to predict which patients with NSCLC will benefit from EGFR TKIs.

Matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry (MS) is a rapid, inexpensive, and simple technique for analyzing complex biologic samples, such as serum, urine, and tissue (14,15). Peaks in the mass spectrum correspond to ions formed from relatively abundant species in the sample, predominantly peptides and proteins. In this study, we tested whether mass spectrometric analysis of pretreatment peripheral blood could assist in the identification of patients who will benefit from treatment with gefitinib and erlotinib. To do so, we developed a prediction algorithm on a training set that comprised three patient cohorts and tested it on two independent validation patient cohorts and three independent control patient cohorts. We also examined the concordance of mass spectra independently acquired at two institutions to assess the reproducibility of the approach.

Methods

Patients and Samples

The training set included 139 patients with NSCLC who were treated systemically with gefitinib and from whom sera had been collected before treatment. These patients primarily had advanced stage disease, but a few were medically inoperable or refused surgery, as shown in Table 1. There were three training cohorts: from

CONTEXT AND CAVEATS

Prior knowledge

Some patients with non-small-cell lung cancer respond to treatment with the epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) gefitinib or erlotinib, but others do not. Clinical parameters alone are not sufficient to identify which patients are likely to benefit.

Study design

Matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry (MS) analysis of a training set of patients was used to develop an algorithm to classify patients as having “good” or “poor” outcomes after EGFR TKI treatment. The algorithm was then tested in several independent validation and control cohorts.

Contribution

The algorithm was able to classify patients in the validation cohorts in terms of their outcomes after treatment with gefitinib or erlotinib. In one validation set, the patients classified as “good” survived for a median of 306 days, whereas those classified as “poor” survived for a median of 107 days. The algorithm did not predict outcomes in control cohorts of patients who were not treated with EGFR TKIs.

Implications

The algorithm was able to classify patients according to their outcomes after EGFR TKI treatment. This classification algorithm, if confirmed in other cohorts, may help to identify appropriate subgroups of non-small-cell lung cancer patients for treatment with EGFR TKIs.

Limitations

Some studies have shown poor reproducibility of MALDI MS profiling, although this study reported that the profile was reproducible in different institutions. The identity of the proteins that make up the MALDI MS features in the classification algorithm is not known.

Scientific Institute Hospital San Raffaele, Milan, Italy (n = 70, “Italian A”); from Kanazawa University, Kanazawa, Japan (n = 26, Japan A); and from the Japanese Foundation for Cancer Research, Tokyo, Japan (n = 43, Japan B). There were two validation cohorts. One was an independent sequential cohort of patients with late-stage or recurrent NSCLC from the Scientific Institute Hospital San Raffaele (n = 67, “Italian B”) from whom sera was obtained before treatment with single-agent gefitinib. The second validation cohort included patients with NSCLC who were treated with first-line erlotinib on Eastern Cooperative Oncology Group (ECOG) protocol E3503, a single-arm phase II study (n = 96). Pretreatment samples of both serum and plasma were available from 73 of the patients in the ECOG study; only pretreatment serum was available for 13 patients, and only pretreatment plasma samples were available for the remaining 10 patients. Sera were also collected from three additional cohorts of NSCLC patients who did not receive treatment with EGFR TKIs. Two of these control cohorts consisted of patients with unresectable disease: Azienda Ospedaliera di Perugia, Perugia, Italy (n = 32, “Italian C”), and Vanderbilt-Ingram Cancer Center, Nashville, TN (n = 61, “VU”). The third control cohort included patients with resectable disease from the Medical

Table 1. Characteristics of the NSCLC patient sets used in this analysis*

Characteristic	Training set	Validation sets		Control sets		
	Italian A/Japan A and B (n = 139)	Italian B (n = 67)	ECOG (n = 96)	Italian C (n = 32)	VU (n = 61)	Polish early-stage (n = 65)
Sex, No. (%)						
Male	95 (68.3)	48 (71.6)	41 (42.7)	23 (71.9)	42 (68.9)	48 (73.8)
Female	44 (31.7)	19 (28.4)	55 (57.3)	9 (28.1)	19 (31.1)	17 (26.2)
Age, y						
Median	65	73	69	59	65	63
Range	36–90	38–91	41–93	37–74	40–84	37–77
Stage, No. (%)						
IA	0	0	0	0	0	12 (18.5)
IB	2 (1.4)	0	0	0	0	28 (43.1)
IIA	2 (1.4)	2 (3.0)	0	0	0	0
IIB	0	0	0	0	0	25 (38.5)
IIIA	2 (1.4)	2 (3.0)	0	4 (12.5)	0	0
IIIB	18 (12.9)	5 (7.4)	9 (9.4)	13 (40.6)	28 (45.9)	0
IV	102 (86.3)	58 (86.6)	67 (69.8)	15 (46.9)	33 (54.1)	0
Postoperative recurrence	13 (9.4)	0	20 (20.8)	0	0	0
Histology, No. (%)						
Adenocarcinoma	98 (70.5)	38 (56.7)	62 (64.6)	5 (15.6)	29 (47.5)	17 (26.2)
Squamous-cell carcinoma	22 (15.8)	15 (22.4)	11 (11.5)	7 (21.9)	19 (31.1)	38 (58.5)
Large-cell carcinoma	3 (2.2)	0	1 (1.0)	6 (18.8)	2 (3.3)	3 (4.6)
NSCLC, NOS	15 (10.8)	14 (20.9)	22 (22.9)	14 (43.8)	11 (18.0)	7 (10.8)
ECOG PS, No. (%)						
0	63 (45.3)	20 (29.8)	29 (30.2)	13 (40.6)	N/A	N/A
1	50 (36.0)	31 (46.3)	42 (43.8)	19 (59.4)	N/A	N/A
2	25 (18.0)	16 (23.9)	25 (26.0)	0	N/A	N/A
3	1 (0.7)	0	0	0	N/A	N/A
Smoking history, No. (%)						
No	31 (22.3)	11 (16.4)	N/A	3 (9.4)	4 (6.6)	4 (6.2)
Current or former	106 (76.3)	54 (80.6)	N/A	29 (90.6)	57 (93.4)	61 (93.8)
Previous chemotherapy†, No. (%)						
0	51 (36.7)	13 (19.4)	96 (100.0)	N/A	0	65 (100)
1	44 (31.7)	26 (38.9)	0	N/A	0	0
2	39 (28.1)	15 (22.4)	0	N/A	0	0
≥3	4 (2.9)	4 (6.0)	0	N/A	0	0
RECIST, No. (%)						
Partial response	35 (25.2)	N/A	N/A	1 (3.1)	N/A	N/A
Stable disease	42 (30.2)	N/A	N/A	4 (12.5)	N/A	N/A
Progressive disease	62 (44.6)	N/A	N/A	16 (50.0)	N/A	N/A

* For some clinical attributes, we did not have complete data; only the available data are shown, with the percentage of the total available calculated. NSCLC = non–small-cell lung cancer; ECOG = Eastern Cooperative Oncology Group; VU = Vanderbilt University; NOS = not otherwise specified; PS = performance status; N/A = not available; RECIST = Response Evaluation Criteria for Solid Tumors.

† Number of prior chemotherapy regimens.

University of Gdansk, Poland (n = 65). The clinical characteristics of the patients in the study are shown in Table 1. Samples were obtained after patients provided written informed consent, and all analyses were performed under protocols approved by the local institutional review boards. Sera were separated by centrifugation at 1000g (Japan B) or 2000g (all other cohorts) for 10 minutes at 4 °C, separated into aliquots, and frozen at –80 °C. Duplicate samples were shipped on dry ice to Vanderbilt University (VU) and to the University of Colorado at Denver and Health Sciences Center (UCDHSC), where they were stored at –80 °C until analysis.

Mass Spectrometry

Mass spectra for all training samples, the Italian and Vanderbilt control samples, and the Italian B test samples were generated

independently at both VU and UCDHSC on a Voyager DE-STR MALDI-TOF mass spectrometer and a Voyager DE-PRO MALDI-TOF mass spectrometer, respectively (Applied Biosystems, Framingham, MA). The ECOG and Polish samples were analyzed only at UCDHSC. Serum or plasma samples were thawed on ice and diluted 1 : 10 in deionized water. One microliter of each diluted sample was spotted at a unique location on the MALDI target (in triplicate), and 1 μL of matrix solution (35 mg/mL sinapinic acid; Sigma, St Louis, MO) in 50% acetonitrile (Burdick & Jackson, Muskegon, MI) and 0.1% trifluoroacetic acid (Sigma) was then added. The solutions were mixed by drawing the mixture into the pipette tip and then expelling it five times. Plates were allowed to dry at room temperature. Positive ion mass spectra were then acquired in linear

mode in an automated manner. Results from 500 to 525 independent spectrum acquisitions for each sample were averaged to generate each spectrum. All mass spectra were output as two-column text files of intensity versus m/z . Spectra were calibrated externally with mixtures of pure, well-characterized protein standards. At UCDHSC, a mixture of insulin (bovine), thioredoxin (*Escherichia coli*), and apomyoglobin (equine) (Applied Biosystems) was used; at VU, calibration was based on a mixture of insulin, cytochrome C, myoglobin, and ubiquitin (Bruker Daltonics, Bremen, Germany).

Spectral Preprocessing

Raw spectra were sent electronically to Bodesix (Steamboat Springs, CO) from each institution for analysis. Mass spectra generated from the same sample but by different personnel, institutions, and instruments can exhibit variations. To enable analysis of these spectra, we applied a suite of preprocessing procedures (16–20) and developed some additional procedures. In brief, the background was estimated and then subtracted from each spectrum based on local noise estimators (16–20), and peaks were detected using a signal-to-noise ratio cutoff of 3.0, which was found to be a good compromise between over-detection and sensitivity. To account for day-to-day and interinstrument variations in the m/z axis scale, spectra were aligned by using a set of common peaks (m/z 6434.5, 6632.1, 11686.9, 12864.8, 15131.1, 15871.5, and 28102.5). Normalization of spectra was complicated by the observed large variability of some intense peaks between individual samples, which can suppress other signals if total ion current is used for normalization. To overcome this problem, we used partial ion current normalization techniques (21), which are based on the union of the m/z ranges (6100–7500), (8500–10700), and (13300–16400). The entire preprocessing procedure was optimized by using the training set data and was held fixed for the classification of testing sets. Preprocessing gave comparable spectra for mass spectra generated at the two source institutions.

Training and Classifier Optimization

Each spectrum was characterized by a set of features. Features were defined as integrated, background-subtracted, and normalized intensities over a chosen m/z range containing a peak. A classifier was then constructed to map a (sub)set of these features to “good” or “poor” outcome (survival or time to progression) defined from the clinical data. The classification algorithm we used is a straightforward implementation of a k -nearest neighbor (KNN) algorithm (22). The KNN algorithm requires as parameters a set of representative and labeled “instances” (i.e., a list of selected feature values). In brief, to classify a new spectrum, the KNN algorithm first calculates the Euclidean distance of the feature values of the new spectrum to those of its representative spectra. This calculation yields a list of distances from the test spectrum to each representative spectrum. For the k nearest neighbors (those with the k smallest distances) the labels are compared. The calculated label is a simple majority vote over the KNN labels.

To optimize parameters for the KNN algorithm, representative spectra need to be selected from training set patients that are themselves representative of the clinical groups. For this analysis we initially examined five clinical groups: progressive disease–early

(i.e., disease progression in <1 month), progressive disease (i.e., disease progression in 1–3 months), partial response, stable disease–short (i.e., stable disease for ≤ 6 months), and stable disease–long (i.e., stable disease for >6 months). Visual inspection of all available training spectra showed that the most spectrally distinct clinical groups were those from patients with disease progression in less than 1 month and those with stable disease for more than 6 months (Supplementary Fig. 1, A and B; available online). We chose 13 total spectra from these groups. The remaining spectra in the training set were used to optimize parameters (the value of k , preprocessing parameters, and the integration range for feature values) and to select the features that are the most discriminating. As an optimization criterion, we used the leave-one-out cross-validation (LOOCV) error. Candidate features for the classification algorithm were identified as differentially expressed m/z values from spectra from patients with rapid progressive disease (training label “poor”) and from spectra from patients with long-term stable disease (training label “good”) by using univariate testing (Mann–Whitney U test). The list of features used in the classification is shown in the Supplementary Table 1 (available online).

Classification Procedure

After training and optimization, all parameters were frozen. No changes in the classification algorithm were allowed during the analysis of the validation and control sets. Classification labels for two groups, “good” (i.e., closest to the stable disease–long pattern) and “poor” (i.e., closest to the progressive disease–early pattern), were determined by KNN analysis using the following procedure: mass spectra were generated in triplicate; the spectra were preprocessed using the fixed, predetermined parameters for alignment and normalization; for each spectrum, the required feature values for the eight features defined in Supplementary Table 1 (available online) were determined; and for each replicate spectrum, these feature values were presented to the fixed KNN classifier ($k = 7$), which then returned a label, either “good” or “poor,” or gave a message that the values were unclassifiable. If the labels for all replicate samples were the same, that label was assigned to the patient sample; if the replicate labels disagreed or if no designation could be made, the patient sample was labeled “undefined.”

Statistical Analyses

Time to progression and overall survival were calculated using the Kaplan–Meier method, and graphs were generated using GraphPad Prism software (GraphPad Software Inc, San Diego, CA). Association of clinical variables, stage, sex, age, performance status, smoking status (not available for the ECOG validation cohort), and histology with survival was evaluated in univariate analyses and in multivariable analyses using Cox proportional hazards regression modeling. Proportionality was checked visually by examining plots of log–log survival curves and of Schoenfeld residuals. The data for all cohorts except ECOG are available as supplemental data online, and access to the clinical data for the ECOG set is available by contacting Robert Gray at the ECOG biostatistical office. SAS/JMP software (<http://www.jmp.com>, Cary, NC) and R (<http://www.r-project.org>, Boston, MA) were used. All P values are two-sided.

Table 2. Concordance between classification labels from two different institutions*

Classification label	UCDHSC		
	Good	Poor	Undefined
VU			
Good	139	1	0
Poor	2	59	0
Undefined	3	0	2

* Samples from the combined training set and Italian B validation set (n = 206) were analyzed independently at Vanderbilt University (VU) and the University of Colorado at Denver and Health Sciences Center (UCDHSC).

Results

Development and Assessment of the Prediction Algorithm

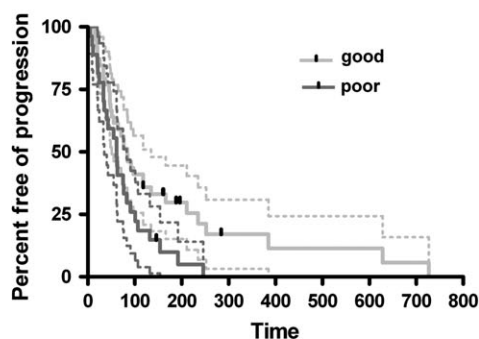
The set of eight discriminating features indicated in Supplementary Table 1 gave the least LOOCV error in the classification of patients in the training set (data not shown). Although the feature at m/z 5843.2 is the doubly charged form of the peak at m/z 11685 and is, therefore, not a completely independent feature, its inclusion improved the performance of the classification algorithm (data not shown).

The interlaboratory reproducibility of these eight MALDI MS features was determined, and Supplementary Fig. 2 (available online) shows a graph of the feature values from the spectra obtained at VU against those obtained at UCDHSC for two of these features. Good agreement in the feature values was observed across the two institutions. The concordance of the classification results of the MALDI mass spectral data obtained at the two institutions using these eight features is shown in Table 2. The overall concordance with which the 206 samples constituting the training set and the Italian B validation set were labeled as “good,” “poor,” or “undefined” was 97.1%. Thus, the spectral preprocessing techniques that we adopted enabled the generation of similar MALDI mass spectra (i.e., with consistent m/z values and amplitudes) across different institutions and nearly identical patient classification.

Validation of the Classification Algorithm

The classification algorithm was then validated in an independent cohort of 67 sequential NSCLC patients from Italy treated with second- or greater line gefitinib (validation set Italian B, see Table 1). This validation was performed in a blinded manner in that MALDI MS analysis was performed and classifications generated before the clinical outcome data were made available to the investigators. One of the 67 samples did not yield interpretable spectra and was excluded from the analysis. In Kaplan–Meier analysis (Fig. 1, A), patients classified as being in the “good” group had a statistically significantly longer time to progression than the patients in the predicted “poor” group (medians of 84 versus 61 days, respectively), with a univariate hazard ratio (HR) of progression of 0.56 (95% confidence interval [CI] = 0.28 to 0.89, log-rank $P = .02$). A statistically significant difference in overall survival was also observed between the “good” and “poor” groups (medians of 207 versus 92 days; HR of death = 0.50, 95% CI = 0.24 to 0.78, log-rank $P = .0054$) (Fig. 1, B).

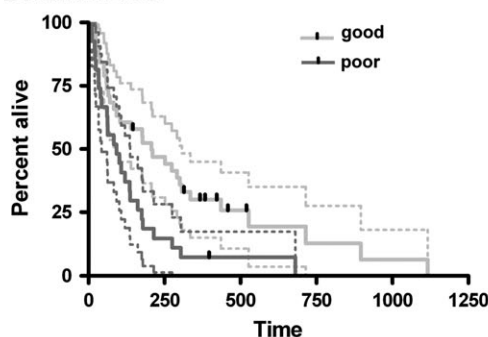
A. Italian B TTP



Number of patients at risk:

Time [days]	0	50	100	150	200	250	300
good	39	27	17	13	8	6	4
poor	27	17	8	4	2	1	0

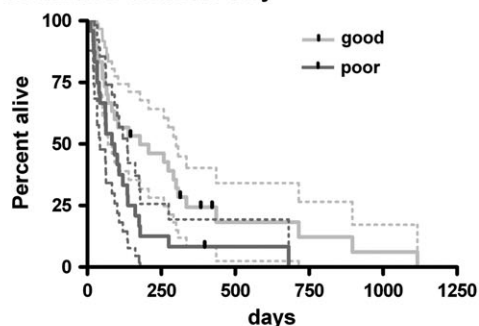
B. Italian B OS



Number of patients at risk:

Time [days]	0	50	100	150	200	250	300	350	400	700
good	38	32	25	22	20	18	14	11	9	4
poor	27	19	14	9	6	5	4	3	2	1

C. Italian B Smokers only



Number of patients at risk:

Time [days]	0	50	100	150	200	250	300	350	400	700
good	30	25	19	16	15	14	10	7	6	4
poor	24	17	12	6	4	4	3	3	2	1

Fig. 1. Kaplan–Meier analysis of outcomes in the Italian B validation cohort. These patients with non–small-cell lung cancer received second- or later-line treatment with gefitinib alone. **A)** Time to progression. **B)** Overall survival (n = 67, one sample was undefined, one patient had no survival data). **C)** Overall survival among smokers in the Italian B validation cohort (n = 54). **Solid lines** = event-free fraction; **dashed lines** = 95% confidence intervals; **tick marks** = censored patients.

Multivariable Analysis of the Italian B Validation Set

A Cox multivariable analysis of overall survival was performed using data from the Italian B validation cohort to compare the

Table 3. Outcomes in the patient sets included in this analysis*

Outcome	Training set	Validation sets			Control sets	
	Italian A/Japan A and B (n = 139)	Italian B (n = 67)	ECOG (n = 96)	Italian C (n = 32)	VU (n = 61)	Polish early stage (n = 65)
Classification from MALDI MS algorithm, No. (%)						
Good	105 (75.5)	39 (58.3)	69 (71.9)	20 (62.5)	41 (67.2)	44 (67.7)
Poor	33 (23.7)	27 (40.3)	27 (28.1)	12 (37.5)	20 (32.8)	21 (32.3)
Undefined	1	1	0	0	0	0
Overall survival						
HR (95% CI)	0.45 (0.19 to 0.63)	0.5 (0.24 to 0.78)	0.4 (0.24 to 0.70)	0.74 (0.3 to 1.6)	0.81 (0.4 to 1.6)	0.9 (0.4 to 1.9)
Log-rank <i>P</i>	<.001	.0054	<.001	.42	.54	.79
Median time to death, days (good/poor)	441/148	207/92	306/107	163/141	729/312	1430/1233
Time to progression						
HR (95% CI)	0.5 (0.23 to 0.74)	0.56 (0.28 to 0.9)	0.53 (0.33 to 0.85)	N/A	N/A	N/A
Log-rank <i>P</i>	.0031	.02	.007	N/A	N/A	N/A
Median time to progression, days (good/poor)	161/63	84/61	98/58	N/A	N/A	N/A
Multivariable analysis of overall survival†						
HR (95% CI)	ND	0.74 (0.55 to 0.99)	0.53 (0.30 to 0.94)	ND	ND	ND
Wald <i>P</i>	ND	.048	.03	ND	ND	ND

* ECOG = Eastern Cooperative Oncology Group; VU = Vanderbilt University; MALDI = matrix-assisted laser desorption ionization; MS = mass spectrometry; HR = hazard ratio; CI = confidence interval; N/A = not available; ND = not done.

† In the multivariable analysis, the cofactors included were performance status (0–5), sex (male/female), histology (adenocarcinoma, squamous cell carcinoma, large cell carcinoma, or not otherwise specified), smoking history (no versus current or former), and MALDI MS classification (good versus poor) in the Italian B set and performance status (0–5), number of involved sites (1–5), prior weight loss ($\geq 5\%$ or $< 5\%$), histology, and MALDI MS classification (good versus poor) in the ECOG validation set.

classification obtained from the MALDI MS–based test with classification according to clinical parameters that have been previously associated with responsiveness to EGFR TKIs, i.e., being a never smoker and adenocarcinoma histology (3). The analysis (Table 3) showed that only performance status and the MALDI MS result were independently associated with survival benefit. Patients who were classified in the “good” group had a statistically significantly lower risk of death than patients classified in the “poor” group (HR = 0.74, 95% CI = 0.55 to 0.99, log-rank *P* = .048).

We also asked whether it was possible to use the classifier to identify subgroups of patients with improved outcomes in clinical groups known to have low response rates to EGFR TKIs. Indeed, even in current or former smokers in the Italian B validation set (54 of the 67 patients), the group identified as “good” by the classification algorithm had statistically significantly better median survival (178 days) than the group identified as “poor” (88 days, HR = 0.52, 95% CI = 0.25 to 0.87, log-rank *P* = .017) (Fig. 1, C).

Eastern Cooperative Oncology Group Validation Cohort

We also applied the prediction algorithm to a second validation cohort. This cohort consisted of 96 previously untreated patients participating in ECOG protocol E3503, a phase II trial of erlotinib, for whom blinded samples of pretreatment serum, plasma, or both were available. Plasma and serum samples gave similar values for all eight MALDI MS features used in our classification algorithm (Supplementary Fig. 3, available online). Each sample was then classified using the eight-feature classification algorithm, and the results were sent to the ECOG biostatistical office for correlation with the clinical data (data not shown). Using just the 73

patients for whom both serum and plasma were available, classification of patients into good and poor groups in terms of overall survival was equally powerful whether based on analysis of serum or plasma (data not shown). Consequently, we classified all 96 ECOG patients, using spectra from serum if available (n = 86) or plasma if not (n = 10). The patients classified in the “good” outcome group indeed had better survival than those classified in the “poor” outcome group (Fig. 2; median survivals of 306 versus 107 days, HR = 0.41, 95% CI = 0.24 to 0.70, log-rank *P* < .001). With the available

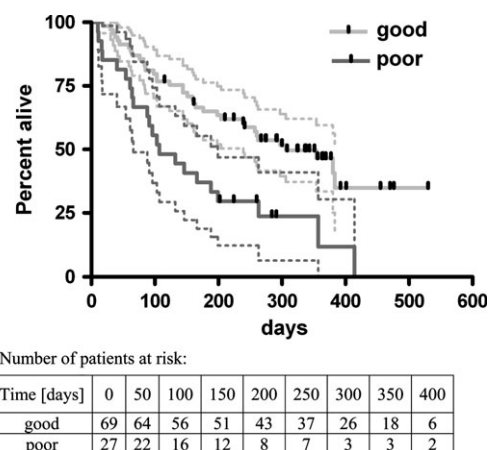


Fig. 2. Kaplan–Meier analysis of overall survival in the Eastern Cooperative Oncology Group validation cohort (n = 96). These patients had advanced non–small-cell lung cancer and had been treated first line with erlotinib alone. **Solid lines** = event-free fraction; **dashed lines** = 95% confidence intervals; **tick marks** = censored patients.

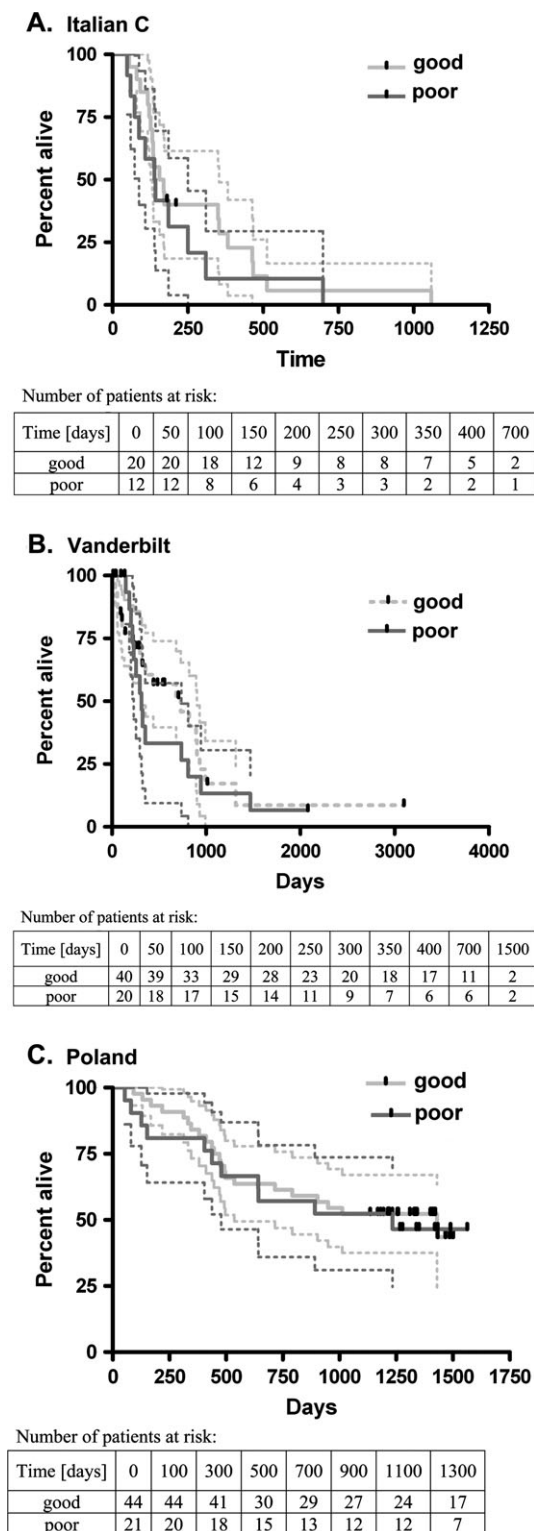


Fig. 3. Kaplan–Meier analysis of overall survival in control cohorts of patients with non–small-cell lung cancer. **A)** Overall survival in the Italian C control cohort. These patients received chemotherapy alone and no epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) ($n = 32$). **B)** Overall survival in the Vanderbilt control cohort. These patients received chemotherapy alone and no EGFR TKIs ($n = 61$, one sample undefined). **C)** Overall survival in the Polish control cohort. These patients received surgery alone and no EGFR TKIs ($n = 65$). **Solid lines** = event-free fraction; **dashed lines** = 95% confidence intervals; **tick marks** = censored patients.

follow-up (the median follow-up for time to progression was 6.7 months for those eight of the 96 patients for whom no progression was observed), patients classified in the “good” group had statistically significantly longer time to progression than those classified in the “poor” group (median time to progression was 3.2 and 1.9 months, respectively; HR = 0.53, 95% CI = 0.33 to 0.85, log-rank $P = .007$; Table 3). In a Cox multivariable analysis that included the parameters that were most statistically significant in the univariate model—performance status (0, 1, or 2), number of involved sites (≤ 3 or > 3), and prior weight loss ($< 5\%$ or $\geq 5\%$)—the MALDI MS classification algorithm was independently statistically significant (HR = 0.53, 95% CI = 0.30 to 0.94, Wald $P = .03$; Table 3). Smoking status was not available in these patients, and, thus, this important cofactor could not be analyzed.

Predictive or Prognostic?

Because the preceding analyses were all based on outcomes in patients treated with gefitinib or erlotinib, it was important to show that the classifications of survival outcomes from the MALDI MS algorithm were not simply prognostic but rather identified patients who would benefit from therapy with EGFR TKIs. For this analysis, we examined outcomes in three separate cohorts of patients with advanced NSCLC, none of whom received treatment with EGFR TKIs. In the first cohort, a group of 32 NSCLC patients from Perugia (Italian C) from whom serum was collected immediately before second-line chemotherapy, no statistically significant differences were seen in the overall survival curves of patients classified in the “good” and “poor” groups (Fig. 3, A; HR = 0.74, 95% CI = 0.33 to 1.6, log-rank $P = .42$). Because this set was so small, we performed a permutation analysis to investigate the possibility that such a result could have been obtained by chance, using the Italian B set as a reference population (see Supplementary Fig. 4, available online). There was only a 6.6% chance that these results could have been obtained by chance.

Similarly, in the second control cohort, of 61 patients with advanced NSCLC from VU, no difference in survival was observed between patients classified in the “good” and “poor” groups (Fig. 3, B; HR = 0.81, 95% CI = 0.4 to 1.6, log-rank $P = .54$). Finally, in the third control cohort, of 65 patients with resected early-stage (i.e., pathologic stage IA–IIB) NSCLC from Gdansk, Poland, again survival was the same in the “good” and “poor” groups (Fig. 3, C; HR = 0.90, 95% CI = 0.43 to 1.89, log-rank $P = .79$). Thus, the classification algorithm did not accurately classify patient outcomes among patients not treated with EGFR TKIs.

Discussion

In this study, we developed a classification algorithm based on MALDI MS analysis of pretreatment serum and plasma that could identify subgroups of NSCLC patients with improved time to progression and overall survival after treatment with the EGFR TKIs gefitinib and erlotinib. On multivariable testing in two independent validation cohorts, this algorithm retained its predictive value independent of clinical factors associated with sensitivity to EGFR TKIs. The classifier thus performed well for both gefitinib and erlotinib. However, it did not perform well for traditional chemotherapy or surgery, based on its inability to identify patients with

poor outcomes in the control cohorts. In addition, the MALDI MS algorithm also performed well for the same classifier peaks in both plasma and serum.

The best studied predictive tumor markers for benefit from treatment with EGFR TKIs in NSCLC are EGFR protein expression, specific EGFR mutations, and EGFR gene copy number. Specifically, EGFR protein expression, as assessed by immunohistochemical analysis, has been shown to be modestly associated with overall survival benefit after EGFR TKI treatment in some studies (3,9). Activating mutations in the EGFR tyrosine kinase domain (6–8) have been shown to be associated with dramatic responses to EGFR TKIs, and in some retrospective single-arm studies mutations have been associated with improved survival of patients treated with EGFR TKIs compared with those without mutations (23–27). However, in other studies (3,9), including a prospective randomized study of erlotinib compared with placebo (3), EGFR mutations were not associated with survival benefit from EGF TKI treatment. In addition, an observed survival benefit with erlotinib was also found in groups less likely to have EGFR mutations, such as males, smokers, and patients with squamous cell carcinoma (13). A predictive classification algorithm would have added value in the identification of patients who would benefit from relatively nontoxic treatment with EGFR TKIs. One molecular feature of tumors, amplification or high polysomy of the EGFR gene (i.e., FISH positivity), was associated with improved survival of lung cancer patients treated with EGFR TKIs in a multivariable analysis (9) (HR = 0.44), a finding that has been confirmed in a similar study of adenocarcinoma patients (HR = 0.50) (10) and in randomized studies of erlotinib and gefitinib compared with placebo (3,11).

The classification ability of the MALDI MS test described in this study appears to be similar to that of this tumor tissue-based assay. The univariate hazard ratio for death from any cause for predicted “good” compared with predicted “poor” groups in the Italian B validation set (HR = 0.50, 95% CI = 0.24 to 0.78, $P = .008$) was similar to that for EGFR FISH-positive compared with FISH-negative patients (HR = 0.44, 95% CI = 0.23 to 0.82) (9,10). The multivariable analysis showed that the MALDI MS algorithm provides information over and above the clinical parameters proposed to be predictive of response to EGFR TKIs, specifically sex, smoking history, and histology. The algorithm even identified subgroups of smokers with statistically significantly improved survival after gefitinib treatment. Therefore, even in patients with clinical features associated with low response rates to EGFR TKIs as a group, it was possible to use the algorithm to identify a subset with a substantial predicted survival benefit.

Both FISH and mutation analysis are tumor-based assays that require well-preserved biopsy material, are technically difficult, have a substantial cost, and have a slow turnaround time. By contrast, the MALDI MS method that we have described can be performed on less than 1 μ L of pretreatment serum, at low cost, and rapidly, and the method can easily be fully automated. It is thus much more readily applied in a clinical setting than the other assays.

An important and appropriate criticism of many previous studies using MALDI MS profiling of serum is lack of reproducibility (28,29). However, here we have demonstrated that processed mass spectra independently obtained from two institutions on two

different instruments can yield highly reproducible classification by using appropriate preprocessing methods. The observed concordance of 97.1% between the two institutions that generated the MALDI MS data for our study compares favorably with interlaboratory variability of well-established tests, such as immunohistochemistry or FISH testing for HER2 (30).

In the clinical development of biomarkers for the individualization of therapy, it is important to distinguish between biomarkers that can accurately classify patients according to whether they will benefit from an intervention and those that simply portend a favorable or unfavorable prognosis, independent of the planned intervention. Biomarkers predictive for survival benefit from an intervention are much more useful for guiding management. The discriminating features that we have identified in the mass spectra of serum and plasma are unlikely to represent markers of poor prognosis, given the lack of prognostic significance of the classification algorithm when it was used to analyze three independent cohorts of patients with NSCLC who did not receive EGFR TKIs. Moreover, in multivariable analysis, the MALDI MS test was predictive of survival independent of performance status, which also suggests that it was not merely prognostic.

One limitation of our analysis is the lack of smoking data in the ECOG cohort, because smoking is a clear predictive factor for response to EGFR TKIs. However, our classifier predicted outcomes independent of smoking status in the Italian B validation cohort. Moreover, in a US-based trial of first-line treatment for advanced disease such as the ECOG study, the number of never smokers is likely to be too low to account for a substantial portion of the discriminatory power of our signature. Another limitation is the unknown biology underlying the ability of these features to predict benefit. The identification and analysis of the informative peaks might lead to important insights into the mechanism of the association, and these studies are under way.

This study represents the first comprehensive and rigorously validated attempt to use MALDI MS methods to classify patients for their clinical benefit from a molecularly targeted anticancer agent. MALDI MS analysis of pretreatment serum performed in parallel at two institutions and based on samples from three continents offers a robust and reproducible method. In two blinded validation studies, in patients receiving both first- and second-line treatment and using both gefitinib and erlotinib, the test had a classification ability similar to that of tumor-based assays and independent of other clinical parameters associated with response to EGFR inhibitors. It will be important to confirm the clinical value of this strategy in randomized trials with larger cohorts of patients treated with EGFR TKIs.

References

- (1) Arteaga CL. ErbB-targeted therapeutic approaches in human cancer. *Exp Cell Res* 2003;284:122–30.
- (2) Fukuoka M, Yano S, Giaccone G, Tamura T, Nakagawa K, Douillard JY, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial) [corrected]. *J Clin Oncol* 2003;21:2237–46.
- (3) Tsao MS, Sakurada A, Cutz JC, Zhu CQ, Kamel-Reid S, Squire J, et al. Erlotinib in lung cancer—molecular and clinical predictors of outcome. *N Engl J Med* 2005;353:133–44.

- (4) Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med* 2005;353:123–32.
- (5) Thatcher N, Chang A, Parikh P, Rodrigues Pereira J, Ciuleanu T, von Pawel J, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 2005;366:1527–37.
- (6) Lynch TJ, Bell DW, Sordella R, Gurubhagavata S, Okimoto RA, Brannigan BW, 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;350:2129–39.
- (7) Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–500.
- (8) Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, et al. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A* 2004;101:13306–11.
- (9) Cappuzzo F, Hirsch FR, Rossi E, Bartolini S, Ceresoli GL, Bemis L, et al. Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *J Natl Cancer Inst* 2005;97:643–55.
- (10) Hirsch FR, Varella-Garcia M, McCoy J, West H, Xavier AC, Gumerlock P, et al. Increased epidermal growth factor receptor gene copy number detected by fluorescence in situ hybridization associates with increased sensitivity to gefitinib in patients with bronchioloalveolar carcinoma subtypes: a Southwest Oncology Group Study. *J Clin Oncol* 2005;23:6838–45.
- (11) Hirsch FR, Varella-Garcia M, Bunn PA Jr. Molecular predictors of outcome with gefitinib in a phase III placebo-controlled study in advanced non-small cell lung cancer. *J Clin Oncol* 2006;24:5034–42.
- (12) Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, et al. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med* 2005;2:e17.
- (13) Clark GM, Zborowski DM, Santabarbara P, Ding K, Whitehead M, Seymour L, et al. Smoking history and epidermal growth factor receptor expression as predictors of survival benefit from erlotinib for patients with non-small-cell lung cancer in the National Cancer Institute of Canada Clinical Trials Group Study BR.21. *Clin Lung Cancer* 2006;7:389–94.
- (14) Schwartz SA, Weil RJ, Thompson RC, Shyr Y, Moore JH, Toms SA, et al. Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. *Cancer Res* 2005;65:7674–81.
- (15) Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003;362:433–9.
- (16) Amann JM, Chaurand P, Gonzalez A, Mobley JA, Massion PP, Carbone DP, et al. Selective profiling of proteins in lung cancer cells from fine-needle aspirates by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Cancer Res* 2006;12:5142–50.
- (17) Chaurand P, Norris JL, Cornett DS, Mobley JA, Caprioli RM. New developments in profiling and imaging of proteins from tissue sections by MALDI mass spectrometry. *J Proteome Res* 2006;5:2889–900.
- (18) Cornett DS, Mobley JA, Dias EC, Andersson M, Arteaga CL, Sanders ME, et al. A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. *Mol Cell Proteomics* 2006;5:1975–83.
- (19) Meistermann H, Norris JL, Aerni HR, Cornett DS, Friedlein A, Erskine AR, et al. Biomarker discovery by imaging mass spectrometry: transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat. *Mol Cell Proteomics* 2006;5:1876–86.
- (20) Roder H, Grigorieva J, Tsybin M. The use of mass spectra for cancer biomarker detection. Bodesix; 2005. Available at: <http://www.bodesix.com/Documents/MarkerWhitePaper.pdf>.
- (21) Tsybin M, Grigorieva J, Rothhoff E, Roder H. A061107. Partial ion current normalization of mass spectra and biomarker reproducibility. Proceedings of the 54th ASMS Conference on Mass Spectrometry and Allied Topics, Seattle, WA, May 28–June 1, 2006; p. 121S.
- (22) Webb A. Statistical pattern recognition. Chichester (U.K.): Wiley; 2002.
- (23) Cortes-Funes H, Gomez C, Rosell R, Valero P, Garcia-Giron C, Velasco A, et al. Epidermal growth factor receptor activating mutations in Spanish gefitinib-treated non-small-cell lung cancer patients. *Ann Oncol* 2005;16:1081–6.
- (24) Han SW, Kim TY, Hwang PG, Jeong S, Kim J, Choi IS, et al. Predictive and prognostic impact of epidermal growth factor receptor mutation in non-small-cell lung cancer patients treated with gefitinib. *J Clin Oncol* 2005;23:2493–501.
- (25) Mitsudomi T, Kosaka T, Endoh H, Horio Y, Hida T, Mori S, et al. Mutations of the epidermal growth factor receptor gene predict prolonged survival after gefitinib treatment in patients with non-small-cell lung cancer with postoperative recurrence. *J Clin Oncol* 2005;23:2513–20.
- (26) 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.
- (27) Taron M, Ichinose Y, Rosell R, Mok T, Massuti B, Zamora L, et al. Activating mutations in the tyrosine kinase domain of the epidermal growth factor receptor are associated with improved survival in gefitinib-treated chemorefractory lung adenocarcinomas. *Clin Cancer Res* 2005;11:5878–85.
- (28) Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool. *Mol Cell Proteomics* 2004;3:367.
- (29) Diamandis EP. Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? *Clin Chem* 2003;49:1272.
- (30) Perez EA, Suman VJ, Davidson NE, Martino S, Kaufman PA, Lingle WL, et al. HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 intergroup adjuvant trial. *J Clin Oncol* 2006;24:3032–8.

Notes

F. Taguchi and B. Solomon contributed equally to this work.

Supported by grants from National Cancer Institute SPORES in Lung Cancer CA90949 (to D. P. Carbone) and CA58187 (to P. A. Bunn); cancer center core grants to Vanderbilt and Colorado CA68485 and CA046934, respectively; International Association for the Study of Lung Cancer fellowship to R. Dziadziuszko; ECOG support for J. Schiller, J. Brahmer, and R. Gray; and unrestricted funds from Genentech to David Johnson, supporting F. Taguchi. OSI Pharmaceuticals provided some funding for sample collection and processing for the E3503 ECOG study. H. Roder is the Chief Technical Officer of Bodesix, which is developing tools for MS-based clinical diagnostics. J. Brahmer is conducting research sponsored by AstraZeneca (manufacturer of gefitinib). F. R. Hirsch is doing research in collaboration with AstraZeneca, OSI Pharmaceuticals, and Genentech under research agreements with the University of Colorado. F. R. Hirsch has served on advisory boards, and P. A. Bunn has consulted for OSI/Genentech/Roche (manufacturers of erlotinib) and AstraZeneca. J. Schiller has served on advisory boards for Genentech. Genentech and the other study sponsors played no role in the design of the study; the collection, analysis, or interpretation of the data; and the preparation of the manuscript or the decision to submit the manuscript for publication.

Funding to pay the Open Access publication charges for this article was provided by Bodesix, Inc, Steamboat Springs, CO.

Manuscript received November 20, 2006; revised March 26, 2007; accepted April 13, 2007.