Detection of Epidermal Growth Factor Receptor Variations by Partially Denaturing HPLC

TAN MIN CHIN,1 DIYANAH ANUAR,2 ROSS SOO,1 MANUEL SALTO-TELLEZ,2,3,4 WEI QI LI,5 BAIDAH AHMAD,2 SOO CHIN LEE,1 BOON CHER GOH,1 KAZUYUKI KAWAKAMI,6 AMANDA SEGAL,7 BARRY IACOPETTA,5 AND RICHIE SOONG2,3*

Background: Epidermal growth factor receptor gene (EGFR) variants may be useful markers for identifying responders to gefitinib and erlotinib, small-molecule tyrosine kinase inhibitors of EGFR; therefore, sensitive and cost-effective assays are needed to detect EGFR variants in routine clinical samples. We have developed a partially denaturing HPLC (pDHPLC) assay that is superior to direct sequencing with respect to detection limits, costs, and time requirements.

Methods: Primers, temperatures, and buffer conditions were optimized for PCR-pDHPLC analysis of EGFR exons 18–21. We evaluated the detection limits of pDHPLC and direct sequencing by analyzing mixtures of wild-type and variant EGFR DNA and screened 192 lung cancer samples to examine the diversity of pDHPLC-detectable variants. To assess amenability to routine analysis, we tested lung and pleural tissue specimens from 14 lung cancer patients treated with gefitinib.

Results: The detection limits for variant alleles were 1:100 for pDHPLC and 1:5 for direct sequencing. pDHPLC analysis detected 26 unique EGFR variants, including the common deletions in exon 19 and substitutions in codons 787 and 858. Direct sequencing could not identify 30% (18 of 60) of the variant amplicons identified by pDHPLC. We identified these 18 amplicons by fraction collection after pDHPLC analysis. Analysis of a limited series of lung biopsy samples detected EGFR variants more frequently in gefitinib responders than in nonresponders. pDHPLC analysis was 56% less expensive and 39% faster than direct sequencing.

Conclusions: pDHPLC-based analysis detects EGFR variations in routine clinical samples with a better detection limit and lower cost and time requirement than direct sequencing.

© 2007 American Association for Clinical Chemistry

Population Genetics

Variation in the tyrosine kinase–encoding domain of epidermal growth factor receptor (EGFR) was discovered in 2004 (1, 2). Since then, both in vitro functional studies and analyses of clinical samples have shown EGFR tyrosine kinase variants to be potentially useful markers for identifying likely responders to the small-molecule EGFR tyrosine kinase inhibitors (TKIs) gefitinib and erlotinib (3, 4). More than 10 independent studies have reported findings of significantly higher EGFR variant frequencies in TKI responders than in nonresponders. Some studies (5–8), but not others (9, 10), have associated EGFR variants with improved survival in recipients of EGFR TKI treatment.

The prospect that EGFR variations could identify individuals with a likely improved outcome after EGFR TKI treatment has generated keen interest in assays that detect such variants. After studies of BRCA1 and BRCA2 variants for familial breast cancer, MLH1 and MSH2 variants

Nonstandard abbreviations: EGFR, epidermal growth factor receptor; TKI, tyrosine kinase inhibitor; pDHPLC, partially denaturing HPLC; FFPE, formalin-fixed, paraffin-embedded tissue.

Human genes: EGFR, epidermal growth factor receptor [erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian]; BRCA1, breast cancer 1, early onset; BRCA2, breast cancer 2, early onset; MLH1 mutl homolog 1, colon cancer, nonpolyposis type 1 (E. coli); MSH2, mutS homolog 2, colon cancer, nonpolyposis type 2 (Escherichia coli); PYD, diphosphoryridimidine dehydrogenase; KRAS, v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog.
for hereditary nonpolyposis colorectal cancer, and DPYD variants for predicting toxicity to fluoropyrimidine treatment, partially denaturing HPLC (pDHPLC) has become recognized as an optimal method for detecting diverse variants distributed potentially throughout the genome (11). The technique discriminates between wild-type and variant alleles by their distinctive chromatographic properties at partially denaturing temperatures and pH. Comparative studies have shown pDHPLC to be one of the most sensitive DNA-screening techniques available (12); pDHPLC can detect variants in mixtures with the wild-type allele at ratios as low as 1:100 (13). By screening out the often large proportions of samples with wild-type alleles, pDHPLC also offers significant savings in costs and time compared with direct sequencing. Moreover, available pDHPLC instrumentation features precisely regulated temperature control, liquid handling, sample injection, processing, and analysis.

The ability of pDHPLC to detect a large range of sequence variants with high sensitivity and low detection limits for minority alleles in an inexpensive and standardized manner suggests it as a useful approach for routine detection of EGFR variants. Recently, investigators used a DHPLC instrument and non-denaturing size analysis of DNA digested with a DNA mismatch–recognizing enzyme (SURVEYOR) to detect EGFR variants (14); however, the feasibility of pDHPLC for EGFR variant analysis has not been reported. We undertook this study to optimize a pDHPLC-based assay system for detecting EGFR variants in small biopsy samples of lung tumors and to assess its potential for routine screening.

**Materials and Methods**

**DNA SAMPLES**

DNA from 192 lung cancer samples was obtained from 3 different institutions: 86 samples extracted from frozen tissues from the Department of Surgery, Kanazawa University School of Medicine, Kanazawa, Japan; 51 samples extracted from formalin-fixed, paraffin-embedded tissue (FFPET) from Pathwest, Sir Charles Gairdner Hospital, Perth, Australia; and 55 samples extracted from FFPET in the Department of Pathology, National University Hospital, Singapore. DNA from small biopsy and cytology FFPET samples from lung and adjacent pleura of 14 lung cancer patients treated with gefitinib was also obtained from the Department of Pathology, National University Hospital, Singapore. DNA was extracted from frozen tissue (15) and FFPET (16) as described previously. In Singapore, DNA was extracted from FFPET sections 5 μm thick with the Puregene DNA purification reagent set (Gentra Systems). Response to gefitinib treatment was assessed with standard protocols per the criteria of the Response Evaluation Criteria in Solid Tumors Group (17). All samples were obtained according to institutionally approved protocols.

**PCR-pDHPLC ANALYSIS**

PCR was performed with the FastStart Taq DNA Polymerase reagent set (Roche Diagnostics). In brief, 50 ng of DNA was amplified in a 25-μL reaction volume containing 1× PCR buffer, 2 mol/L MgCl₂, 400 nmol/L PCR primers, 200 μmol/L deoxynucleoside triphosphates, and 1 U of FastStart Taq polymerase. Thermal cycling conditions were as follows: heating at 95 °C for 4 min; 40 cycles of 95 °C for 30 s, the specific annealing temperature for 30 s, and 72 °C for 1 min; and a final step of 72 °C for 1 min. PCR products were electrophoresed on an agarose gel to ensure that correctly sized amplicons were obtained. The PCR product was then denatured at 95 °C for 4 min and cooled to 25 °C at a rate of 1.2 °C/min to allow formation of heterozygote DNA. PCR product (10 μL) was then loaded for pDHPLC analysis on the WAVE 3500HT High Sensitivity System instrument (Transgenomic) and eluted with a gradient mixture of Buffers A (0.1 mol/L triethylammonium acetate) and B (0.1 mol/L triethylammonium acetate, 250 mL/L acetonitrile), obtained from Transgenomic. The respective forward and reverse PCR primer sequences, PCR annealing temperatures, pDHPLC running temperatures, and buffer B gradients were as follows: TTCCAGCATGTTAGGG and ACACGTG-CAAGCACTCT (exon 18), 60 °C, 61.2 °C, 53.5%–62.5%; AGCATGTGCCACCATCCTC and AGACATGAGAAGA- GTGTTG (exon 19), 60 °C, 57.7 °C, 51.9%–60.9%; CATGT-GCCCTCACCACTGC and CATGCCGCAAGCAGA (exon 20), 55 °C, 61.7 °C, 55.4%–64.4%; and AATTCG-GATGCAGCTT and TACAGCTAGTGGGAAGGC (exon 21), 60 °C, 61.2 °C, 54.5%–63.5%.

**SEQUENCING**

We purified PCR products for sequencing with the ExoSAP-IT reagent for PCR product clean-up (USB). DNA was sequenced with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with the same primers used for the PCR. The 3100 Genetic Analyzer (Applied Biosystems) was used for capillary electrophoresis and sequence analysis. All variants were sequenced bidirectionally.

**FRACTION COLLECTION AND SEQUENCING**

We reloaded the post-pDHPLC PCR product for pDHPLC analysis on the WAVE instrument and collected eluate fractions within times corresponding to those of aberrant peaks on the instrument systems. We precipitated DNA from the eluate with an equal volume of 10 mol/L ammonium acetate and washed the precipitate with 700 mL/L ethanol. The precipitate was resuspended in 20 μL of water; 2 μL was used for PCR and subsequent sequencing.

**PLASMID PREPARATION**

PCR product from samples with wild-type or mutant DNA was cloned into pGEM-T Easy Vectors by means of the pGEM-T Easy reagent set (Promega). After ligation,
transformation, and propagation, plasmids were purified with the QIAprep Spin MiniPrep system (Qiagen). Bidirectional sequencing confirmed the identity of plasmid DNA used in the mixing experiments to determine detection limits.

Results

Optimization of the PCR-pDHPLC System
Various PCR primer sequences, reagent concentrations, cycling conditions, pDHPLC temperatures, and buffer gradients were tested empirically to determine optimal conditions for detecting sequence variants within the entire coding region of 4 EGFR exons (18–21 inclusive). To minimize the risk of contamination and artifact generation, the PCR consisted of a single round of ~3 h. pDHPLC analysis of each amplicon required ~8 min.

Comparison of Variant Detection Limits for pDHPLC and Sequencing
We compared the detection limits for pDHPLC and direct sequencing by analyzing plasmid samples containing both variant (746B, 15-bp deletion) and wild-type sequences in various proportions from 1:1 to 1:1000. pDHPLC clearly revealed a peak corresponding to the mutant allele above the background signal in the 1:100 sample (Fig. 1). With direct sequencing, the same deletion variant was reliably apparent only in the 1:5 sample. Similar results were obtained with plasmid mixtures containing 1 variant (858B: T→G) and the wild-type sequence.

Analysis of Lung Cancer Samples from Japan, Australia, and Singapore
Of the 192 samples, 150, 7, 18, 12, and 5 samples provided pDHPLC-readable results for 4, 3, 2, 1, and 0, respectively, of the 4 analyzed amplicons. The 42 samples (22%) with at least 1 amplicon unavailable for analysis were excluded from further consideration. Of the remaining cases, 10% (60 of 600) of the amplicons from 33% (50 of 150) of the cases exhibited pDHPLC chromatograms that differed from those of known wild-type samples (Fig. 2). Forty-one cases, 8 cases, and 1 case displayed aberrant chromatograms for 1, 2, and 3 amplicons, respectively.

DNA sequencing identified 70% (42 of 60) of the variant amplicons identified by pDHPLC and 70% (35 of 50) of the cases with aberrant pDHPLC chromatograms, corresponding to 23% (35 of 150) of the eligible cases overall (Table 1). Two variants were concurrently identified in 3 amplicons. We sequenced 40 amplicons determined to be wild-type by pDHPLC (10 for each exon), and all showed the wild-type sequence.

To identify the variants in the 18 amplicons that were not identifiable by direct sequencing, we developed a fraction-collection protocol for sequencing enriched variants (Fig. 3). We identified sequence variants in all 18 amplicons, including 7 different variant types (Table 1). We sequenced fraction-collected samples from wild-type pDHPLC peaks for 8 amplicons (2 for each exon) and obtained only wild-type sequences.
Twenty-three different variants were identified, including deletions in exon 19 (25 of 63; 40%) and single-base changes in codons 719 (7 of 63; 11%), 787 (15 of 63; 24%), and 858 (5 of 63; 8%).

**ANALYSIS OF SMALL BIOPSY SAMPLES FROM LUNG CANCER PATIENTS TREATED WITH GEFITINIB**
To investigate whether pDHPLC is suitable for routine clinical samples, we used pDHPLC to analyze DNA extracted from small biopsy samples from 14 lung cancer patients treated with gefitinib (Table 2). We analyzed all of these samples by both pDHPLC and direct sequencing. With both methods, 71% of the samples (10 of 14) gave results suitable for analysis, and 80% (8 of 10) contained a variant. Complete concordance between the 2 techniques was observed for the amplicons identified as having DNA variants. Three additional unique variants were detected in this series: 734B: A→G (Glu→Gly); a silent 759C: C→A (Ile→Ile) substitution; and an intronic substitution (IVS19+53: C→A) (the letters A, B, and C after the codon number indicate the 1st, 2nd, and 3rd nucleotide positions of the codon, respectively). Not considering silent and intronic variants, we observed a higher frequency of EGFR variants in responders (4 of 5; 80%) than in nonresponders (2 of 5; 40%). This result is consistent with previous reports; however, the small numbers of cases in this series make it difficult to extrapolate the observed variant incidences to those of responders vs nonresponders in general.

**COST AND TIME OF VARIANT SCREENING**
We compared the pDHPLC-based system and direct sequencing with respect to the costs and time required for EGFR variant screening. The cost per sample unit for the 2 methods was based on reagent costs in Singapore and on optimal throughput analysis with 4 × 96-well plates. This system allows the analysis of 4 amplicons (1 per exon) from 94 test samples, 1 wild-type sample, and 1 negative-control sample. Processing times for the respective workflows were also estimated. The calculations showed that pDHPLC provided 56% cost savings and 39% time savings compared with direct sequencing (Fig. 4). The major savings with the pDHPLC method derived...
from eliminating the need to sequence large numbers of wild-type alleles.

**Discussion**

Since their initial characterization, EGFR variants have been associated with improved patient response to EGFR TKIs (1, 2, 18), improved survival in patients treated with (6-8) and without (10, 19) EGFR TKIs, female sex (1, 2), a history of never smoking (18), Asian heritage (2), bronchoalveolar-type histology (20), and absence of KRAS variants (21). The EGFR variant encoding Thr790Met has been associated with resistance to EGFR TKIs (22) and a familial risk of lung cancer (23). Its apparent association with clinical outcome demonstrates the strong clinical relevance for the detection of EGFR variants. EGFR variants also appear to define a distinct lung cancer subtype, and the investigation of such variants in this context may provide insights into the biology of this disease.

Our results suggest that pDHPLC analysis may be a useful approach for detecting EGFR variants in clinical tissue samples, with a number of potential advantages over current methods. pDHPLC detected variants in mutant–wild-type mixtures of 1:100, compared with 1:5 by direct sequencing (Fig. 1). Reports have described detection limits of 1:100 for SURVEYOR mismatch enzyme analysis (14) and 1:10 for both fragment length polymorphism/capillary electrophoresis (24) and TaqMan-based allele-specific assays (25), but no report has described detection limits for assays based on single-strand conformation variation (20) and melting curves (26). Our results indicate that pDHPLC has a capacity for detecting EGFR variants that is equivalent to or better than other currently available methods.

Our detection of 26 unique EGFR variants in 164 lung cancer samples (Tables 1 and 2) demonstrates the ability of pDHPLC to detect a variety of EGFR variants. We detected all the frequent EGFR variants—variations at codons 719, 787, and 858 and deletions in exon 19. In contrast, Marchetti et al. (20) detected only 8 unique variants in exons 18, 19, and 21 in 860 lung cancer samples in a study based on single-strand conformation variation, and in the largest direct-sequencing series to date, Shige-matsu et al. identified only 28 unique variants in 1136 lung cancer samples (27).

The ability of pDHPLC to detect such diverse EGFR variants from just 4 amplicons has advantages over many allele-specific methods, such as melting curve (26), restriction fragment length polymorphism (24), and TaqMan (25) analyses. Only 58% of the variants detected in the current study comprised the single-nucleotide variants at codons 719 and 858 and deletions in exon 19 that such methods commonly detect. To detect the same number of EGFR variants as with our approach, allele-specific methods would require large numbers of expensive primers, probes, and reactions and would consume large amounts of often scarce sample material. Methods based on real-time PCR analysis may be more rapid than pDHPLC and may become the method of choice for small numbers of samples and the short turnaround times needed for routine analysis. Our knowledge of the functional significance of most EGFR variants is limited, however, and the benefits from faster analysis are unlikely to compensate for the risk of false negatives arising from limited variant screening. That we detected the 734B: A→G variant in a patient responding to gefitinib that current allele-specific methods would not have identified emphasizes this point (Table 2).

The amenability of pDHPLC analysis to fraction collection allowed identification of variants beyond the detection limit of direct sequencing (Fig. 3). The frequency of such amplicons was 29% (18 of 63) in our study. Marchetti et al. (20) and Pan et al. (24) identified additional variants by single-strand conformation variation (5 of 39; 13%) and restriction fragment length polymorphism (4 of 29; 14%) methods that direct sequencing did not detect. By enrich-
ing for tumor content with laser microdissection, Takano et al. (5) found variants in an additional 26% of samples (11 of 43) that direct sequencing did not detect in nonenriched samples. The high frequency of these cases (13%–29%) and the possibility that the variants identified in these cases could explain responses to EGFR TKIs in the apparent absence of EGFR variants identifiable by direct sequencing highlight the importance of the superior detection limit of pDHPLC and its amenability to fraction collection. That these fraction-collection capabilities could also decrease the need for time-consuming and labor-intensive microdissection is another benefit.

Recently, Janne et al. (14) reported a similar fraction-collection procedure; however, the workflow comprised DNA digestion by SURVEYOR enzymes, nondenaturing size analysis on the DHPLC instrument, repeat PCR analysis of variant amplicons, pDHPLC determination of aberrant peak retention times, reinjection of the PCR product, and then fraction collection. By requiring only the last 3 steps, our method decreases sample and time requirements.

pDHPLC analysis is 56% less expensive and 39% faster than direct sequencing (Fig. 4). These differences could make pDHPLC feasible for many laboratories and health maintenance systems.

Our finding of the wild-type sequence only in 40 randomly selected amplicons from lung cancer samples previously determined by pDHPLC to lack variants dem-

Table 2. \textit{EGFR} variants detected by pDHPLC and direct sequencing analysis in small biopsy samples from patients treated with gefitinib.

<table>
<thead>
<tr>
<th>Case</th>
<th>Site</th>
<th>Type</th>
<th>Sex</th>
<th>Smoker</th>
<th>pDHPLC: variant exons</th>
<th>Direct sequencing: nucleotide (protein) change$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pericardium</td>
<td>Aspirate</td>
<td>Female</td>
<td>No</td>
<td>19</td>
<td>746A–750C: del 15 bp (del Glu746-Ala750)</td>
</tr>
<tr>
<td>2</td>
<td>Bronchus</td>
<td>Aspirate</td>
<td>Female</td>
<td>No</td>
<td>19 + 21</td>
<td>734B: A→G (Glu734Gly) + 858B: T→G (Leu858Arg)</td>
</tr>
<tr>
<td>3</td>
<td>Lung</td>
<td>Aspirate</td>
<td>Female</td>
<td>No</td>
<td>19</td>
<td>759C: C→A (Ile759Leu)</td>
</tr>
<tr>
<td>4</td>
<td>Lung</td>
<td>Aspirate</td>
<td>Female</td>
<td>No</td>
<td>18</td>
<td>719B: G→A (Gly719Ser)</td>
</tr>
<tr>
<td>5</td>
<td>Lung</td>
<td>Biopsy</td>
<td>Female</td>
<td>No</td>
<td>19</td>
<td>746A–750C: del 15 bp (del Glu746-Ala750)</td>
</tr>
<tr>
<td>6</td>
<td>Pleura</td>
<td>Aspirate</td>
<td>Female</td>
<td>No</td>
<td>Wild-type</td>
<td>Wild-type</td>
</tr>
<tr>
<td>7</td>
<td>Pleura</td>
<td>Aspirate</td>
<td>Female</td>
<td>No</td>
<td>19</td>
<td>IVS19 + 53: C→A</td>
</tr>
<tr>
<td>8</td>
<td>Pleura</td>
<td>Aspirate</td>
<td>Male</td>
<td>Yes</td>
<td>19</td>
<td>746A–750C: del 15 bp (del Glu746-Ala750)</td>
</tr>
<tr>
<td>9</td>
<td>Lung</td>
<td>Biopsy</td>
<td>Male</td>
<td>Yes</td>
<td>Wild-type</td>
<td>Wild-type</td>
</tr>
<tr>
<td>10</td>
<td>Pleura</td>
<td>Biopsy</td>
<td>Male</td>
<td>No</td>
<td>21</td>
<td>858B: T→G (Leu858Arg)</td>
</tr>
</tbody>
</table>

$^*$ Letters A, B, and C after codon numbers denote 1st, 2nd, and 3rd nucleotide positions, respectively. Del, deletion.
onstrates this method’s specificity. Furthermore, we found no sequence variants in 8 fraction-collected ampli-
cons and 31 amplicons from gefitinib-treated patients that pDHPLC determined to be wild-type. We could have further investigated sensitivity, specificity, and positive and negative predictive values by comparing pDHPLC and direct sequencing with respect to the detection fre-
quencies of \( \text{EGFR} \) variants for the entire series of clinical samples; however, limitations in finances and sample availability precluded such an investigation.

Finally, the detection of a higher frequency of \( \text{EGFR} \) variants in small biopsy samples from responders to gefitinib treatment than in samples from nonresponders (Table 2) indicates that pDHPLC can provide clinically relevant results from clinically relevant samples. This achievement is important because until very recently (28), \( \text{EGFR} \) variant analyses had been performed only on relatively large, surgically resected samples. Resection samples, however, are primarily obtained only from ear-
ly-stage lung cancers, whereas EGFR TKI therapies are currently indicated for late-stage refractory disease. Hence, the demonstration that small biopsy specimens are amenable to \( \text{EGFR} \) variant analysis should significantly enhance the clinical value of the pDHPLC assay.

In summary, the present results indicate that pDHPLC can be an inexpensive and fast system, with a good detection limit for detecting \( \text{EGFR} \) variants in clinical tissue samples. The assay has a number of advantages over current methods that make it amenable to routine \( \text{EGFR} \) variant analysis. The benefits from pDHPLC, however, are contingent on being able to meet the high start-up costs and the availability of the necessary expertise. Nevertheless, in situations with large numbers of samples, the decreases in costs and time and the superior detection limit of pDHPLC should eventually compensate for the initial outlay. The next challenge may well be the interpretation of results. pDHPLC detected a number of

---

**Fig. 4. Comparisons of costs and time for \( \text{EGFR} \) variant analysis by direct sequencing (left) and pDHPLC (right).**

Calculations are based on optimal analysis of 4 × 96-well plates (1 for each of the 4 exons). Costs are in Singaporean dollars and include only reagent costs. Total costs per reaction for each procedure were calculated by itemizing reagents and adding individual reagent costs per reaction. Times do not include preparation time and replicate analysis. A breakdown of the costs can be obtained from the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol53/issue1.
intronic and silent sequence variants in the present study. Moreover, some novel variants were detected in gefitinib responders, whereas exon 19 deletions and the Leu858Arg variant were detected in nonresponders to this treatment. The latter findings have also been reported for other studies (3, 4). These novel variants may represent either novel functional changes or sequencing artifacts; thus, these results call for additional in vivo and in vitro studies to improve our understanding of the functional significance of individual EGFR variants. Our results suggest that pDHPLC can be an efficient assay to facilitate such investigations.

This work was supported by the National Medical Research Council, the National University of Singapore Office of Life Sciences, the National University of Singapore Oncology Research Institute Translational Interface, and Prof. Yoshiaki Ito. We thank Dr. Jeffrey Settleman for reading the manuscript and for helpful comments.

References

