

Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing

Roman K Thomas^{1,2,12}, Elizabeth Nickerson^{3,12}, Jan F Simons^{3,11,12}, Pasi A Jänne¹, Torstein Tengs^{1,2}, Yuki Yuza¹, Levi A Garraway^{1,2,4}, Thomas LaFramboise^{1,2}, Jeffrey C Lee^{1,2}, Kinjal Shah^{1,2}, Keith O'Neill², Hidefumi Sasaki⁵, Neal Lindeman⁶, Kwok-Kin Wong¹, Ana M Borras⁷, Edward J Gutmann⁸, Konstantin H Dragnev⁹, Ralph DeBiasi^{1,2}, Tzu-Hsiu Chen^{1,2}, Karen A Glatt¹, Heidi Greulich^{1,2}, Brian Desany³, Christine K Lubeski³, William Brockman², Pablo Alvarez², Stephen K Hutchison³, J H Leamon³, Michael T Ronan³, Gregory S Turenchalk³, Michael Egholm³, William R Sellers^{1,2}, Jonathan M Rothberg³ & Matthew Meyerson^{1,2,10,11}

The sensitivity of conventional DNA sequencing in tumor biopsies is limited by stromal contamination and by genetic heterogeneity within the cancer. Here, we show that microreactor-based pyrosequencing can detect rare cancer-associated sequence variations by independent and parallel sampling of multiple representatives of a given DNA fragment. This technology can thereby facilitate accurate molecular diagnosis of heterogeneous cancer specimens and enable patient selection for targeted cancer therapies.

The targeting of pivotal genetic alterations has led to remarkable successes in cancer care, such as the treatment of chronic myeloid leukemia (CML) and gastrointestinal stromal cell tumors (GIST) with the kinase inhibitor imatinib¹. Where diverse and specific mutations are a major determinant of the response to targeted therapies, DNA sequencing is likely to provide the most effective analytic and diagnostic approach, in contrast to mutant-specific genotyping, which can detect only known sequence variations. For example, the response to imatinib diverges among GISTs bearing distinct *KIT* and *PDGFR* mutations²; multiple secondary imatinib-resistance alleles can be detected in GIST and CML³, and multiple primary and secondary mutations in the epidermal growth factor receptor gene (*EGFR*) in non-small-cell lung cancer (NSCLC) have been found to predict response to the tyrosine kinase inhibitors gefitinib and erlotinib^{1,4–8}. Selection of both primary therapy and targeted inhibitors for relapsed individuals^{8–10} can benefit from DNA sequence analysis.

Although commonly used in many clinical settings, dideoxynucleotide chain termination (or 'Sanger') sequencing¹¹ of PCR products often lacks sufficient sensitivity for detecting mutant alleles in tumor biopsies, where the failure rate has reached 75% in some cases¹². Gain-of-function oncogenic mutations are frequently heterozygous events or may represent a single allele of an amplified gene; thus, the signal for mutated residues is typically reduced relative to neighboring bases. Moreover, the ability to detect single base mutations or small insertions or deletions in biopsy material by Sanger sequencing depends heavily on sample purity (for example, the extent of contaminating stromal DNA) and genomic DNA integrity. Furthermore, resistance to kinase inhibitors may correlate with low-frequency second-site mutations^{1,7,8}. These observations underscore the challenges for accurate mutation detection in cancer specimens.

A massively parallel sequencing-by-synthesis approach, 'picotiter plate pyrosequencing,' provides a new alternative to Sanger sequencing. This approach relies on emulsion PCR-based clonal amplification of a DNA library adapted onto micron-sized beads and subsequent pyrosequencing-by-synthesis¹³ of each clonally amplified template in a picotiter plate (**Supplementary Fig. 1 and Supplementary Note** online), generating over 200,000 unique clonal sequencing reads per experiment¹⁴. Sequence variants that represent a fraction of a complex sample can be vastly oversampled, thus enabling statistically meaningful quantification of low-abundance variants (**Supplementary Fig. 2 and Supplementary Note** online).

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. ²The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ³454 Life Sciences, 20 Commercial Street, Branford, Connecticut 06405, USA.

⁴Melanoma Program in Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA. ⁵Department of Surgery 2, Nagoya City University Medical School, Kawasumi 1, Mizuho-cho, Mizuho-ku, Nagoya 467-8601, Japan. ⁶Department of Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA. ⁷Translational Research Laboratory, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA.

⁸Department of Pathology and ⁹Hematology/Oncology, Department of Medicine, Dartmouth-Hitchcock Medical Center, Norris Cotton Cancer Center, One Medical Center Drive, Lebanon, New Hampshire 03756, USA. ¹⁰Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

¹¹Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA. ¹²These authors contributed equally to this work. Correspondence should be addressed to M.M. (matthew_meyerson@dfci.harvard.edu).

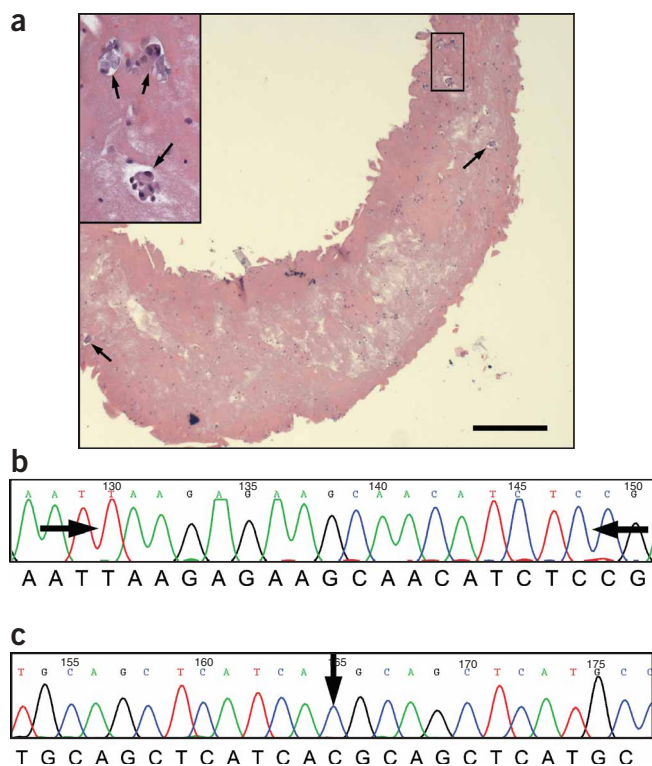


Figure 1 Failure of Sanger sequencing to detect clinically relevant *EGFR* mutations in a malignant pleural effusion specimen with low tumor content. (a) Photomicrograph of a hematoxylin and eosin-stained section of a paraffin-embedded fibrin clot from the pleural effusion fluid obtained at time of relapse. Four clusters of tumor cells showing features of adenocarcinoma, including rudimentary gland formation (arrows; <50 total cells), are found within a mix of benign inflammatory and mesothelial cells. Scale bar, 250 μ m. Sanger sequencing of exons 19 (b) and 20 (c) of *EGFR* from DNA isolated from the sample in a. Arrows indicate the sites of the mutations revealed by picotiter plate sequencing but not visible in the electropherograms.

of the mutations detected by pyrosequencing, but these were computationally indistinguishable from experimental noise (**Supplementary Fig. 4** online). The low percentage of mutation in samples that had been estimated to contain high tumor content most likely relates to difficulties in tumor content evaluation, especially as the tissue section that we had analyzed microscopically does not correspond precisely to the tissue from the same specimen, from which DNA was extracted. The newly discovered P772_H773insV mutation is likely to be oncogenic, as it was able to transform NIH-3T3 cells (**Supplementary Fig. 5** and **Supplementary Note** online).

Formalin-fixed, paraffin-embedded (FFPE) tissue specimens are a widespread source of clinically available cancer samples. We therefore analyzed ten FFPE lung cancer specimens, five of which were recently found to carry mutations in the *EGFR* kinase domain¹⁵. We carried out picotiter plate sequencing in a blinded fashion, which accurately detected all five mutants at frequencies ranging from 4.1 to 43.5% (**Supplementary Table 1** and **Supplementary Note** online), suggesting that this approach might be useful for FFPE cancer specimens.

The ability of picotiter plate pyrosequencing to reveal low-abundance mutations prompted us to analyze a clinical specimen with low tumor content in which *EGFR* mutations had been undetectable by Sanger sequencing. The individual's medical history, however, suggested a high probability of *EGFR* mutation (patient 12.3, **Supplementary Note** online); he was a nonsmoker with lung adenocarcinoma and had a strong partial response to erlotinib treatment before relapsing after 12.5 months with pleural effusions.

Histological analysis of a cell block, derived from pleural fluid obtained upon relapse, showed an estimated 1–10% tumor cells (**Fig. 1a**). Sanger sequencing of pleural effusion-derived DNA showed wild-type *EGFR* with no apparent sequence noise, as illustrated by representative traces of exons 19 and 20 (**Fig. 1b,c**). In contrast, picotiter plate pyrosequencing of 11 PCR products covering exons 18–22 of *EGFR* amplified from the same sample generated flowgrams representing a deletion mutation in exon 19 of *EGFR*, admixed with flowgrams representing the wild-type sequence at the same position (**Fig. 2a,b**). We found this 18-bp deletion, encoding the amino acid deletion-substitution L747_S752del_P753S (Del-4)⁴, at a frequency of approximately 3% of 11,367 reads (**Fig. 2c**). In addition, flowgrams showed that this specimen harbored a nucleotide substitution encoding the T790M mutation in exon 20 (**Fig. 2d**), associated with clinical resistance to *EGFR* inhibitors^{7,8}. The T790M mutation, present in approximately 2% of 136,776 reads, was admixed with wild-type exon 20 sequence (**Fig. 2e,f**). We also confirmed these mutations and their relative representation within the sample by subcloning of PCR products and subsequent Sanger sequencing (**Supplementary Fig. 6** and **Supplementary Note** online).

In a pretreatment cell block in which tumor cells were exceptionally rare, picotiter plate pyrosequencing showed the same exon 19 deletion (Del-4) at a frequency of approximately 0.3% (**Supplementary Fig. 7** online) but no T790M mutation, consistent with the original TKI

RESULTS

Performance of picotiter plate pyrosequencing

To test the performance of this parallel sequencing approach in mutation detection, we performed a dilution experiment, mixing PCR amplicons containing either a single base substitution or an exon 19 deletion mutation of *EGFR* with the corresponding wild-type amplicons. We subjected these mixed amplicons to emulsion PCR and subsequent picotiter plate pyrosequencing. In this experiment, we were able to detect the mutated allele in a linear manner at proportions as low as 0.2% at 65,000–110,000-fold oversampling (**Supplementary Fig. 3** and **Supplementary Note** online).

Mutation analysis in primary tumor specimens

To extend this approach to primary tumor specimens, 11 fragments of approximately 100 bp each, covering exons 18–22 of the *EGFR* gene, were individually amplified by PCR from DNA of 22 lung adenocarcinoma specimens. We pooled the resulting PCR products from each sample and subjected them to subsequent emulsion PCR and picotiter plate pyrosequencing. We had previously observed mutated *EGFR* in 9 of these samples and wild-type *EGFR* in 13 samples by Sanger sequencing⁴, each chosen for an estimated 70% tumor content by histological analysis.

In addition to validating the previously detected *EGFR* mutations, picotiter plate pyrosequencing showed *EGFR* mutations in 2 of the 13 samples previously defined as wild type by Sanger sequencing (**Supplementary Table 1** online). Newly detected mutations included a previously unknown insertion mutation, P772_H773insV (frequency, 11% of 619 reads; **Supplementary Fig. 4** online) and a known deletion mutation not detected by Sanger sequencing, E746_A750del5 (Del-1a; frequency, 9% of 4,488 reads; **Supplementary Fig. 4** online). Using retrospective inspection of the Sanger sequence, we identified additional faint peaks at the positions

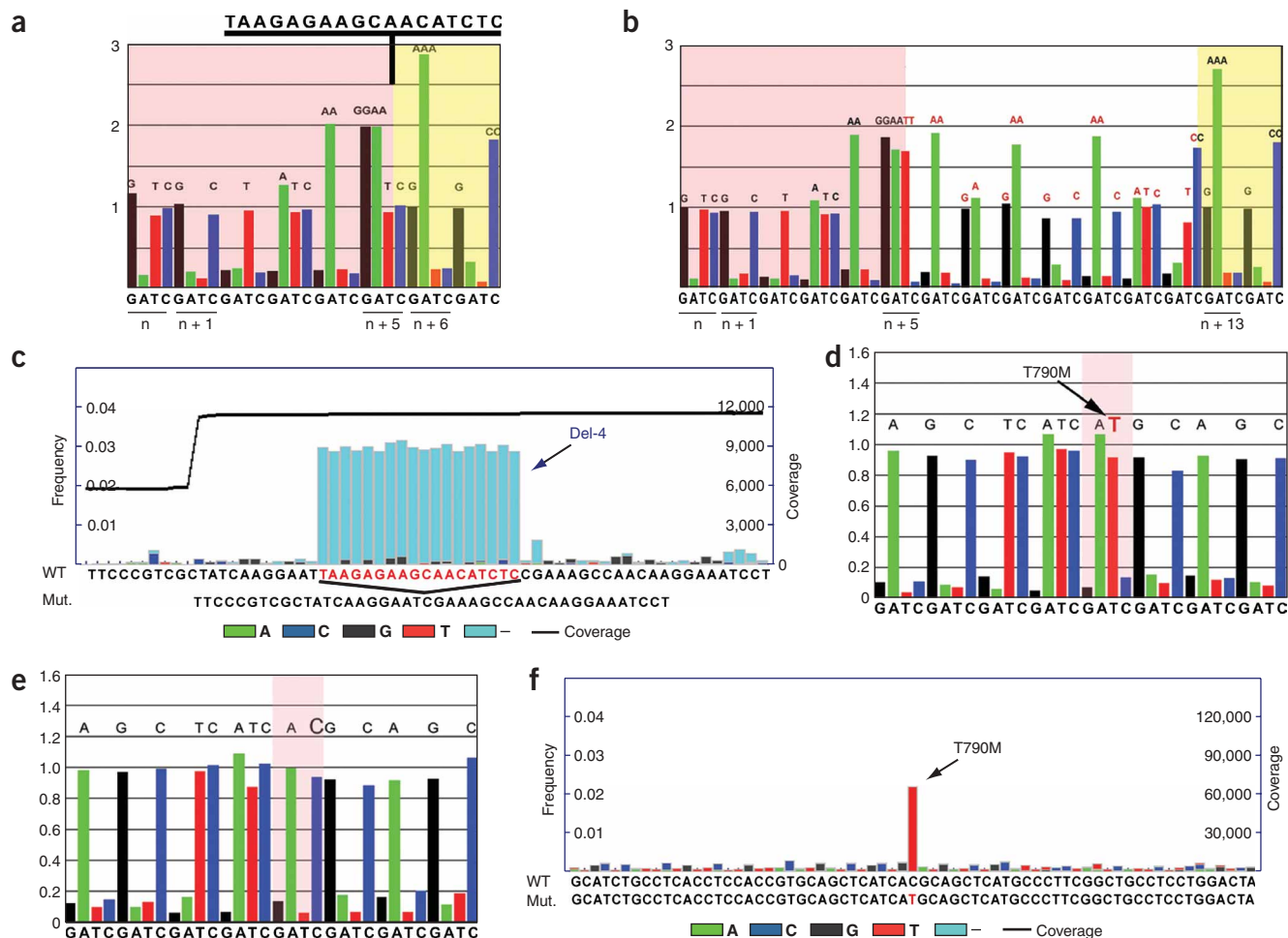


Figure 2 Detection of clinically relevant *EGFR* kinase domain mutations by array-based pyrosequencing in a malignant pleural effusion specimen with low tumor content. **(a,b)** Raw flowgrams from individual reads (wells) showing relative luminescence signal (y-axis) obtained with each sequentially flowed nucleotide from DNA extracted from the pleural effusion obtained from individual 12.3 at the time of relapse. The number of nucleotides in a homopolymer is proportional to the luminescent signal. **(a)** Raw flowgram (exon 19) revealing a deletion, Del-4 (L747_S752del_P753S). The altered nucleotide sequence leads to a shift in the flow cycle, indicated by the short lines below the nucleotide. The deleted sequence is shown above the panel; the black bar indicates the position of the deleted sequence. **(b)** Raw flowgram of nonmutant sequence of *EGFR* exon 19 in the same sample. The nucleotides deleted in **a** are indicated by a white background. **(c)** Variation plot analysis (sequencing coverage, right; mutation frequency, left) of *EGFR* exon 19 from picotiter plate sequencing of pleural fluid obtained from individual 12.3 at the time of relapse, revealing the mutation Del-4 at a frequency of ~3%. The wild-type and mutant sequences are given below the variation plot. **(d)** Raw flowgram (exon 20) from DNA extracted from pleural effusion obtained from individual 12.3 at time of relapse showing a substitution mutation, T790M. **(e)** Raw flowgram of a nonmutant sequence of *EGFR* exon 20 from the same sample. **(f)** Variation plot analysis of *EGFR* exon 20 of DNA extracted from the pleural fluid sample upon relapse. A substitution mutation in exon 20, T790M, is present at a relative allele frequency of approximately 2%.

sensitivity and subsequent relapse. Furthermore, we found these mutations to functionally recapitulate the individual's clinical course *in vitro*, as Ba/F3 cells transformed with the Del-4 mutant were sensitive to erlotinib, whereas cells transformed with the Del-4/T790M were resistant (**Supplementary Fig. 5** and **Supplementary Note**).

DISCUSSION

In summary, picotiter plate pyrosequencing enabled detection of low-abundance oncogene mutations in complex samples with low tumor content for which conventional Sanger sequencing was not informative. Although other technologies that can detect rare mutations have been proposed for this purpose, methods such as allele-specific genotyping are limited to known mutations, whereas methods that depend on subcloning of PCR products, in conjunction with conventional sequencing, are limited by bacterial cloning artefacts and time constraints.

Highly parallel sequencing approaches could make it feasible to monitor the molecular composition and evolution of tumor subtypes even in settings of low or impure tumor content without the need for laborious tumor cell enrichment methods. Picotiter plate pyrosequencing might thereby help, for example, to resolve the current controversy on the power of *EGFR* mutations as predictors of response and survival of patients treated with *EGFR* TKIs^{16–18}. The application of picotiter plate pyrosequencing could impact cancer diagnostics and therapeutics by affording redundant and therefore highly accurate mutation discovery in clinical cancer specimens.

METHODS

Tumor samples. We obtained tumor samples from 33 individuals with lung adenocarcinoma after obtaining their informed consent (**Supplementary Note** online). Twenty-two individuals were from Japan, ten were whites from the

United States and one was Vietnamese. The institutional review board of the Dana-Farber Cancer Institute approved the study. We extracted DNA using standard procedures.

Picotiter plate pyrosequencing. We generated PCR products using primers designed to cover exons 18–22 of the *EGFR* gene and adapted with 5' overhangs to facilitate emulsion polymerase chain reaction (emPCR) and sequencing (Supplementary Note online). emPCR and picotiter plate sequencing-by-synthesis were performed as recently described¹⁴. In brief, to favor single-template amplification during emPCR, we performed annealing with an average of 0.5 DNA molecules per bead and 450,000 beads per reaction. After amplification by emPCR, we isolated DNA-carrying beads. For most samples, we sequenced approximately 20,000 beads (10,000 each direction), yielding 8,000–12,000 sequencing reads on average per sample (~1,000 reads per amplicon per sample) using the GS20 picotiter plate pyrosequencing sequencing instrument provided by 454 Life Sciences through Roche. In some experiments, we performed additional confirmatory sequencing on 115,000–360,000 beads per sample, yielding from ~40,000 to ~150,000 sequencing reads per run.

Data analysis. We performed base calling as previously described¹⁴. To facilitate contiguous BLAST hits and to facilitate detection of large intragenic deletions and insertions, we performed a BLASTN analysis for mutation detection using permissive gapping parameters (Supplementary Note online). We filtered the top-scoring BLAST hit for each sequence read by several quality metrics for inclusion in a multiple alignment with the reference. We next used the qualifying top-hit sequence read segments to construct an alignment with respect to the reference sequence. We determined counts of bases in the alignment deviating from the reference for each reference position.

Functional experiments. We introduced *EGFR* cDNAs carrying the mutation Del-4 either alone or in combination with the T790M as well as the P772_H773insV mutation into pBABE-puro retroviral vectors (Supplementary Note online). We transduced interleukin (IL)-3-dependent BA/F3 cells with the retroviruses and transformed them to become IL-3 independent, after which we treated pooled stable cells with erlotinib. We determined viability using the MTS assay. Similarly, we transduced NIH-3T3 cells with mutant *EGFR* retroviruses, and we suspended pooled stable cells in soft agar as described to assay for their ability to grow in an anchorage-independent fashion¹⁹.

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

ACKNOWLEDGMENTS

This work was supported by the Deutsche Krebshilfe through a Mildred-Scheel Fellowship for Cancer Research to R.K.T. W.R.S. and M.M. are supported by the Novartis Research Foundation, the Claudia Adams Barr Foundation and the Charles A. Dana Human Cancer Genetics Program of the Dana-Farber Cancer Institute, the Poduska Family Foundation, the Damon-Runyon Cancer Research Foundation, Joan's Legacy, the American Cancer Society (RSG-03-240-01-MGO), the Flight Attendant Medical Research Institute and the US National Cancer Institute (R01CA098185). We are greatly indebted to J. Prensner and J. Baldwin for comments and advice.

AUTHOR CONTRIBUTIONS

R.K.T., E.N., J.F.S., M.E., W.R.S., J.M.R. and M.M. designed the research. R.K.T., E.N., J.F.S., T.T., Y.Y., T.L., J.C.L., K.S., K.O'N., R.D., T.-H.C., K.A.G., H.G., B.D., C.K.L., W.B., P.A., S.K.H., J.H.L., M.T.R. and G.S.T. performed research and analyzed data. P.A.J., H.S., N.L., K.-K.W., A.M.B., E.J.G. and K.H.D. provided samples and patient information. R.K.T., E.N., J.F.S., L.A.G., M.E., W.R.S. and M.M. wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the Nature Medicine website for details).

Published online at <http://www.nature.com/naturemedicine/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Sawyers, C. Targeted cancer therapy. *Nature* **432**, 294–297 (2004).
2. Heinrich, M.C. *et al.* Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J. Clin. Oncol.* **21**, 4342–4349 (2003).
3. Shah, N.P. *et al.* Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* **2**, 117–125 (2002).
4. Paez, J.G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
5. Pao, W. *et al.* EGFR 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* **101**, 13306–13311 (2004).
6. Lynch, T.J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
7. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, e73 (2005).
8. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).
9. Shah, N.P. *et al.* Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science* **305**, 399–401 (2004).
10. Kwak, E.L. *et al.* Irreversible inhibitors of the EGF receptor may circumvent acquired resistance to gefitinib. *Proc. Natl. Acad. Sci. USA* **102**, 7665–7670 (2005).
11. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467 (1977).
12. van Ommen, G.J., Bakker, E. & den Dunnen, J.T. The human genome project and the future of diagnostics, treatment, and prevention. *Lancet* **354** Suppl 1, S15–10 (1999).
13. Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
14. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
15. Janne, P.A. *et al.* A rapid and sensitive enzymatic method for epidermal growth factor receptor mutation screening. *Clin. Cancer Res.* **12**, 751–758 (2006).
16. Tsao, M.S. *et al.* Erlotinib in lung cancer - molecular and clinical predictors of outcome. *N. Engl. J. Med.* **353**, 133–144 (2005).
17. Bell, D.W. *et al.* Epidermal growth factor receptor mutations and gene amplification in non-small-cell lung cancer: molecular analysis of the IDEAL/INTACT gefitinib trials. *J. Clin. Oncol.* **23**, 8081–8092 (2005).
18. Cortes-Funes, H. *et al.* Epidermal growth factor receptor activating mutations in Spanish gefitinib-treated non-small-cell lung cancer patients. *Ann. Oncol.* **16**, 1081–1086 (2005).
19. Greulich, H. *et al.* Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Med.* **2**, e313 (2005).

Erratum: Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing

Roman K Thomas, Elizabeth Nickerson, Jan F Simons, Pasi A Jänne, Torstein Tengs, Yuki Yuza, Levi A Garraway, Thomas LaFramboise, Jeffrey C Lee, Kinjal Shah, Keith O'Neill, Hidefumi Sasaki, Neal Lindeman, Kwok-Kin Wong, Ana M Borras, Edward J Gutmann, Konstantin H Dragnev, Ralph DeBiasi, Tzu-Hsiu Chen, Karen A Glatt, Heidi Greulich, Brian Desany, Christine K Lubeski, William Brockman, Pablo Alvarez, Stephen K Hutchison, J H Leamon, Michael T Ronan, Gregory S Turenchalk, Michael Egholm, William R Sellers, Jonathan M Rothberg & Matthew Meyerson

Nat. Med. 12, 852–855 (2006); published online 25 June; corrected after print 13 September 2006

In the version of this article initially published, it should have been acknowledged that Jan F. Simons, in addition to Roman K. Thomas and Elizabeth Nickerson, contributed equally to this work. The error has been corrected in the HTML and PDF versions of the article.