nature medicine

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], Pasi A Jänne[1], Torstein Tengs[1,2], Yuki Yuza[1], Levi A Garraway[1,2,4], Thomas LaFramboise[1,2], Jeffrey C Lee[1,2], Kinjal Shah[1,2], Keith O'Neill[2], Hidefumi Sasaki[5], Neal Lindeman[6], Kwok-Kin Wong[1], Ana M Borras[7], Edward J Gutmann[8], Konstantin H Dragnev[9], Ralph DeBiasi[1,2], Tzu-Hsiu Chen[1,2], Karen A Glatt[1], Heidi Greulich[1,2], Brian Desany[3], Christine K Lubeski[3], William Brockman[2], Pablo Alvarez[2], Stephen K Hutchison[3], J H Leamon[3], Michael T Ronan[3], Gregory S Turenchalk[3], Michael Egholm[3], William R Sellers[1,2], Jonathan M Rothberg[3] & Matthew Meyerson[1,2,10,11]

© 2006 Nature Publishing Group http://www.nature.com/naturemedicine

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 imatinib[1]. 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 *PDGFRA* mutations[2]; multiple secondary imatinib-resistance alleles can be detected in GIST and CML[3], 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') sequencing[11] of PCR products often lacks sufficient sensitivity for detecting mutant alleles in tumor biopsies, where the failure rate has reached 75% in some cases[12]. 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-synthesis[13] 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 experiment[14]. 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).

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. [2]The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [3]454 Life Sciences, 20 Commercial Street, Branford, Connecticut 06405, USA. [4]Melanoma Program in Medical Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA. [5]Department of Surgery 2, Nagoya City University Medical School, Kawasumi 1, Mizuho-cho, Mizuho-ku, Nagoya 467-8601, Japan. [6]Department of Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA. [7]Translational Research Laboratory, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA. [8]Department of Pathology and [9]Hematology/Oncology, Department of Medicine, Dartmouth-Hitchcock Medical Center, Norris Cotton Cancer Center, One Medical Center Drive, Lebanon, New Hampshire 03756, USA. [10]Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [11]Center for Cancer Genome Discovery, Dana-Farber Cancer Institue, 44 Binney Street, Boston, Massachusetts 02115, USA. [12]These authors contributed equally to this work. Correspondence should be addressed to M.M. (matthew_meyerson@dfci.harvard.edu).

Received 29 November 2005; accepted 16 May 2006; published online 25 June 2006; doi:10.1038/nm1437

NEOGENOMICS EX.1105.001

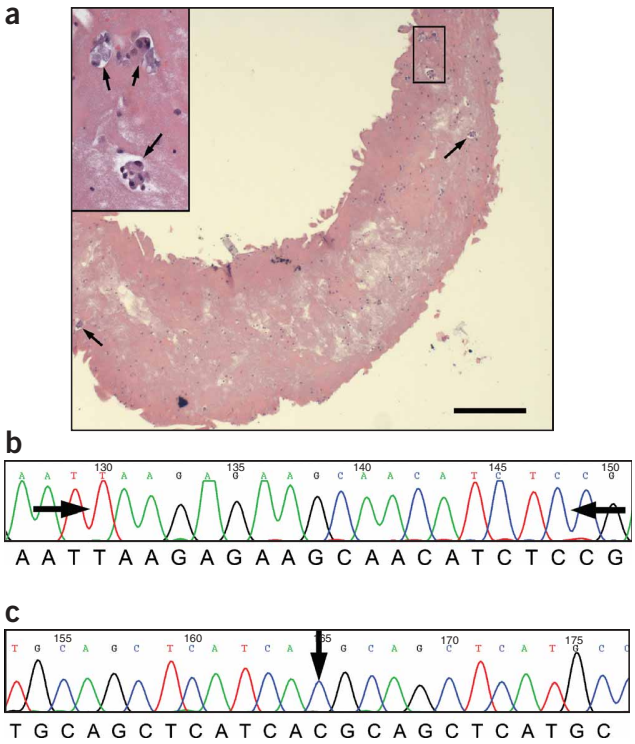


Figure 1 Failure of Sanger sequencing to detect clinically relevant *EGFR* mutations in a malignant pleural effusion specimen with low tumor content. (**a**) Photomicrograph of a hematoxylin and eosin–stained section of a paraffin-embedded fibrin clot from the pleural effusion fluid obtained at time 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 μm. 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.

© 2006 Nature Publishing Group http://www.nature.com/naturemedicine

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 sequencing[4], 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 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 domain[15]. 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

NEOGENOMICS EX.1105.002

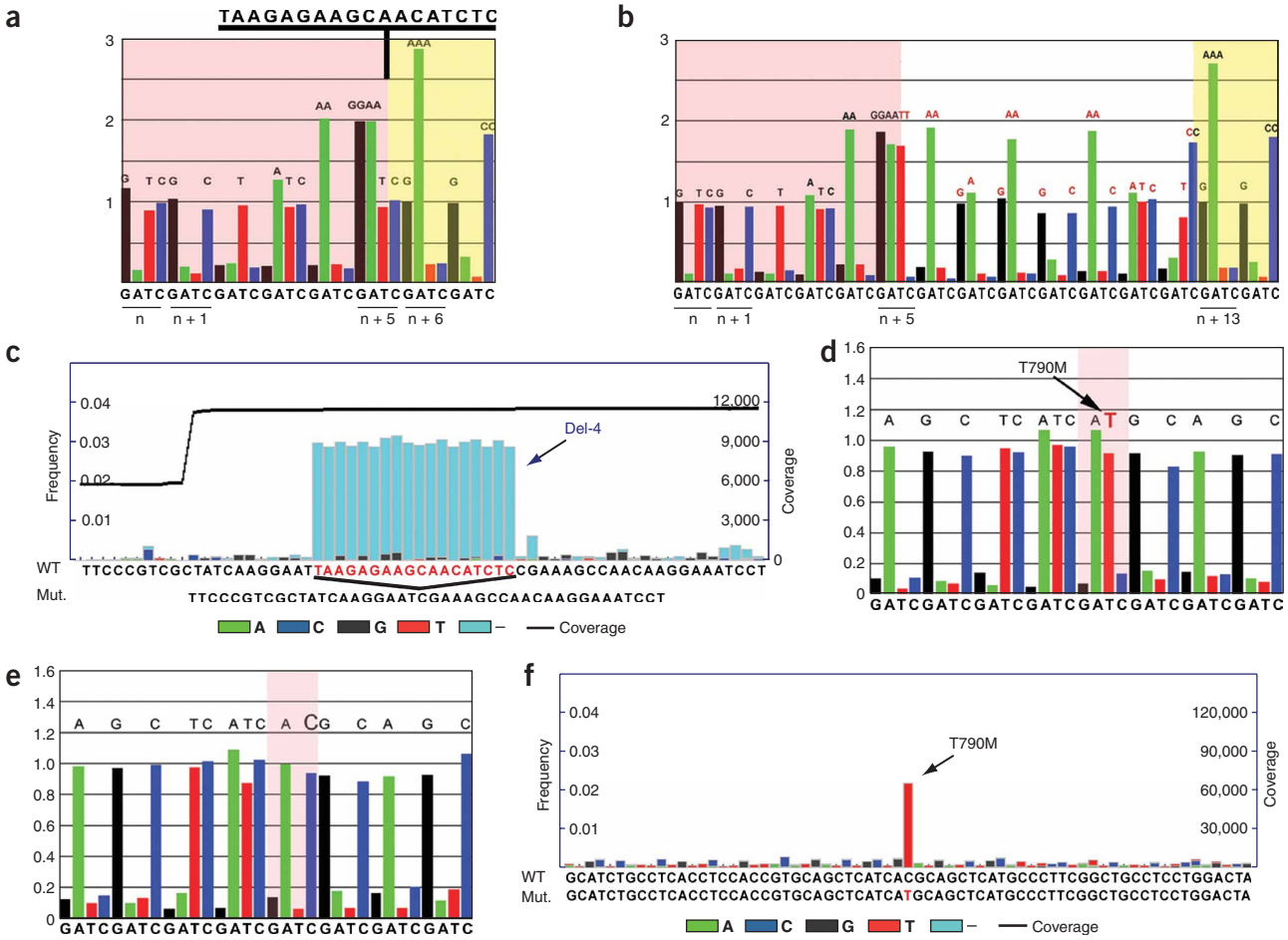


Figure 2 Detection of clinically relevant *EGFR* kinase domain mutations by array-based pyrosequencing in a malignant pleural effusion specimen with low tumor content. (**a**,**b**) Raw flowgrams from individual reads (wells) showing relative luminescence signal (*y*-axis) obtained with each sequentially 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%.

© 2006 Nature Publishing Group http://www.nature.com/naturemedicine

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

NEOGENOMICS EX.1105.003

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 described[14]. 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 ($\sim$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 $\sim$40,000 to $\sim$150,000 sequencing reads per run.

Data analysis. We performed base calling as previously described[14]. 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 fashion[19].

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/

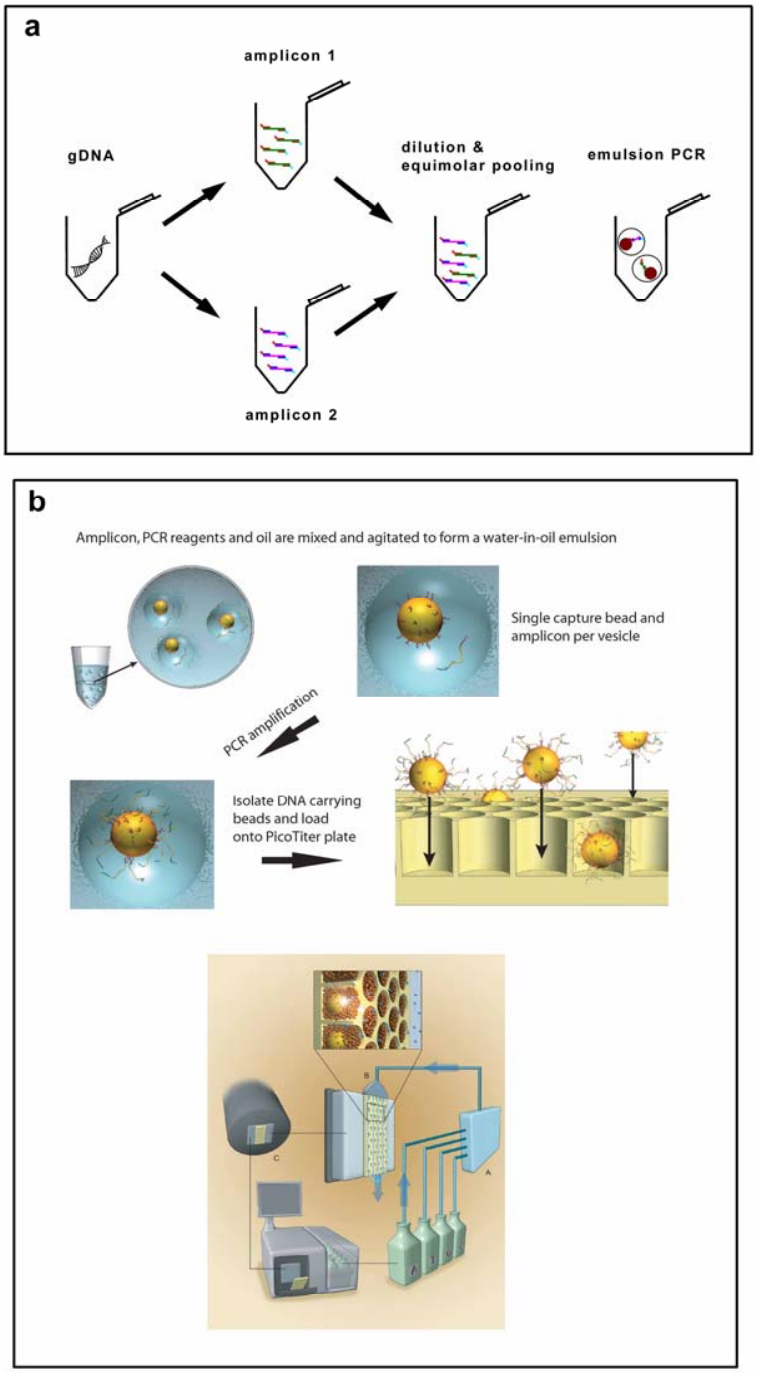
1. Sawyers, C. Targeted cancer therapy. *Nature* **432**, 294–297 (2004).
2. Heinrich, M.C. *et al.* Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J. Clin. Oncol.* **21**, 4342–4349 (2003).
3. Shah, N.P. *et al.* Multiple BCR-ABL kinase domain mutations 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.* EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA* **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, SI5–10 (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).

© 2006 Nature Publishing Group http://www.nature.com/naturemedicine

View publication stats

NEOGENOMICS EX.1105.004

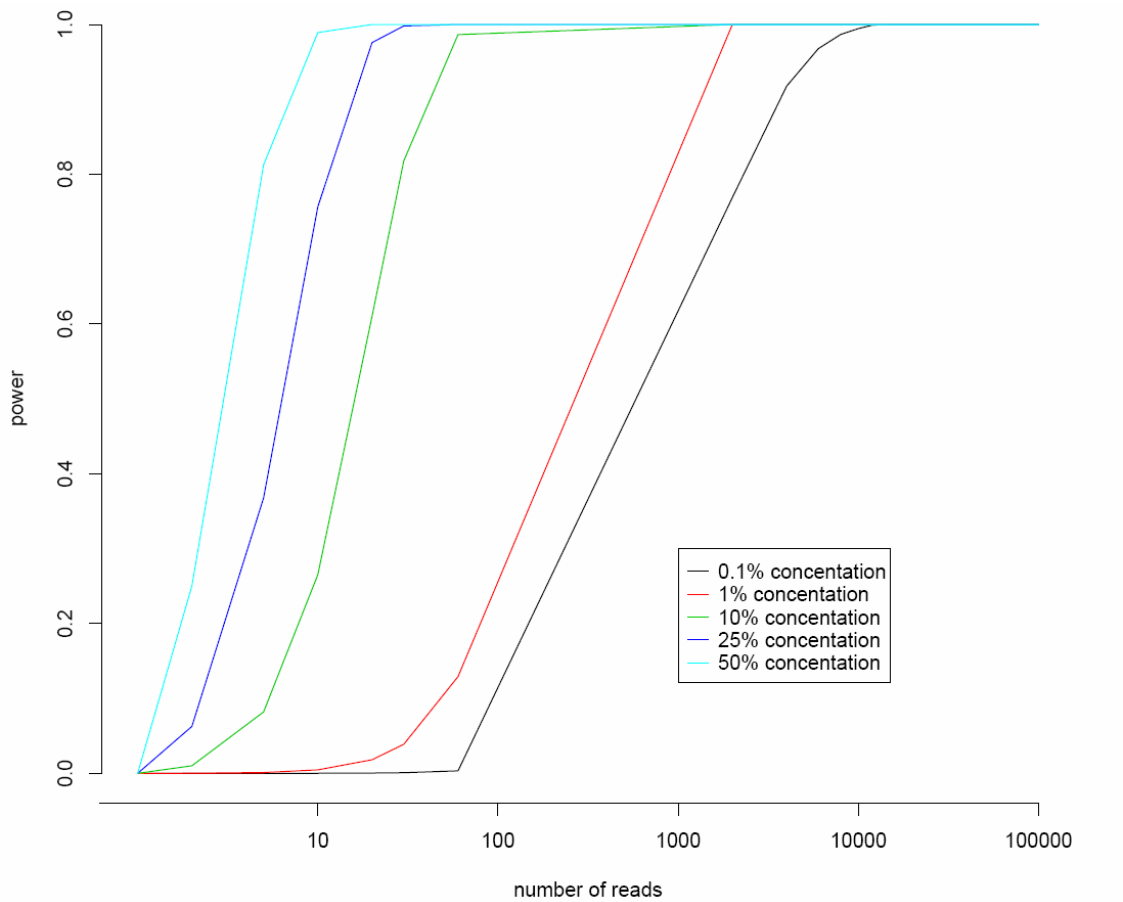
Supplementary Figure 1



Supplementary Figure 1. Schema of experimental approach for PCR-based exon-re-sequencing. Detailed explanations are given in the text. **a** Target amplicons are amplified from genomic DNA using specific primers. PCR products are pooled and subjected to emulsion PCR. **b** Bead-bound, clonally amplified single PCR molecules are recovered from the emulsion, deposited onto a picotiter plate and subjected to picotiter plate sequencing. Figure adopted from Margulies et al[1].

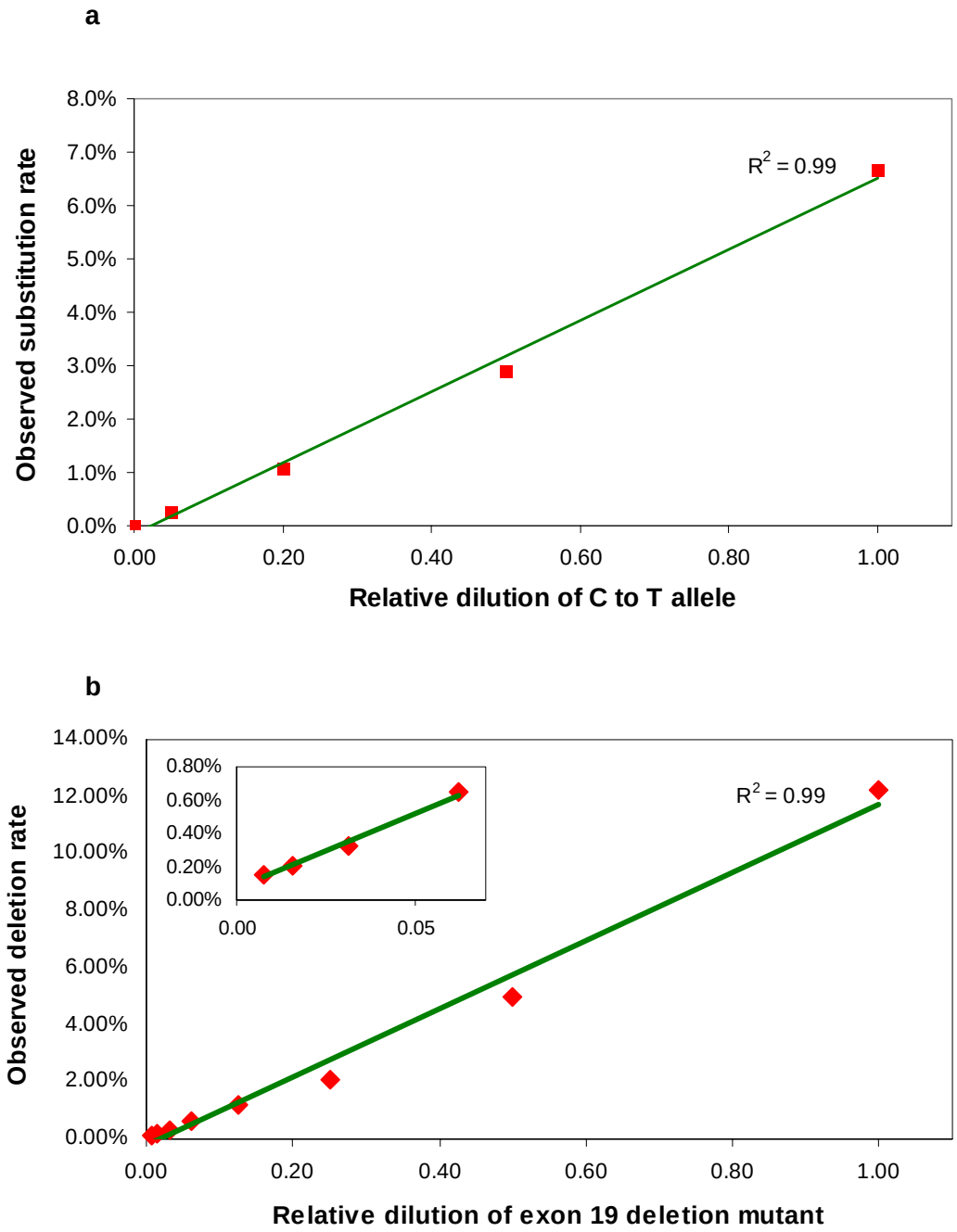
NEOGENOMICS EX.1105.005

Supplementary Figure 2



Supplementary Figure 2. Power as a function of number of reads for various concentrations of mutations in the sample. The term concentration here refers to the percentage of DNA fragments harboring the mutation. For example, in a sample with all cells carrying the mutation heterozygously, concentration would be 50%. As can be seen from the figure, we achieve essentially full power for concentrations as low as 1% for numbers of reads typical in our experiments. Even for concentrations less than 1%, very high power is attainable if the number of reads is increased.

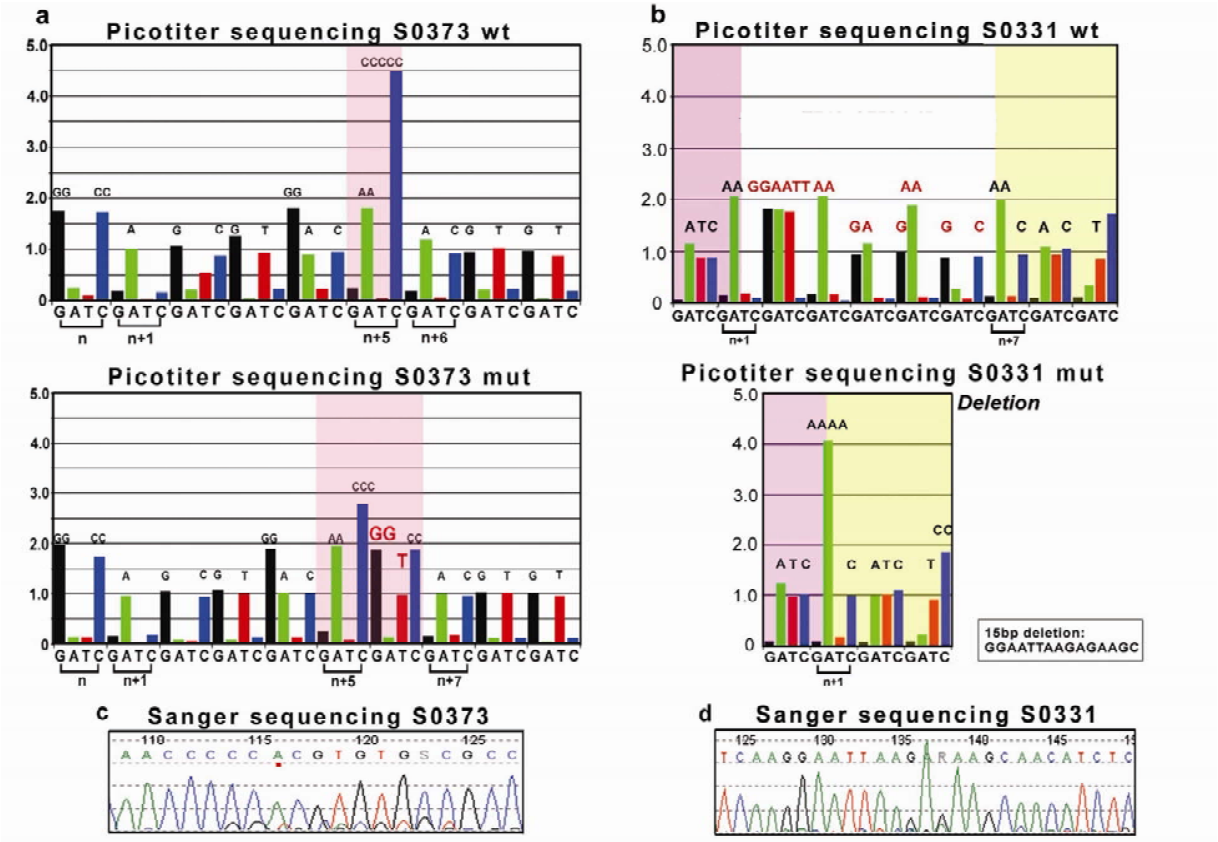
Supplementary Figure 3



Supplementary Figure 3. Serial dilution of amplicons containing a single base substitution (**a**) or a 12-bp deletion (**b**) into the corresponding wild-type backgrounds, followed by high-coverage sequencing for low level mutation detection. The inset in **b** shows a magnification of the lower part of the curve.

NEOGENOMICS EX.1105.007

Supplementary Figure 4



Supplementary Figure 4. Picotiter plate sequencing of samples S0373 and S0331.

a Top panel: raw flowgram from individual reads (wells) showing relative luminescence signal (Y-axis) obtained with each sequentially flowed nucleotide from sample S0373. The number of nucleotides in a homopolymer is proportional to the luminescent signal. flow order. Shown is an example of a non-mutant pyrogram ("wt")

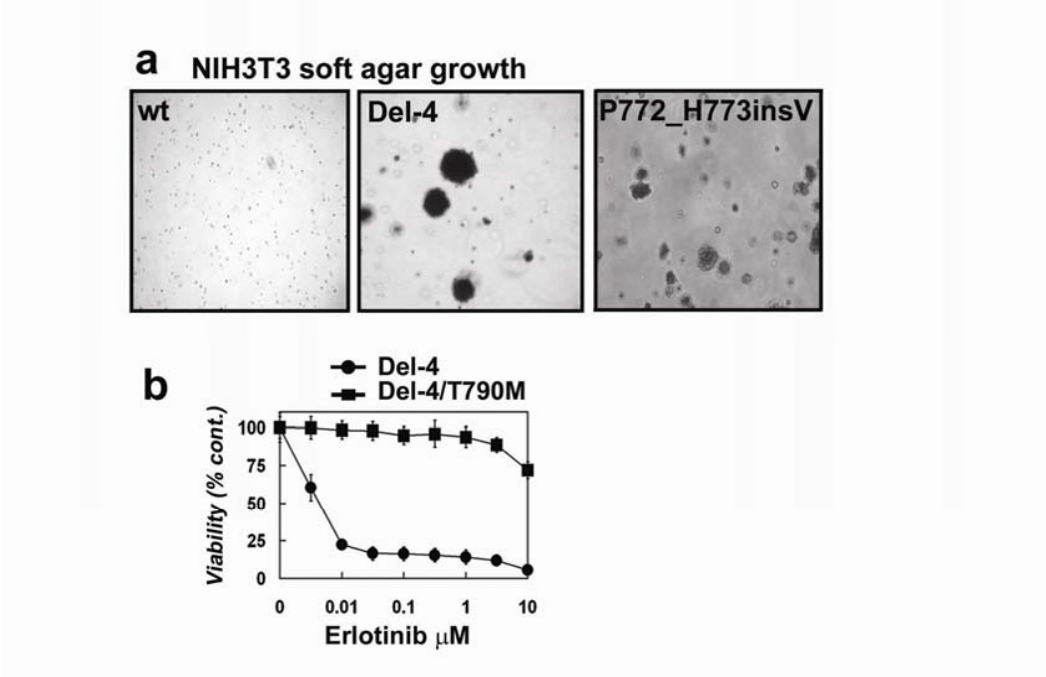
Bottom panel: raw flowgram from sample S0373. Note the insertion GGT in this mutant pyrogram ("mut", red nucleotides). The altered nucleotide sequence leads to a shift in the flow cycle, indicated by the square brackets below the nucleotide

b Top panel: non-mutant ("wt") raw flowgram from sample S0331. Bottom panel: mutant ("mut") raw flowgram from sample S0331. Note the 15-bp deletion (E746_A750del5, Del-1a) in this sample leading to a shift in the flow cycles.

c Corresponding Sanger sequencing traces of sample S0373.

d Corresponding Sanger sequencing traces of sample S0331.

Supplementary Figure 5

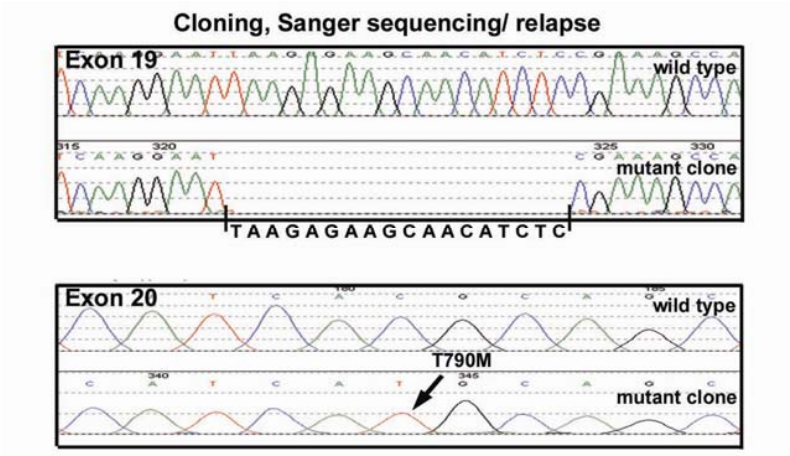


Supplementary Figure 5. Oncogenic transformation by patient-derived EGFR mutations and differential sensitivity of Del-4 with and without T790M mutation

a Murine NIH3T3 fibroblasts, infected with a retrovirus expressing the EGFR Del-4, the P773_H773insV mutation or a wild-type EGFR control (wt), were assayed for their ability to form colonies in soft agar.

b Ba/F3 cells, rendered IL3-independent by infection with a retrovirus carrying either Del-4 EGFR or Del-4 + T790M EGFR, were treated with the reversible EGFR TKI erlotinib in the absence of IL3. Cell viability was determined using the MTS assay. Values are depicted as percentages of the untreated control. Solid circles: Ba/F3 cells expressing Del-4 EGFR. Solid squares: Ba/F3 cells expressing Del-4/T790M EGFR.

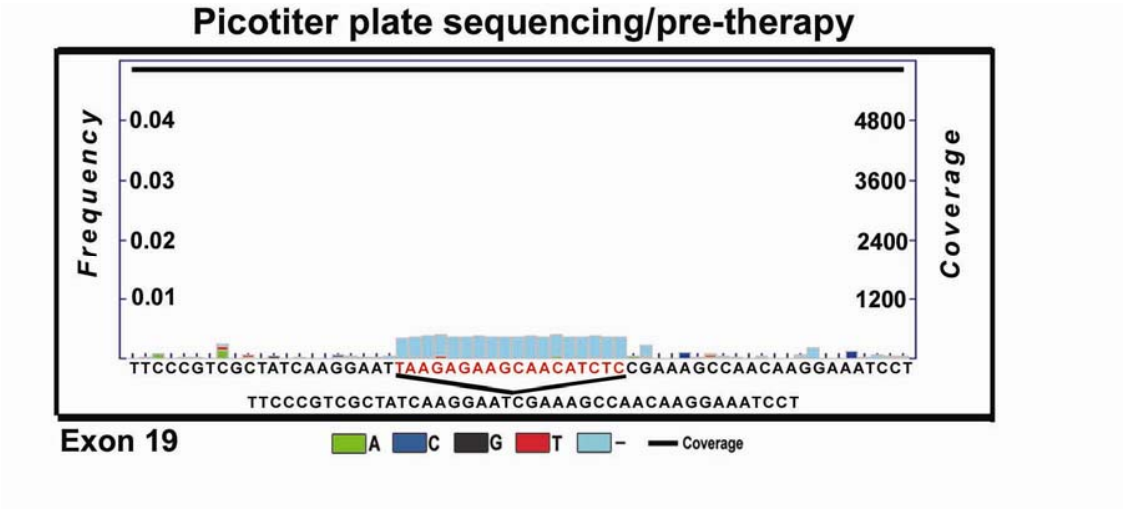
Supplementary Figure 6



Supplementary Figure 6. Validation of mutations detected by Picotiter plate sequencing using extensive cloning and sequencing

Exons 19 and 20 were amplified from DNA obtained from patient 12.3 at time of relapse, ligated into plasmid vectors and cloned. At least 768 clones were isolated per amplicon and each clone was sequenced bidirectionally. Sanger traces show either the reference sequence ("wild type") or the sequence from an exemplary clone carrying the respective mutation ("mutant clone"). Shown are examples of clones harboring the Del-4 (upper panel), and of clones carrying the T790M mutation (lower panel) as detected by Mutation Surveyor software.

Supplementary Figure 7



Supplementary Figure 7 Picotiter plate sequencing of a paraffin-embedded pleural effusion sample obtained before onset of therapy from patient 12.3. Variation plot analysis within *EGFR* exon 19 showing an 18-bp deletion (Del-4) at an allele frequency of ~0.3%, with the wild-type and mutant sequences shown below.

NEOGENOMICS EX.1105.011

Supplementary Table 1 Mutations detected by picotiter sequencing

Sample	Run#	Exon	Amino acid change	Frequency (%)	Coverage
S0377*		18	G719S	17.90	1,231
S0377*$		**18**	**E709A**	**16.00**	**755**
S0380*		19	E746_A750del, Del-1a	15.20	1,566
S0399*		19	E746_A750del, Del-1a	47.00	621
S0353*		19	E746_A750del, Del-1a	28.00	1,169
S0405*		19	E746_A750del, Del-1b	44.50	2,314
S0361*		21	L858R	15.60	519
S0388*		21	L858R	46.90	303
S0392#					
S0326#					
S0335#		19	E746_A750del, Del-1b	26.50	1,975
S0300#					
S0426#					
S0431#		20			
S0439#		19	E752_I759del, Del-2	28.20	340
S0373#		**20**	**P772_H773insV**	**11.60**	**619**
S0381#					
S0396#					
S0337#					
S0331#		**19**	**E746_A750del**	**8.70**	**4,488**
S0350#					
S0427#					
S0409#					
2		*18*	*G719C*	*7.2*	*96,838*
4		*19*	*L747_S752del_P753S, A755G*	*43.5*	*9,849*
8		*19*	*E746_A750del, Del-1a*	*5.8*	*81,890*
10					
11		*21*	*L858R*	*40*	*4,2034*
14					
15					
19					
21					
27		*21*	*L858R*	*4.1*	*52,952*
12.3 pre-therapy	**1**	**19**	**L747_752del_P753S, Del-4**	**0.29**	**5,843**
12.3 relapse	**1**	**19**	**L747_752del_P753S, Del-4**	**2.9**	**11,367**
12.3 relapse	**2**	**20**	**T790M**	**2.17**	**136,776**

Legend: Bold fonts indicate mutations detected by picotiter sequencing only. Asterisks indicate samples analyzed in the first, non-blinded experiment. Pound symbols indicate samples analyzed in the second, blinded experiment. Italics indicate results obtained from DNA samples extracted from formalin-fixed, paraffin-embedded tissue.

$ Note that the E709A mutation was found in a sample where Sanger sequencing had revealed a G719S mutation but not the E709A mutation.

NEOGENOMICS EX.1105.012

Supplementary information

Picotiter plate sequencing

The picotiter plate sequencing instrument as well as the reagents were provided by 454 Life Sciences (Branford, CT; www.454.com) through Roche. The recently developed picotiter plate sequencing technology involves two critical steps, clonal amplification of a single DNA molecule onto micron-sized beads, and sequencing-by-synthesis in a microfabricated plate containing 1,600,000 wells sized to contain only one of the beads[1]. In a first step, a highly diluted DNA library that can be either a sheared genomic DNA library or a PCR product is bound to an excess of beads via specific adaptors (**Supplementary Fig. 1a**) that are added as 3' tails to the PCR primers or attached by ligation. This yields beads with single DNA molecules on their surface. Next, the beads together with reagents for PCR amplification are emulsified to form a water-in-oil emulsion which then is subjected to thermal cycling to achieve million-fold clonal amplification of the single bead-bound DNA molecules. Following amplification the beads are removed from the emulsion and DNA-carrying fraction of beads are isolated thorough a biotin-streptavidin-mediated enrichment step. DNA-carrying beads are deposited onto the picotiter sequencing plate by centrifugal force, one bead per well. Sequencing-by-synthesis is performed by flowing each triphospho-nucleotides in repeated cycles across the wells of the picotiter plate. Upon incorporation of a particular nucleotide into the emerging strand, the liberation of the pyrophosphate group is being used to generate ATP in a reaction catalyzed by sulfurylase. Luciferase in conjunction with ATP catalyzes the generation of light through oxidation of luciferin. Light emitted from each well is being captured by a high-resolution CCD camera. A nucleotide sequence output read ("flowgram") is generated by integrating the light signal for each

flown nucleotide and correlating the measured signal to dynamically determined thresholds for mono- and multi-mers of each nucleotide.

Our experimental approach involves PCR amplification of target exons. A total of 11 primer pairs were designed to cover exons 18-22 of the EGFR gene, each adapted with 5' overhangs to facilitate emulsion PCR and sequencing. For each patient all PCR products were generated in individual wells, quantified and, to allow parallel processing, pooled in equimolar ratios before emulsion PCR to allow parallel analysis. Emulsion PCR and picotiter plate sequencing were performed as recently described[1]. In brief, to favor single template amplification during emulsion PCR, annealing was performed with an average of 0.5 DNA molecules per bead and 450,000 beads per reaction. After amplification by emulsion PCR, DNA-carrying beads were isolated. For most samples, 8,000-12,000 sequencing reads were generated across the eleven analyzed amplicons (~1,000 reads per amplicon per sample). For patient 12.3 tumor DNA additional confirmatory sequencing was performed on select exons yielding from ~40,000 to ~150,000 sequencing reads per template.

NEOGENOMICS EX.1105.014

Primers used for target region amplification:

EGFR_e18_360F **5'-GCCTCCCTCGCGCCATCAG**GACCCTTGTCTCTGTGTTCTTG-3'

EGFR_e18_360R **5'-GCCTTGCCAGCCCGCTCAG**CCTCAAGAGAGCTTGGTTGG-3'

EGFR_e18_399F **5'-GCCTCCCTCGCGCCATCAG**AGCCTCTTACACCCAGTGGA-3'

EGFR_e18_399R **5'-GCCTTGCCAGCCCGCTCAG**CCTTATACACCGTGCCGAAC-3'

EGFR_e18_460F **5'-GCCTCCCTCGCGCCATCAG**TGAATTCAAAAAGATCAAAGTGC-3'

EGFR_e18_460R **5'-GCCTTGCCAGCCCGCTCAG**CCCCACCAGACCATGAGA-3'

EGFR_e19_110F **5'-GCCTCCCTCGCGCCATCAG**TCACAATTGCCAGTTAACGTCT-3'

EGFR_e19_110R **5'-GCCTTGCCAGCCCGCTCAG**GATTTCCTTGTTGGCTTTCG-3'

EGFR_e19_156F **5'-GCCTCCCTCGCGCCATCAG**TCTGGATCCCAGAAGGTGAG-3'

EGFR_e19_156R **5'-GCCTTGCCAGCCCGCTCAG**GAGAAAAGGTGGGCCTGAG-3'

EGFR_e20_122F **5'-GCCTCCCTCGCGCCATCAG**CCACACTGACGTGCCTCTC-3'

EGFR_e20_122Rb **5'-GCCTTGCCAGCCCGCTCAG**ATGAGCTGCGTGATGAGCT-3'

EGFR_e20_204F **5'-GCCTCCCTCGCGCCATCAG**GCATCTGCCTCACCTCCAC-3'

EGFR_e20_204R **5'-GCCTTGCCAGCCCGCTCAG**GCGATCTGCACACACCAG-3'

EGFR_e20_254F **5'-GCCTCCCTCGCGCCATCAG**GGCTGCCTCCTGGACTATGT-3'

EGFR_e20_254R **5'-GCCTTGCCAGCCCGCTCAG**GGATCCTGGCTCCTTATCTCC-3'

EGFR_e21_289F **5'-GCCTCCCTCGCGCCATCAG**TCTTCCCATGATGATCTGTCCC-3'

EGFR_e21_289R **5'-GCCTTGCCAGCCCGCTCAG**GACATGCTGCGGTGTTTTC-3'

EGFR_e21_381F **5'-GCCTCCCTCGCGCCATCAG**GGCAGCCAGGAACGTACT-3'

EGFR_e21_381R **5'-GCCTTGCCAGCCCGCTCAG**ATGCTGGCTGACCTAAAGC-3'

EGFR_e22_183F **5'-GCCTCCCTCGCGCCATCAG**CACTGCCTCATCTCTCACCA-3'

EGFR_e22_183R **5'-GCCTTGCCAGCCCGCTCAG**CCAGCTTGGCCTCAGTACA-3'

Data analysis

Each individual read was base called as described[1] with an average accuracy of each read of 98% or higher over 100 bases. To verify known mutations and identify new variant alleles, a BLASTN analysis of sequencing reads was carried out using permissive gapping parameters (blastall version 2.2.2 with parameters '–F F', '-G1' and

NEOGENOMICS EX.1105.015

‘–E1’) to encourage contiguous blast hits and facilitate the detection of multi-base pair deletions. The top scoring blast hit for each sequence read was filtered by several quality metrics for inclusion in a multiple alignment with the reference. Qualifying sequence reads were required to be at least 30 bp long, and the width of the top blast match alignment (High-scoring Segment Pair or HSP) also had to be at least 30 bp or more. To discourage short repeat matches, at least 50% of a sequence read needed to be represented in the top HSP match. Furthermore, for the passed reads, the quality of the match was considered such that either the percent identity or the percent accuracy of the hit had to be at least 95%. The qualifying top-hit HSP sequence read segments were used to construct an alignment with respect to the reference. Counts of bases in the alignment deviating from the reference were determined for each reference position. To facilitate visual identification of mutations, graphs (“variation plot analysis”) were produced from the alignment counts data where deviations from the reference sequence were presented as stacked column bars and coverage by position was displayed as a curve.

Statistical considerations of mutation calling

Suppose that we are analyzing a tumor sample in which proportion q of the cells harbor some mutation of interest. The proportion q may be less than 1 due to either stromal contamination or tumor heterogeneity. Therefore, the mutation appears with concentration p in the sample’s DNA, where $p = q/2$ (assuming the mutation is heterozygous). We may think of each read as randomly sampling a fragment from this pool of DNA, with probability p that the fragment harbors the mutation. The rate at which one would recurrently see a false substitution mutation as an artifact of the technology is no more than $e = 0.04\%$. Note that the probability of recurrently erroneously detecting, such as a specific multi-base deletion would be far smaller, so as to be considered

NEOGENOMICS EX.1105.016

negligible. (**Supplementary Fig. 2**). Therefore, the probability of any single read showing a mutation is approximately $p' = p + e\,(1 - p)$. Statistically, we may therefore think of the outcome of each read as a realization of a Bernoulli(p') random variable ("success" being detection of the mutation of interest).

In this framework, consider an experiment with n reads, and let X denote the number of reads that detect the mutation. Under the null hypothesis of no mutation (i.e. solely false positives), X will have a Binomial(n, e) distribution. Owing to our large sample sizes and the small value of e, the variance of X is, in practice, essentially 0. Thus, we can statistically call a valid mutation if the value of X is greater than what one would expect as the "background rate" of recurrently erroneous mutation calls. That is, we may consider a mutation valid when it appears in more than 0.04% of the reads.

Using this rule for calling mutations, we can compute power to find mutations for a range of sample sizes and concentrations. Supplementary Figure 2 summarizes these power calculations.

Practical issues of mutation analysis using picotiter plate sequencing

Practical issues and throughput

The entire sequencing procedure, including initial PCR, emulsion PCR, sample loading onto the picotiter plate and operation of the sequencer, can routinely be performed by one technician. During an average picotiter plate sequencing run (duration app. 5 hours) at least 200,000 high-quality reads can be generated. Considering an average read length of 100 bases, one run produces at least 20 million bases of high quality. A maximum of four runs can be run on one machine within 24 hours totaling a maximum of 80 million bases per day and machine.

NEOGENOMICS EX.1105.017

Cost issues

Taking into account the above considerations and given a minimum number of 200,000 high-quality reads from a single run of one plate, it is conceivable that a number of 20 individual patient samples be analyzed in one run. This approach would allow for analysis of the 20 most commonly affected amplicons in this particular cancer type with a mean coverage of 500X per sample and amplicon. Using this metric, samples with a degree of stromal contamination of >95% (**Supplementary Fig. 2**) could reliably be analyzed for the presence of somatic mutations. With the current price of one sequencing run and a need for a re-confirmation run in the case of a mutation call, a highly accurate molecular diagnosis for each patient (regardless of stromal contamination) would become feasible at a cost comparable to that of magnetic resonance imaging.

Limitations of mutation detection using picotiter plate sequencing

The most important problem connected with this approach lies in its extreme sensitivity. Therefore, any contamination with DNA from another sample, with PCR products or expression constructs carrying cDNAs of the gene that is being sequenced will be represented as a fraction of all alleles in the final analysis. Consequently, DNA should be extracted and handled with extreme caution prior to analysis. The use of a designated pre-PCR area that is spatially separate from the other areas (post-PCR and emulsion PCR areas) is in our view mandatory. Consequently, the only false-positives (n=3) detected in our study were related to cross-contamination with expression constructs carrying EGFR cDNA. We were able to identify these as sample-related false positives because they could be successfully cloned and sequenced and revealed expression construct-specific cDNA sequences (data not shown).

NEOGENOMICS EX.1105.018

An additional potential source of error might be the appearance of mutation artifacts using formalin-fixed, paraffin-embedded samples. As pointed out recently[2], these artifacts can easily be called by any method as artifactual mutations. We therefore suggest running a repeat experiment including PCR amplification. In our hands as well as in those of others[2], this approach reliably yielded high quality mutation calls. Additionally, the risk of paraffin-related artifacts can be minimized by using a high amount of input DNA into the original PCR reaction and by treating the sample with uracil-N-glycosylase[2]. We obtained robust results using >40 ng of input DNA.

Another potential source of error comes from the sequencing process itself, which is prone to single base insertions and deletions, particularly near long homopolymer stretches. Thus, "false" mutations consisting of adjacent insertions and deletions may appear. This is an unlikely event, except in the neighborhood of a long homopolymer stretch (n>7). This problem can be compensated for by increasing coverage in these areas. Since accurate mutation detection in stromally admixed clinical tumor specimens critically depends on high coverage, these stretches might be amenable to picotiter plate sequencing. However, we suggest an increase in coverage for those amplicons where a mutation might be located adjacent to a homopolymer stretch. However, it should be noted that homopolymer stretches of length >7 represent 0.7% of the human genome (human genome, build35) questioning the overall relevance of this obstacle in clinical cancer mutation detection.

Sensitivity of mutation detection using picotiter plate sequencing

To assess the detection sensitivity within the confinement of controlled dilution experiments, either a 99 base-pair HLA-derived amplicon with a single nucleotide polymorphism or a 144 base-pair amplicon covering a 12 base-pair deletion within exon 19 of *EGFR* were diluted with the corresponding wild-type amplicons. In each case the

NEOGENOMICS EX.1105.019

wild-type and mutant amplicons were amplified separately and after quantification used for serial dilutions. Samples were sequenced to a coverage depth of 65,000-110,000 reads and the observed frequency of the respective mutations were plotted against the relative degree of dilution (**Supplementary Fig. 3**). In both cases a strictly linear correlation (R^2=0.99) was observed between measured and expected mutant frequencies down to allele frequencies below 0.5%, with the small deviations observed most likely attributable to experimental inaccuracies in DNA quantification and pipeting during the serial dilution.

Lung cancer samples

DNA samples from 33 patients with lung adenocarcinoma were studied. 22 patients (those studied in the pilot experiments) were from Japan. Patient 12.3 (see above) as well as nine of the ten patients whose paraffin-embedded tumor specimens were studied were Caucasian patients from the US. One of the tumors of the paraffin collection was from a Vietnamese patient. The study was conducted under institutional review board approval. All patients had provided informed consent. DNA was extracted from frozen biopsy specimens (Japanese samples), from freshly obtained pleural effusion fluid and debris (relapse sample of patient 12.3), from paraffin-embedded pleural effusion material obtained at the time of initial diagnosis (pre-treatment sample of patient 12.3), or from paraffin-embedded biopsy specimens (paraffin collection, n=10) following standard procedures. DNA from the Japanese samples was whole-genome amplified as described recently[3]. Candidate mutations were reconfirmed using newly generated PCR products from non-whole-genome amplified DNA. Paraffin embedded specimens from 10 NSCLC patients treated with gefitinib were analyzed. Tumor specimens were obtained prior to treatment with gefitinib. DNA was prepared from whole tumor sections without gross or microdissection as previously described.[4] All patients

NEOGENOMICS EX.1105.020

have previously been analyzed using either by Sanger Sequencing, Surveyor analyses or both and published.[4,5]

Case report

A 47-year old male Caucasian (patient 12.3) presented with a large pleural effusion in the right chest. Drainage, cytological and immunohistochemical examination of the pleural fluid led to diagnosis of lung adenocarcinoma confirmed by cytological evaluation of bronchoscopically obtained bronchial washings. Staging revealed a tumor mass in the right lower lobe of the lung, bilateral nodular opacities and multiple bone metastases. Treatment with erlotinib and bexarotene as part of a clinical trial resulted in significant symptomatic improvement and an objective partial response as determined by a greater than 50% reduction of the tumor burden. The patient remained in good clinical condition for one year until he relapsed with a recurrent pleural effusion. He was treated initially with higher dose erlotinib, followed by vinorelbine and bevacizumab. Response was not achieved and for progression of disease treatment was subsequently switched to carboplatin, docetaxel and bevacizumab. Despite aggressive treatment, patient 12.3 deceased one month later.

Functional experiments

The EGFR cDNA was introduced into pBABE-puro retroviral vectors using the Gateway system (Invitrogen). The mutations Del-4 and T790M were introduced either alone or together into the EGFR construct using site-directed mutagenesis (Quikchange II XL, Stratagene) and verified by sequencing as described recently[6]. Likewise, the P772_H773insV mutation of EGFR was made in the same manner. Recombinant retrovirus was obtained by transfection of the construct into a packaging cell line. NIH3T3 cells were transduced with the mutant or wt retroviruses and selected in

NEOGENOMICS EX.1105.021

puromycin-containing medium. After selection, stable pooled cells were suspended in soft agar to assess their ability to grow in an anchorage-independent fashion[6] (**Supplementary Fig. 5a**).

IL-3 dependent BA/F3 cells, derived from murine pro-B cells, become independent from IL-3 when transformed by activated oncogenes, such as the BCR-ABL tyrosine kinase[7]. Ba/F3 cells were infected with the retroviruses and selected in puromycin-containing medium. After outgrowth of polyclonal stable cell lines IL-3 was withdrawn from the medium to select for transformed cells. Del-4 alone rendered Ba/F3 cells IL-3 independent, an observation that is widely accepted as a sign of transformation. Cells were plated in 96-well plates and cultured overnight. Cells were then incubated with various doses of erlotinib (WuXi PharmaTech) at the following concentrations: 0, 0.0033, 0.01, 0.033, 0.1, 0.33, 1, 3.3, 10 µM. Impact on cell viability was determined using the MTS assay following the recommendations of the manufacturer (Promega). Incubation of Del-4-transformed Ba/F3 cells led to a strong reduction of cell viability with an IC_{50} (concentration at which 50% of the cells are being killed) of less than 3.3 nM (**Supplementary Fig. 5b**). This value resembles those obtained with cells carrying transgenes for other EGFR KD mutations, such as G719S or L858R (data not shown). In a subsequent experiment Ba/F3 cells were infected with retrovirus carrying both Del-4 and T790M. Cells were selected as before and impact of erlotinib on cell survival was measured using the MTS assay (**Supplementary Fig. 5b**). Incubation of the double-mutant cells with erlotinib showed a complete resistance to this treatment even when concentrations of up to 10 µM were used.

PCR cloning and sequencing

Del-4 and T790M from patient 12.3 were validated by cloning of PCR products covering whole exons 19 and 20 using primers and PCR conditions described recently[5].

NEOGENOMICS EX.1105.022

PCR products were ligated into PCR2.1-Topo vectors (Invitrogen) and transformed into E.coli. After transformation, bacteria were plated onto selection plates and grown overnight. At least two 384-well plates of colonies were isolated per mutation using a colony picking robot (QPix2, Genetix Limited). Colonies were grown overnight and bidirectionally sequenced using the production sequencing facility at the Broad Institute. Sequence traces were analyzed using Mutation Surveyor software (SoftGenetic Inc.). Del-4 could be detected in approximately 3% of the exon 19-derived clones analyzed (**Supplementary Fig. 6**). T790M was present in approximately 2% of the exon 20-derived clones (**Supplementary Fig. 6**). These frequencies match the ones detected in the sample using picotiter plate pyrosequencing.

References

1. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80 (2005).
2. Marchetti, A., Felicioni, L. & Buttitta, F. Assessing EGFR mutations. *N Engl J Med* **354**, 526-8; author reply 526-8 (2006).
3. Paez, J.G. et al. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* **32**, e71 (2004).
4. Janne, P.A. et al. A rapid and sensitive enzymatic method for epidermal growth factor receptor mutation screening. *Clin Cancer Res* **12**, 751-8 (2006).
5. Paez, J.G. et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497-500 (2004).
6. Greulich, H. et al. Oncogenic transformation by inhibitor-sensitive and resistant EGFR mutants. *PLoS Med* **2**, in press (2005).
7. Azam, M., Latek, R.R. & Daley, G.Q. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **112**, 831-43 (2003).

NEOGENOMICS EX.1105.023