Virchows Arch (2013) 462:409–419
DOI 10.1007/s00428-013-1376-6

ORIGINAL ARTICLE

Increased detection rates of *EGFR* and *KRAS* mutations in NSCLC specimens with low tumour cell content by 454 deep sequencing

Evgeny A. Moskalev · Robert Stöhr · Ralf Rieker · Simone Hebele · Florian Fuchs · Horia Sirbu · Sergey E. Mastitsky · Carsten Boltze · Helmut König · Abbas Agaimy · Arndt Hartmann · Florian Haller

Received: 22 October 2012 / Revised: 16 December 2012 / Accepted: 23 January 2013 / Published online: 7 March 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Detection of activating *EGFR* mutations in NSCLC is the prerequisite for individualised therapy with receptor tyrosine kinase inhibitors (TKI). In contrast, mutant downstream effector *KRAS* is associated with TKI resistance. Accordingly, *EGFR* mutation status is routinely examined in NSCLC specimens, but the employed methods may have a dramatic impact on the interpretation of results and, consequently, therapeutic decisions. Specimens with low tumour cell content are at particular risk for false-negative *EGFR* mutation reporting by sequencing with Sanger chemistry. To improve reliability of detecting clinically relevant mutant variants of *EGFR* and *KRAS*, we took full advantage of 454 deep sequencing and developed a two-step amplification protocol for the analysis of *EGFR* exons 18–21 and *KRAS* exons 2 and 3. We systematically addressed the sensitivity, reproducibility and specificity of the developed assay. Mutations could be reliably identified down to an allele frequency of 0.2–1.5 %, as opposed to 10–20 % detection limit of Sanger sequencing. High reproducibility (0–2.1 % variant frequency) and very low background level (0.4–0.8 % frequency) further complement the reliability of this assay. Notably, re-evaluation of 16 NSCLC samples with low tumour cell content ≤40 % and *EGFR* wild type status according to Sanger sequencing revealed clinically relevant *EGFR* mutations at allele frequencies of 0.9–10 % in seven cases. In summary, this novel two-step amplification protocol with 454 deep sequencing is superior to Sanger sequencing with significantly increased sensitivity, enabling reliable analysis of *EGFR* and *KRAS* in NSCLC samples independent of the tumour cell content.

Keywords NSCLC · EGFR · KRAS · 454 Deep sequencing · Sensitivity

Electronic supplementary material The online version of this article (doi:10.1007/s00428-013-1376-6) contains supplementary material, which is available to authorized users.

E. A. Moskalev · R. Stöhr · R. Rieker · S. Hebele · A. Agaimy · A. Hartmann · F. Haller
Department of Pathology, University Medical Center, Erlangen, Germany

F. Fuchs
Department of Gastroenterology, Pneumology and Endocrinology, University Medical Center, Erlangen, Germany

H. Sirbu
Department of Thoracic Surgery, University Medical Center, Erlangen, Germany

S. E. Mastitsky
Division of Theoretical Bioinformatics, German Cancer Research Center, Heidelberg, Germany

C. Boltze
Institute of Pathology, SRH-Waldklinikum, Gera, Germany

H. König
Practice for Pathology, Ingolstadt, Germany

F. Haller (✉)
Diagnostic Molecular Pathology, Institute of Pathology, University Medical Center Erlangen, Krankenhausstr. 8-10, 91054, Erlangen, Germany
e-mail: florian.haller@uk-erlangen.de

Introduction

The observation of activating mutations in the tyrosine kinase domain of the epidermal growth factor receptor (*EGFR*) gene in ~10 % of all non-small cell lung cancers (NSCLC) has opened the possibility of targeted therapy with

receptor tyrosine kinase inhibitors (TKIs) directed against mutant *EGFR* [1–3]. Since clinical phase III trials have demonstrated the benefit of TKI application for patients whose tumours harbour activating *EGFR* mutations [4, 5], mutation analysis of *EGFR* is suggested to be routinely performed in NSCLC specimens [6]. In contrast, activating mutations in the v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*) gene, a member of the Ras family of small GTPases, were present in 27 % of NSCLCs (all adenocarcinomas) in a recently published sequencing study, and these two mutations occur mutually exclusive [7]. Importantly, mutant *KRAS* is located downstream in the signalling cascade of *EGFR* and consequently associated with resistance to TKI therapy [8]. Therefore, mutation analysis of both *EGFR* and *KRAS* is vital for individualised therapeutic decisions.

Several issues exist, however, which hamper employment of *EGFR* mutation detection as a reliable diagnostic tool. First, significant discrepancy of *EGFR* mutation frequencies (6.8–25.9 %) and, hence, reporting of *EGFR* mutation status has been revealed in a recent inter-laboratory comparison in routinely processed NSCLC samples [9]. This raises the question of methodical problems in this therapeutically relevant testing. Second, in patients with extensive disease (stage IV), only small biopsies or cytological specimens are usually available with limited amount of tumour cells. This may represent a serious obstacle for mutation detection by routinely used sequencing with Sanger chemistry. For example, a recent study has demonstrated that ~30 % of NSCLC specimens in a large clinical cohort contained less than 40 % tumour cells, the minimal threshold needed for a reliable detection of *EGFR* mutations using Sanger sequencing [10]. Other techniques frequently used for detection of *EGFR* mutations are based on real-time polymerase chain reaction (PCR) or pyrosequencing methods, with several commercially available kits. However, while these methods have a better sensitivity of ~1–5 % compared to Sanger sequencing, they will not identify 5–10 % of the currently known *EGFR* mutations according to their targeted approach. Collectively, these obstacles underscore the need for alternative analytical principles that achieve more accurate diagnostic results.

Next-generation sequencing techniques allow massively parallel, or deep, sequencing of target regions with >1,000 reads per sample, thereby enabling detection of mutations at much lower allele frequencies compared to Sanger sequencing. For example, 100 % of mutations were detected in clinical responders to TKI therapy by 454 massively parallel sequencing in a comparative study on 18 *EGFR*-mutated NSCLC specimens, compared to only 89 % and 67 % detection rates of mutations by pyrosequencing and Sanger sequencing, respectively [11]. Despite an obviously lower detection limit, no systematic analysis of the sensitivity, reproducibility and specificity of 454 deep sequencing regarding *EGFR* and *KRAS* mutations analysis has been reported as yet. The unique possibility of detection of clinically relevant mutations at very low allele frequencies in the range of 1–10 % is associated with the risk of considering technical errors, which are introduced by DNA polymerase during amplicon library preparation or through base-calling process as low-frequency variants [12]. Therefore, a reliable threshold for background variants is desirable for discrimination of noise and low-frequency variants.

Given the fact that clinical samples are almost exclusively available as formalin-fixed and paraffin-embedded (FFPE) tissue specimens with often low-quality DNA, a special procedure for amplicon library preparation is needed to maximize the number of informative patient specimens [13]. Since complex PCR primers are commonly used for amplicon library preparation, which include 5′-overhangs of adapter sequences for binding to the DNA capture beads and barcode sequences for identity of different patient samples, the percentage of efficiently amplified DNA samples may be even lower. In the current study, we established a two-step DNA amplification protocol with subsequent 454 deep sequencing of *EGFR* and *KRAS* genes, which is capable of successful amplification of FFPE NSCLC samples with low DNA quality. We systematically evaluated its sensitivity, reproducibility and specificity and provided reliable thresholds for the lower detection limit of mutations (sensitivity), variation of allele frequencies (reproducibility) and background levels (specificity). We next applied this assay to re-evaluate clinical NSCLC samples with low tumour cell content (≤40 %) that were *EGFR* wild type according to conventional Sanger sequencing and identified *EGFR* mutations in a significant proportion of these cases. In summary, this study demonstrates the much higher sensitivity of the developed 454 deep sequencing assay compared to Sanger sequencing and strongly argues for its wide application in routine molecular diagnostic analysis of clinical FFPE NSCLC samples with low tumour content.

Materials and methods

Patient samples and cell culture

A total of 21 FFPE specimens were obtained from the Institutes of Pathology in Erlangen, Gera and Ingolstadt (Germany). The samples included cell block preparations from cytological specimens (pleural effusions, n=4; fine needle aspirations, n=3), small endoscopic biopsies (n=11) and resection specimens (n=3). Tumour specimens were inspected by a pathologist to estimate the tumour cell content and the histomorphological pattern.

The lung adenocarcinoma cell lines NCI-H1650 (*EGFR* exon 19 deletion E746_A750del), NCI-H1975 (*EGFR* exon

Table 1 Sequences of the PCR primers for two-step amplification protocol and 454 deep sequencing

Primer ID	Primer sequences (5′-3′)	Amplicon length, bp[a]
Sequences of the template specific 3′-portion of the fusion PCR primers[b]		
EGFR-ex18-F	TGGAGCCTCTTACACCCAGT	179
EGFR-ex18-R	CCCCACCAGACCATGAGA	
EGFR-ex19-F	CATGTGGCACCATCTCACA	179
EGFR-ex19-R	CCACACAGCAAAGCAGAAAC	
EGFR-ex20-F	CTCCAGGAAGCCTACGTGAT	180
EGFR-ex20-R	CACACCAGTTGAGCAGGTACT	
EGFR-ex21-F	CCTCACAGCAGGGTCTTCTC	182
EGFR-ex21-R	TGCCTCCTTCTGCATGGTAT	
KRAS-ex2-F	AAGGCCTGCTGAAAATGACT	170
KRAS-ex2-R	AGAATGGTCCTGCACCAGTAA	
KRAS-ex3-F	AAAGGTGCACTGTAATAATCCAGAC	171
KRAS-ex3-R	AAAGAAAGCCCTCCCCAGT	
Sequences of the outer PCR primers used for two-step library preparation		
EGFR-ex18-pre-amp-F	GCTGAGGTGACCCTTGTCTC	246
EGFR-ex18-pre-amp-R	ACAGCTTGCAAGGACTCTGG	
EGFR-ex19-pre-amp-F	GCTGGTAACATCCACCCAGA	247
EGFR-ex19-pre-amp-R	GAGAAAAGGTGGGCCTGAG	
EGFR-ex20-pre-amp-F	CACACTGACGTGCCTCTCC	250
EGFR-ex20-pre-amp-R	TATCTCCCCTCCCCGTATCT	
EGFR-ex21-pre-amp-F	GCAGAGCTTCTTCCCATGAT	247
EGFR-ex21-pre-amp-R	GGAAAATGCTGGCTGACCTA	
KRAS-ex2-pre-amp-F	TTAACCTTATGTGTGACATGTTCTAA	262
KRAS-ex2-pre-amp-R	TCATGAAAATGGTCAGAGAAACC	
KRAS-ex3-pre-amp-F	TCAAGTCCTTTGCCCATTTT	253
KRAS-ex3-pre-amp-R	TGGCAAATACACAAAGAAAGC	

F PCR forward, *R* PCR reverse

[a] Template-specific portion of each amplicon is shown. Given additional auxiliary 35-mer sequences at both ends, real amplicon size is 70-bp longer

[b] The complete PCR primer sequence includes an additional 35-mer portion at 5′-end. It serves for binding to the DNA capture beads and annealing the emPCR amplification primers and the sequencing primer. Additionally, it contains the sequencing tetranucleotide key and 10-mer multiple identifier (MID) for demultiplexing reads. A total of 12 MIDs (MID1–12) were used from the 454 Standard MID set

20 missense mutation T790M; EGFR exon 21 missense mutation L858R), NCI-H460 (*KRAS* exon 3 missense mutation Q61H), NCI-H1299 (*EGFR* and *KRAS* wildtype status) as well as the colorectal adenocarcinoma cell line HCT-116 (*KRAS* exon 2 missense mutation G13D) were purchased from the American Type Culture Collection (ATCC, USA). All the cells were grown in a medium consisting of 90 % Roswell Park Memorial Institute Medium 1640 supplemented with 300 mg/L L-glutamine (Invitrogen, Carlsbad, USA) and 10 % foetal bovine serum (Invitrogen).

DNA isolation and amplicon library preparation

DNA was extracted from the FFPE samples using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) and from cell lines by DNeasy Blood & Tissue Kit (Qiagen) as suggested by the manufacturer. The concentration of DNA from the FFPE samples was measured in a ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA). Quantification of DNA from cell lines for preparation of a dilution series of mutated DNA in wild-type DNA was performed by using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen) and Synergy 2 Multi-Mode Microplate Reader (Biotek, Winooski, USA). These control DNA samples represented distinct percentages of mutant variants: 50 %, 10 %, 7.5 %, 5 %, 2.5 %, 1 % and 0.5 %.

A two-step amplification protocol (nested PCR) included a pre-amplification step with outer PCR primers followed by re-amplification of diluted amplicons by using fusion primers with inner template specific sequences (Table 1). Pre-amplification was carried out in 25-μL reactions that contained 50 ng of cell line DNA or a variable quantity (50–250 ng) of DNA from FFPE samples, 1.5 mM $MgCl_2$, 200 mM dNTP, 500 nM primers and 1 unit Phusion Hot Start Flex DNA polymerase (New England Biolabs, Ipswich, USA). An amplification programme was started by an initial activation of the enzyme at 98 °C for 30 s. The initial amplification cycle was denaturation at 98 °C for 10 s, annealing at 72 °C for 30 s and elongation at 72 °C for 30 s. Amplification was continued for ten cycles, reducing the annealing temperature by 1 °C each cycle, followed by 40 cycles of 10 s denaturation at 98 °C, 30 s annealing at 62 °C and 30 s elongation at 72 °C. PCR products were diluted $1:10^6$ or $1:10^3$ for *EGFR* or *KRAS* amplicons, respectively, and

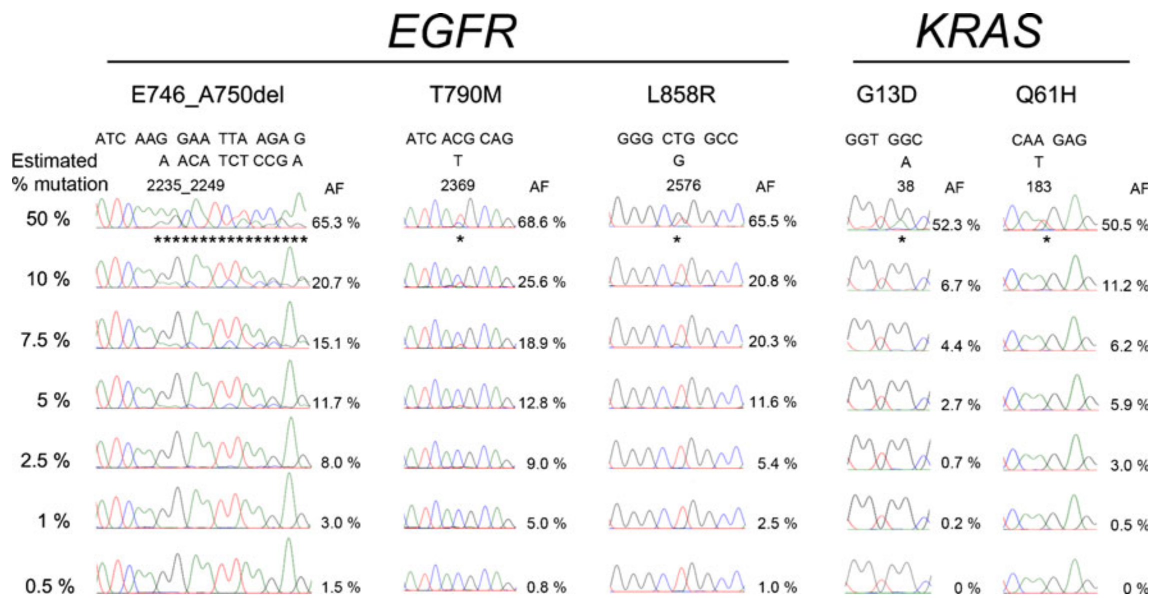

Fig. 1 Comparative analysis of sensitivity of 454 deep sequencing and Sanger sequencing by using dilution series of mutant *EGFR* and *KRAS* variants in wild-type DNA from cell lines. Analysis of mutation frequencies in artificial dilutions of DNA from cell lines with known mutations (estimated mutation frequency, leftmost percentages) in *EGFR* (E746_A750del, T790M and L858R) and *KRAS* (G13D and Q61H) by Sanger sequencing (chromatograms) and 454 sequencing (allele frequency (*AF*)). Reference sequence and positions of corresponding mutations are shown, with *asterisks* indicating the respective changes in the chromatograms. While mutations were reliably detected by 454 deep sequencing in all dilutions, these were no more detectable by Sanger sequencing at allele frequencies <10–20 %. Notably, NSCLC cell lines with *EGFR* mutations (E746_A750del, T790M and L858R) displayed a low level amplification of the mutated allele, with observed allele frequencies of ~65 % in the undiluted DNA

Table 2 Characteristics of two-step amplification protocol and 454 deep sequencing assay for detection of *EGFR* and *KRAS* mutations

Sensitivity and linearity of PCR amplification						
Dilution	E746_A750del	T790M	L858R	G13D	Q61H	
50 %	65.3 %	68.6 %	65.5 %	52.3 %	50.5 %	
10 %	20.7 %	25.6 %	20.8 %	6.7 %	11.2 %	
7.5 %	15.1 %	18.9 %	20.3 %	4.4 %	6.2 %	
5 %	11.7 %	12.8 %	11.6 %	2.7 %	5.9 %	
2.5 %	8.0 %	9.0 %	5.4 %	0.7 %	3.0 %	
1 %	3.0 %	5.0 %	2.5 %	0.2 %	0.5 %	
0.5 %	1.5 %	0.8 %	1.0 %	0 %	0 %	
R^2	0.995	0.990	0.994	0.998	0.996	
Reproducibility						
	E746_A750del	T790M	L858R	G13D	Q61H	
Mean AF	66.0 %	70.9 %	64.4 %	51.1 %	100.0 %	
Std. dev.	1.5 %	1.6 %	1.0 %	2.1 %	0.0 %	
Specificity						
	Exon 18	Exon 19	Exon 20	Exon 21	Exon 2	Exon 3
With homopolymer stretches						
99.9 %	1.1 %	2.4 %	2.4 %	2.5 %	0.8 %	0.6 %
After exclusion of homopolymer stretches						
99.9 %	0.6 %	0.8 %	0.6 %	0.7 %	0.7 %	0.4 %

AF allele frequency, *Std. dev.* standard deviation

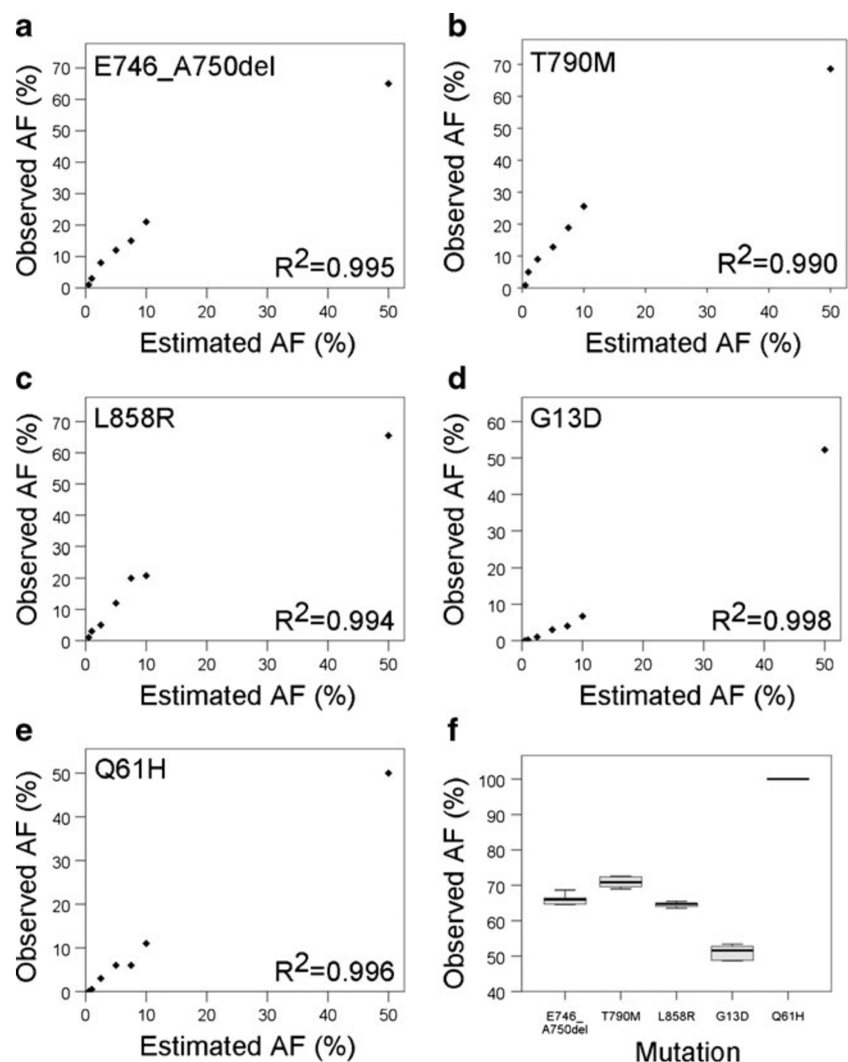

Fig. 2 Demonstration of linear PCR amplification and evaluation of reproducibility of *EGFR* and *KRAS* mutation detection using a novel two-step amplification and 454 deep sequencing. **a**–**e** Demonstration of linear PCR amplification in artificial dilutions of DNA from cell lines with known mutations in *EGFR* (**a** E746_A750del, **b** T790M, **c** L858R) and *KRAS* (**d** G13D, **e** Q61H) by 454 sequencing (allele frequency (*AF*)). **f** Allele frequencies of known *EGFR* and *KRAS* mutant variants (*x*-axis) analysed from cell line DNA in sixfold replicates are plotted as *boxplots*, with high reproducibility of mutation detection independent of mutation type or its localisation. Note the low-level amplification of mutated alleles in lung cancer cell lines with *EGFR* mutations (E746_A750del, T790M, L858R) as well as homozygous status of another lung cancer cell line with *KRAS* mutation (Q61H)

employed for re-amplification reaction. It was started by an initial activation of the enzyme at 98 °C for 30 s. Each amplification cycle included denaturation at 98 °C for 20 s, annealing and elongation at 72 °C for 40 s. This procedure was continued for 40 cycles. Negative control PCR reactions supplemented with equal amount of water instead of DNA were included for each amplicon on the same PCR plate. About 10 μL of each reaction was examined on 3 % agarose gels. Amplicons were purified by using Agencourt AMPure XP kit (Beckman Coulter, Beverly, USA), quantified by fluorometry in triplicates using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen) and Synergy 2 Multi-Mode Microplate Reader (Biotek) as directed by the manufacturers. Finally, the library was pooled in equimolar ratios, and the concentration was adjusted to 10^7 molecules/microlitre.

Analysis by deep sequencing

Deep sequencing was performed using the GS Junior Titanium chemistry according to the standard protocols of Roche (Basel, Switzerland). A total of ~500,000 beads were loaded on the picotiter plate yielding a total of 101,109 high-quality reads per run on average and average coverage of 1,451 reads/amplicon. All reads were processed, aligned to the human reference sequences of *EGFR* and *KRAS* and analysed for mutation frequencies by using the Amplicon Variant Analyser software v. 2.5 from Roche.

Sanger sequencing

EGFR and *KRAS* PCR products for direct sequencing with Sanger chemistry were amplified in 50 μL reactions that contained 50 ng cell line DNA or about 100–250 ng DNA isolated from FFPE tissue, 1.5 mM $MgCl_2$, 200 mM dNTP, 500 nM primers and 2.5 units Taq Polymerase S (Genaxxon BioScience GmbH, Ulm, Germany). Initial denaturation step was 94 °C for 3 min. Each amplification cycle included denaturation at 94 °C for 30 s, annealing at 70 °C for 30 s and elongation at 72 °C for 40 s. This process was continued for ten cycles, reducing the annealing temperature by 1 °C

each cycle, followed by 30 cycles of 30 s denaturation at 94 °C, 30 s annealing at 60 °C and 40 s elongation at 72 °C. The amplicons were purified with the QIAquick PCR Purification kit (Qiagen) and sequenced bidirectionally at an external facility (Seqlab–Sequence Laboratories Göttingen GmbH, Göttingen, Germany). The sequencing data were visualised using the FinchTV 1.4.0 software (Geospiza, Inc., Seattle, USA; http://www.geospiza.com). The sequences of the PCR and sequencing primers employed are listed in Supplementary Table S1.

Results

Novel 454 sequencing assay exhibits high sensitivity, reproducibility and specificity of *EGFR* and *KRAS* mutation detection

Sensitivity

To determine the sensitivity of our novel two-step amplification protocol and to confirm unbiased (linear) PCR amplification of mutant alleles, a dilution series of cell line DNAs with known mutant variants in wild-type DNA was employed, with defined mutant allele frequencies of 50 %, 10 %, 7.5 %, 5 %, 2.5 %, 1 % and 0.5 %. DNA aliquots were individually amplified and analysed by 454 deep sequencing and Sanger sequencing (Fig. 1). While expected *EGFR* and *KRAS* mutations were reliably detected by our assay down to 0.5 % and 1 % estimated mutation frequency, respectively, they were no more detectable by Sanger chemistry at allele frequencies of ≤10–20 % (Fig. 1). Notably, low-level amplification of the three mutant *EGFR* alleles was reproducibly detected in undiluted DNA from both cell lines NCI-H1650 (E746_A750del) and NCI-H1975 (T790M and L858R), with allele frequencies of 65.3 %, 68.6 % and 65.5 %, respectively, as opposed to an expected value of 50 % for heterozygous variants (Table 2). Accordingly, observed allele frequencies in the 0.5 % dilution step were similarly increased, with 1.5 %, 0.8 % and 1.0 %. Regression analysis of allele frequencies determined by 454 sequencing revealed linear amplification by our novel amplification protocol, with regression coefficients $R^2>0.99$ (Fig. 2a–e; Table 2).

Reproducibility

To ascertain reproducibility of mutation quantification by our deep sequencing assay, we performed sequencing analysis of each of the mutated cell lines in six analytical replicates (individual PCR amplification reactions). For

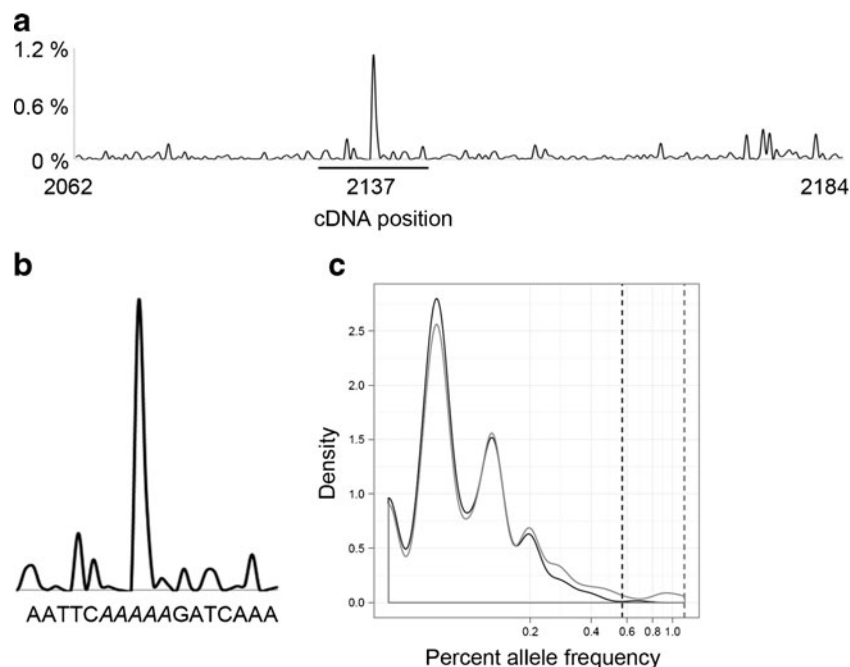

Fig. 3 Evaluation of specificity and background levels of *EGFR* and *KRAS* mutation detection using a novel two-step amplification and 454 deep sequencing. **a** Background level of unspecific variants for *EGFR* exon 18 as determined by 454 sequencing of sixfold DNA replicates from *EGFR* wild-type cell line that were separately amplified by PCR. **b** Note the higher allele frequency of background variants due to base-calling errors of 454 pyrosequencing chemistry at positions c.2137_2141 with a homopolymer stretch of five adenosines. **c** Probability density estimation of the background variants was conducted using the Gaussian algorithm as implemented in the function density in the R statistical software (R Development Core Team 2012). The 99.9 % specificity values were calculated using the quantile function of R for all variants before (*grey line*) and after (*black line*) exclusion of homopolymer stretches, with background allele frequencies of 1.1 % and 0.6 %, respectively

different mutations, mean allele frequencies of 51.1–100 % with standard deviations of 0–2.1 % were observed (Table 2), with no significant difference regarding mutation type (deletion vs. missense mutation) or localisation along the gene (e.g. *EGFR* exon 20 vs. exon 21; Fig. 2f).

Specificity

The unique possibility to detect very low allele frequencies of *EGFR* and *KRAS* mutations raises the question of considering technical errors (background) as low-frequency variants. To determine the background level or "noise", represented by the occurrence of unexpected variants (single nucleotide changes), we deep-sequenced cell line DNA for each amplicon in six analytical replicates (individual PCR amplification reactions), and postulated that variants which deviate from the reference sequence and occur at very low allele frequencies are more likely to be artefacts—e.g. DNA polymerase or base-calling errors—than true biological variants. Indeed, we observed background variants at low frequencies which were unequally distributed among each amplicon (Fig. 3a). There were "hotspots" with clearly higher frequency of artefacts, and these were located at stretches of homopolymers (e.g. single base insertions at *EGFR* exon 18 cDNA positions c.2137_2141 corresponding to a homopolymer stretch of five adenosines; Fig. 3b). A total of 99.9 % of all background variants (99.9 % specificity) were present at an allele frequency below 0.8–2.5 %, while this threshold could be reduced down to 0.4–0.8 % after exclusion of variants associated with homopolymer stretches (Fig. 3c; Table 2).

Higher sensitivity of *EGFR* mutation detection in clinical NSCLC samples using the novel two-step amplification protocol and 454 deep sequencing compared to Sanger sequencing

We next applied our novel two-step amplification protocol to analyse *EGFR* and *KRAS* mutations in 21 NSCLC samples available as FFPE tissue. Five NSCLC specimens with high tumour cell content >40 % were included that harboured different *EGFR* mutations previously detected by Sanger sequencing (cases 8–12), while the other 16 specimens with a tumour cell content of ≤40 % were wild type for *EGFR* according to previous Sanger sequencing (Table 3). The samples included cell block preparations derived from pleural effusions (n=4) or fine needle aspirations (n=3), small endoscopic bronchial biopsies and core needle biopsies (n=11) and resection specimens (n=3). The DNA of all samples was amplified using our two-step amplification protocol and analysed for *EGFR* and *KRAS* mutations by 454 deep sequencing.

Remarkably, nine unequivocal *EGFR* mutations with clinical implications were observed in seven cases with tumour cell contents of 5–30 %, with allele frequencies ranging from 0.9–10 % (Table 3; cases 1–7). The mutations included six missense mutations and three deletions at typical positions, with two cases harbouring two mutations each. The critical re-evaluation of the initial Sanger sequencing results confirmed that these mutations were not reliably detectable in the respective chromatograms (Fig. 4). Five other samples harboured *KRAS* mutations at allele frequencies of 1.4–30.5 % (cases 13–17) while the remaining four samples with tumour cell contents of 30–35 % were of wild type status for *EGFR* and *KRAS* (cases 18–21). Notably, eight of 12 samples with *EGFR* mutations displayed a micropapillary, papillary or lepidic growth pattern or a combination thereof (Fig. 4) while four of the five samples with *KRAS* mutations were of acinar growth pattern. The comparison of tumour cell content vs. observed allele frequency revealed a significant correlation (R^2= 0.898; Fig. 4l).

Discussion

Systematic evaluation of sensitivity, specificity and reproducibility of *EGFR* and *KRAS* mutation detection by the developed 454 deep sequencing assay

In the current study, we established a novel deep sequencing assay for detection of clinically relevant *EGFR* and *KRAS* mutant variants, which is especially feasible for poor quality DNA from FFPE tissue. A key feature of the assay—a two-step PCR-amplification protocol for amplicon library preparation—enables a substantial increase in the number of FFPE specimens which can be efficiently processed for detection of *EGFR* and *KRAS* mutations. Systematic analysis of its performance was evaluated using artificial dilutions of mutant *EGFR* and *KRAS* alleles in bulk of wild-type cell line DNA. We confirmed linear amplification of mutant alleles with regression coefficients >0.99. The assay was able to detect as little as 0.2–1.5 % mutation, an enormous increase in sensitivity compared to Sanger sequencing with a lower detection limit of 10–20 %. To identify the limit of reliable detection, specificity of the assay was studied by considering unexpected variants when sequencing wild-type DNA as technical errors (background). The background is an integral measure of errors that are introduced by DNA polymerase and base-calling process. Dependent on the different amplicons, 99.9 % of background variants occurred with a frequency below 0.6–2.4 %. Notably, errors were unequally distributed along the sequence, with "hotspots" associated with homopolymer stretches, the known issue of pyrosequencing chemistry [11]. Exclusion of homopolymer-associated errors resulted in 99.9 % specificity of 0.4–0.8 %. Thus, we conclude that our assay can

Table 3 Patient characteristics and mutation analysis findings

Case/specimen	Age/gender/smoking history	Histology/mucin production	Grade/TTF1	Tumor cell content	*EGFR* Sanger	cDNA	Protein	Allele frequency
EGFR mutation								
1/fna	39/m/u	Micropapillary/–	G3/+	5 %	Wild type	c.2318A > T	p.H773L	0.9 %
2/pe	70/f/u	Micropapillary/–	G2/+	5 %	Wild type	c.2235_2249del15	p.E746_A750del	1.6 %
3/fna	73/f/u	Micropapillary/–	G3/+	10 %	Wild type	c.2240_2254del15; c.2392C > G	p.L747_T751del; p.L798V	4.3 %
4/b	74/m/u	Micropapillary/–	G1/+	10 %	Wild type	c.2527G > A	p.V843I	8.1 %
5/b	66/m/–	Solid/–	G3/+	15 %	Wild type	c.2575G > A	p.A859T	2.5 %
6/b	63/m/+	Solid/–	G3/+	15 %	Wild type	c.2237_2255_del19insT	p.E746_S752delinsV	10.0 %
7/b	55/f/u	Acinar/+	G2/+	30 %	Wild type	c.2281G > A; c.2369C > T	p.D761N; p.T790M	4.0 %
8/r	74/m/+	Solid/–	G2/+	40 %	Mutated	c.2237_2255_del19insT	p.E746_S752delinsV	31.2 %
9/b	48/m/+	Papillary/–	G2/+	40 %	Mutated	c.2573T > G	p.L858R	52.6 %
10/fna	81/f/u	Micropapillary/–	G2/+	60 %	Mutated	c.2155G > A; c.2126A > C	p.E709A; p.G719S	39.5 %
11/r	64/f/u	Lepidic/–	G1/+	70 %	Mutated	c.2235_2249del15	p.E746_A750del	43.3 %
12/pe	73/m/+	Micropapillary/–	G1/+	70 %	Mutated	c.2582T > A	p.L861Q	90.6 %
KRAS mutation								
13/pe	75/m/u	Acinar/–	G1/+	5 %	Wild type	c.169G > A	p.D57N	1.4 %
14/b	61/m/u	Solid/–	G2/+	35 %	Wild type	c.35G > A	p.G12D	17.6 %
15/b	78/f/u	Acinar/+	G3/+	35 %	Wild type	c.35G >T	p.G12V	28.0 %
16/b	80/f/+	Acinar/+	G2/+	40 %	Wild type	c.38G > A	p.G13D	29.9 %
17/b	81/f/u	Acinar/+	G2/+	40 %	Wild type	c.182A > G	p.Q61R	30.5 %
EGFR and *KRAS* wild type								
18/r	67/f/–	Acinar/–	G2/+	30 %	Wild type	–	–	–
19/pe	97/f/u	Micropapillary/–	G3/+	35 %	Wild type	–	–	–
20/b	72/m/+	Acinar/–	G3/+	35 %	Wild type	–	–	–
21/b	51/f/+	Solid/–	G3/+	35 %	Wild type	–	–	–

fna fine needle aspiration, *pe* pleural effusion, *b* biopsy, *r* resection, *u.* unknown, *m* male, *f* female

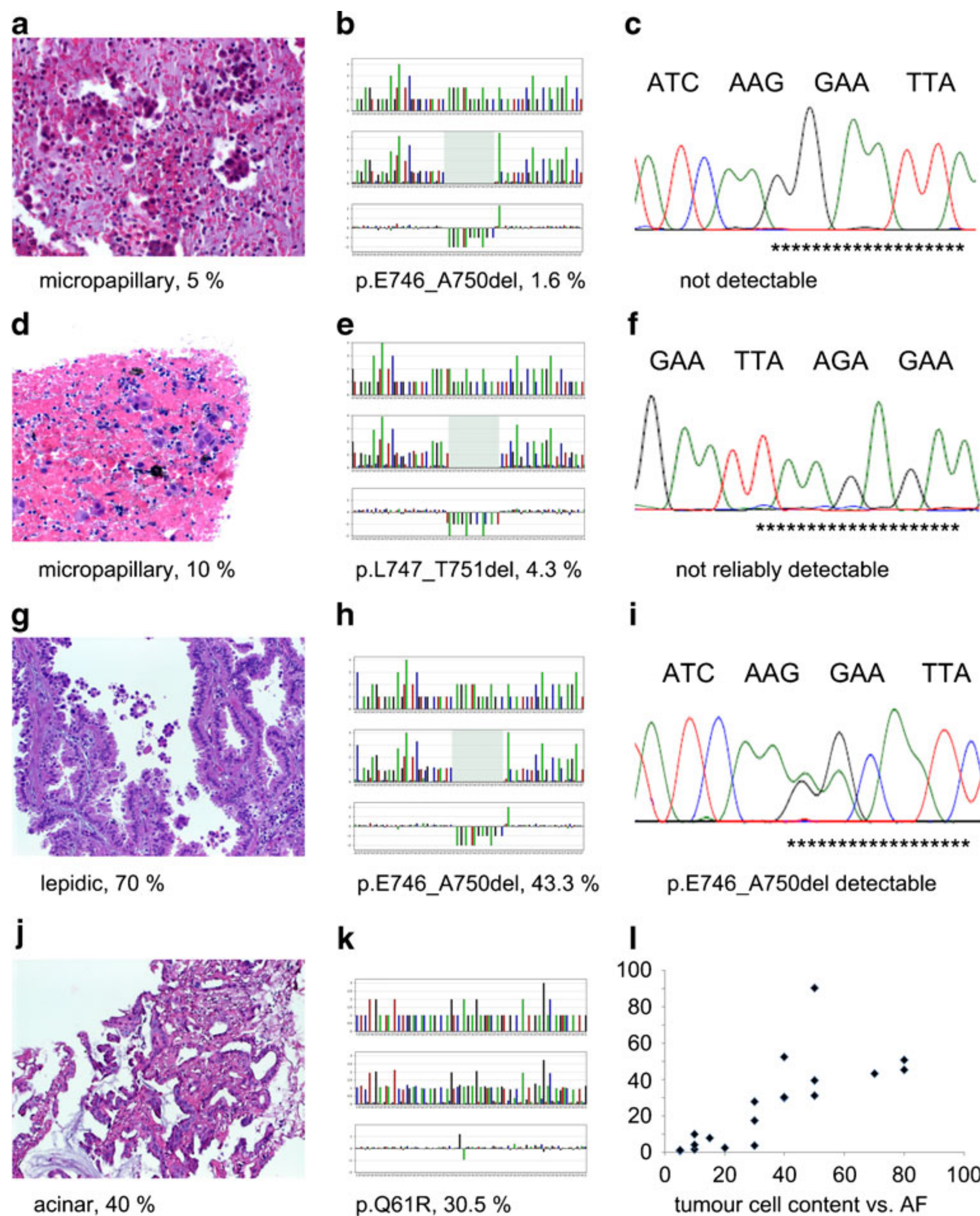

Fig. 4 Representative examples of the higher sensitivity of *EGFR* mutation detection by 454 deep sequencing compared to Sanger sequencing in clinical NSCLC samples with low tumour cell content. **a**–**c** Case 2. Cell block preparation from pleural effusion specimen with micropapillary growth pattern and low tumour cell content of 5 % (**a**) and *EGFR* exon 19 deletion p.E746_A750del detected with 454 deep sequencing at allele frequency of 1.6 % (**b**), compared to false-negative wild type sequence with Sanger sequencing (**c**). The corresponding region in the chromatogram is indicated with *asterisks*. **d**–**f** Case 3. Cell block preparation from fine needle aspiration specimen with micropapillary growth pattern and low tumour cell content of 10 % (**d**) and *EGFR* exon 19 deletion p.L747_T751del detected with 454 deep sequencing at allele frequency of 4.3 % (**e**), which cannot be reliably detected with Sanger sequencing (**f**). **g**–**i** Case 11. Resection specimen with lepidic growth pattern and high tumour cell content of 70 % (**g**), and *EGFR* exon 19 deletion p.E746_A750del detectable both with 454 deep sequencing at allele frequency of 43.3 % (**h**) and with Sanger sequencing (**i**). **j**, **k** Case 17. Resection specimen with mucin-producing acinar growth pattern and high tumour cell content of 40 % (**j**), and *KRAS* exon 3 missense mutation p.Q61R detectable by 454 deep sequencing at allele frequency of 30.5 % (**k**). **l** Significant correlation of tumour cell content (*x*-axis) and observed allele frequency (AF, *y*-axis) in 17 NSCLC specimens with *EGFR* and *KRAS* mutations determined by 454 deep sequencing (R^2=0.898)

reliably detect *EGFR* and *KRAS* mutations down to at least 1 % allele frequency, whereas background variants occur at a maximal frequency of 0.8 %. Finally, high reproducibility of mutant variant detection was achieved with a mean standard deviation of 0–2.1 %. Taken together, our results demonstrate highest sensitivity, specificity and reproducibility of quantitative detection of *EGFR* and *KRAS* mutations by 454 deep sequencing.

Higher sensitivity of 454 deep sequencing compared to Sanger sequencing in FFPE NSCLC samples: clinical implications for stratification of TKI therapy

In a recent study, it has been demonstrated that Sanger sequencing has a lower threshold of ≤40 % tumour cell content for reliable detection of *EGFR* mutations in clinical NSCLC specimens [10]. Notably, ~30 % of NSCLC samples in that large series contained ≤40 % tumour cells and were therefore non-informative for *EGFR* mutation status. This issue is particularly relevant for patients with extensive disease (stage IV), where frequently only small biopsies or even cytological preparations from pleural effusions are available, and non-informative *EGFR* mutation analysis will lead to re-biopsy with additional risk for the patient as well as additional costs for the clinic. Moreover, a false-negative *EGFR* mutation analysis due to low tumour cell content will delay or even preclude targeted therapy with TKI in these patients, with potentially negative impact on survival. Therefore, we aimed to demonstrate the significant impact of higher sensitivity of the developed 454 deep sequencing assays compared to Sanger sequencing in 16 clinical NSCLC samples with *EGFR* wild type status according to Sanger sequencing and low tumour cell content ≤40 %. Remarkably, we observed seven cases that harboured clinically relevant *EGFR* mutations at allele frequencies of 0.9–10 %, which were not detectable by Sanger sequencing. Five other cases displayed *KRAS* mutations, and four samples were of wild type status for both genes. In the control set of five samples with tumour cell contents of >40 %, *EGFR* mutations were detectable both with 454 deep sequencing and Sanger sequencing. This observation is concordant with an earlier study which concludes that NSCLC samples with a tumour cell content of ≤40 % cannot be reliably analysed by Sanger sequencing given unacceptably high probability of false-negative reports [10]. In contrast to another study [11], we found a clear correlation between the tumour cell content and the observed allele frequency of mutations analysed by 454 deep sequencing, and this observation further emphasizes the need of a careful evaluation of the tumour cell content in clinical samples before choosing the adequate technique for molecular analysis. Of note, the quantification of the tumour cell content in the samples further helps to interpret the clinical relevance of mutations found at low allele frequencies. In samples with low tumour cell content, a concurring low allele frequency of a mutation likely reflects the large amount of non-neoplastic cells with wild-type alleles, comparable to the artificial dilution series of cell line DNA employed in this study. Accordingly, the authors suggest that these tumours should be regarded and treated similar to those tumours with high tumour cell content, where mutations can be detected with less sensitive techniques (e.g. Sanger sequencing). In contrast, the high sensitivity of deep sequencing will also identify variants at low allele frequencies in samples with high tumour cell content, and the biological and therapeutic consequences of these findings are less clear. Until further ongoing studies will clarify this important question, these cases should be discussed together with the clinician, and if a targeted therapy with TKI is administered, it might be helpful to closely monitor the therapy effect in these patients.

Notably, there was also a higher prevalence of *EGFR* mutations in NSCLC samples with predominant micropapillary, papillary or lepidic growth pattern while *KRAS* mutations were restricted to tumours with acinar or solid growth pattern [14]. These different phenotypes in molecularly defined subgroups are increasingly recognized in different tumour entities and underline the impact of critical interpretation of genetic results in the context of clinicopathological characteristics. Thus, it should be emphasized that reporting of a wild-type genotype in NSCLC specimens displaying one of the characteristic *EGFR* phenotypes should alert to the possibility of false-negative results and the need for more sensitive analytical methods. In contrast, a tumour with acinar or solid growth pattern in combination with the finding of an unusual mutation at low allele frequency in a sample with high tumour cell content may indicate less responsiveness to TKI therapy.

While the much higher sensitivity of deep sequencing in clinical NSCLC samples is clearly demonstrated in the publication by Thomas and colleagues [11] and also in the current study, reagent costs and time for sample preparation are also critical factors for diagnostic molecular pathology laboratories in daily routine diagnostics and may influence the decision which technique is used. In direct comparison to Sanger sequencing, the reagent costs for this assay are two- to threefold higher, and the time for sample preparation is more labour intensive. Thus, only in samples with low tumour cell content ≤40 %, the advantage of higher sensitivity outweighs the higher costs and longer time of deep sequencing at the current stage. However, the predictable need for simultaneous analysis of multiple genetic aberrations within one sample (e.g. mutation analysis of *BRAF*, *PI3K*, *MEK1*, *HER2*) [7] in the near future will likely shift this balance to next generation parallel sequencing methods, as reagent costs and time for sample preparation will be less important when larger gene panels are analysed.

In conclusion, our protocol reliably and specifically detects *EGFR* and *KRAS* mutations present at allele frequencies as low as 1 % while 99.9 % of the background variants in non-homopolymer regions occur at allele frequencies <0.8 %. Therefore, this assay is superior to Sanger sequencing with dramatically increased sensitivity, and can be applied to NSCLC specimens independent of the tumour cell content or tumour cell amount (small endoscopic biopsies, core needle biopsies, cell block preparations from pleural effusions).

Regarding the dramatic consequences of a false-negative *EGFR* mutation analysis, NSCLC samples with a low tumour cell content ≤40 % cannot be reliably analysed by Sanger sequencing, and especially in these cases the increased sensitivity of deep sequencing will outweight its higher costs and longer sample preparation time. As flexibility of 454 deep sequencing easily allows the inclusion of further DNA-based predictive targets, simultaneous multi-gene panel testing for optimised therapy-related stratification of NSCLC patients can be envisioned in the near future and will help to establish 454 deep sequencing or similar next generation parallel sequencing methods in daily routine diagnostics.

Conflict of Interest The authors state that there is no conflict of interest. All authors have read and approved the final version of the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA (2004) 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
2. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304:1497–1500
3. Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, Varmus H (2004) EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A 101:13306–133011
4. Sequist LV, Martins RG, Spigel D, Grunberg SM, Spira A, Jänne PA, Joshi VA, McCollum D, Evans TL, Muzikansky A, Kuhlmann GL, Han M, Goldberg JS, Settleman J, Iafrate AJ, Engelman JA, Haber DA, Johnson BE, Lynch TJ (2008) First-line gefitinib in patients with advanced non-small-cell lung cancer harboring somatic EGFR mutations. J Clin Oncol 26:2442–2449
5. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, Porta R, Cobo M, Garrido P, Longo F, Moran T, Insa A, De Marinis F, Corre R, Bover I, Illiano A, Dansin E, de Castro J, Milella M, Reguart N, Altavilla G, Jimenez U, Provencio M, Moreno MA, Terrasa J, Muñoz-Langa J, Valdivia J, Isla D, Domine M, Molinier O, Mazieres J, Baize N, Garcia-Campelo R, Robinet G, Rodriguez-Abreu D, Lopez-Vivanco G, Gebbia V, Ferrera-Delgado L, Bombaron P, Bernabe R, Bearz A, Artal A, Cortesi E, Rolfo C, Sanchez-Ronco M, Drozdowskyj A, Queralt C, de Aguirre I, Ramirez JL, Sanchez JJ, Molina MA, Taron M, Paz-Ares L (2012) Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 13:239–246
6. Keedy VL, Temin S, Somerfield MR, Beasley MB, Johnson DH, McShane LM, Milton DT, Strawn JR, Wakelee HA, Giaccone G (2011) American Society of Clinical Oncology provisional clinical opinion: epidermal growth factor receptor (EGFR) Mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. J Clin Oncol 29:2121–2127
7. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansén S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Jänne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150:1107–1120
8. Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, Zakowski MF, Heelan RT, Kris MG, Varmus HE (2005) KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. PLoS Med 2:e17
9. Penzel R, Sers C, Chen Y, Lehmann-Mühlenhoff U, Merkelbach-Bruse S, Jung A, Kirchner T, Büttner R, Kreipe HH, Petersen I, Dietel M, Schirmacher P (2011) EGFR mutation detection in NSCLC–assessment of diagnostic application and recommendations of the German Panel for Mutation Testing in NSCLC. Virchows Arch 458:95–98
10. Warth A, Penzel R, Brandt R, Sers C, Fischer JR, Thomas M, Herth FJ, Dietel M, Schirmacher P, Bläker H (2012) Optimized algorithm for Sanger sequencing-based EGFR mutation analyses in NSCLC biopsies. Virchows Arch 460:407–414
11. Querings S, Altmüller J, Ansén S, Zander T, Seidel D, Gabler F, Peifer M, Markert E, Stemshorn K, Timmermann B, Saal B, Klose S, Ernestus K, Scheffler M, Engel-Riedel W, Stoelben E, Brambilla E, Wolf J, Nürnberg P, Thomas RK (2011) Benchmarking of mutation diagnostics in clinical lung cancer specimens. PLoS One 6:e19601
12. Zagordi O, Klein R, Däumer M, Beerenwinkel N (2010) Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Res 38:7400–7409
13. Gallegos Ruiz MI, Floor K, Rijmen F, Grünberg K, Rodriguez JA, Giaccone G (2007) EGFR and K-ras mutation analysis in non-small cell lung cancer: comparison of paraffin embedded versus frozen specimens. Cell Oncol 29:257–264
14. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Beer DG, Powell CA, Riely GJ, Van Schil PE, Garg K, Austin JH, Asamura H, Rusch VW, Hirsch FR, Scagliotti G, Mitsudomi T, Huber RM, Ishikawa Y, Jett J, Sanchez-Cespedes M, Sculier JP, Takahashi T, Tsuboi M, Vansteenkiste J, Wistuba I, Yang PC, Aberle D, Brambilla C, Flieder D, Franklin W, Gazdar A, Gould M, Hasleton P, Henderson D, Johnson B, Johnson D, Kerr K, Kuriyama K, Lee JS, Miller VA, Petersen I, Roggli V, Rosell R, Saijo N, Thunnissen E, Tsao M, Yankelewitz D (2011) International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol 6:244–285