Supplementary Information

Cross-Platform Comparison of Four Leading Technologies for Detecting *EGFR* Mutations in Circulating Tumor DNA from Non-Small Cell Lung Carcinoma Patient Plasma

Ting Xu[1]*, Xiaozheng Kang[2]*, Xiaofang You[1], Liang Dai[2], Dequan Tian[1], Wanpu Yan[2], Yongbo Yang[2], Hongchao Xiong[2], Zhen Liang[2], Grace Q.Zhao[3], Shengrong Lin[3], Ke-Neng Chen[2]✉, Guobing Xu[1]✉

1. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Clinical Laboratory, Peking University Cancer Hospital & Institute, Beijing, China.

2. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Thoracic Surgery I, Peking University Cancer Hospital & Institute, Beijing, China.

3. AccuraGen Holding Incorporated, Shanghai, China.

*Ting Xu and Xiaozheng Kang contributed equally to this study.

✉ Ke-Neng Chen and Guobing Xu jointly supervised this work.

Corresponding authors:

Ke-Neng Chen, chenkeneng@bjmu.edu.cn

Guobing Xu, 18813186407@bjmu.edu.cn

Table of Contents

Supplementary Material and Methods

Identification of *EGFR* alterations and germline variants in tumor tissue and white blood cell samples by standard NGS

Tumor tissue and white blood cell genomic DNA was sonicated into fragments with a peak length at 200bp using a Covaris S2 instrument. 100ng of fragmented tumor tissue and white blood cell genomic DNA were used for standard NGS library construction following KAPA sequencing library construction kit (Kapa Biosystems, Boston, MA, USA) manufacturer's protocol. Genomic DNA NGS library was captured by Accu-Act pan cancer panel (AccuraGen Inc., Shanghai, China), and followed by sequencing with 100 bp paired-end runs on an Illumina HiSeq 2500 system (Illumina, San Diego, CA, USA). The average coverage depth for all probes was at least 500X. Germline and somatic variants are called with AccuraGen's NGS pipeline. Briefly, sequence reads collected from the HiSeq were aligned to the hg19/GRCh37 reference genome using the Burrows-Wheeler Alignment tool (http://bio-bwa.sourceforge.net/bwa.shtml). Duplicate removal, and single nucleotide polymorphism and indel calling were performed using SAMBLASTER [1] and FreeBayes (http://arxiv.org/abs/1207.3907). Variants in low complexity regions were removed, and annotated with GEMINI [2].

Qualitative and quantitative detection of *EGFR* alterations in plasma samples

The cfDNA concentration, as determined by Qubit, ranged from 21.0ng/mL to 84.4ng/mL, with a mean of 35.1 ± 17.6ng/mL. The detailed concentration and quantity of cfDNA is shown in Supplementary Table S3.

ADx-ARMS

In parallel to the processing of tissue DNA using ADx-ARMS, two AmoyDx® *EGFR*

29 Mutations Detection Kits were utilized for plasma *EGFR* detection according to the manufacturer's instructions. For each sample, 40ng of plasma DNA was used. Real-time PCR was performed on an ABI StepOne plus Platform (Applied Biosystems, Foster City, CA, USA). The cycling conditions for the mutation assays were as follows: 5 min at 95°C, followed by 15 cycles at 95°C for 25s, 64°C for 20s, and 72°C for 20s, then 31 cycles at 93°C for 25s, 60°C for 35s, and 72°C for 20s. Fluorescence was measured at 60°C. Data on each mutation was interpreted according to the kit manual after curve analysis and calculation of ΔCt values.

Cobas-ARMS

Cobas-ARMS (cobas® *EGFR* Mutation Test v2) also served as a real-time PCR test for the qualitative detection of ctDNA and identified a larger number of *EGFR* loci than the AmoyDx® *EGFR* 29 Mutations Detection Kit, as detailed in Supplementary Table S4 and Table S5. Experiments using cell line DNA, mechanically sheared to an average size of 220bp with a wild type DNA background of approximately 100,000 copies/mL plasma, showed that the cobas-ARMS was capable of detecting mutations in *EGFR* exons 18, 20 and 21 with 100 copies of mutant DNA per mL of plasma and *EGFR* exon 19 with 75 copies of mutant DNA per mL of plasma using a standard input of 25μL of DNA stock per reaction well.

For each sample, 75ng of cfDNA was used for each PCR reaction according to the manufacturer's instructions. Real-time PCR was conducted on a cobas z480 analyzer with *EGFR* Plasma Analysis Package Software version 1.0. The software included the default PCR run method and automatically generated a PDF file containing the plasma *EGFR* mutation status of each sample after the end of the reaction.

ddPCR

ddPCR has not been approved by the FDA or CFDA, but is widely used for clinical trials because of its unparalleled sensitivity and ability to detect ctDNA in a quantitative manner. But it can only detect one locus per reaction well at a time.

EGFR L858R, T790M, E746_A750del and G719X mutations in plasma were quantitatively detected by ddPCR from Bio-Rad (Bio-Rad Laboratories Inc., Hercules, CA, USA) in this study. The reagents and consumables used in ddPCR analysis were purchased from Bio-Rad and summarized in Supplementary Table S6. L858R and exon 19 deletion (most frequently E746_A750del) account for approximately 85% of all *EGFR* mutations [3]. Both mutations are associated with sensitivity to gefitinib or erlotinib [4], which are alternative choices for therapy in NSCLC patients with these particular *EGFR* mutations. Patients carrying the uncommon p.G719X point mutation, comprising less than 4% of cases overall [5], are also sensitive to TKIs. However, tumors harboring these TKI-sensitive mutations often acquire resistance to TKIs within two years of the initiation of treatment with these drugs [6, 7]. The most common mechanism of resistance involves a secondary *EGFR* T790M mutation, which has a prevalence of 40%–60% [8, 9]. Once a tumor has developed resistance to TKIs, alternative therapies are necessary for ongoing therapy.

The QX200 droplet generator and QX200 droplet reader (Bio-Rad Laboratories Inc., Hercules, CA, USA) were used for gene testing according to the manufacturer's instructions. For each sample, approximately 20ng of cfDNA was used for each *EGFR* locus. Four PCR reactions (L858R, T790M, E746_A750del and G719X mutations)

were performed for every patient. cfDNA was amplified in a conventional calibrated thermal cycler as follows: 1 cycle of 95°C for 10 min; 40 cycles of 94°C for 30s and 55°C for 1 min, and a final extension at 98°C for 10 min. The product was stored at 4°C. When cycling was complete, the plate was loaded into the QX200 droplet reader and the allele frequency was analyzed using QuantaSoft v1.6 (Bio-Rad Laboratories Inc., Hercules, CA, USA).

Firefly NGS

NGS-based techniques for *EGFR* mutation profiling have not been approved by the FDA or CFDA; however, they are widely used for research purposes due to their multiplexing capability and ability to detect ctDNA in a quantitative manner. AccuraGen (AccuGen Inc.) recently developed a highly sensitive NGS-based assay, and it allows detection of 0.02% mutated DNA (http://www.accuragen.com/en-coretechnology.html).

For each sample, cfDNA was circularized using CircLigase II single strand DNA ligase (EpiCentre, Madison, WI, USA) with 10-30ng of DNA per 20μL reaction. After ligation, samples were treated with exonuclease (NEB, Ipswich, MA, USA) to remove uncircularized DNA. The circularized DNA was then amplified in a rolling circle amplification reaction using the Phi29 DNA polymerase (NEB, Ipswich, MA, USA) and exonuclease-resistant random primers using manufacturer's instructions with modifications.

Whole genome amplification product was purified by AMPure beads (Beckman Coulter, Brea, CA, USA) and sonicated to short fragments that were suitable for NGS library

construction using a Covaris sonicator. NGS sequencing libraries were generated from 100ng of amplified cfDNA using the KAPA sequencing library construction kit (Kapa Biosystems, Boston, MA, USA) per manufacturer's instructions. Briefly, double strand DNA was subjected to end repair and A-tailing reactions for 30 min at 20°C and 30°C, respectively. After 1.8X bead purification, adaptor was added through ligated product at 20°C for 15 min. After 1X bead purification, PCR amplification with Illumina primers P5 and P7 was performed. Amplified genomic DNA was cleaned up by 0.8X bead prior to panel capture.

A panel (Accu-Act Pan Cancer, Accuragen, Inc.) consisting of 61 cancer related genes of clinical significance (Supplementary Table S7) including *EGFR* exons 18, 19, 20, 21 (>15kb), was used for capture. NGS libraries were sequenced on an Illumina HiSeq 2500 (Illumina, San Diego, CA, USA), and the unique sequencing reads were determined by using an AccuraGen proprietary algorithm. The average coverage depth for all probes in plasma was approximately 7,000X.

Similar to the variant calls from genomic DNA, sequence reads were aligned to the hg19/GRCh37 human reference sequence, and background noise introduced by random NGS error was removed by AccuraGen proprietary algorithms. True variants were then identified, and the allele frequency was calculated by comparing the number of unique reads containing variants to the total number of sequencing reads that mapped to the position of the variant. The ctDNA and tumor genomic DNA sequencing data were cross-checked with the germline variants from white blood cell genomic DNA to identify somatic mutations.

Data analysis

Determination of the *EGFR* mutant and wild type loci and coincidence rate

When calculating the total number of *EGFR* mutant loci and positive coincidence rates between any two platforms, we defined the locus as *EGFR* mutant if a positive result was obtained from any one of the platforms. Otherwise, when calculating the total number of *EGFR* wild type loci and negative coincidence rate between any two platforms, we defined the locus as *EGFR* wild type if a negative result was obtained from any one of the platforms. The positive and negative coincidence rates were calculated as follows:

The positive coincidence rate = the number of *EGFR* mutant variants detected using both methods /the number of variants detected using either method.

The negative coincidence rate = the number of *EGFR* wild type loci detected using both methods / the number of loci detected using either method.

Supplementary Tables

Supplementary Table S1. Clinical data of twenty NSCLC patients.

ID	Date of Surgery	Date of Recurrence	Data of Blood Sampling	Date of Response Evaluation	Response Evaluation (RECIST)	Stage at Initial Evaluation	Stage at Blood Sampling	Site of Recurrence	Systemic Therapy (Duration, Regime, RECIST)
S1	2009/7/30	2010/10/20	2015/11/3	2015/12/4	SD	IIA	IV(r)	Brain, Bone	2009/8/21-2009/10/3, PEM+DDP, PD; 2013/5/14-2016/7/20, EGFR-TKI[a], SD.
S2	2012/7/2	/	2015/10/30	2015/11/13	PD	IV	IV	/	2012/7/22-2013/12/3, EGFR-TKI, PD; 2014/3/5-2014/5/13, PEM+NDP, PD; 2014/10/14-2014/10/23, TXT, PD; 2014/12/3-2016/3/1, IT[b], PD.
S3	2013/7/1	2014/8/13	2015/11/12	2015/11/20	PR	IIIA	IV(r)	Lung, Bone	2013/12/25-2014/3/13, PEM+NDP, PD; 2014/8/23-2016/3/1, EGFR-TKI, PD.
S4	2015/10/19	/	2015/11/23	2016/2/24	SD	IV	IV	/	2015/10/27-2015/11/27, PEM+DDP, SD.
S5	2012/9/2	2015/10/27	2015/11/3	2015/12/15	PD	IV	IV(r)	Pleura	2013/11/5-2015/11/3, EGFR-TKI, PD; 2015/12/26-2016/3/18, PEM+NDP, PD; 2016/7/8-2016/7/10, TXT, SD.
S6	2015/1/22	2015/8/5	2015/11/30	2016/1/5	SD	IA(m)	IV(m,r)	Pleura , Mediastina Lymph Nodes	2015/2/28-2015/4/21, PTX+DDP, PD; 2015/8/5-2016/7/20, EGFR-TKI, PR.
S7	2013/7/3	2014/6/9	2015/11/14	2016/2/22	SD	IIIA	IV(r)	Lung	2013/9/19-2013/10/16, PTX+DDP, PD; 2014/11/3-2016/6/12, EGFR-TKI, PD.
S8	2013/6/26	2014/2/10	2015/11/5	2015/12/1	PR	IIIB	IV(r)	Lung	2013/6/28-2013/9/6, PEM+DDP, PD; 2014/2/15-2016/2/1, EGFR-TKI, PD.
S9	2013/2/27	2015/3/17	2015/11/20	2015/12/30	SD	IIIB	IV(r)	Brain	2015/3/24-2016/7/20, EGFR-TKI, PD.
S10	2012/7/5	2015/11/17	2015/12/1	2016/3/9	SD	IB	IV(r)	Lung	2015/11/1-2016/7/11, EGFR-TKI, SD.
S11	2012/5/17	2012/8/18	2015/11/5	2015/11/23	SD	IIA	IV(r)	Lung	2012/6/10-2012/7/1, PEM+NDP, PD; 2012/8/20-2016/7/20, EGFR-TKI, PR.

S12	2015/9/2	/	2015/11/2	2015/11/30	PR	IV	IV	/	2015/10/31-2016/7/20, EGFR-TKI, PR.
S13	2012/2/2	2014/4/8	2015/11/10	2015/12/3	SD	IIIA	IB(r)	Mediastina Lymph Nodes, Lung, Brain	2012/2/29-2012/3/21, PTX+DDP, PD; 2014/4/23-2016/6/5, EGFR-TKI, PR.
S14	2002/12/5	2015/10/9	2015/11/4	2015/11/8	SD	IIIA	IA(r)	Lung	none
S15	2015/9/10	/	2015/11/18	2016/6/1	SD	IA(m)	IA(m)	/	2015/11/20-2016/2/21, PTX+NDP, SD.
S16	2015/10/22	/	2015/11/2	2015/12/21	SD	IV	IV	/	2015/11/5-2016/2/25, PEM+CAP, PR.
S17	2010/11/1	2015/4/13	2015/11/2	2015/11/20	SD	IB	IV(r)	Pleura	2015/8/20-2016/7/20, EGFR-TKI, PD.
S18	2015/9/24	/	2015/11/14	2016/11/16	PR	IV	IV	/	2015/10/27-2015/11/27, PTX+DDP, PR.
S19	2013/5/13	2014/11/17	2015/11/23	2016/2/22	PD	IB	IV(r)	Mediastina Lymph Nodes, Brain	2016/3/15-2016/4/12, GEM, PD; 2016/5/10-2016/6/8, TXT, PD.
S20	2015/11/30	/	2015/12/1	2016/1/27	SD	IV	IV	/	2015/12/12-2016/4/11, PEM+DDP, SD.

7th edition AJCC/UICC TNM staging system for NSCLC was applied accordingly. a, EGFR-TKIs that patients took in this study were first generation EGFR-TKI; m, multiple primary lung cancer; r, recurrence; none, no medical treatment. b, IT denotes interventional therapy via bronchoscopy; TT, targeted therapy; CT, chemotherapy; PEM, pemetrexed; PTX, paclitaxel; DDP, cisplatin; GEM, gemcitabine; NDP, nedaplatin; TXT, docetaxel; CAP, carboplatin; RECIST, Response Evaluation Criteria in Solid Tumors; PR, partial response; SD, stable disease; PD, progressive disease.

Supplementary Table S2. Performance of four ctDNA test platforms for detection of three *EGFR* driver mutations (L858R, exon 19 deletion and G719X) in 20* plasma samples.

	ADx-ARMS	cobas-ARMS	ddPCR	Firefly NGS
Sensitivity	26.7% (4/15)	53.3% (8/15)	53.3% (8/15)	53.3% (8/15)
Specificity	100% (5/5)	100% (5/5)	100% (5/5)	100% (5/5)
Total coincidence rate	45% (9/20)	65% (13/20)	65% (13/20)	65% (13/20)

*Seventeen patients with stage IV disease at blood sampling were enrolled to determine the total coincidence rates of three *EGFR* driver mutations (L858R, exon 19 deletion and G719X) profiles between tissue and plasma. For patients S5 and S6, their tissue test results from ADx-ARMS and standard NGS were both used as the reference, but one case with two different *EGFR* mutation types using ADx-ARMS and standard NGS was considered as two cases with only one *EGFR* mutation type each. For the other fifteen patients, the tissue test results obtained from ADx-ARMS were used as the reference. However, one patient with double mutations (S3) was also considered as two cases with only one *EGFR* mutation type each, the other patient with double mutations (S4) was considered as one case with the exon 19 deletion because T790M mutation was excluded from the table. As a result, twenty cases were analyzed.

Supplementary Table S3. Concentration of cfDNA isolated from plasma of twenty patients using QIAamp Circulating Nucleic Acid Kit.

ID	Plasma Volume (mL)	cfDNA yield (ng/mL)
S1	10.4	26.4
S2	10.0	31.2
S3	10.0	24.4
S4	9.0	23.6
S5	10.8	25.9
S6	10.4	22.7
S7	8.4	49.1
S8	8.4	26.5
S9	8.0	46.5
S10	10.0	84.4
S11	8.4	35.4
S12	9.6	41.4
S13	10.0	29.6
S14	10.0	21.0
S15	10.0	21.8
S16	10.2	22.9
S17	9.6	23.3
S18	9.0	32.8
S19	10.0	76.8
S20	8.0	35.4

Supplementary Table S4. *EGFR* mutations detected in AmoyDx® *EGFR* 29 Mutations Detection Kit.

Exon	*EGFR* Mutation	*EGFR* Nucleic Acid Sequence	COSMIC ID
18	G719A	2156G>C	6239
	G719S	2155G>A	6252
	G719C	2155G>T	6253
19	Exon 19 deletion	2235_2249del15	6223
		2236_2250del15	6225
		2240_2257del18	12370
		2235_2252>AAT(complex)	13551
		2236_2253del18	12728
		2237_2251del15	12678
		2237_2254del18	12367
		2237_2255>T(complex)	12384
		2238_2255del18	6220
		2238_2248>GC(complex)	12422
		2238_2252>GCA(complex)	12419
		2239_2247del9	6218
		2239_2253del15	6254
		2239_2256del18	6255
		2239_2248TTAAGAGAAG>C(complex)	12382
		2239_2258>CA(complex)	12387
		2240_2251del12	6210
		2240_2254del15	12369
		2239_2251>C(complex)	12383
20	S768I	2303G>T	6241
	T790M	2369C>T	6240
	Exon 20 Insertion	2319_2320insCAC	12377
		2310_2311insGGT	12378
		2307_2308insGCCAGCGTG	12376
21	L858R	2573T>G	6224
	L861Q	2582T>A	6213

Supplementary Table S5. *EGFR* mutations detected in cobas® *EGFR* Mutation Test v2.

Exon	*EGFR* Mutation	*EGFR* Nucleic Acid Sequence	COSMIC ID
18	G719A	2156G>C	6239
	G719S	2155G>A	6252
	G719C	2155G>T	6253
19	Exon 19 deletion	2240_2251del12	6210
		2239_2247del9	6218
		2238_2255del18	6220
		2235_2249del15	6223
		2236_2250del15	6225
		2239_2253del15	6254
		2239_2256del18	6255
		2237_2254del18	12367
		2240_2254del15	12369
		2240_2257del18	12370
		2239_2248TTAAGAGAAG>C	12382
		2239_2251>C	12383
		2237_2255>T	12384
		2235_2255>AAT	12385
		2237_2252>T	12386
		2239_2258>CA	12387
		2239_2256>CAA	12403
		2237_2253>TTGCT	12416
		2238_2252>GCA	12419
		2238_2248>GC	12422
		2237_2251del15	12678
		2236_2253del18	12728
		2235_2248>AATTC	13550
		2235_2252>AAT	13551
		2235_2251>AATTC	13552
		2253_2276del24	13556
		2237_2257>TCT	18427
		2238_2252del15	23571
		2233_2247del15	26038
20	S768I	2303G>T	6241
	T790M	2369C>T	6240
	Exon 20 Insertion	2307_2308ins9GCCAGCGTG	12376
		2319_2320insCAC	12377
		2310_2311insGGT	12378

		2311_2312ins9GCGTGGACA	13428
		2309_2310AC>CCAGCGTGGAT	13558
21	L858R	2573T>G	6224
		2573_2574TG>GT	12429
	L861Q	2582T>A	6213

Supplementary Table S6. Reagents used in ddPCR (Bio-Rad) analysis.

Reagents	Catalog Number
EGFR WT for p.T790M and *EGFR* p.T790M	10040782
EGFR WT for p.L858R and *EGFR* p.L858R	10040783
EGFR WT for p.E746_A750del and *EGFR* p.E746_A750del	10041170
EGFR WT for p.G719S and *EGFR* p.G719S	10041172
ddPCR Supermix for Probes (no dUTP)	1863023
Droplet generation oil, includes 10 7 mL bottles	1863005
Droplet reader oil, includes 2 1L bottles	1863004

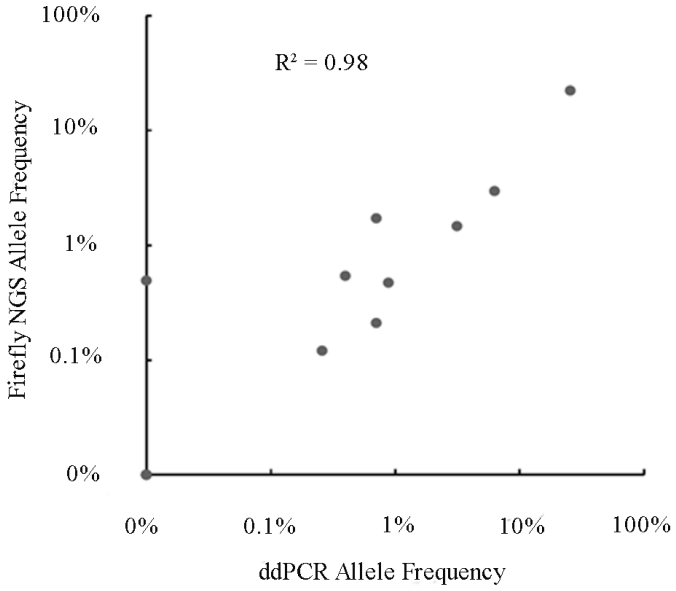
WT, wild type.

Supplementary Table S7. Accu-Act Panel gene list for Firefly NGS analysis.

AR	***CDK6***	*EZH2*	*GNAS*	***KRAS***	***PDGFRA***	*SMAD4*
ABL1	*CDKN2A*	*FBXW7*	*HNF1A*	*MAP2K1*	***PIK3CA***	*SMARCB1*
AKT1	*CSF1R*	*FGFR1*	*HRAS*	***MET***	*PTCH1*	*SMO*
ALK	*CTNNB1*	*FGFR2*	*IDH1*	*MLH1*	*PTEN*	*SRC*
APC	*DDR2*	*FGFR3*	*IDH2*	*MPL*	*PTPN11*	***TERT***
ATM	*DNMT3A*	*FLT3*	*JAK2*	*MSH6*	*RB1*	***TP53***
BRAF	***EGFR***	*FOXL2*	*JAK3*	*NOTCH1*	*RET*	*TSC1*
CDH1	***ERBB2***	*GNA11*	*KDR*	*NPM1*	*ROS1*	
CDK4	*ERBB4*	*GNAQ*	***KIT***	***NRAS***	*STK11*	

****Bold** indicates the gene include all exons*

Supplementary Figures



Supplementary Figure S1. Correlation of *EGFR* mutation abundance between ddPCR and Firefly NGS assays in twenty NSCLC patients ($R^2 = 0.98$).

References

1. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 2014; 30: 2503-5.
2. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. PLoS Comput Biol. 2013; 9: 1-8.
3. Baek JH, Sun JM, Min YJ, Cho EK, Cho BC, Kim JH, et al. Efficacy of EGFR tyrosine kinase inhibitors in patients with EGFR-mutated non-small cell lung cancer except both exon 19 deletion and exon 21 L858R: a retrospective analysis in Korea. Lung Cancer. 2015; 87: 148-54.
4. Gazdar AF. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene. 2009; 28 (Suppl 1): S24-S31.
5. Massarelli E, Johnson FM, Erickson HS, Wistuba, II, Papadimitrakopoulou V. Uncommon epidermal growth factor receptor mutations in non-small cell lung cancer and their mechanisms of EGFR tyrosine kinase inhibitors sensitivity and resistance. Lung Cancer. 2013; 80: 235-41.
6. Diaz LA, Jr., Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature. 2012; 486: 537-40.
7. Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. N Engl J Med. 2013; 368: 842-51.
8. Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. Clin Cancer Res. 2013; 19: 2240-7.
9. Yang X, Zhuo M, Ye X, Bai H, Wang Z, Sun Y, et al. Quantification of mutant

alleles in circulating tumor DNA can predict survival in lung cancer. Oncotarget. 2016; 7: 20810-24.