



A transforming MET mutation discovered in non-small cell lung cancer using microarray-based resequencing

Torstein Tengs^{a,b,*}, Jeff C. Lee^{a,b}, J. Guillermo Paez^{a,c}, Xiaojun Zhao^a, Thomas LaFramboise^{a,b}, Georgia Giannoukos^b, Roman K. Thomas^{a,b,**}

^aDepartment of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA

^bThe Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

^cMerck & Co., Inc., 33 Avenue Louis Pasteur, Boston, MA 02115, USA.

Received 9 July 2005; received in revised form 26 July 2005; accepted 3 August 2005

Abstract

We have designed resequencing microarrays to test the performance of this platform when interrogating a large number of exons (164 total) from genes associated with cancer. To evaluate false positive and negative rates, dideoxy sequencing was done for 335,420 bases interrogated by the arrays. From the array data, calls could be made for ~97.5% of the bases, and false positive rates were very low with only a single mutation reported from the array dataset for which the corresponding dideoxy trace had a clean wildtype sequence. For the nucleotide positions where array calls were made, false negative rates were 1.41% for heterozygous mutations. All the homozygous mutations were detected, but 8.11% were erroneously reported as heterozygous changes from the reference sequence by the array analysis software. In addition, 20 non-small cell lung cancer (NSCLC) samples were analyzed using the arrays, and both somatic and germline mutations were found. The most interesting findings were two MET mutations that have recently been implemented in NSCLC. Large scale MALDI-TOF genotyping indicated that one of these mutations (T1010I) might represent a true cancer-causing genotype, whereas the other (N375S) appears to be a common germline polymorphism.

© 2005 Elsevier Ireland Ltd. All rights reserved.

Keywords: Oncogene; Tumor suppressor gene; Gene mutation; Microarray; Dideoxy sequencing; Resequencing.

1. Introduction

The completion of the human genome project has opened avenues for systematic investigation of cancer-associated gene mutations. Using such unbiased systematic screens, we have recently identified EGFR mutations in non-small cell lung cancer (NSCLC) patients that strongly predict clinical response to the tyrosine kinase inhibitor gefitinib

* Corresponding author. Address: Torstein Tengs, National Veterinary Institute, Ullevaalsveien 68, 0454 Oslo, Norway.

Tel.: +47 23216243; fax: +47 23216202.

** Corresponding author. Address: The Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA.

Tel.: +1 6173240804; fax: +1 6172580901.

E-mail addresses: torstein.tengs@vetinst.no (T. Tengs), rthomas@broad.mit.edu (R.K. Thomas).

[1]. Additional discoveries from similar resequencing projects include a novel mutation in the FLT3 gene in acute myeloid leukemia (AML) [2], BRAF mutations in melanoma, colorectal carcinoma and lung cancer [3,4], PIK3CA mutations in colorectal cancer [5] and ERBB2 mutations in lung cancer [6]. Most of these projects have relied on analyses of sequence traces from dideoxy-based sequencing (‘Sanger sequencing’) to search for genetic changes.

Recently, single color microarray-based resequencing technology was introduced [7,8]. GeneChip CustomSeq Resequencing arrays (Affymetrix, Santa

Clara, CA 95051) rely on hybridization of fragmented and end labeled DNA to sets of 25 bp probes that have the central base as either A, C, G or T. These probes are complementary to the reference sequence to be resequenced, with the exception of the central base where only one of the probes corresponds to the wildtype allele for that particular nucleotide position. Probes are tiled throughout the sequence, with each set shifted one base upstream of the previous. Probes corresponding to both sense and anti-sense strand are synthesized on the array, and signal intensities from the 8 probes can be analyzed to give homozygous or

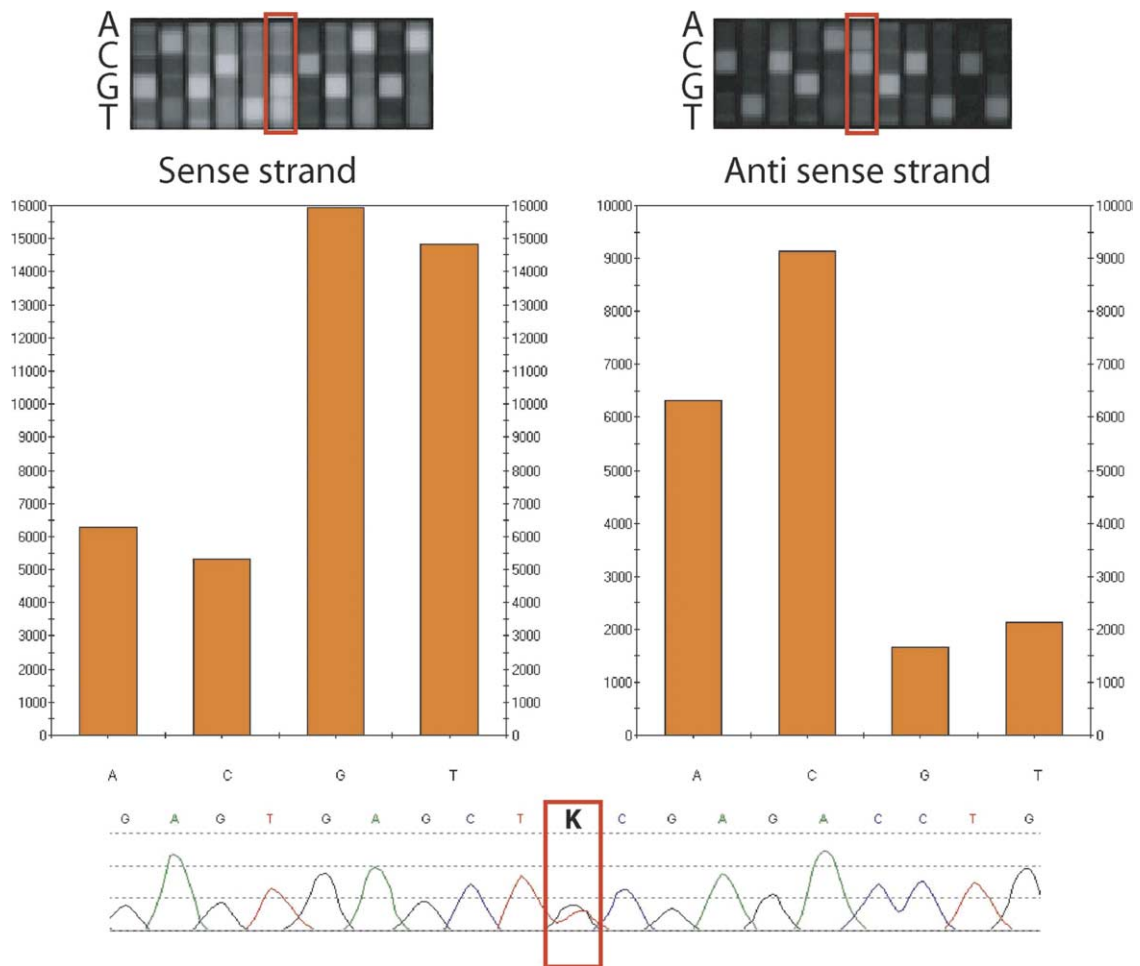


Fig. 1. Example of a heterozygous call using resequencing arrays and dideoxy traces. Top, 25 bp probe sets complementary to the sense (forward) and anti-sense (reverse) strand of the DNA to be resequenced. A, C, G and T refers to the central base in each probe; Middle, signal intensities from the central, heterozygous nucleotide position; Bottom, dideoxy trace from the same locus with the heterozygous nucleotide as a double peak (T and G, IUPAC code: K).

heterozygous base calls in IUPAC format for all loci (Fig. 1) or a no-call (*n*) where the genotype cannot be determined.

We have designed resequencing arrays to search for mutations in a set of 164 exons (23,966 bases) from genes associated with cancer. DNA from 20 lung tumor samples and their matched normals was analyzed, and dideoxy sequencing was performed on a subset of these exons in order to evaluate the performance of the arrays.

2. Materials and methods

2.1. Specimens and DNA preparations

Genomic DNA was extracted from tumor and non-tumoral ('normal') lung tissue using the QIAGEN DNeasy kit (QIAGEN, Valencia, CA 91355) and subjected to whole-genome amplification (WGA) as

described previously [9]. The identity of all tumor-normal pairs was conformed by genotyping 20 single-nucleotide polymorphism loci in each sample (data not shown). Twenty primary NSCLC samples were included in the study: 16 adenocarcinomas (1 with bronchioloalveolar carcinoma (BAC) features) and 4 squamous-cell carcinomas (Table 1). The tumors were either from US (Brigham and Women's Hospital, Boston, MA 02115, USA) or from Japanese sources ($n=5$, all adenocarcinomas, provided by Dr Hidefumi Sasaki, Department of Surgery, Nagoya City University Medical School, Nagoya 467-8601, Japan).

2.2. PCR

PCR primers covering ARAF, BRAF, CDK4, CDK6, CDKN2A, KLF6, HRAS, KRAS, MET, NRAS, PTEN, RAF1, RB1, RET and TP53 (164 exons total, exon reference sequences and primer

Table 1
Patient data

Sample	Histology	Differentiation	Age	Gender	T ^a	N ^a	M ^a	Source
S0392T	NSC ^b Adeno ^c	Poor	50	M	4	0		Japan
S0397T	NSC Adeno	Moderate	70	M	3	2		Japan
S0405T	NSC Adeno	Moderate	79	F	4	0		Japan
S0410T	NSC Adeno	Well	71	M	4	0		Japan
S0376T	NSC Adeno/ BAC ^d	Well	70	F	1	0		Japan
S0537T	NSC Adeno	Poor	63	F	2	0	X	U.S.
S0464T	NSC Adeno	Poor	69	M	2	1	X	U.S.
S0488T	NSC Adeno	Moderate/Poor	65	F	2	0	X	U.S.
S0498T	NSC Adeno	Moderate	63	M	2	0	1	U.S.
S0500T	NSC Adeno	Moderate/Poor	53	F	1	0	X	U.S.
S0516T	NSC Adeno	Poor	60	F	2	0	X	U.S.
S0518T	NSC Adeno	Moderate	47	F	1	2	X	U.S.
S0522T	NSC Adeno	Poor	50	F	1	0	1	U.S.
S0530T	NSC Adeno	Well	71	F	1	0	X	U.S.
S0535T	NSC Adeno	Unknown	73	F	2	2	X	U.S.
S0539T	NSC Adeno	Poor	72	F	3	1	X	U.S.
S0446T	NSC Squamous ^e	Poor	60	M	2	0	X	U.S.
S0461T	NSC Squamous	Moderate	70	M	2	X	X	U.S.
S0480T	NSC Squamous	Poor	74	F	1	0	X	U.S.
S0515T	NSC Squamous	Moderate	35	M				U.S.

^a Tumor, lymph-node and metastases staging according to the TNM classification.

^b Non-small cell lung carcinoma.

^c Adenocarcinoma.

^d Bronchioloalveolar carcinoma features.

^e Squamous cell carcinoma.

sequences available upon request) with amplicon lengths 400–700 bases were designed using the Primer3 software (http://www-genome.wi.mit.edu/genome_software/other/primer3.html). PCR amplifications were performed using 2.5 ng of WGA reaction product per reaction and the HotStarTaq Master Mix Kit (QIAGEN), with 5% DMSO and the following cycle: 15 min/95 °C, 40 cycles of 30 s/94 °C, 30 s/60 °C, 45 s/72 °C and a final extension step of 1 min/72 °C. PCR products were quantified using PicoGreen (Molecular Probes, Eugene, OR 97402) and a 96-well SpectraMax Gemini XPS microplate spectrofluorometer (Molecular Devices, Sunnyvale CA 94089). All amplicons were diluted to a final concentration of 0.011 pmol/μl, and 10 μl from each dilution was pooled to give a total of approximately 8 μg of DNA per sample. Pools of amplicons were purified using QIAquick PCR Purification columns (QIAGEN).

2.3. Amplicon preparation, array scan and analysis

For each sample, 4 μg of eluted DNA (corresponding to approximately 0.055 pmoles of each amplicon) was fragmented and end labeled using the Affymetrix ReSequencing Assay Kit (Affymetrix, Santa Clara, CA 95051) according to the manufacturer's recommendations. Arrays were scanned using a GeneChip Scanner 3000 and GCOS (version 1.2, Affymetrix). Arrays were then analyzed for substitution mutations using GDAS version 3.0.1 (Affymetrix) in batch mode with default parameters. GDAS uses an algorithm based on the work of Cutler et al. [10]. Briefly, given a fixed nucleotide position, eleven models—representing the ten different possible genotypes (specific; A, C, G and T, or ambiguous; W, R, K, Y, S and M) plus one 'no-call' model—are fit to signal intensity data from the corresponding position on the array. The algorithm's genotype call is the call whose corresponding model best fits the data.

During the course of the study, 22 nucleotide positions were found that gave high numbers of heterozygous calls and/or no-calls, but where the corresponding dideoxy traces showed a clean wild type sequence. These probe sets were thus deemed unreliable, and excluded from our analyses.

2.4. Dideoxy sequencing

All exons from ARAF, BRAF, MET, PTEN, RAF1, RET and TP53 were amplified using DNA from our set of 20 NSCLC samples as described above, but with M13-tailed primers. PCR products were treated with ExoSAP (Amersham Biosciences, Piscataway, NJ 08855) in order to degrade unincorporated nucleotides and PCR primers. ExoSAP treated PCR products were then sequenced directly using a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA 94404) and BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). High quality dideoxy traces (>90% of exonic bases phred>30) were aligned and analyzed using Mutation Surveyor version 2.50 (SoftGenetics, State College, PA 16803).

2.5. Genotyping

The Sequenom MassArray system at the Broad Institute was used to genotype for selected mutations in 361 lung tumor samples (265 primary tumors, Japanese origin: $n = 163$; US origin: $n = 102$; 96 lung cancer cell lines) as described elsewhere [11]. Healthy control DNA samples from 90 Caucasian and 90 Asian individuals were also included as controls. In brief, primers and probes were designed for each candidate mutation by the SpectroDesign software (Sequenom, San Diego, CA 92121) (FASTA sequences of primers and probes are available upon request). After Primer extension was carried out on PCR products, primer-extension reactions were loaded onto a matrix pad (3-hydroxypicolinic acid) of a SpectroCHIP (Sequenom). SpectroCHIPS were analyzed using a Bruker Biflex III MALDI-TOF mass spectrometer (SpectroREADER, Sequenom). Raw spectra were processed using SpectroTYPER (Sequenom).

3. Results and discussion

A total of approximately 1 Mb of sequence information was generated from the arrays, and overall coverage was estimated by calculating the number of base calls made and comparing this number to the number of possible calls. Coverage was

Table 2
Coverage and accuracy of resequencing arrays when compared to dideoxy sequencing.

Resequencing arrays	
Total number of bases interrogated	958,640
Total number of 'no calls' made by GDAS	24,211 (2.53%)
Coverage	934,429 (97.47%)
Exons also covered by dideoxy sequencing	
Total number of bases interrogated	335,420
Total number of 'no calls' made by GDAS	8,283 (2.47%)
Coverage	327,137 (97.53%)
Number of homozygous mutations found by dideoxy sequencing in loci where GDAS made calls	37
Number of heterozygous mutations found by dideoxy sequencing in loci where GDAS made calls	71
Total number of 'no calls' made by GDAS in mutated loci	11 (9.24%)
Homozygous mutations called correctly by GDAS	34 (91.89%)
Heterozygous mutations called correctly by GDAS	70 (98.59%)
Homozygous mutations called as wild type by GDAS	0
Heterozygous mutations called as wild type by GDAS	1
False positive homozygotes calls by GDAS	0
False positive heterozygotes calls by GDAS	1
Homozygous mutations called as heterozygous by GDAS	3
Heterozygous mutations called as homozygous by GDAS	0
Total number of correct calls by GDAS in loci covered	327,132 (99.99%)

consistent between arrays, with an overall fraction of 97.5% of bases called and a no-call rate of 2.5% (Table 2). A total of 35 homozygous and 87 heterozygous changes from the reference sequence were reported by the GDAS software (complete list of mutations is available upon request).

To determine the accuracy of the array calls, dideoxy electropherograms ('traces') were generated for a large fraction of the exons covered by the arrays (335,420 bases overlap, Table 2). 37 homozygous and 71 heterozygous sequence variants could be identified in the dideoxy traces for loci where calls were made from the array data (Table 2). Of these variants, 34 (91.89%) of the homozygous changes and 70 (98.59%) of the heterozygous loci were identified correctly by automated calling in the microarray dataset. The remaining incorrect calls for the homozygously variant loci were all homozygous changes erroneously reported as heterozygous loci. A single false positive call was identified. For the mutated loci identified by the dideoxy sequencing, the rate of no-calls was somewhat higher than the overall rate; 9.24% of the nucleotides that were shown to be variable by Sanger traces were not called by the array analysis.

Several amino acid changing mutations were detected by the arrays (Table 3, more details and list

of silent mutations found is available upon request). Among these, mutations in the TP53 gene were most frequently detected, followed by mutations in the RET proto-oncogene. Additional coding mutations were found in PTEN, CDKN2A, NRAS and KRAS2. Importantly, we discovered two mutations in the MET proto-oncogene that have been shown to be mutated in a small fraction of NSCLC during the preparation of this manuscript [12]. One of these mutations (A1311G) leads to the substitution of N375 with serine in the sema domain of the protein, and the other (C3162T) to the substitution of T1010 with isoleucine. The latter mutation (T1010I) was recently discovered as a somatic mutation in small-cell lung cancer (SCLC), and, furthermore, it was shown to transform IL-3 dependent BA/F3 cells [13]. Thus, this mutation might also be involved in malignant transformation in NSCLC.

Neither of the two mutations appeared to be somatic in our dataset based on resequencing array results from matched normal DNA. To determine the distribution of these genotypes in tumor and normal cells, we performed MALDI-TOF genotyping. Genotyping success rate was >98% for both assays. The N375S mutation was found in 22 lung tumor specimens (12 adenocarcinomas, 5 squamous-cell carcinomas and 5 small-cell carcinomas) as well as

Table 3

Amino acid changing mutations detected in the 20 NSCLC samples included in this study

Gene	Refseq	Nucleotide change	Amino acid change	Origin ^a	Heterozygous/ homozygous
CDKN2A	NM_000077	G654A	A148T	Germline	2/0 ^b
CDKN2A	NM_000077	T556A	V115E	Germline	1/0
KRAS2	NM_004985	G216A	G12D	Somatic	2/1
KRAS2	NM_004985	G216T	G12V	Somatic	1/0
MET	NM_000245	A1311G	N375S	Germline	3/0
MET	NM_000245	C2646T	P814S	Germline	1/0
MET	NM_000245	C3162T	T1010I	Germline	1/0
NRAS	NM_002524	A435T	Q61L	Germline	1/0
PTEN	NM_000314	G1266A	A79T	Germline	1/0
RET	NM_020630	G1645A	D489N	Germline	1/0
RET	NM_020630	G2251A	G691S	Germline	8/0
RET	NM_020630	C3124T	R982C	Germline	1/0
TP53	NM_000546	G1075T	C275F	Somatic	1/0
TP53	NM_000546	G984T	G245C	Somatic	1/0
TP53	NM_000546	G714T	R158L	Somatic	2/0
TP53	NM_000546	G775A	R175H	Somatic	1/0
TP53	NM_000546	G997T	R249M	Somatic	1/0
TP53	NM_000546	C1167T	R306Stop	Somatic	1/0
TP53	NM_000546	G466C	R72P	Germline	4/1
TP53	NM_000546	G>T	splice-site	Somatic	1/0
TP53	NM_000546	G1065T	V272L	Somatic	1/0
TP53	NM_000546	A739G	Y163C	Somatic	1/0

^a Determined using results from resequencing array analysis of matched normal DNA.

^b Numbers refer to the total number of mutations found (from both resequencing arrays and dideoxy traces). All mutations were scored in at least two independent experiments.

in 24 healthy individuals (9 Caucasians and 15 Asians), indicating that this is a common germline polymorphism rather than a cancer-causing mutation. We detected the T1010I mutation in 5 lung cancer samples (2 adenocarcinomas and 3 small-cell lung cancers), but we did not find this genotype in any of the 180 healthy individuals. Although this difference in prevalence of this mutation in tumors (5/361 tumors vs. 0/180 normals) was not statistically significant ($P=0.132$), the fact that we did not detect this mutation in any normal DNA sample may indicate a role in malignant transformation in lung cancer. Additionally, none of the samples that harbored T1010I showed a mutation in the EGFR kinase domain (data not shown), suggesting that at least in these samples, these mutations are mutually exclusive. A genome-wide search for alterations in gene copy numbers and loss of heterozygosity (LOH) using 100K-SNP arrays has been performed by Zhao et al. [14] for a large fraction of the 20 NSCLC samples included in this study, including the patient with the

T1010I mutation. The SNP array analysis did not reveal any aberrant copy numbers at the MET locus for any of the overlapping samples.

Our results show that resequencing microarrays can be used as a tool for cancer mutation detection and discovery. Furthermore, this is the first report of the use of GeneChip CustomSeq Resequencing arrays for multi-gene analyses in a cancer study, although similar array-based resequencing technologies using two-color fluorescence analysis of co-hybridized normal and sample DNA with different fluorescent labeling have been used to investigate exons from multiple genes previously [15]. The overall performance of the platform is comparable to traditional Sanger-based sequencing with a very high concordance rate (327,132 out of 327,137 bases called consistently; >99.99% concordance), and given the potential to significantly increase probe density and lower production costs, microarrays should be considered as an alternative technology for gene mutation detection and discovery in the future, for

instance as tools to screen for mutations predictive of response to anticancer drugs, or to study human diseases of complex genetic etiology in general.

Furthermore, we have found the transforming MET mutation T1010I in NSCLC to be present in a small fraction of lung tumors. Since MET inhibitors are currently being evaluated in lung cancer, it is tempting to speculate that they might prove beneficial in a subset of lung tumors with activated MET tyrosine kinase.

Acknowledgements

We are indebted to Dr Matthew Meyerson for continuous support. We thank Ralph deBiasi and Timothy Peck for technical assistance, Daniel Richter for help with bioinformatics and Liuda Ziaugra for help with MALDI-TOF genotyping. This work was supported by: Claudia Adams Barr Foundation and Novartis Foundation to Matthew Meyerson and William Sellers, and the Deutsche Krebshilfe through a Mildred-Scheel fellowship for Cancer Research to RKT.

References

- [1] J.G. Paez, P.A. Janne, J.C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F.J. Kaye, et al., EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy, *Science* 304 (2004) 1497–1500.
- [2] J. Jiang, J.G. Paez, J.C. Lee, R. Bo, R.M. Stone, D.J. DeAngelo, I. Galinsky, B.M. Wolpin, et al., Identifying and characterizing a novel activating mutation of the FLT3 tyrosine kinase in AML, *Blood* 104 (2004) 1855–1858.
- [3] H. Davies, G.R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, J. Teague, H. Woffendin, et al., Mutations of the BRAF gene in human cancer, *Nature* 417 (2002) 949–954.
- [4] K. Naoki, T.H. Chen, W.G. Richards, D.J. Sugarbaker, M. Meyerson, Missense mutations of the BRAF gene in human lung adenocarcinoma, *Cancer Res.* 62 (2002) 7001–7003.
- [5] Y. Samuels, Z. Wang, A. Bardelli, N. Silliman, J. Ptak, S. Szabo, H. Yan, A. Gazdar, et al., High frequency of mutations of the PIK3CA gene in human cancers, *Science* 304 (2004) 554.
- [6] P. Stephens, C. Hunter, G. Bignell, S. Edkins, H. Davies, J. Teague, C. Stevens, S. O'Meara, et al., Lung cancer: intragenic ERBB2 kinase mutations in tumours, *Nature* 431 (2004) 525–526.
- [7] M. Chee, R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, et al., Accessing genetic information with high-density DNA arrays, *Science* 274 (1996) 610–614.
- [8] J.A. Warrington, N.A. Shah, X. Chen, M. Janis, C. Liu, S. Kondapalli, V. Reyes, M.P. Savage, et al., New developments in high-throughput resequencing and variation detection using high density microarrays, *Hum Mutat.* 19 (2002) 402–409.
- [9] J.G. Paez, M. Lin, R. Beroukhim, J.C. Lee, X. Zhao, D.J. Richter, S. Gabriel, P. Herman, et al., Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification, *Nucleic Acids Res.* 32 (2004) e71.
- [10] D.J. Cutler, M.E. Zwick, M.M. Carrasquillo, C.T. Yohn, K.P. Tobin, C. Kashuk, D.J. Mathews, N.A. Shah, et al., High-throughput variation detection and genotyping using microarrays, *Genome Res.* 11 (2001) 1913–1925.
- [11] S.B. Gabriel, S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, et al., The structure of haplotype blocks in the human genome, *Science* 296 (2002) 2225–2229.
- [12] P.C. Ma, R. Jagadeeswaran, S. Jagadeesh, M.S. Tretiakova, V. Nallasura, E.A. Fox, M. Hansen, E. Schaefer, et al., Functional expression and mutations of c-Met and its therapeutic inhibition with SU11274 and small interfering RNA in non-small cell lung cancer, *Cancer Res.* 65 (2005) 1479–1488.
- [13] P.C. Ma, T. Kijima, G. Maulik, E.A. Fox, M. Sattler, J.D. Griffin, B.E. Johnson, R. Salgia, c-MET mutational analysis in small cell lung cancer: novel juxtamembrane domain mutations regulating cytoskeletal functions, *Cancer Res.* 63 (2003) 6272–6281.
- [14] X. Zhao, B.A. Weir, T. LaFramboise, M. Lin, R. Beroukhim, L. Garraway, J. Beheshti, J.C. Lee, et al., Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis, *Cancer Res.* 65 (2005) 5561–5570.
- [15] S.M. Lipkin, L.S. Rozek, G. Rennert, W. Yang, P.C. Chen, J. Hacia, N. Hunt, B. Shin, et al., The MLH1 D132H variant is associated with susceptibility to sporadic colorectal cancer, *Nat Genet.* 36 (2004) 694–699.