Short Communication   **Open Access**

Targeting Cancer Related Genes by Multiplex PCR Followed by High Throughput Parallel Sequencing

Vollbrecht C[1]*, Mairinger F[2], Schweighofer CD[3], Heukamp L[1], Merkelbach-Bruse S[1], Büttner R[1] and Margarete O[1]*

[1]*Institute of Pathology, University Hospital Cologne, Germany*
[2]*Institute of Pathology, University Hospital Essen, University of Duisburg-Essen, Germany*
[3]*Medical Clinic I, University Hospital Cologne, Germany*

Abstract

The detection of a wide range of genomic alterations plays an important role in the diagnostics improving individual therapeutic approaches of cancer patients. Technologies that help to identify therapeutic relevant targets in tumor samples are a major factor on the way to a personalized medicine. The number of predictive and prognostic markers that influence the therapeutic outcome is continuously increasing. Therefore, parallel sequencing also named next generation sequencing (NGS), allowing the simultaneous analysis of numerous cancer related hotspots in many patients starting from a limited amount of DNA, is urgently needed to be established in cancer diagnostics. Different methods of target library preparation are commercially available and offer the opportunity to sequence tumor relevant hotspot mutations in a panel of genes.

In the present study, a multiplex PCR approach, targeting a lung cancer and a leukemia gene panel, each consisting of 20 disease and therapy relevant genes, was investigated. Twelve formalin fixed and paraffin embedded lung tumors and twelve native Chronic Lymphocytic Leukemia (CLL) samples, respectively, were analyzed. Samples were sequenced on the MiSeq sequencer platform.

The results showed a very high quality of each run. In spite of the low DNA input, each multiplex approach allowed a simultaneous analysis of 20 genes, in total covered by around 1,000 amplicons, in up to twelve cancer samples.

Thus, the application of NGS on amplicon targets revealed an excellent performance in detecting a wide range of genetic alterations, combined with a high sensitivity.

Introduction

With the discovery of biomarkers and the development of small molecule inhibitors for an individual treatment of cancer patients, the determination of the mutational status of these cancer patients becomes more and more important [1,2]. Thus, the number of predictive and prognostic markers for each cancer entity is steadily growing [2]. In order to meet the novel diagnostic requirements in molecular pathology, technical approaches with a high sensitivity and high capacity have to be established [1,3]. In particular, a more cost- and time efficient procedure than Sanger sequencing is needed in molecular diagnostics of different cancer subtypes, such as lung cancer or leukemia, for which comprehensive panels of potentially aberrant genes are known and of diagnostic interest.

Parallel sequencing, also known as Next Generation Sequencing (NGS) has been recently established and is currently the hottest topic in the field of human research. These new technical approaches are significantly more sensitive than conventional techniques used in clinical practice. They allow the mutational analysis of multiple genes starting from a limited amount of DNA [4-6].

Due to the demands to analyze many different genes per tumor sample, different approaches have been designed to assess tumor relevant hotspot mutations in a panel of genes.

Parallel sequencing needs the generation of a target specific library covering diagnostic relevant loci. This first step can be carried out by means of hybridization capture or by template-specific multiplex PCR [7,8].

Customized capture probe panels representing the genes of interest require at least 200 ng of DNA input [3,9-11]. If low amounts of DNA are available, a whole genome amplification step has to be performed before subsequent capture hybridization [12]. Alternatively, multiplex target amplification by PCR can be applied, allowing a minimum DNA input as low as 10 ng [13]. Due to the low quantity and accessibility of DNA, extracted from formalin-fixed embedded (FFPE) tissue, multiplex PCR approaches are the preferred library construction method in NGS [6].

In the present study, a multiplex PCR-based NGS approach using primers and protocols developed by Qiagen, each targeting 20 genes known to be potentially mutated in lung cancer or leukemia was investigated on FFPE lung tumor and primary chronic lymphocytic leukemia (CLL) specimen.

*Corresponding author: Margarete Odenthal, Institute of Pathology, University Hospital Cologne Kerpener Str. 62, 50937 Cologne, Germany, Tel: +49 (0) 221 478 6367; E-mail: m.odenthal@uni-koeln.de

Claudia Vollbrecht, Institute of Pathology, University Hospital Cologne, Kerpener Str. 62, 50937 Cologne, Germany, Tel: +49 (0) 221 478 88936; E-mail: claudia.vollbrecht@uk-koeln.de

Component	Volume [µl]	
Nuclease-free Water	7.4	
Primer (Fwd), 10nM (5' ATG GAT GCC AAG GAG TTC GAA CC 3')	0.8	19µl/well
Primer (Rev), 10nM (5' GCC ATA ATT ACC TCC TCA GGC AC 3')	0.8	
iScript™ One-Step RT-PCR Kit with SYBR® Green (Bio-Rad Laboratories Inc.,Hercules, CA, USA)	10.0	
Total	1.9	
	+	
gDNA	1.0	

Table 1: Components for *HFE* qPCR.

Stage	Temperature	Time
Hold	94°C	5 min
55 Cycles	94°C	30 sec
	60°C	30 sec
	72°C	30 sec
Plate read and melting curve		

Table 2: Run parameter of HFE qPCR.

Methods

NGS library construction

Twelve DNA samples from routinely processed and macro dissected FFPE lung tumor samples as well as twelve DNA samples from B-lymphocyte from patients with CLL isolated from EDTA-peripheral blood were obtained. Mutation status of *KRAS* exon 2 and 3 or the *EGFR* mutations in exon 18, 19 and 21 were determined previous either by conventional Sanger or by 454 sequencing.

Mutation hotspots were amplified by multiplex PCR using the Gene Read DNA seq Lung Cancer Gene Panel and Gene Read DNA seq Leukemia Gene Panel (Qiagen, Hilden, Germany), respectively. NGS sample preparation was performed according to the manufacturer instructions (GeneRead DNAseq Gene Panel Handbook 11/2012, Appendix B, Qiagen, Hilden, Germany; NEBNext DNA Library Prep Master Mix Set for Illumina, New England BioLabs Inc. Ipswich, MA, USA). For normalization of an equal input of DNA, a qPCR for *HFE* gene was performed using i Script™ One-Step RT-PCR Kit with SYBR® Green (Bio-Rad Laboratories Inc., Hercules, CA, USA) (Tables 1 and 2). A total of 80 ng of purified genomic DNA was then amplified in four separate PCR reactions per sample and purified afterwards. Primer pools contained more than thousand primer sets covering all exons of the following 20 leukemia related genes: *ABL1, ASXL1, CDKN2A, CEBPA, FLT3, GATA1, GATA2, IDH1, IDH2, JAK2, KIT, KRAS, MPL, NPM1, NRAS, PTPN11, RUNX1, TET2, TP53, WT1* and the following lung cancer related genes: *AKT, ALK, BRAF, CDKN2A, CTNNB1, EGFR, ERBB2, HRAS, KIT, KRAS, MET, MTOR, NRAS, PDGFRA, PIK3CA, PTEN, PTGS2, RB1, STK11, TP53.*

Subsequently, PCR amplicons of each sample were pooled and purified with the Agencourt AMPure XP Beads (Beckman Coulter, Brea, CA, USA). After adapter ligation (NEBNext Multiplex Oligos for Illumina, Index Primer 1-12; New England BioLabs Inc.), purification and size selection of the fragments was performed with Agencourt AMPure XP Beads (Beckman Coulter). A final PCR was performed to amplify adapter-carrying fragments and amplicons were re-purified with Agencourt AMPure XP Beads (Beckman Coulter).

Finally, the concentration of the amplicon library was determined according to the manufacturer's protocol by qPCR (Gene Read DNA seq Library Quant Array, Protocol 3, Qiagen). Afterwards, samples were diluted to a final concentration of 2nM and pooled (sample library pool).

Illumina MiSeq sequencing

In order to denaturize the DNA, 0.2N NaOH was added to the sample pool followed by incubation for five minutes. A PhiX spike-in control (Illumina, San Diego, CA, USA) was denaturized in the same manner. Both, the sample pool and the PhiX control were then diluted to a final concentration of 8pM (run 1) and 10pM (run 2 and 3), respectively. 1% (run 1) and 10% (run 2 and 3) PhiX was added to the sample pool. Next, 600µl of the finalized sample pool was applied to the MiSeq cartridge (llumina) according to the manufacturer´s instructions. Sequencing was performed using a MiSeq instrument (llumina) and the v2 chemistry as recommended by Illumina.

Data analysis

Fastq.gz files generated by the MiSeq Reporter program (Illumina) were uploaded into cloud space for automatically variant analysis (Qiagen). Somatic analysis and paired-end read mode was chosen. The primary, so called preliminary alignment of the raw data, including the full read set, was done using Bowtie2. This first alignment was followed by trimming of primer sequences and quality filtering, which excluded reads with an untrimmed length of less than 45 bp. In the final alignment the trimmed reads were mapped against the reference genome. Alignment parameters were identical to those used in the preliminary alignment. The results of the final read alignment were used further for the variant calling, which was performed with the GATK Lite version 2.1-8 (GATK Unified Genotyper program, Broad Institute Cambridge, USA). Variant filtering was done automatically in two steps: first, variants that failed some of the thresholds for variant calling were marked, and second, single nucleotide polymorphisms (SNP) with less than 4% as well as insertions and deletions (indels) with less than 20% variant allele frequencies were removed.

Results

Sequencing quality

The three runs which contained 12 (run 1) and 6 samples (run 2 and 3) each, produced 12.97×10^6, 14.74×10^6 and 22.16×10^6 reads. Equimolar sample pooling according to the final library quantification by qPCR resulted in a low deviation between samples (Figure 1). All three runs showed a very good read quality with Q30 scores greater than 83% (Table 3).
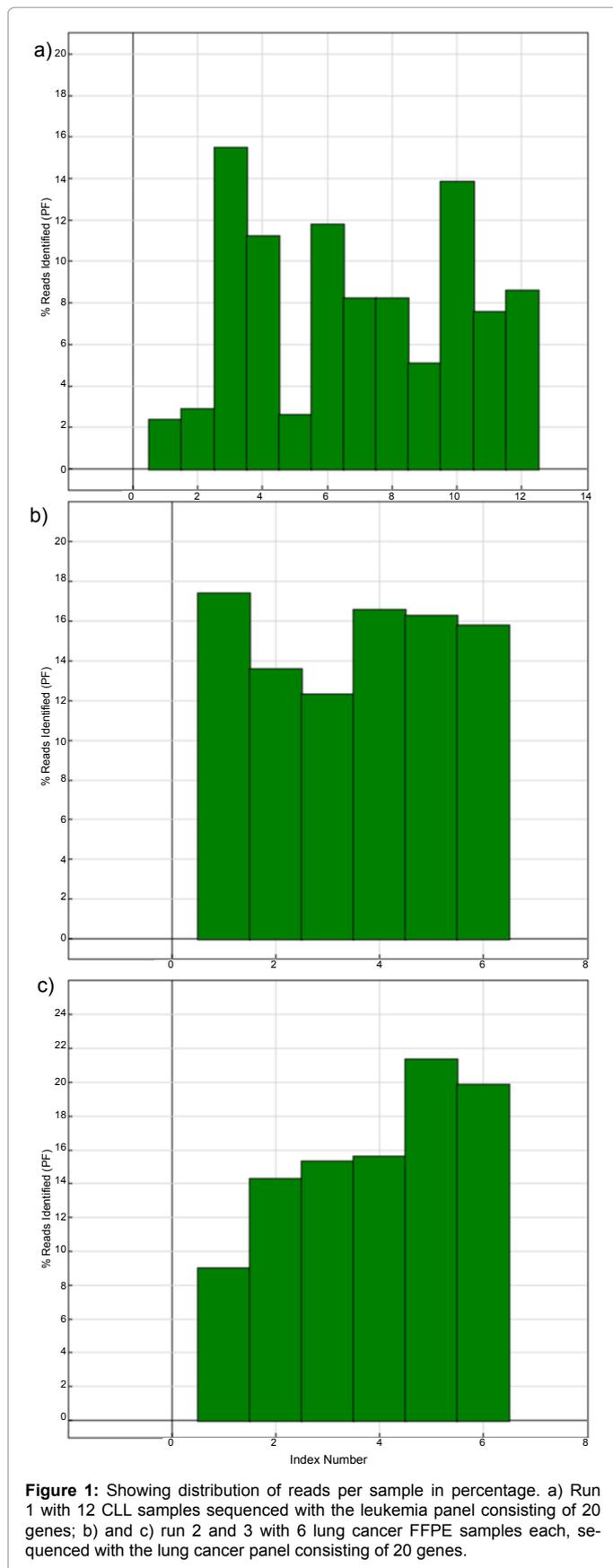
The coverage statistics were comparable between both runs (Table 3). The mean coverage per gene in each run was 11,028, 12,622 and 19,159 reads, respectively.

No difference was observed concerning the quality and quantity of the constructed libraries, derived from native B-cell DNA in comparison to FFPE lung cancer DNA. However, three genes in the leukemia panel were clearly under-represented, namely *CEBPA*, *RUNX1* and *GATA2* (Figure 2). These genes showed a mean coverage of 2120, 1653 and 2744 reads per gene and run.

In the lung panel, the genes *HRAS*, *STK11* and *RB1* had a significant lower median coverage than the other analyzed genes (Figure 2).

Variant detection

83% of the detected variants in the leukemia panel were point

Figure 1: Showing distribution of reads per sample in percentage. a) Run 1 with 12 CLL samples sequenced with the leukemia panel consisting of 20 genes; b) and c) run 2 and 3 with 6 lung cancer FFPE samples each, sequenced with the lung cancer panel consisting of 20 genes.

mutations. By far, most variants were found in the genes *TET2* (16%), *FLT3* (15%), *KRAS* (13%), *ASXL1* (10%), *TP53* (10%) and *WT1* (8%).

In the lung panel, mutations occurred most frequently in the genes *ALK* (13%), *KRAS* (11%), *EGFR* (10%), *PDGFRA* (10%), *TP53* (7%), *NRAS* (6%), *MET (6%)* and *mTOR* (5%). 89% of the detected variants were point mutations. Most indels were found in *BRAF* (72% of *BRAF* variants).

Concordance with data shown by conventional sanger and 454 sequencing

Interestingly, we found two discrepancies in sample 3 and 10, in which an *EGFR* exon 19 deletion was demonstrated by previous 454 sequencing or Sanger sequencing, whereas the result of MiSeq sequencing was a wild type (Table 4). However, data interpretation by IGV (Integrative Genomic Viewer, Broad Institute) clearly proved the mutation in the sequence derived from the MiSeq system (Figure 3).

Discussion

An increasing number of anticancer therapeutic strategies focus on specific hotspot mutations in genes, coding for proteins that are expressed by many different tumor types [2]. Current technologies to determine the mutation status of patients allow only the simultaneous analysis of one or a few hotspots [2]. The NGS methods can overcome these limitations and facilitate the process enabling the stratification of a high number of genes for example by cancer panels.

In the present study, we showed that targeted NGS using Qiagen Gene Read panels provides information about multiple genes starting from a limited amount of tumor DNA. In spite of the low DNA input, the panels allowed a simultaneous analysis of 20 genes covered by around 1,000 amplicons in up to twelve samples.

All three runs showed very good read quality with Q30 Scores greater than 83%. Even if the libraries are pooled according to the results of a final qPCR, the runs can result in an unequal distribution of reads per sample. Furthermore, the runs showed a heterogeneous median coverage per gene that is reflected by the clear underrepresentation of some genes in both panels (in the leukemia panel: *RUNX1*, *GATA2*, *CEBPA;* in the lung panel: *HRAS*, *STK11*, *RB1*). This fact can be potentially due to the amplification of the targets by multiplex PCR which can generate a bias. Some targets are preferentially amplified whereas others seem to be not, probably depending on the targeted sequences and primer annealing efficiency. The fact that the primer sets produce overlapping PCR products and the non-adjacent primer sets are divided into four tubes can help to minimizes the risk of a heterogeneous coverage of the targets and decreases non-specific amplifications. In order to avoid problems resulting from multiplex PCR amplifications, targets can be tested to be enriched by other methods like probe capture based techniques or single micro-droplet PCR [14].

The approach to sequence tumor relevant genes seems to be interesting especially for translational diagnostic research. In contrast, the analysis of whole genes including also exons which have no diagnostic relevance seems not to be useful for daily routine cancer diagnostics, as it occupies a lot of sequencing capacity [3]. Thus, in diagnostic approaches panels covering hotspot regions of therapeutic relevant targets are preferred and more cost-effective [3].

The Gene Read Panels of Qiagen are very flexible for research approaches as they allow to choose targets out of 127 genes and two different sequencing platforms (at time of accomplishment). This can

	Run 1	**Run 2**	**Run 3**	
Entity	CLL	Lung	Lung	**Average**
Samples	12	6	6	
Panel	Qiagen Leukemia	Qiagen Lung	Qiagen Lung	
Amplicons [n]	1,272	1,337	1,337	
Cluster Density [K/mm²]	665	826	1,312	934
Cluster PF	93.00%	93.30%	89.17%	91.82%
Reads	12.97M	15.80M	24.84M	17.87M
Reads PF	12.17M	14.47M	22.16M	16.27M
Total Yield	3.7Gb	4.5Gb	6.8Gb	5Gb
>= Q30 Score	86.50%	91.00%	84.00%	87.17%
PF… passed filter; Q30 Score… 0.1% chance of wrong base call; M… million; Gb… gigabases				

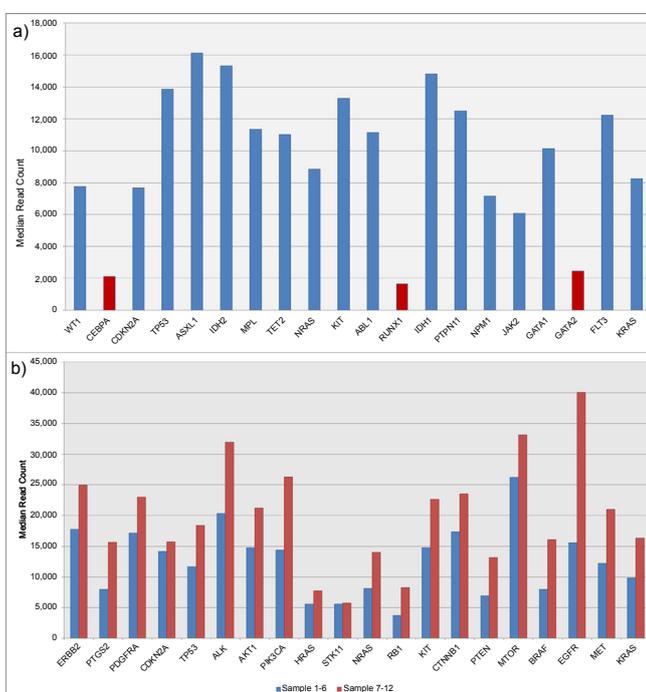**Table 3:** Overview of MiSeq run parameter.

Figure 2: Showing median coverage per gene and run. a) 20 genes of the leukemia panel sequenced with 12 CLL samples; b) 20 genes of the lung panel, marked in blue lung FFPE samples 1-6, marked in red lung FFPE samples 7-12.

Sample	Gene	454 Sequencing		MiSeq	
1	*EGFR*	wt	wt	wt	wt
2	*EGFR*	wt	wt	wt	wt
3	*EGFR*	c.2235_2249del	p.E746_A750del (20%)	wt	wt
4-1	*EGFR*	c.2238_2248del	p.L747_A750del (25%)	c.2236_2244del	p.E746_R748del (24%)
4-2	*EGFR*	c.2238_2248del	p.L747_A750del (25%)	c.2236_2244del	p.E746_R748del (22%)
5-1	*EGFR*	c.2573T>G	p.L858R (27%)	c.2573T>G	p.L858R (20%)
5-2	*EGFR*	c.2573T>G	p.L858R (27%)	c.2573T>G	p.L858R (27%)
6	*EGFR*	c.2573T>G	p.L858R (83%)	c.2573T>G	p.L858R (87%)
		Sanger Sequencing		MiSeq	
7	*KRAS*	c.35G>T	p.G12V	c.35C>A	p.G12V (30%)
8	*KRAS*	c.34G>T	p.G12V	c.34C>A	p.G12C (19%)
9	*KRAS*	c.183T>A	p.Q61H	c.183T>A	p.Q61H (41%)
10	*EGFR*	c.2235_2249del	p.L746_A750del	wt	wt
wt… wild type					

Table 4: Concordance of MiSeq Data and other sequencing methods.
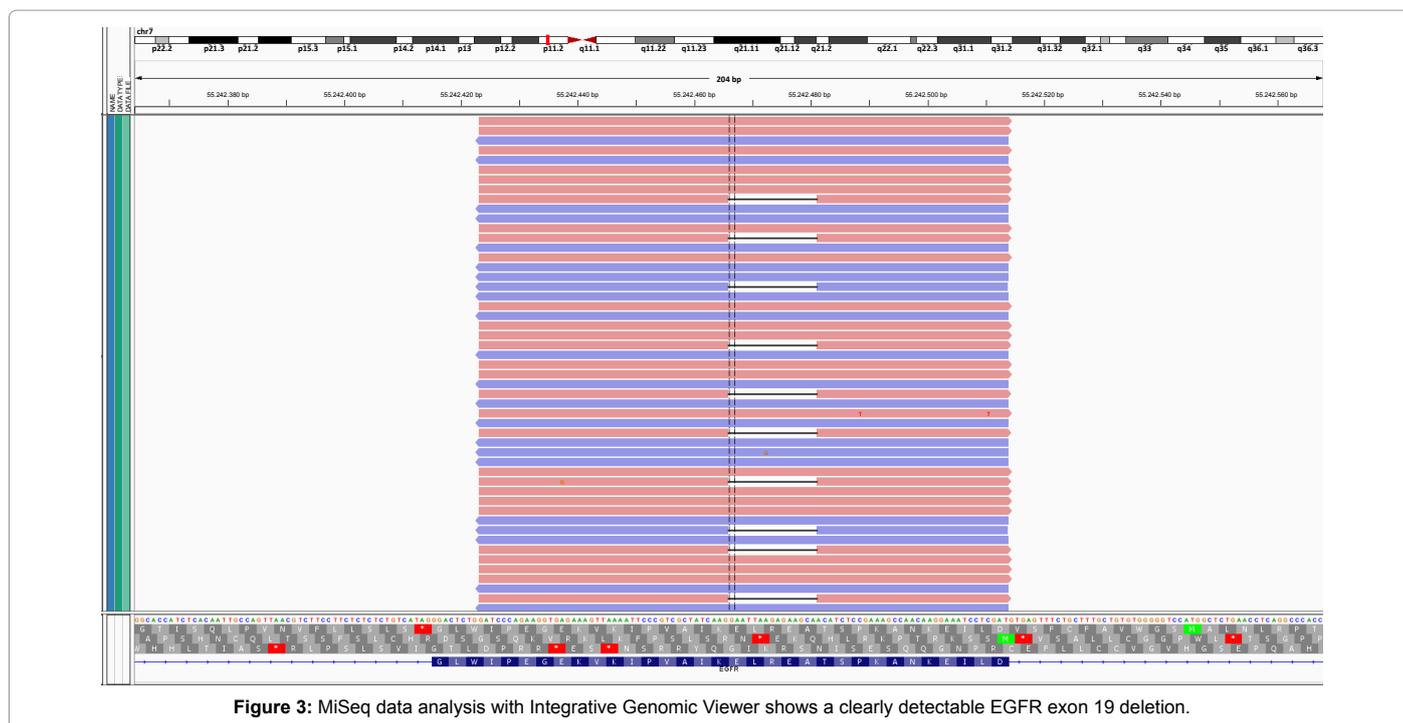
Figure 3: MiSeq data analysis with Integrative Genomic Viewer shows a clearly detectable EGFR exon 19 deletion.

be an advantage for exome sequencing of genetic discovery panels, as the panel design offers the opportunity of sequencing genes without knowledge of their responsibility for a disease [3].

The provided data analysis program helps with the alignment and variant calling process of the data, but still it needs an evaluation process to identify and separate real mutations from background noise and to determine which variants have clinical significance [1,14].

The fact that we could detect two deletions in the IGV but not in the final output results that had been listed in an excel file, seems to be caused by the very stringent variant filtering. The advantage of the stringent variant filtering is the avoidance of false positive variants but it should be mentioned that alignments have to be visually verified with other tools like the IGV (Broad Institute). Especially, when FFPE material is analyzed, one should be aware of the risk of an increased background due to the formalin fixation [14].

Our application of the lung and the leukemia cancer panel in routine pathology needs a validation on larger cohorts of cases [14]. However, hereby we present its performance in detecting a wide range of genetic alterations with a high sensitivity and show that the approach can help to assess tumor-specific therapeutic susceptibility and individual prognosis [5]. The upcoming challenge will be in the reliable identification of an ultimate cancer-specific multigene panel in order to significantly improve the care of cancer patients [5].

References

1. Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 11: 685-696.

2. Mac Conaill LE, Van Hummelen P, Meyerson M, Hahn WC (2011) Clinical implementation of comprehensive strategies to characterize cancer genomes: opportunities and challenges. Cancer Discov 1: 297-311.

3. Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, et al. (2013) Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Hum Mutat 34: 1035-1042.

4. Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52: 413-435.

5. Scarpa A, Sikora K, Fassan M, Rachiglio AM, Cappellesso R, et al. (2013) Molecular typing of lung adenocarcinoma on cytological samples using a multigene next generation sequencing panel. PLoS One 8: e80478.

6. Becker K, Vollbrecht C, Koitzsch U, Koenig K, Fassunke J, et al. (2013) Deep ion sequencing of amplicon adapter ligated libraries: a novel tool in molecular diagnostics of formalin fixed and paraffin embedded tissues. J ClinPathol 66: 803-806.

7. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. ProcNatlAcadSci U S A 106: 19096-19101.

8. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27: 182-189.

9. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD, et al. (2013) Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J Mol Diagn 15: 623-633.

10. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53-59.

11. Seyed F, Grunenewald H, Caruccio N (2009) Optimized library preparation method for next-generation-sequencing. Nature Methods App. Note 6: I-II.

12. ElSharawy A, Warner J, Olson J, Forster M, Schilhabel MB, et al. (2012) Accurate variant detection across non-amplified and whole genome amplified DNA using targeted next generation sequencing. BMC Genomics 13: 500.

13. Koenig K, Heukamp LC. manuscript in preparation.

14. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, et al. (2012) Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol 30: 1033-1036.