Roche

Cancer Research Application Note No. 7

Next-Generation Amplicon-Based Deep Sequencing and Its Application to Characterize Hematological Malignancies

Alexander Kohlmann*, Vera Grossmann*, Susanne Schnittger, Claudia Haferlach, and Torsten Haferlach

MLL Munich Leukemia Laboratory, Munich, Germany
*A.K. and V.G. contributed equally

Correspondence: Dr. rer. nat. Alexander Kohlmann, MLL Munich Leukemia Laboratory GmbH, Max-Lebsche-Platz 31, 81377 München, Germany; Tel: +49 89 990 17 380, Fax: +49 89 990 17 389; e-mail: alexander.kohlmann@mll-online.com

1 Introduction

Recently, next-generation sequencing (NGS) platforms have evolved to provide an accurate and comprehensive means for the detection of molecular mutations.[1;2] Massively parallel pyrosequencing in picoliter-sized wells is an innovative technique that allows the detection of low-abundance oncogene aberrations in complex research samples, even with low tumor content. Moreover, this approach may offer the possibility of monitoring the molecular composition and evolution of tumor subtypes delineating a hierarchical model of pathogenesis.

Here, we used an amplicon-based deep sequencing strategy to demonstrate the power of deep sequencing for the molecular characterization of hematological malignancies.[3] We analyzed 95 cases by applying next-generation sequencing from 454 Life Sciences, a Roche Company, to investigate 7 candidate genes.

For life science research only.
Not for use in diagnostic procedures.

2 Materials and Methods

We analyzed 95 cases of distinct hematological malignancies: acute myeloid leukemia (AML), chronic myelomonocytic leukemia (CMML), essential thrombocythemia (ET), and myelodysplastic syndrome (MDS). All cases were sent to the MLL Munich Leukemia Laboratory for diagnostic assessment between 10/2005 and 09/2008. Genomic DNA research samples in this study were obtained from untreated individuals. DNA was extracted from purified mononuclear cells and diluted to 20 ng/µl for subsequent PCR amplification.

We applied next-generation pyrosequencing using GS FLX System amplicon sequencing to investigate 5 candidate genes at known mutational hotspot regions: *CBL* (exons 8 and 9), *JAK2* (exons 12 and 14), *MPL* (exon 10), *NRAS* (exons 2 and 3), and *KRAS* (exons 2 and 3). Additionally, complete coding regions were analyzed for *RUNX1* (beta isoform) and *TET2*. The genes were selected based on recent data on molecular aberrations in leukemias and myeloproliferative neoplasms.

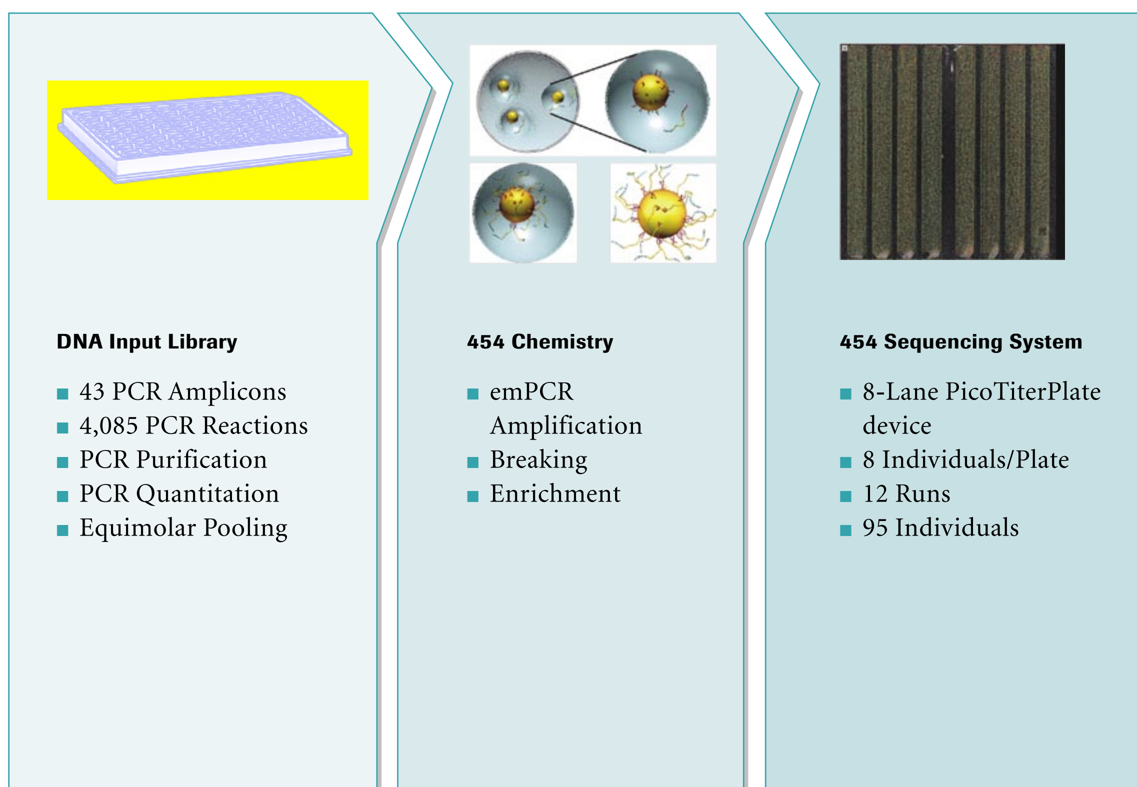

Figure 1: Overview of the next-generation amplicon-based deep sequencing assay. For each individual, a specific PCR amplicon library was prepared and processed using the recommended GS FLX System protocols. Sequencing was performed on an 8-lane PicoTiterPlate device. Each individual was processed on a single sequencing lane. A total cohort of 95 individuals was sequenced in a consecutive series of 12 runs.

For each of the 95 individuals, 43 amplicons were processed for deep sequencing, representing, in total, 4085 individual PCR products (see Figure 1). After generating the 4085 amplicons, all PCR products were purified using the MinElute kit (Qiagen, Hilden, Germany) and quantified using the Quant-iT PicoGreen kit (Invitrogen, Carlsbad, CA). For each individual, the 43 amplicons were pooled in an equimolar ratio – 300 ng for each amplicon – to generate one single library. Figure 2 demonstrates a representative Bioanalyzer quality result of such an individual-specific pool of PCR amplicons. Subsequent 454 emulsion amplification and amplicon sequencing was performed according to the manufacturer's recommendations (Roche Applied Science, Penzberg, Germany). Each separate individual library was sequenced in one region of an 8-lane PicoTiterPlate device. Sequencing data was analyzed using the GS Amplicon Variant Analyzer software. For the detection of variances, filters were set to display sequence variances occurring in >3% of bidirectional reads per amplicon in at least one individual. All detected variances were first compared with published SNP data (dbSNP; http://www.ncbi.nlm.nih.gov/projects/SNP). Variances within introns were not scored. The melting curve analyses using the LightCycler® Instrument to detect mutations in *JAK2* were performed as described earlier by Schnittger *et al* [5].

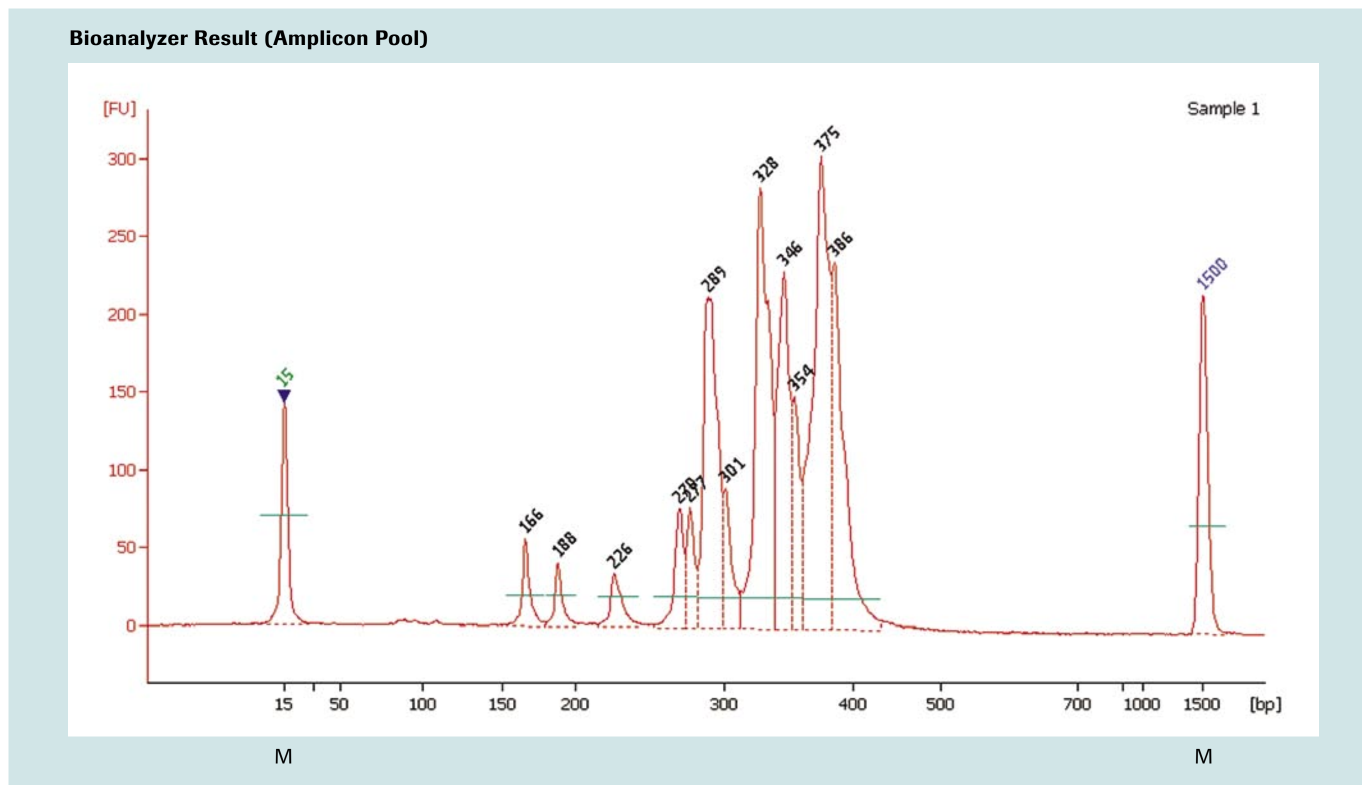

Figure 2: PCR amplicon library assessment. For each individual, 43 distinct PCR amplicons were generated and pooled in an equimolar ratio to cover genomic DNA regions for mutation screening. A representative Bioanalyzer DNA 1000 chip result indicated the distinct amplicon peaks for the individual-specific pool of PCR amplicons. A lower marker (15 bp) and an upper marker (1500 bp) indicate the analysis range of the DNA 1000 chip.

3 Results and Discussion

On average, we obtained per individual approximately 80,000 Key Pass Wells, yielding in median 40,087 high-quality sequencing reads (Passed Filter Wells). As shown in Figure 3, an exemplary distribution of amplicon reads for one individual clearly resembles the input library of pooled PCR molecules. The median length of reads per individual ranged between 242 bp and 248 bp. The median number of base pairs sequenced per individual was 9.24 Mb.

Figure 4 demonstrates the distribution of overall amplicon coverage for the 95 individuals. Box-plots are given for each of the corresponding 43 PCR products representing the amplicons from the 7 target genes *CBL, JAK2, KRAS, MPL, NRAS, RUNX1,* and *TET2*. The median number of generated reads per amplicon was 846 (coverage range: 502-fold to 1977-fold). Importantly, forward and reverse reads were homogeneously distributed (see Figure 5). Therefore, the generated data allowed highly sensitive detection of underlying variances in these cancer genomes.

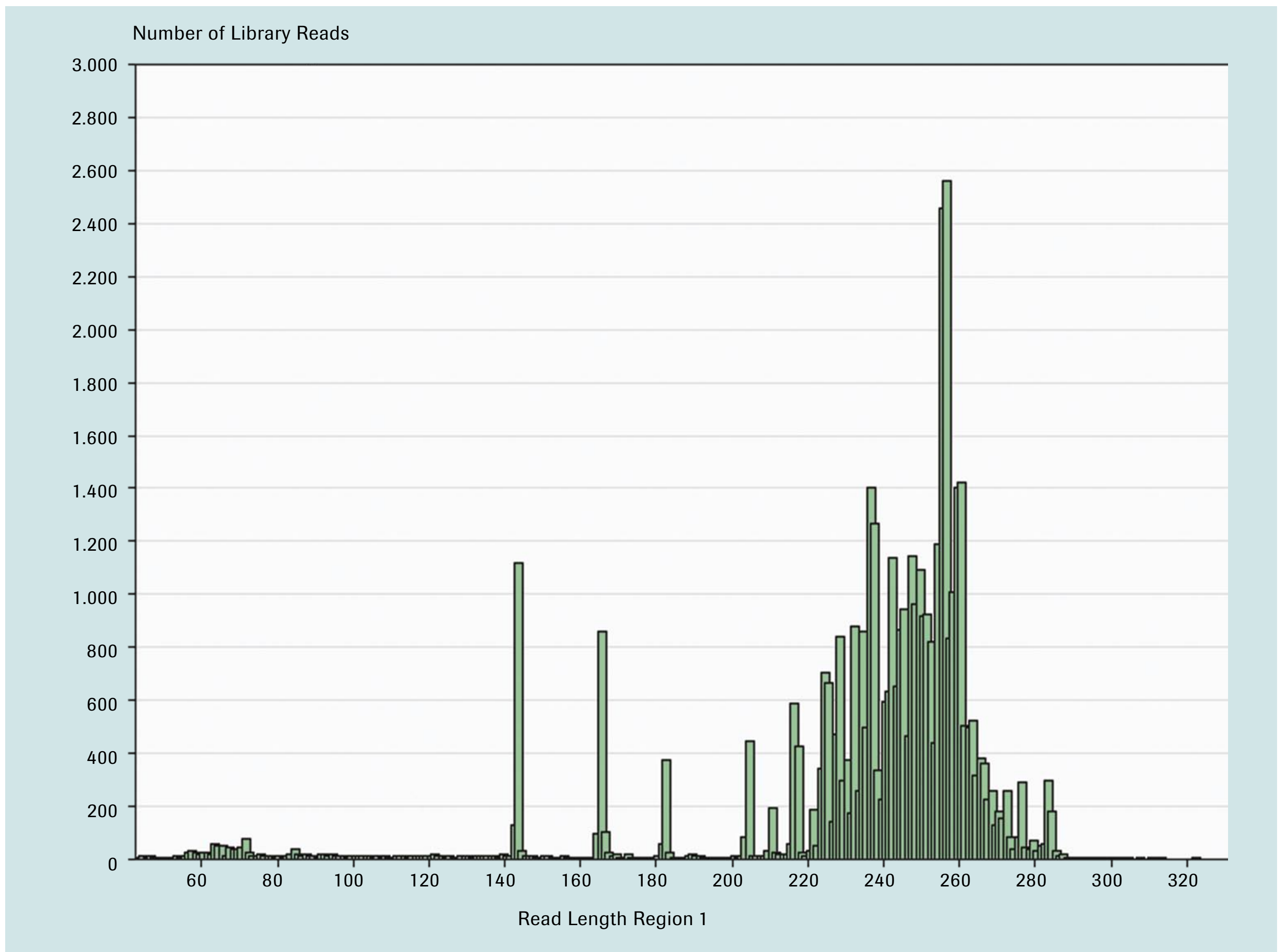

Figure 3: Next-generation amplicon sequencing data output. For an exemplary individual sample, a distribution of the obtained read length (x-axis) and the cumulative number of generated sequence reads for the 43 amplicons (y-axis) is represented.

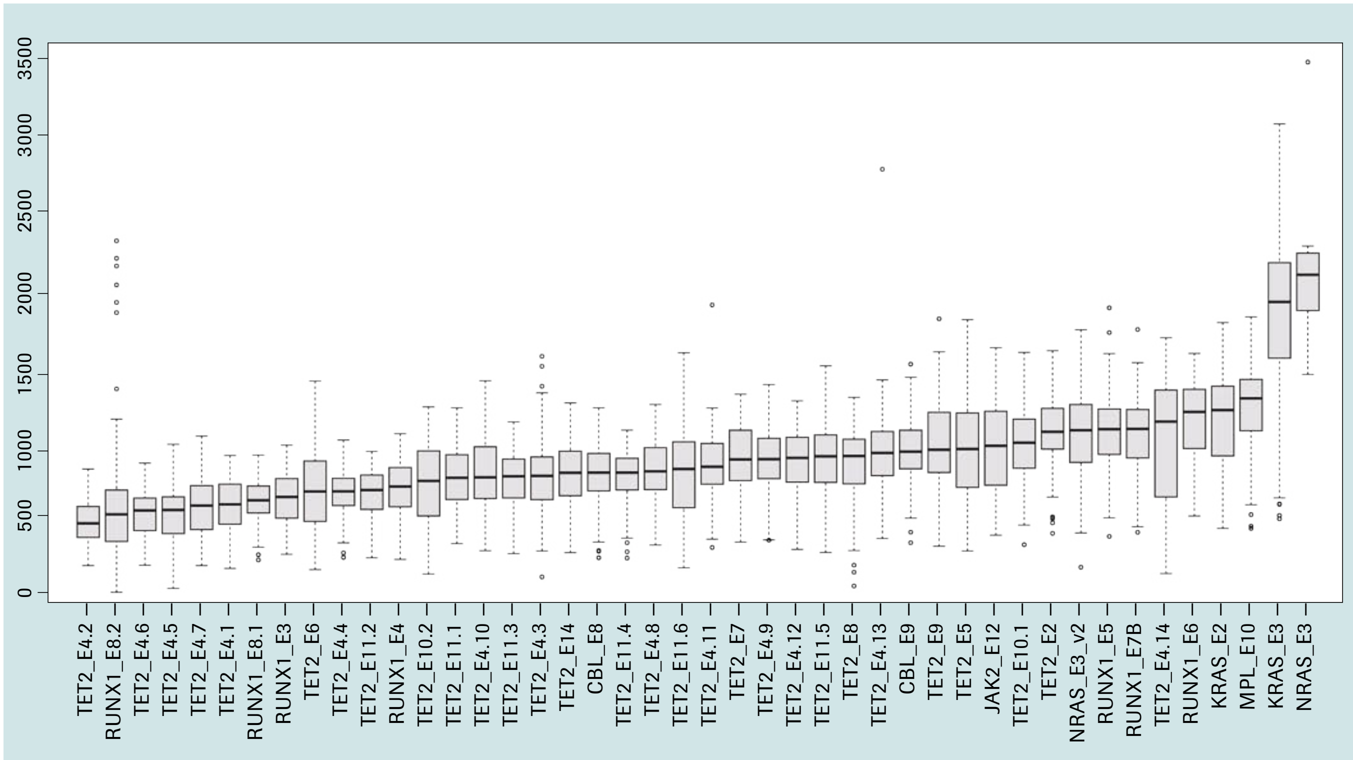

Figure 4: Next-generation amplicon-based deep sequencing coverage for 95 individual samples. For each of the 43 amplicons (x-axis; median coverage in ascending order), the range of generated sequence reads is represented by box-and-whisker plots (y-axis).

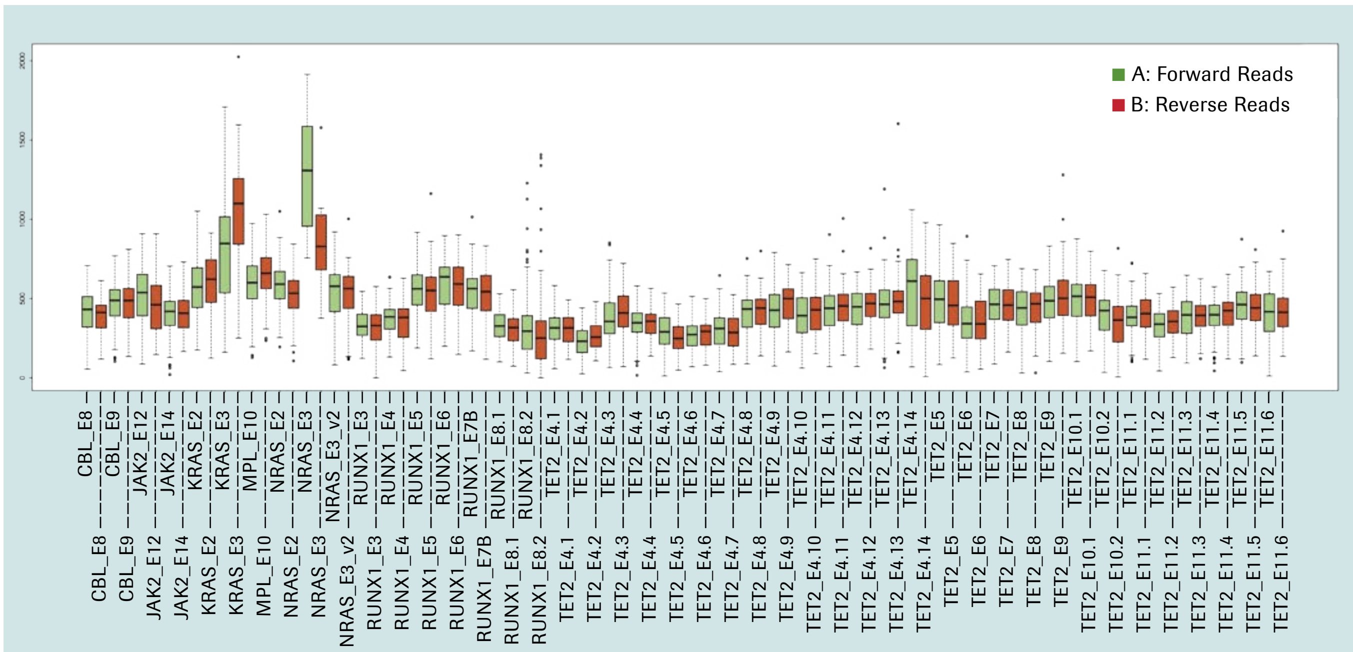

Figure 5: Forward and reverse reads sequencing coverage. For each of the 43 amplicons (x-axis; genes in alphabetic order), the range of generated sequence reads according to forward strand and reverse strand data is represented by box-and-whisker plots (y-axis). A reads are shown in green, and B reads are shown in red.

As demonstrated in Figure 6, amplicon sequencing allowed identification of molecular mutations with high sensitivity. Interestingly, in some cases mutations were observed with a frequency lower than 10%, which is below the currently accepted cut-off value of conventional Sanger-based sequencing. Thus, this method outperforms conventional single-marker mutation screening and reveals novel molecular insights.

Given this specific study design, in median, an 846-fold coverage was generated. Therefore, 3% of reads would represent 25 distinct reads harboring a detected variance. However, it remains to be further investigated what an ideal coverage would be in a clinical setting, and consensus guidelines need to be defined, in particular with respect to the heterogeneity of tumor specimens.

	Reference	Variant	Max	Sample_13	Sample_14	Sample_16	Sample_17	Sample_18	Sample_19	Sample_20	Sample_21	Sample_22
	CBL_E8	86:G/C	4.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	91:A/G	71.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	102:T/C	42.34	0.00	0.00	0.00	0.00	1.39	0.00	0.00	0.00	0.00
	CBL_E8	102:T/G	4.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	130:T/C	83.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	141:T/C	84.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	84.57	0.00
	CBL_E8	177:T/C	34.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	184:A/G	87.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	190:T/G	6.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.30	0.00
	CBL_E8	193:G/A	81.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	202:G/A	32.01	32.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CBL	CBL_E8	214:G/C	52.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	215:G/T	32.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	219:G/T	45.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	221:A/T	53.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E8	222:C/T	29.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E9	54:T/C	30.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E9	55:G/A	10.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E9	61:T/C	48.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E9	67:G/A	60.02	33.54	60.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E9	67:G/T	89.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.57
	CBL_E9	93-110:DEL(18)	5.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	CBL_E9	95:C/T	47.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
JAK2	JAK2_E14	180:G/T	58.89	0.00	0.00	0.00	0.00	0.00	3.06	0.00	4.21	0.00
	KRAS_E2	87:T/A	17.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E2	88:C/A	17.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E2	107:C/T	24.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E2	110:C/T	▾ 44.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KRAS	KRAS_E2	111:C/A	9.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E2	111:C/G	28.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E3	59:C/T	42.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E3	62:G/T	▾ 27.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	KRAS_E3	65:G/A	79.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.84
MPL	MPL_E10	148:G/T	27.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MPL_E10	174:C/T	43.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NRAS_E3_v2	47:G/T	3.52	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NRAS_E3_v2	149:T/G	35.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NRAS	NRAS_E3_v2	150:T/C	14.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NRAS_E3_v2	151:G/T	33.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NRAS_E3_v2	153:C/T	20.02	0.00	1.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NRAS_E3_v2	159:G/A	7.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RUNX1_E5	95:G/A	5.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RUNX1_E5	103:C/T	30.61	30.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RUNX1_E5	104:G/A	47.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RUNX1	RUNX1_E5	112:T/C	▾ 82.26	0.00	0.00	0.00	▾ 82.26	0.00	0.00	0.00	0.00	0.00
	RUNX1_E6	99-112:DEL(14)	46.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RUNX1_E6	122:C/T	34.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RUNX1_E8.2	137:G/C	49.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RUNX1_E8.2	252:G/A	50.29	▾ 0.00	0.00	0.00	0.00	▾ 0.00	0.00	0.00	0.00	0.00

Figure 6: Analysis of molecular mutations. The screenshot depicts a representative output of the GS Amplicon Variant Analyzer software. Individuals are shown in columns and the detected variances are shown in rows. Each row represents a single variance found. The matrix indicates the reads with observed variances for the genes *CBL, JAK2, KRAS, MPL, NRAS* and *RUNX1*. Reads with a percentage <20% (indicating a cut-off for Sanger-based sequencing data) are highlighted by blue circles.

In total, 130/133 mutations detected by NGS were in addition evaluated by either conventional Sanger sequencing or LightCycler® melting curve analyses. In 110/130 analyses (84.6%), concordant results were obtained. In the remaining 20/130 tests (15.4%), discrepancies were observed. Of these 20 cases, in 18 (90.0%) samples, mutational reads were detected with a frequency <10% and were thus below the detection limit of conventional Sanger sequencing.[4] As presented in Figure 7 in an exemplary individual, 454 Life Sciences ultra-deep sequencing detected the characteristic *JAK2* V617F mutation with 1.16% sensitivity.

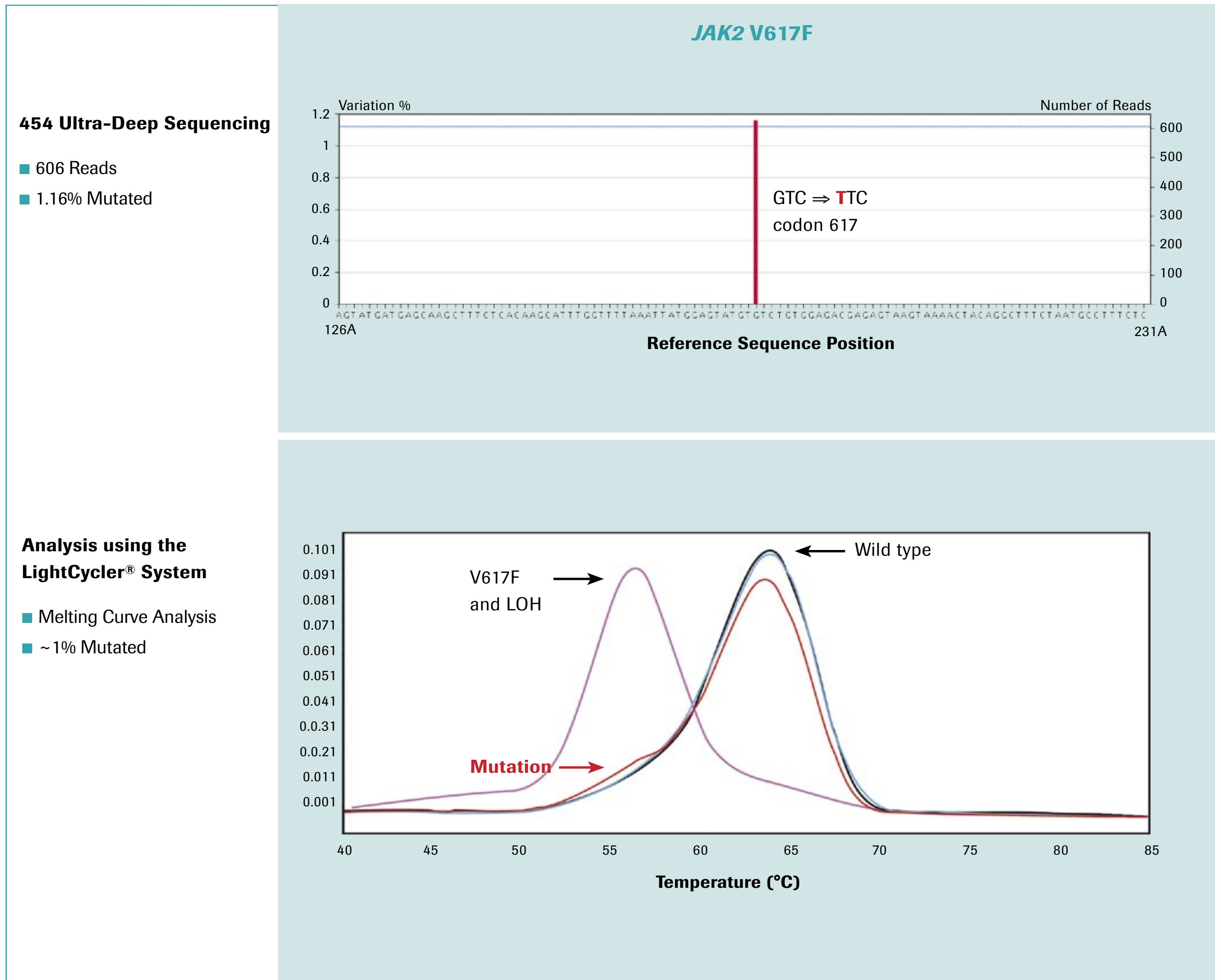

Figure 7: Validation of molecular mutations. The *JAK2* V617F mutation is a characteristic mutation in hematopoietic malignancies. Here, in an exemplary individual, ultra-deep sequencing with the GS FLX System detected the mutation in 1.16% of reads. Data from a sensitive LightCycler® melting curve analysis confirmed the mutation in approximately 1% of cells.[5]

With the recent introduction of the GS FLX Titanium chemistry for amplicon sequencing, it is now possible to increase the amplicon size to close to 400 bp. Importantly, with GS FLX Titanium amplicon sequencing, it is also possible to apply 10-base molecular barcode sequences (so-called MIDs, or multiplex identifiers) incorporated into the gene-specific primer sequences (see Figure 8).

With this assay design, it is possible to sequence different samples in the same run and achieve high coverage of PCR amplicons to provide high sensitivity in detecting molecular mutations (see Figure 9).

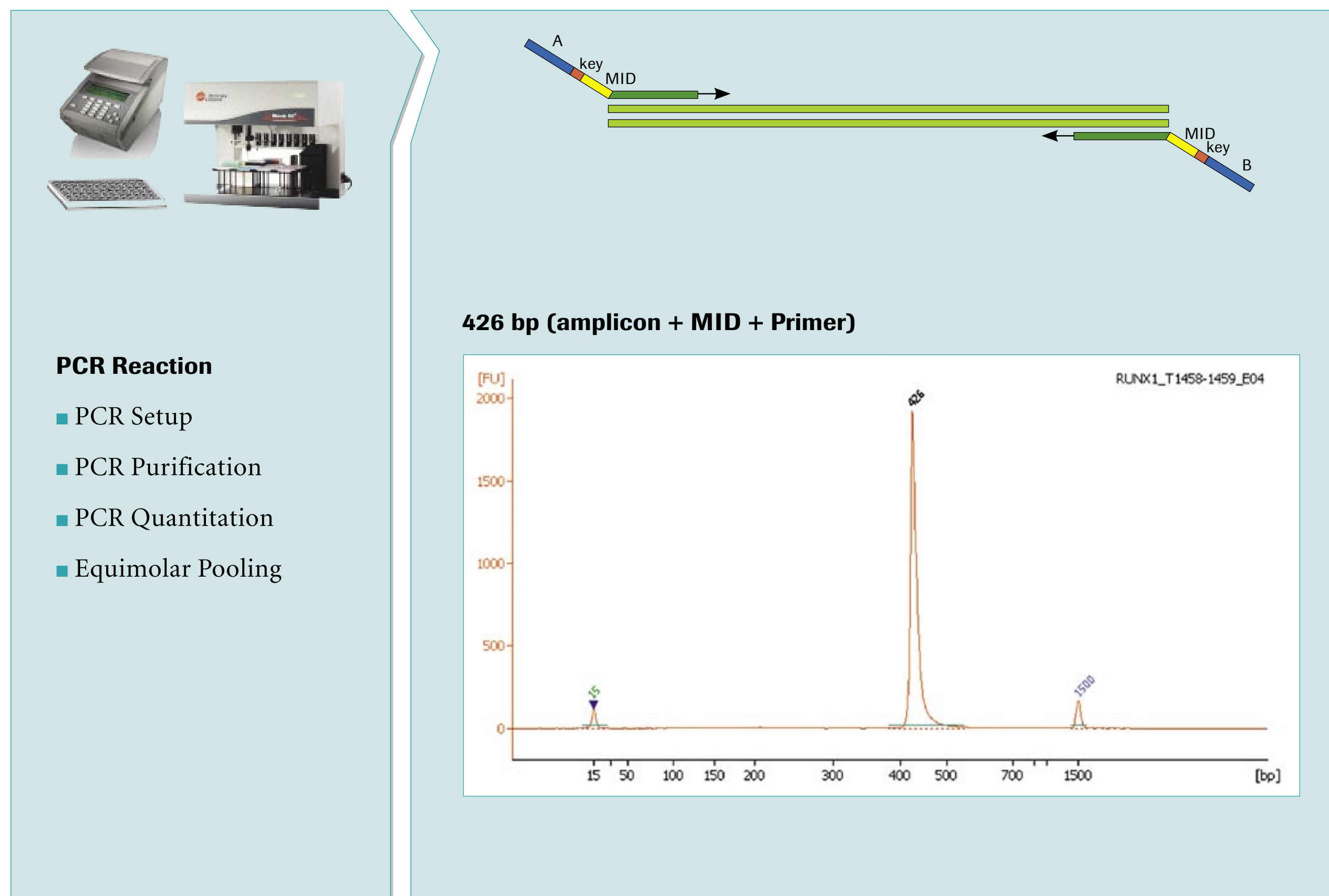

Figure 8: Overview of the amplicon assay workflow based on GS FLX Titanium chemistry. This chemistry allows incorporating both molecular barcodes into primer sequences for multiplexing of indvidual samples per sequencing lane and an increased amplicon size of close to 400 bp for complete bidrectional sequencing.

In conclusion, comprehensive NGS screening using amplicon-based deep sequencing has been successfully applied to genetically characterize cancer genomes. It is anticipated that NGS will evolve as a suitable platform to provide data on molecular mutations with high throughput and accuracy and, as such, will advance into the field of standard molecular diagnostics. It is further expected that with a future increase in read length, it will even be possible to clearly correlate the reads of double mutations to a monoallelic or biallelic status of the mutation.

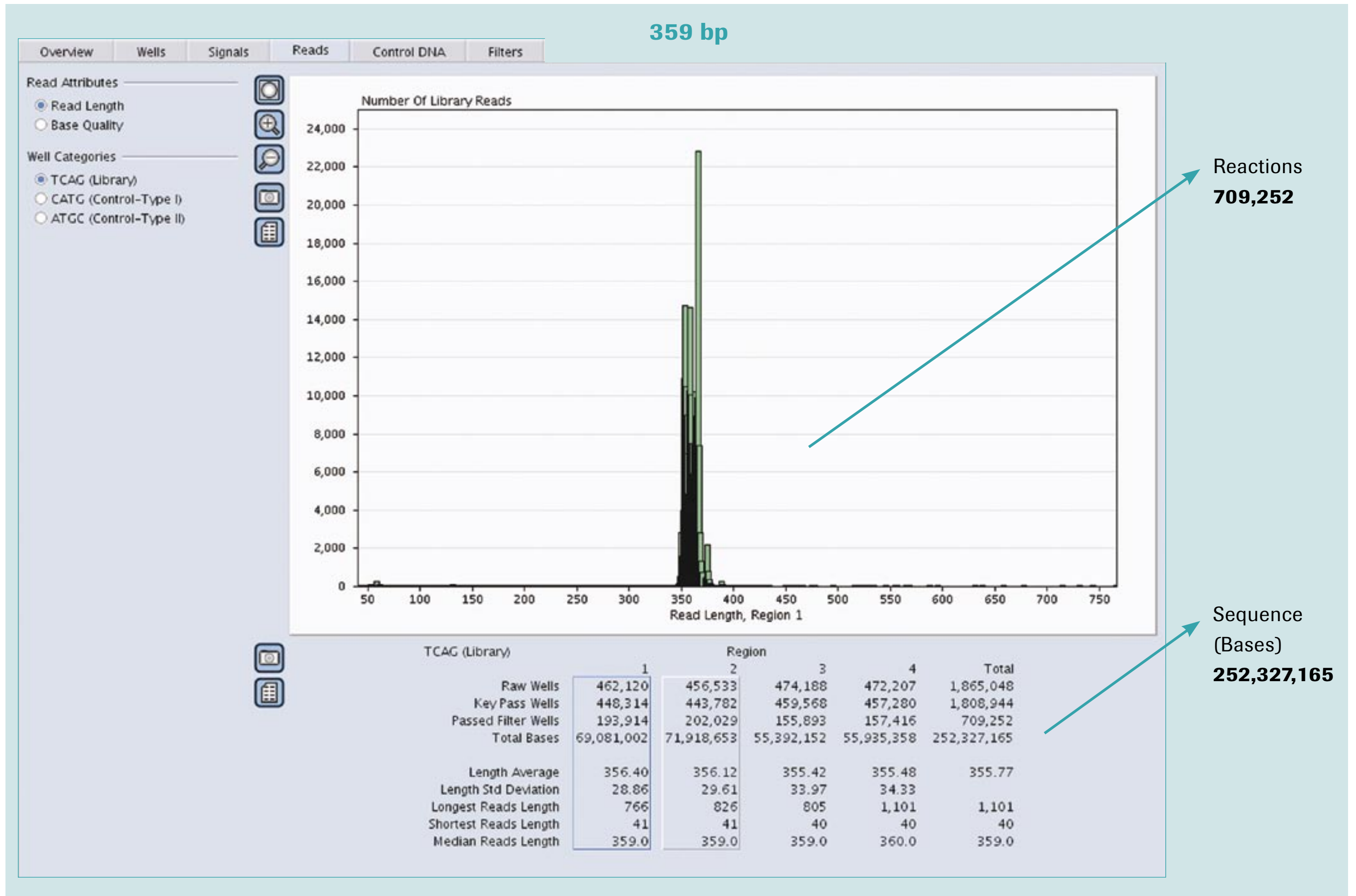

Figure 9: Representative output of an amplicon library processed with GS FLX Titanium chemistry adaptors, including individual-specific molecular barcodes. Here, a 4-lane PicoTiterPlate device is shown, indicating an overall sequencing output of 709,252 Passed Filter Wells. A median peak size of 359 bp indicates the long bidirectional read length of amplicons using the GS FLX Titanium chemistry.

References

1. Thomas RK, Nickerson E, Simons JF *et al.* Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat.Med. 2006;**12:**852-855.

2. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin.Chem. 2009;**55:**641-658.

3. Margulies M, Egholm M, Altman WE *et al.* Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;**437:**376-380.

4. Kohlmann A, Grossmann V, Klein HU *et al.* Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia (CMML) by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. J Clin Oncol. 2010 Aug 20;**28**(24):3858-65. Epub 2010 Jul 19.

5. Schnittger S, Bacher U, Kern W *et al.* Report on two novel nucleotide exchanges in the JAK2 pseudokinase domain: D620E and E627E. Leukemia 2006;**20:**2195-2197.

Ordering Information

Product	Catalog Number	Pack Size
Genome Sequencer FLX Instrument	04 896 548 001	1 instrument plus accessories
GS FLX Titanium SV emPCR Kit (Lib-A)	05 619 165 001	1 kit (for 32 small volume emulsion amplification reactions)
GS FLX Titanium emPCR Filters SV 64pc	05 233 674 001	1 kit (64 filters)
GS FLX Titanium emPCR Breaking Kit LV/MV 12pc	05 233 658 001	1 kit (for 12 LV emulsion-breaking setups)
GS FLX Titanium Sequencing Kit XLR70	05 233 526 001	1 kit (for one Genome Sequencer FLX sequencing run)
GS FLX Titanium PicoTiterPlate Kit 70x75	05 233 682 001	1 plate (with accessories)
FastStart High Fidelity PCR System, dNTPack	04 738 292 001	500 U for up to 1,000 reactions of 50 µl final volume each containing 2.5 U FastStart Taq DNA Polymerase
LightCycler® 2.0 Instrument	03 531 414 001	1 instrument
LightCycler® FastStart DNA Master HybProbes	03 003 248 001 12 239 272 001	96 reactions 480 reactions

For detailed information regarding 454 Sequencing Systems, visit **www.454.com**

For life science research only.
Not for use in diagnostic procedures.

454, 454 SEQUENCING, 454 LIFE SCIENCES, EMPCR, GS FLX, GS FLX TITANIUM, LIGHTCYCLER, and PICOTITERPLATE are trademarks of Roche.

Other brands or product names are trademarks of their respective holders.

Published by
Roche Diagnostics GmbH
Sandhofer Straße 116
68305 Mannheim
Germany

© 2010 Roche Diagnostics. All rights reserved.

www.roche-applied-science.com

06364837001 ① 0910