Transferring the Power of Next-Generation Deep-Sequencing into Clinical Routine Applications in Combination with Robust High-Throughput Variant Detection

Thorsten Kurz,[1] Alexander Kohlmann,[2] Niroshan Nadarajah,[2] Torsten Haferlach,[2] Joachim Strub,[1]

[1]JSI medical systems GmbH, 77971 Kippenheim, Germany. [2]MLL Munich Leukemia Laboratory, 81377 Munich, Germany.

Introduction

The new methods for high-throughput next-generation sequencing (NGS) have enabled completely novel insights into the heritability and pathophysiology of human disease.

Thus, the global genome sequencing technologies market is estimated to be worth ~$2 billion per year and this is expected to increase to >$4 billion by 2017.[1]

The introduction of next-generation sequencing technologies have evolved to provide an accurate and comprehensive means for the robust detection of molecular mutations. Thus, NGS technologies have led to important discoveries in biomedical research and have already been implemented in clinical diagnostics.[2]

Recent studies showed that amplicon deep-sequencing is a technically feasible method that enables an innovative novel sequencing approach in a routine diagnostics environment.[3-6]

It is anticipated that NGS will be integrated into clinical practice as a suitable platform to provide highly quantitative data on a constantly increasing number of molecular mutations in a necessary throughput and accuracy and, as such, will advance into the field of standard molecular diagnostics.[5]

Here, we highlight data from daily clinical diagnostics routine operations based on the amplicon deep-sequencing analysis pipeline from the Munich Leukemia Laboratory (MLL). These data demonstrate the power and utility of the **SeqNext** module from the **JSI SEQUENCE PILOT** software package in transferring the application of next-generation sequencing into clinical routine operations in combination with robust high-throughput variant detection based on the data of the recently published paper from Grossmann et al..[7]

SEQNext
Next Generation sequencing

The Power of

Performance Study

To demonstrate the power of the **SeqNext** module from the **JSI SEQUENCE PILOT** software the linear relationship between distinct mutation loads and the relative frequency of obtained reads carrying a respective mutation was investigated. A series of dilution experiments for a broad range of mutation measurements were performed, including variants with a low frequency as they often occur in heterogeneous cancer specimens (for further details please see Grossmann et al.[7]).

Materials and Methods

Patients and Study Design

In a cohort of 10 cases, the coding region of *RUNX1* was amplified and nine cases were randomly selected according to their known *RUNX1* mutations (three missense, one nonsense, and five frame-shift alterations). A tenth case not harboring any *RUNX1* mutations served as a wild-type control. Each PCR product was individually purified and four serial dilutions (1:2, 1:4, 1:8, 1:16) of each of the nine amplicons were generated.

Furthermore, one case with two *CEBPA* variants in the same amplicon was analyzed in six serial dilutions (1:2, 1:4, 1:8, 1:16, 1:32, and 1:64).

SEQNext
Next Generation sequencing

Next-Generation Sequencing and Data Analysis

454 amplicon deep-sequencing data were generated as reported using the GS FLX Titanium XLR70 Sequencing Kit (Roche Applied Science).[4] The obtained reads for each amplicon were aligned against the human genome reference sequences using the GS Amplicon Variant Analyzer software version 2.5.3 (Roche Applied Science). Mapping results and detected variants were exported to R/Bioconductor for further analyses and visualization.[8,9]

To avoid the utilization of many different software tools as mentioned above the analyses were independently performed by using the third-party software **SEQUENCE PILOT** from **JSI Medical Systems** (Kippenheim, Germany).

The **SeqNext** module from the **JSI SEQUENCE PILOT** software package provides a convenient and powerful workflow to perform all relevant steps (mapping, alignment, visualization, variant detection and interpretation) at once.

Results

SEQNext
Next Generation sequencing

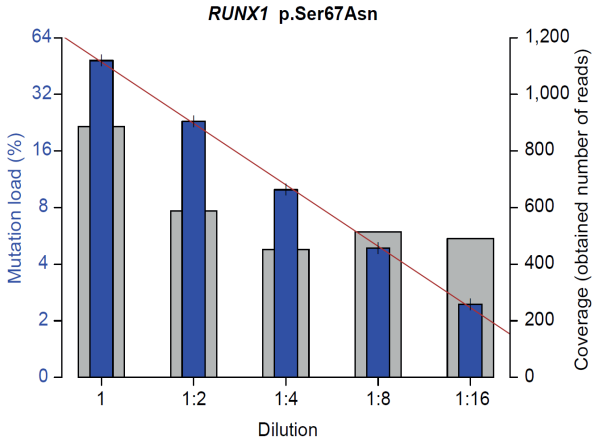

Figure 1: Serial dilution data. The mutation load of the p.Ser67Asn alteration is given on the left y-axis. The number of reads (coverage) is given on the right y-axis. The serial dilution data points are given on the x-axis (figure adopted from Grossmann et al.[7]).

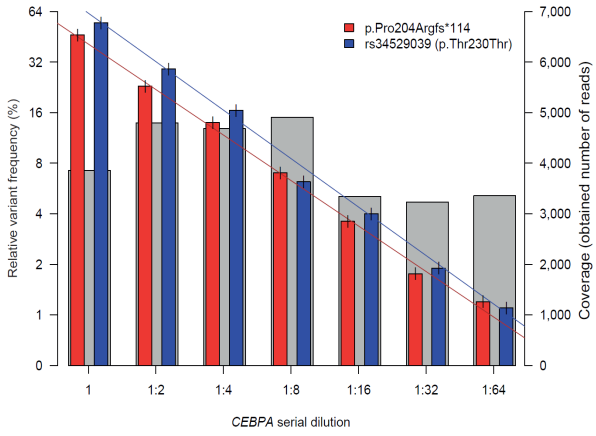

Figure 2: Serial dilution data. The load of the *CEBPA* variants is given on the left y-axis. The number of reads (coverage) is given on the right y-axis. The serial dilution data points are given on the x-axis (figure adopted from Grossmann et al.[7]).

The *RUNX1* p.Ser67Asn undiluted mutation load of 48.4% (886-fold coverage) decreased from 48.4% to 23.0% (1:2 dilution; 588-fold coverage), 10.0% (1:4 dilution; 452-fold coverage), 4.9% (1:8 dilution; 514-fold coverage), and 2.5% (1:16 dilution; 490-fold coverage), respectively. By fitting a regression line to these data, a slope ß of -1.085 (95% CI, ±0.069) was obtained, which indicates a near-perfect decrease of the mutation load of the p.Ser67Asn alteration (Figure 1) and therefore the robust variant detection using the **JSI SeqNext** workflow.

The undiluted *CEBPA* variant load for case 32 of 54.8% and 46.4% decreased in a similar way for both variants using the **JSI SeqNext** workflow. By fitting a linear model to the logarithmized data of case 32, a slope ß of -0.956 (SE, ±0.030) was obtained for the polymorphism and a slope ß of -0.900 (SE, ±0.024) was obtained for the frame-shift mutation (Figure 2), respectively.

Based on the data obtained from the **JSI SeqNext** workflow a linear mixed-effects model was fitted to the data of all nine *RUNX1* samples. The resulting slope ß of 1.000 (95% CI, ±0.046) demonstrate that each dilution step halves the measured mutation load based on the robust variant detection of **JSI SeqNext** workflow. Correlation analyses for all nine samples separately resulted in a range from 0.98 to 0.99 and, thus, were underlining the appropriateness of the linear modeling approach (data not shown, see Grossmann et al.[7]).

Robust High-Throughput Variant Detection in Clinical Routine

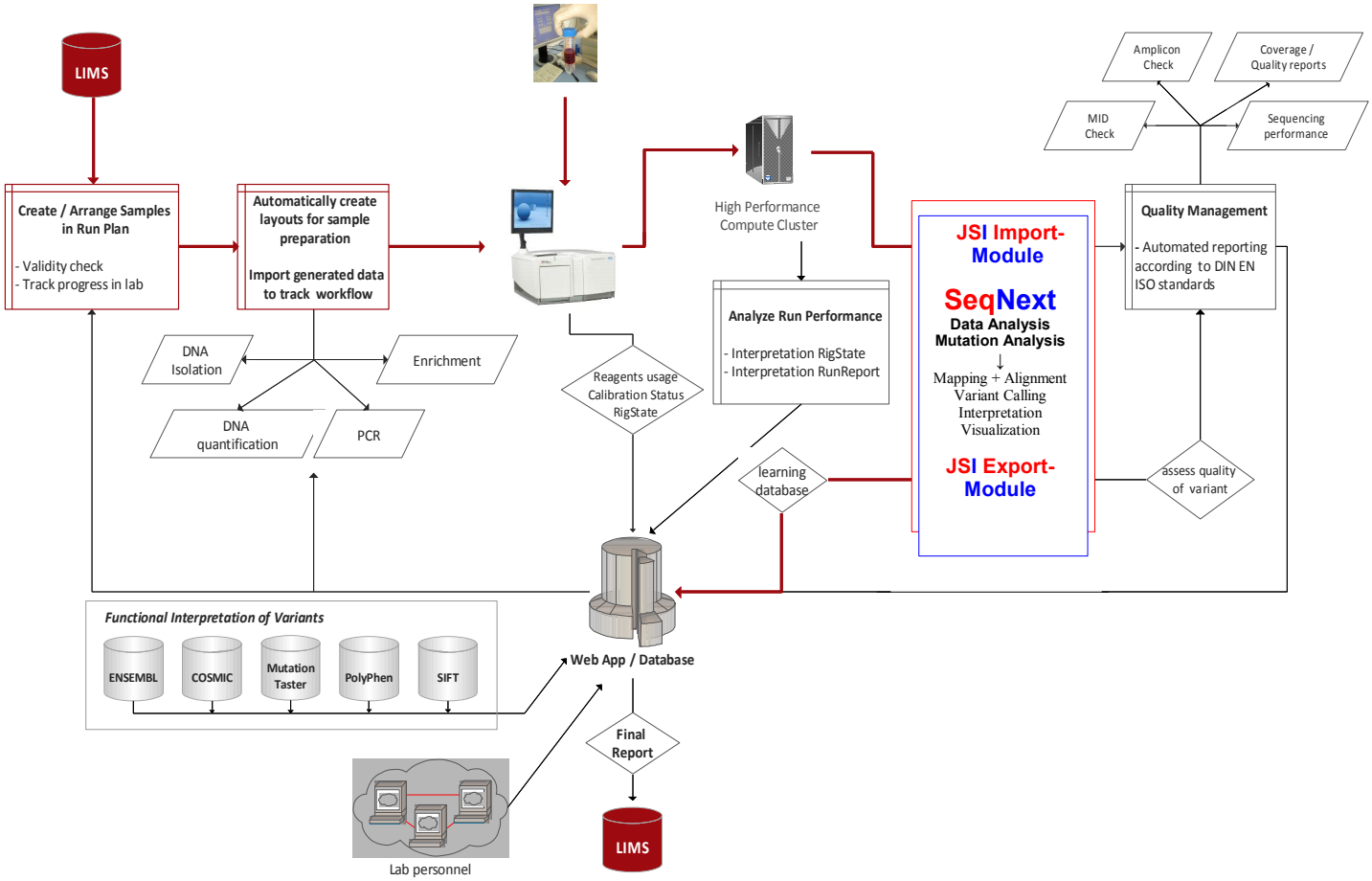

Figure 3: Next-generation sequencing workflow (MLL pipeline). After receiving a sample and entering the relevant patient information into a laboratory information management system (LIMS), individual processes for library preparation, sequencing, data analysis and variant interpretation are performed. Final sequencing report files are then uploaded back into the LIMS to be linked with the requested mutation analysis result for individual patients.

The flow diagram shown above (Figure 3) represents the MLL specific amplicon deep-sequencing analysis pipeline.[5,6] The utilization of the **JSI Import-Module** enables the automatic import of orders from the customer specific LIMS into the **SeqNext** module to automatically perform the subsequent mapping, alignment and variant detection steps. Subsequently, the technical and medical validation is performed within the **SeqNext** module and a customized diagnostic report is generated. The subsequent utilization of the **JSI Export-Module** enables the automatic export of results and reports from the **SeqNext** module back to the customer specific LIMS. Thus, the pipeline shown here enables to address the ever growing demands for robust high-throughput clinical research topics.

Conclusion

Within the framework of the performance study presented here the limit of variant detection was studied in serial dilution experiments based on the **JSI SeqNext** workflow. It could be shown that the limit of detection of the **JSI SeqNext** workflow ranged from 0.25% to 3.5% depending on the local sequence context of the analyzed amplicons (e.g., errors that are introduced during library construction and sequencing).[7] In addition, it turned out that low-level clones (ie, mutation load of 1.76% and 2.60%) which are undetectable by Sanger sequencing are found with the **JSI SeqNext** module because of the higher sensitivity of the next-generation sequencing workflow.[7]

The here shown data convincingly demonstrate that the **SeqNext** module from the **JSI SEQUENCE PILOT** software package is a powerful tool to provide highly quantitative data and consequently to detect variants with low relative frequencies.

Moreover, the data demonstrate the high linearity between the actual mutation load and its measurement (mapped and aligned reads carrying the sequence alteration obtained by using the **JSI SeqNext** workflow) across the broad range of relative mutation frequencies.

Finally, it is very important to emphasize that the utilization and seamless integration of the JSI modules (**SeqNext**, **Import-Module** and **Export-Module**) into NGS pipelines allows to address the ever growing demand for analyzing panels of biomarkers instead of testing individual biomarkers with the necessary throughput and accuracy in a cost-effective manner and fast turnaround time.

Acknowledgement

JSI medical systems would like to acknowledge Dr. Vera Grossmann, Dr. Alexander Kohlmann, Niroshan Nadarajah and Prof. Dr. Dr. Torsten Haferlach from the MLL Munich Leukemia Laboratory (www.mll.com) for sharing the data presented in this application note.

References

1. Sharmarke M, Basharut A.S. (2013) Commercial prospects for genomic sequencing technologies. Nature Reviews Drug Discovery 12:341–342.
2. Frese K, Katus H, Meder B. (2013) Next-Generation Sequencing: From Understanding Biology to Personalized Medicine. Biology 2:378-398.
3. Kohlmann A, Grossmann V, Klein HU et al. (2010) Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. J Clin Oncol. 28(24):3858-65.
4. Kohlmann A, Klein HU, Weissmann S et al. (2011) The Interlaboratory RObustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories. Leukemia 25,1840–1848.
5. Kohlmann A, Grossmann V, Haferlach T. (2012) Integration of next-generation sequencing into clinical practice: are we there yet? Semin Oncol. 39(1):26-36
6. Kohlmann A, Grossmann V, Nadarajah N et al. (2013) Next-generation sequencing - feasibility and practicality in haematology. Br J Haematol. 160(6):736-53.
7. Grossmann V, Roller A, Klein HU et al. (2013) Robustness of amplicon deep sequencing underlines its utility in clinical applications. J Mol Diagn. 15(4):473-84.
8. Gentleman RC, Carey VJ, Bates DM et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5:R80
9. Klein HU, Bartenhagen C, Kohlmann A et al. (2011) R453Plus1Toolbox: an R/Bioconductor package for analyzing Roche 454 Sequencing data. Bioinformatics 27:1162-1163

Germany
Phone: +49 (0) 7825 863620 0
Fax: +49 (0) 7825 863620 20
Mail: mail@jsi-medisys.de
Internet: http://www.jsi-medisys.de/

USA
Phone: +1 949 999 2092
Fax: +1 949 999 2093
Mail : mail-us@jsi-medisys.com
Internet: http://www.jsi-medisys.com/

JSI medical systems