Moving away from Sanger sequencing in diagnostics

Next-Generation Sequencing

A Substitute for Sanger Sequencing

Adinda Diekstra[1], Thorsten Kurz[2], Joachim Strub[2], Kornelia Neveling[1], Marcel R. Nelen[1]

[1]Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands;

[2]JSI medical systems GmbH, 77971 Kippenheim, Germany

Abstract

Many labs have embraced NGS technologies to the point where it is being implemented for routine diagnostic use. Though, until today Sanger sequencing is still regarded as the gold standard approach for sequencing single genes. The development of benchtop NGS instruments allowed the analysis of multiple similar single genes or small gene panels while guaranteeing the turnaround time and even coverage distribution of all analyzed regions, which makes these platforms increasingly competitive with Sanger sequencing.

Since 2010, the department of Human genetics of the Radboud University Medical Center in Nijmegen transferred BRCA1/2, mitochondrial and Noonan testing from Sanger sequencing to a benchtop NGS instrument (Ion Torrent PGM™) in clinical diagnostics. The disadvantage of these implemented approaches is that they are all based on different capture methods, require sample batching, and are time consuming and thus quite expensive to develop. Here, we highlight data from a new project in which researchers from Nijmegen developed a generic sequencing strategy which is able to substitute for Sanger sequencing for almost all the genes (>800 genes) they offer in routine clinical diagnostics.

This newly developed ion semiconductor sequencing (ISS) work flow is based on using the Sanger sequencing primers for the enrichment which enables the same advantages as Sanger sequencing had (i.e., the ability to prevent sample batching and is highly flexible).

The key to the success of this new ISS workflow was the development of an innovative pooling and barcoding strategy which reduces sequencing costs up to 70-80% per amplicon (recently published paper from Diekstra et al.[1]).

In addition, the data of this study [1] demonstrates the power and utility of the SeqNext module from the JSI SEQUENCE PILOT software package in the new generic and fully automated ISS workflow using the Ion Torrent PGM™ in combination with robust and high-throughput variant detection.

Performance Study

The Department of Human Genetics in Nijmegen routine genetic performs testing for over 600,000 Sanger sequencing reactions per year. To process this large number of tests, DNA extraction, amplification, and sequencing have been automated. To further improve efficiency and reduce costs, the new ISS work flow has been incorporated in the fully automated Sanger sequencing workflow (Figure 1).

Materials and Methods

Amplicon generation and pooling

PCRs were performed using conventional Sanger sequencing primers in a fully automated robotic workflow. For ISS, no additional purification, besides the purifications within the library preparation, was necessary following the PCR. To reduce the number of library preparations a barcode per pool of unique amplicons instead of barcoding per sample was used [1]. For this a customized script was developed which assures only unique amplicons to be distributed on a single PCR plate and secondly distributes the recurrent amplicons to different plates (Figure 1).

Library Preparation and Sequencing

The pooled PCR products used in this approach were optimized for Sanger sequencing and varied in size between 200-900bp. For sequencing on the PGM, PCR product lengths needed to be reduced to a mean size of 200-300bp. Therefore, pooled PCR fragments were sheared (Covaris E210) before they undergo an automated library preparation, and sequencing on Ion sequencing chips from Life Technologies.

Study design

Phase 1 – Proof of principle

Unique amplicons representing routine diagnostic requests of a random single day (pool 1: 232 amplicons covering 19 unique genes) and a random single week (pool 2: 1484 amplicons, covering 77 unique genes) were amplified (with the conventional Sanger primers) and pooled to test the ability of ISS to use these Sanger optimized amplicons. These pools were each sequenced on a 316™ sequencing chip by using a barcode per pool.

Phase 2 - Validation

To validate the pooling and barcoding strategy [1] and the newly developed ISS workflow, amplicons from all known positive index cases harboring a disease-causing mutation identified in 2012 were selected. In total, 1224 amplicons (coming from 232 unique genes) were amplified with the conventional Sanger primers, and sequenced on the PGM ™.

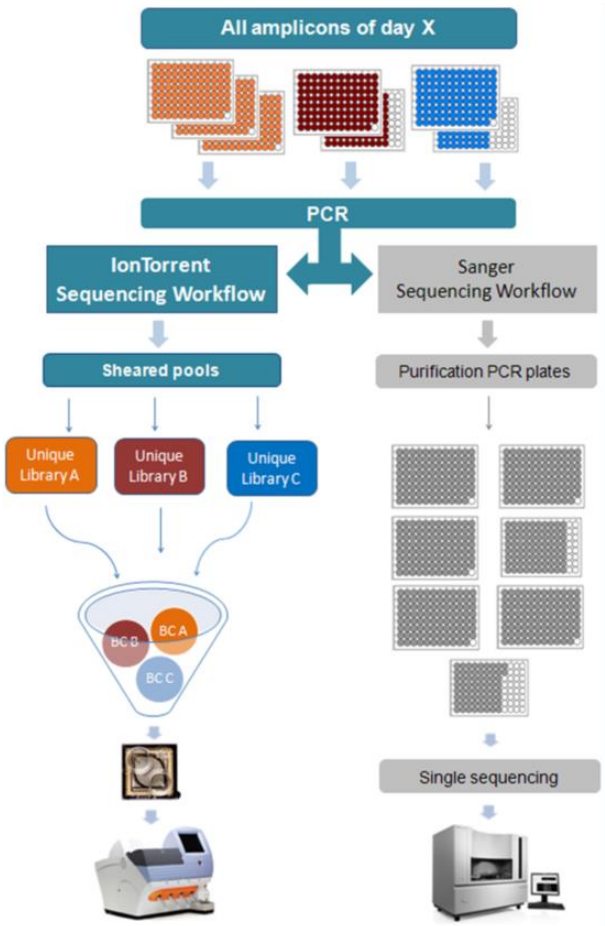

Figure 1: Schematic figure of the generic fully automated semiconductor sequencing workflow.
A generic fully automated workflow using the IonTorrent was constructed with the ability to sequence thousands of amplicons simultaneously. Usage of the conventional Sanger amplicon designs enables to achieve two goals in a single effort (primarily IonTorrent sequencing and Sanger sequencing for mutation confirmation).

Phase 3 – Blind sequencing

To mimic a realistic routine diagnostic scenario, having the new ISS workflow fully automated, a blind sequencing experiment was performed. 100 samples (representing requests of three different routine diagnostic days) were sequenced in parallel using the automated Sanger and ISS workflow, and results were compared.

Data Analysis

The SeqNext module from the JSI SEQUENCE PILOT software was used to perform all the analyses (mapping, alignment, visualization, variant detection and interpretation). Sequencing data in fastq format were automatically sent to the analysis software SeqNext. Within SeqNext, the sequencing reads were mapped to defined ROIs, and variant calling was performed using defined user settings (for further details please see Diekstra et al.[1]). Analysis parameters in combination with selective procedures were used to ensure high coverage and high sensitivity, thus taking specific sequencing technology–based limitations into account (e.g., base call quality, strand bias and homopolymer topics).

Results

Phase 1 – Proof of principle

On average, 2.87 million reads were generated per run, showing an average read length of 176bp and a median coverage of 1490x for pool 1 and 716x for pool 2. Insufficient coverage (below 40x) was observed for 2.7% (pool 1) and 1.2% (pool 2) of the PCR products. Data analysis in SeqNext, using default settings, resulted in the detection of all known variants for both runs.

Phase 2 - Validation

In total, 3.4 million reads were generated, with a median coverage of 808x per amplicon. Insufficient coverage (below 40x) was obtained for 3.1% of the amplicons. This resulted in 1 false negative and 43 false-positive calls. The false positive variants occurred only in homopolymer stretches and appeared exclusively as duplications or deletions. The false negative variant was seen in 50% of the forward reads but was absent in the reverse reads. Due to the optimized software settings (generating the best balance between sensitivity and specificity), requiring a minimum of 15% variant reads per direction, this variant was not reported.

Phase 3 – Blind sequencing

Data analysis with the optimized software setting of phase 2 revealed 58 false positive variants, and one false negative variant compared to Sanger sequencing. The false-positive calling occurred again in homopolymer stretches. The false-negative variant was seen in 77% of reverse reads and in only 5.5% of forward reads.

Discussion

The InDel variations tested in the diagnostic cohort indicated no sign of reduced sensitivity for more difficult insertions, deletions, or duplications. Even larger duplications of 30 nucleotides were accurately detected by the analysis software (Figure 2). In addition, two nonpathogenic variants identified in the ISS data were missed in the Sanger sequencing data (for details please see Diekstra et al.[1]). This underlines the fact that false-negative variants are not only unique to NGS.

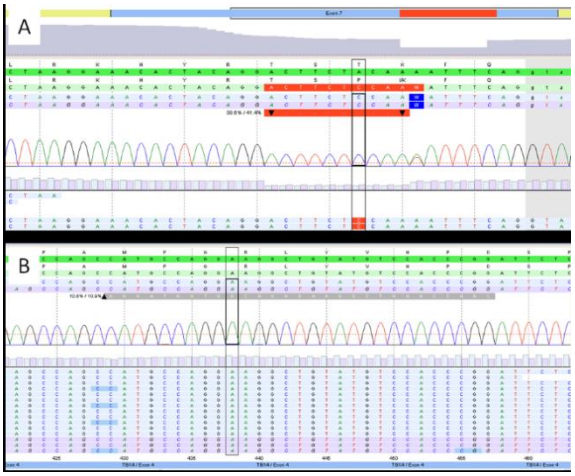

Figure 2: Detected insertions and deletions.

A) Screenshot of SeqNext showing a deletion in gene *CNGB3*. The variant c.886_896delinsT (het) has been correctly annotated by the software.

B) Screenshot of SeqNext showing a 30bp duplication (het) in gene *TBX4*. The duplication is visualized by the grey bar above the artificial Sanger trays.

Sensitivity and Specificity

The described setup based on optimized software settings for this diagnostic cohort have lead to a total analytical sensitivity of 99.61% and a total analytical specificity of 99.98% of the automated ISS workflow compared to Sanger sequencing. To ensure high sensitivity as well as high specificity (balance false positive / false negative) the obtained discrepancies in variant read counts have lead to false negatives based on the used settings (requirement: minimum 15% variant reads per direction). In addition, we suggest to confirm both, suspicious variants detected by ISS as well as amplicons showing a minimal coverage below 40, by Sanger sequencing.

Meanwhile JSI already implemented new features (which can be applied to the two mentioned false negatives in this study) into the new Sequence Pilot Version 4.2. to capture the sequencing platform-based technical limitations. This means if a variant will be found only in one sequencing direction with a defined minimum percentage, the software will be able to switch automatically for this position into the combined mode (deactivation of per direction mode and ratio mode). Furthermore, the new Sequence Pilot Version 4.2 will allow to apply the analysis parameter per individual ROI automatically.

Conclusion

Human Genetics Nijmegen demonstrated the development, validation and implementation of a generic automated ISS work flow for routine genetic testing (ISO15189 accredited). Novel pooling strategies allowed a minimal usage of molecular barcodes, leading to a cost reduction of 70% per sequenced amplicon.

The strong collaboration between Nijmegen and JSI has resulted in data which convincingly show that the SeqNext module from the JSI SEQUENCE PILOT software package is a powerful software tool to provide the needed analytical sensitivity and specificity for ISS in clinical routine diagnostics to make it comparable to those for the gold standard Sanger sequencing.

Acknowledgments

JSI Medical Systems would like to thank Dr. Marcel Nelen, Adinda Diekstra and Konny Neveling for sharing the data presented in this application note.

References

1. Diekstra A, et al. Translating Sanger-Based Routine DNA Diagnostics into Generic Massive Parallel Ion Semiconductor Sequencing. Clin Chem. 2014: [Epub ahead of print] doi:10.1373