TechNote

Improving Whole Exome Sequencing

Introduction

Whole exome sequencing (WES) offers advanced clinical diagnostics and sufficient coverage for rapid, accurate, and reliable patient diagnosis.
At GenomeScan, we work for this application with university medical centers (UMCs) and we process the genetic material of thousands of patients through next generation sequencing (NGS) each year. The data we generate is used by geneticists to diagnose and decide the best-fit treatment and care for the patient. This method is used for prenatal diagnostics, oncology diagnostics, and diagnostics of hereditary diseases.

Our validated procedures contribute to high quality data with short turnaround times. In this context, decreased turnaround time will help to assure a faster path for clinical decisions.

To maximize the quality of the WES sequencing data, GenomeScan tested and implemented a simple modification of the library preparation to improve the coverage of the exome, and consequently, the amount of usable data for clinical interpretation.

Method

Genomic DNA was isolated from whole blood and sheared enzymatically to an average size of 200-500 bp. Library prep was performed on 200 ng genomic DNA using Agilent's SureSelectXT Low Input Reagent Kit. Target capture was performed according to the manufacturer's instructions using Agilent Human All Exon V7 baits. Quality and quantity of capture libraries were determined using the Agilent Fragment Analyzer and Qubit. With this improved version of the sample prep, the temperature of the capture wash was increased by 4°C for optimal capture efficiency.

After completion of the sample preparation, normalized libraries were clustered on-board, and 150 bp paired-end sequencing was performed on the NovaSeq 6000 generating 12Gb of raw data per sample.
After demultiplexing, FASTQ files were processed on our Illumina DRAGEN platform, based on hardware-acceleration of commonly used genomics' analysis algorithms. The platform runs on reconfigurable Field-Programmable Gate Array (FPGA) technology, providing accuracy and speed. It is used for pre-processing the raw data, mapping to the human reference, and calling of variants.
It also includes algorithms for accurate copy number variation (CNV) and structural variant (SV) analyses.

Results

Protocol automation and optimization

The automation of the low input NGS target enrichment workflow and change of the hybridization temperature allowed us to significantly improve the quality of the results as displayed in figure 1.

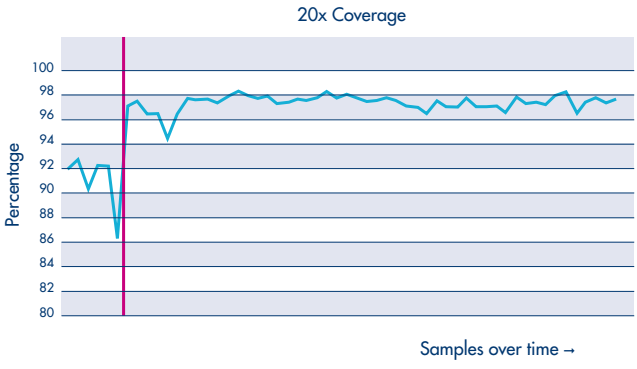

Figure 1. The 20x Coverage percentage over time. The red line marks the moment the modified protocol was introduced.
All samples have a similar amount of sequence data. The change in the protocol results in a higher and more stable 20x Coverage per sample

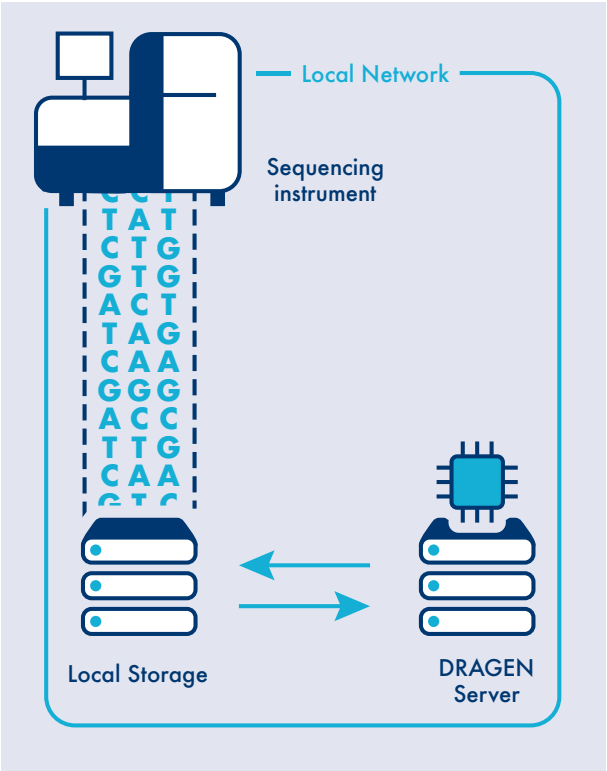

Figure 2. Bio-informatic platform. Image based on original by Illumina

Bio-informatic analysis

The DRAGEN platform (fig. 2) processes the genetics data from the sequencers into a ready-to-use and annotated list of variants (SNPs and indels). With the latest version of the DRAGEN suite v3.8.4, the number of false positives and negatives is significantly reduced in comparison with earlier versions and other variant calling tools.

Our on-premises DRAGEN solution allowed the direct processing of data from our sequencers without latency in transfer of data to external cloud or HPC solutions. VCF files were ready within an hour after the data was generated on the sequencer. Parallelisation of sample processing was achieved by simultaneously deploying multiple instances of DRAGEN.
The throughput time for a WES sample is 3-4 minutes. Compared to our previous workflow this is a huge improvement in efficiency.

The DRAGEN suite offered excellent quality of mapping and variant calling. When benchmarked against the NIST Genome in a Bottle (GIAB) HG001 reference data, we achieved a precision and sensitivity of 0.9961 and 0.9819 for SNPs, respectively. For indels, precision and sensitivity were 0.9476 and 0.9378, respectively.

Conclusion

By optimizing our current automated WES workflow, we increased the percentage of usable data and coverage, and significantly decreased the turnaround time. GenomeScan has validated this service under ISO/IEC 17025.

The laboratory uses less starting material, which allows us to save precious patient DNA. Hence, prenatal diagnostics is now possible. Even with a reduced amount of DNA, the new protocol provides an increased exome coverage with a lower percentage of duplicate reads.

The complete automation of the workflow and the upgrade to the DRAGEN software reduces the turnaround time from 8 to 5 days. Shortening the time of diagnostic exome sequencing is crucial for the patient and healthcare professionals as it may help to avoid or reduce hospitalization.

By actively improving the NGS workflows, GenomeScan ameliorates the analysis of genetic material submitted by the UMCs; it also contributes to a patient-focused care that will ultimately pave the way for personalized medicine.

Future Plans

To provide patients with timely access to a good diagnostic (sensitive, specific, standardized), healthcare professionals rely on high-quality and accurate data, comprehensive reporting, and fast turnaround. In the interest of each patient, GenomeScan continues to innovate and create solutions that enable rapid delivery of results (e.g. workflow automation) and better interpretation of those results. To gradualy ameliorate the standard diagnosis, we also focus on RNA-Sequencing to gain a better insight into the disease and improve the diagnostic yield.

genomescan.nl
+31 71 5681050
info@genomescan.nl