IonTorrent: 2nd generation sequencing in a diagnostic laboratory

B. Dworniczak[1], S. Fleige-Menzen[1], N. Bogdanova-Markov[1], P. Pennekamp[2]
[1]University Hospital Muenster, Institute of Human Genetics, Muenster, Germany,
[2]University Children's Hospital Muenster, Department of General Paediatrics, Muenster, Germany

Problem

Molecular diagnosis of complex human genetic disorders can still be challenging especially if multiple genes harbouring putative deleterious mutations have to be investigated. Currently in most cases Sanger sequencing is applied, however capillary sequencing is time-consuming as well as expensive for the screening of multiple large genes. In 2005 first next- or second-generation sequencers have been introduced and since then, multiple improved NGS platforms are in several laboratories in use. However most of these platforms and in particular accompanying software are especially made for sequencing projects analysing complete genomes and it is not easy to scale down this technology for the very special needs of a diagnostic setting.
In order to adjust technology to those specific needs recently three benchtop high-throughput sequencing instruments have been launched. While the GS Junior (454 Life Sciences) and the MiSeq (illumina) depend on established technology the IonTorrent PGM introduced a complete new sequencing technology based on the detection of hydrogen ions that are released during the polymerization of DNA.
Here we report on our experiences during the last six month setting up a workflow which finally should result in the integration of the IonTorrent PGM into a diagnostic laboratory. Although results are still preliminary our data will be crucial for the decision whether this technology can be implemented for diagnostic purposes.

Goal/First milestone

Adaptation of a medium high throughput sequencing technology to the needs of a diagnostic laboratory; gradually replacement of low throughput capillary electrophoresis.
To compete with our established work flow as first milestone we aimed at the simultaneous analyses of at least ten patients for the breast cancer genes BRCA1, BRCA2 and RAD51c. The cDNA of these three genes cover roughly 16980 bp split into 85 PCR fragments; in addition 4300 bp of flanking region has to be analysed.

Prerequisites to reach these goals:

- Technical set up of the complete sequencing work flow
- Evaluation of data compared to data obtained by capillary electrophoresis

Work flow

The workflow of each second-generation sequencing experiment is divided into 4 independent parts and each part itself is further sub-divided into several sections.

Enrichment of target sequences

Library construction

Sequencing

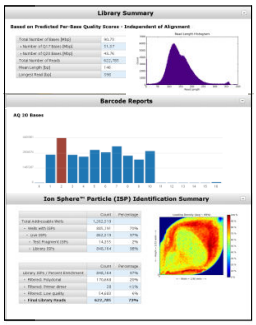
Interpretation of raw data

Results

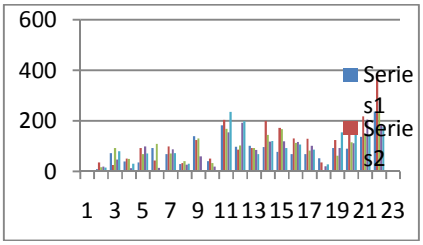
During the last six month we performed nearly 40 runs using the Ion 314 sequencing chip to develop a workflow and to evaluate the system. We exclusively sequenced PCR fragments already checked by capillary electrophoreses. So far we did not change the PCR set up already established for the genes of interest and we did not normalize the amount of PCR fragment used for the analyses. So far we checked more than 15 different genes and for most runs we barcoded the probes; we used between 4 and 12 different barcoded probes per analysis. For post-run data analysis we used:
Soft ware package provided by IonTorrent (Aligner and variant caller)
Gen*search* NGS (version 1.3.7; PhenoSystems SA)
NextGENe (version 2.2; SoftGenetics)

Overall Perfomance

Basic data are represented in the report generated after each run. This report shows that after some training at least $30x10^6$ bp of high quality sequence can be generated using the smallest Ion 314 chip. Assuming an even distribution analysing the three breast cancer genes of ten patients (21500 bp each) each nucleotide could have a sequencing depth of 140. The yield could be increased by improving read length (in this run average 146 bp).



Most software packages allow checking the coverage of each single nucleotide and the comparison below show that even without normalization independent series of PCR reaction yield a reproducible pattern of coverage. In the chard below five independent PCRs for BRCA1 are shown.

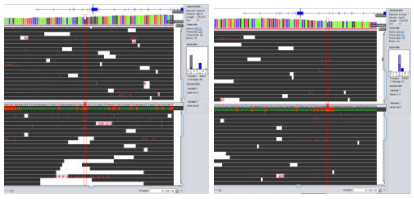


Data analysis

The data package provided by IonTorrent included an SFF file, a data format that contains read sequences and base qualities. There are also BAM files, which include the aligned read sequences, indexed for random access. And there are FASTQ files, which is the common sequence-and-quality input format used by many short read aligners. Using barcodes all files in addition are provided in the specific forms. As plug-in IonTorrent recently introduced a "so-called" variantCaller compiling a list of putative mutations/variants which diver from the reference provided by the user. For independent analyses we used Gen*search* NGS (version 1.3.7; PhenoSystems SA) NextGENe (version 2.2; SoftGenetics). Both software packages were feed with barcode specific FASTQ-files generated by IonTorrent.

IonTorrent and Indels

It is frequently discussed that the IonTorrent technology tends to show small Indels which confuse interpretation of data whereat deletion of single nucleotides are much more frequent compared to insertions. Originally these indels where mostly reported for longer tracks of the same nucleotide. During our evaluation we noticed that beside indels (ore mostly deletions) in tracks deletions frequently occur in a **reproducible sequence-specific** manor (example shown in the figure). However most indels occur strand-specific skewing the balance between the forward read and the reverse read (Value for "balance" will strongly differ from 1).

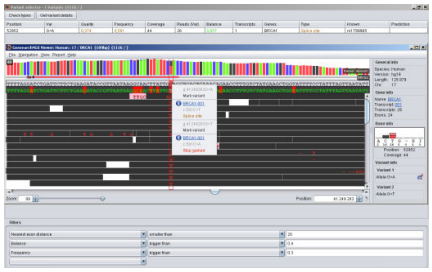


The deletion c.1410_1411delC is only present in the reverse reads which uncovers that this mutation is most likely an artefact (shown by capillary electrophoresis. Shown are two independent series of BRCA1 amplifications.

Reproducibility of sequence specific artefacts and the imbalance of occurrence in the forward and reverse reads allow setting up a validation scheme consisting of:

1. Run specific variant validation filter combining balance, frequency of variant in all reads and distance to the region of interest. Starting from a minimal sequencing depth (30-50) increasing the depth does not further improve the validation and its importance is frequently overestimated.
2. A profile of the known artificial variants as "background" for the gene of interest, based on the analysis of e.g. 10 independent analyses.

An example of efficient filtering reducing the number of variants from 2398 to the only real one can be seen on the following figure. (produced with Gen*search* NGS (version 1.3.7; PhenoSystems SA)

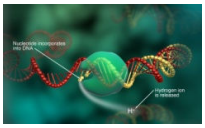


Independent of the quality of the validation scheme detection of real deletions (especially deletions larger than 6-10 bp) on the background of artificial deletions pose a severe problem which still awaits a manageable solution.

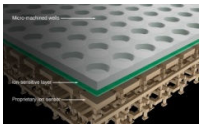
Resümee

2nd generation sequencing in a diagnostic setting using benchtop sequencer - in our case IonTorrent PGM - will be feasible in the near future although soft ware development and adaptation to diagnostic needs lags behind. This may be due to the fact that so far software packages are mainly developed for whole genome/exome sequencing.

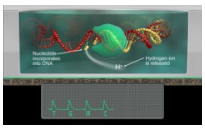
The main principle of the machine



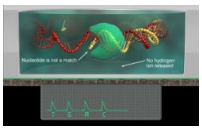
In nature, when a nucleotide is incorporated into a strand of DNA, a hydrogen ion is released as a by-product.



Ion Torrent perform this process in vitro in a massively parallel way using a high-density array of wells. Each well holds a single bead with a DNA template. Beneath the wells is an ion-sensitive layer and beneath that an Ion sensor.



If a nucleotide is added to a DNA template and is incorporated into a strand of DNA, a hydrogen ion will be The change from that ion will change the pH of the solution, which can be detected by the proprietary ion sensor and the sequencer will call the base.



The PGM sequencer then sequentially floods the chip with one nucleotide after another. If the next nucleotide released. that floods the chip does not match, no voltage change will be recorded and no base will be called.
If there are two identical bases on the DNA strand, the voltage will be duplicated, and the chip will record two identical bases called.

Fragments to be sequenced are captured on beads and amplified by emulsion PCR. The beads are then deposited in the microwell such that each well has only one bead carrying a unique amplified fragment. Nucleotides are then added in a predetermined sequence to the wells. A nucleotide complementary to the nucleotide in the fragment being sequenced gets incorporated in the strand being synthesized and a hydrogen ion is released. Upon release of hydrogen ion, the voltage of the solution changes in that well and is detected by the ion sensor. If two nucleotides are incorporated in a cycle, then the voltage is doubled and the sensor records two nucleotides added. If a nucleotide is not added during a cycle then no voltage change is recorded

Enrichment of target sequences

For diagnostic purpose a 100% coverage of the region of interest is essential. An established way to reach this is PCR.
In our pilot studies we used PCR products usually used for capillary sequencing. Without further normalization identical aliquots (e.g. 3µl each) of all 85 PCRs from a single patient were combined and purified using QIAquick PCR purification kit (Qiagen).

Library construction (10 different probes)

This step comprises several steps

- Fragmentation, blunting and phosphorylation of 250ng of the combined PCR products using the Ion Shear™ Plus Reagents Kit resulting in 200bp DNA-fragments
- Ligation with the unique barcode adapter
- Introducing of 10 different patient specific barcode sequences by PCR
- Pooling of the 10 different probes and size selection (3% agarose gel); isolation of the 200bp band (300bp) and DNA recovery (QIAquick Gel Extraction Kit; Qiagen).

After each step probes were purified (Agencourt, AMPure) and yield was quantified by use of the Qubit 2.0 Fluorometer.

- An important step before ePCR is the correct adjustment of the ratio between the quantity of the library (measured in number of molecules) and the amount of Ion sphere particles. Calculation is done by us of the IonTorrent PGM emPCR calculator.
- Emulsion PCR, breakage and enrichment of template- positive spheres (ISPs) is completely automated (Ion OneTouch and OneTouch ES module)

The material can either be sequenced directly or stored for several days.

Performing the sequencing run

- The PGM-Sequencer has to be cleaned and initialized (chemicals are put onto the machine and ph aquilibrated)
- The ISPs are annealed to the sequencing primer and sequencing polymerase is added; after mixing and incubation for 5 min the probe has to be sonicated and can be transferred to the prepared IonChip.
- After loading the chip on the PGM Sequencer the run can be started following the touch screen prompts

Time schedule

Obviously the actual time schedule is dependent of the protocol used and the experience of the user. Protocols and software packages provided by IonTorrent are still changing frequently. However using the most recent set up (using the Ion AmpliSeq Custom panel for enrichment, Ion AmpliSeq Library Kit V2.0, Barcoding-reaction, ePCR using OneTouch and OneTouch ES module, Ion 314 sequencing chip) a complete cycle starting with the enrichment upto data acquisition can be performed by a single person within two working days

UKM
Universitätsklinikum Münster