

Processing Amplicon Sequencing Data from PGM™ Sequencer using NextGENe® Software

May 2011

Phuoc Nguyen, John McGuigan, Megan Manion, Shouyong Ni, Sean Liu, C.S. Jonathan Liu

Amplicon sequencing is a highly accurate and quick method for the screening of known mutations as well as the detection of novel variants in specific targeted regions of the genome. For example, this method can be used to screen for known variants that cause breast cancer. However, current next generation sequencers have some drawbacks when performing amplicon sequencing. They can be too expensive for smaller labs or smaller projects with less throughput. They can also be too slow for clinical applications of amplicon sequencing. Semi-conductor sequencing, developed by Ion Torrent, a part of Life Technologies, is a novel sequencing method that is better suited for this type of analysis. It measures the hydrogen ions released during DNA replication rather than measuring modified nucleotides. The Ion Torrent PGM™ Sequencer offers a quick and inexpensive method to achieve amplicon sequencing while NextGENe software provides quick, accurate, and easy-to-use analysis of PGM™ Sequencer data. NextGENe software includes a unique module for processing data from an Ion Torrent PGM™ sequencer with an easy to use point-and-click interface that takes less than 3 minutes to go from raw data input to variant reporting.

Figure 1: Amplicon Sequencing Results Aligned in NextGENe

BRCA1 and BRCA2 gene were sequenced using Ion Torrent PGM™ Sequencer. Data was imported directly into NextGENe and aligned to a reference sequence for mutation detection. Analysis took seconds with NextGENe.

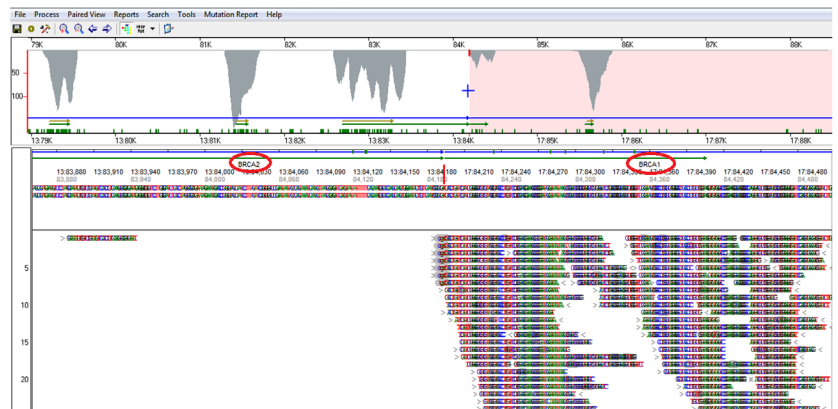


Figure 1

Procedure


1. Open NextGENe's Run Wizard by clicking on the  icon in the main toolbar. (Figure 2)
2. Instrument Type is defaulted to Ion PGM. (Figure 3)
3. Select SNP/Indel Discovery for application type (Figure 3)
4. Click "Next" to load data.
5. Click "load" to browse and load sample files (Figure 4)
Note: If sample file is not in fasta format, use format conversion tool to convert to fasta. Recommended settings are shown in **Figure 5**
6. Click "load" to browse and load reference files (Figure 4)
Note: NextGENe can use fasta or GenBank (.gbk) files for reference.
7. Click "set" to specify output file location and name. (Figure 4)
8. Click Next to specify settings for condensation (if applicable) and alignment. (Figure 6)
9. Click Finish.
10. Select Run NextGENe to begin processing project.



Figure 2

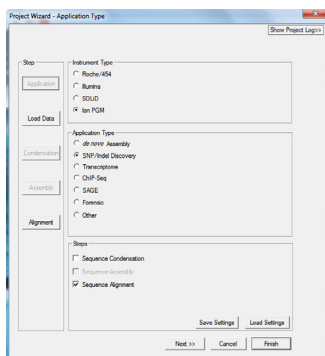


Figure 3

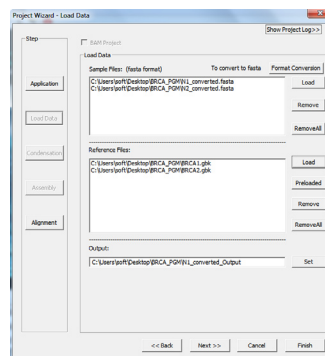


Figure 4

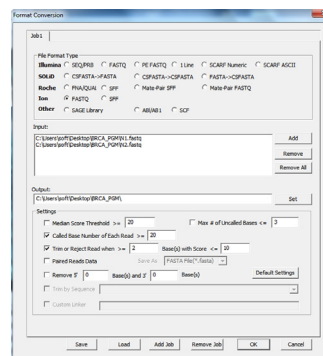


Figure 5

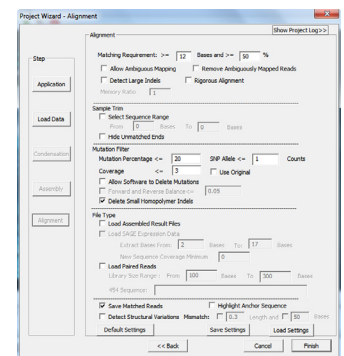


Figure 6

Results

Figure 7 shows the number of matched reads, unmatched reads, matched bases, and other parameters of the analyses.

BRCA1 and BRCA2 genes are well studied and contain many reported variations that are associated with susceptibility to breast cancer. Here, NextGENE is used to analyze the amplicon sequences for BRCA1 and BRCA2 to detect these reported variants using the GenBank reference. By using the GenBank files as a reference, annotation for the mutation is shown in the NextGENE Viewer as shown in Figure 8. Regions of interest can also be specified in the GenBank file in order to filter out mutations that are not in those regions. The ROI is shown by the pink line in the main view. Also in Figure 8, the reported heterozygous variation A->A/G is listed in the mutation report that contains useful quality control information. Along with the coverage, it also displays the individual nucleotides, mutation confidence score, as well as a link to the dbSNP database.

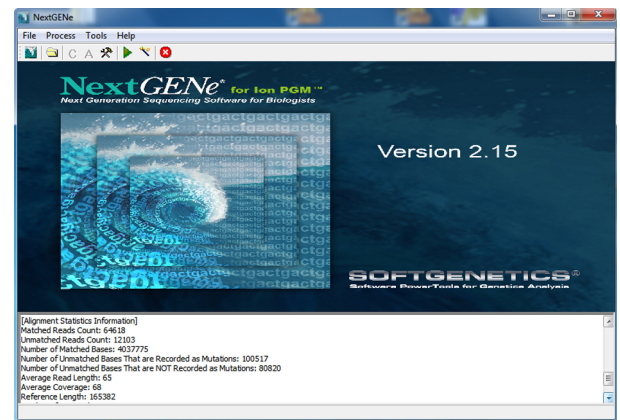


Figure 7: Initial Analysis Statistics



Figure 8: Amplicon Alignment

Discussion

In addition to the quick and accurate analysis, NextGENE offers a variety of reports that can be used for **quality control** of amplicon sequencing projects. In Figure 9, the Distribution Report shows both the directional coverage of the reads as well as the read length. In this view, it is possible to see whether there is any bias for forward or reverse sequences. If the coverage is uniform, there is more confidence in the mutation calls because there is no amplification or coverage bias.

The Coverage Curve is an available report that can assist with determining the amplicon sequencing quality. The red color in Figure 10 indicates regions that fails the user-set coverage threshold. The table below the graph contains detailed information of each highlighted region.

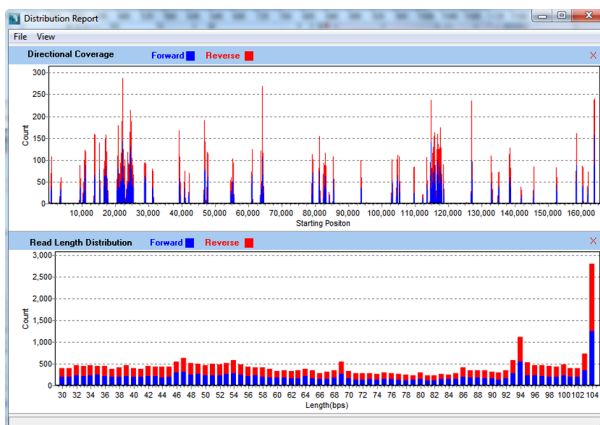


Figure 9: Coverage Distribution

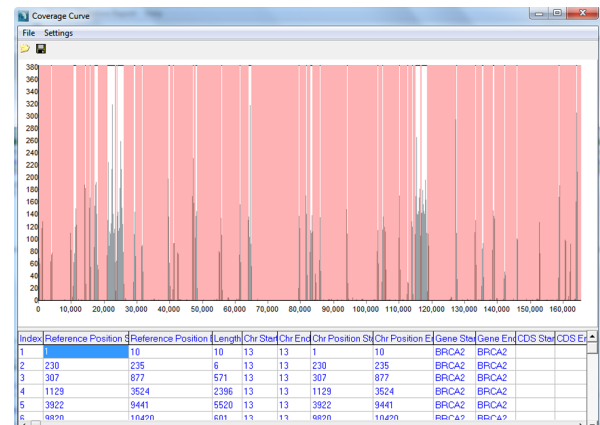


Figure 10: Coverage Curve

The Expression Report for the BRCA1 and BRCA2 data is shown in **Figure 11**, relative to each CDS. This report provides the expression levels and coverage for different regions of the reference genome. This can be displayed for the gene, CDS, or ROI as indicated in the settings dialog box. Statistics, such as minimum depth of coverage per region, are useful for verifying adequate depth of coverage in each amplicon.

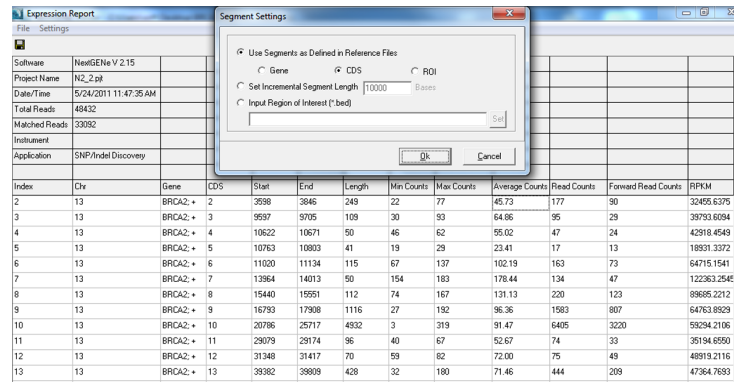


Figure 11: Expression Report

The Variant Comparison tool, as shown in **Figure 12**, is available in NextGENe. The tool can be used to compare multiple projects side-by-side in order to better observe any variations between them. In this case, the tool was used to compare two amplicon sequencing analysis results from two BRCA1 and BRCA2 samples sequenced by a PGM™ Sequencer. The mutation table at the bottom shows all the mutations detected in any sample. Different filtering options are available as shown in **Figure 13**. In this case, only mutations within the CDS regions are shown. In the mutation report, the mutation confidence score, coverage, gene name, chromosome position, and links to the dbSNP database are displayed. Several more annotations and statistics are available for display as well. Double-clicking on a variant shows the alignments at this position in both projects. This tool is very useful for amplicon sequencing because positive controls, negative controls, and multiple undetermined samples can all be compared.

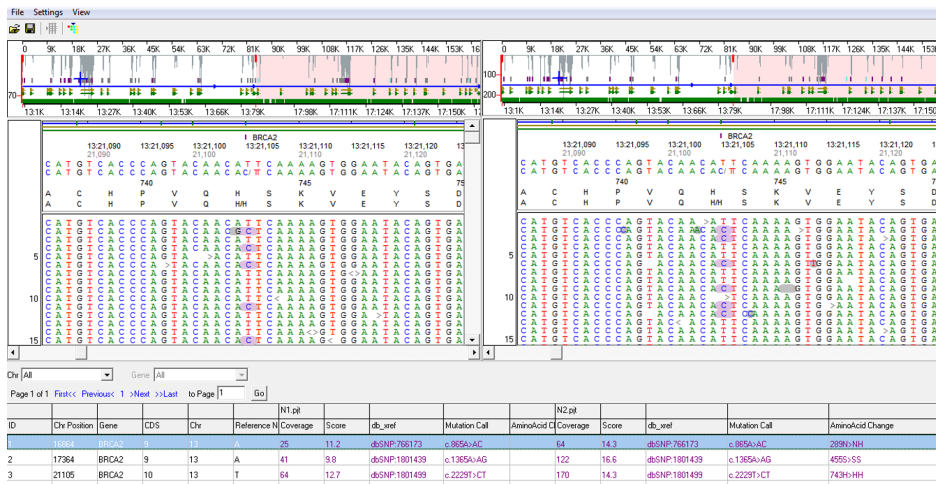


Figure 12: Variant Comparison Tool

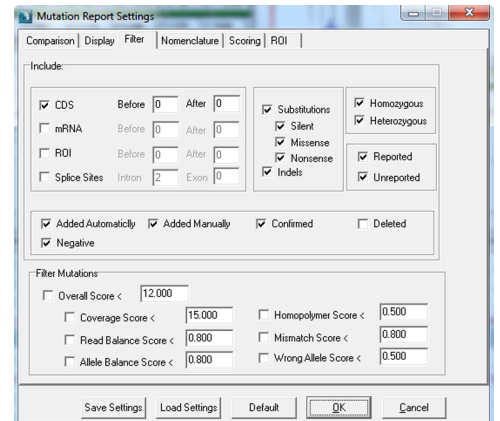


Figure 13: Mutation Report Settings

NextGENe's module for Ion PGM amplicon sequencing analysis outputs quick and accurate results in just a few seconds. In addition to its accuracy and speed, NextGENe also has a variety of tools that can be used for quality control. Once the sequence data is aligned, the Variant Comparison tool can be used to compare samples against one another for discovery of novel variants or to confirm known variations. All of NextGENe's reports are highly flexible and can be exported as tab-delimited text files.

Acknowledgements

Thanks to Halo Genomics for supplying BRCA 1 and BRCA 2 datasets from Ion Torrent PGM™ Sequencer.