Validating a Clinical Cancer Exome Assay: A real world perspective

Natalie Bir, Sean D. McGrath, Ben Kelly, Patrick Brennan, Elaine R. Mardis, Richard K. Wilson, Catherine Cottrell, Vincent Magrini

Institute for Genomic Medicine

The processing profile highlights the need to evaluate sequencing metrics and examine trends associated with sample quality in the pursuit of a standardized clinical process and analytical pipeline with defined pass/fail metrics, while optimizing the sensitivity and specificity of variant detection.

Sample quality, genomic DNA input into the library preparation, library input into the capture, and % loading per lane on the sequencer appear to perform the same with respect to observed % loading and average coverage as seen in Figure 6. If you break down the samples by sample quality we identify three different categories (Figure 7). Although this is a small sample size, this is a good start into understanding how the samples need to be loaded onto the sequencer.

The results from 59 cancer samples highlight the need to evaluate sequencing metrics and examine trends associated with sample quality in the pursuit of a standardized clinical process and analytical pipeline with defined pass/fail metrics, while optimizing the sensitivity and specificity of variant detection.

56 samples were processed using the NEB UltraII FS (enzymatic shearing) library prep method and the IDT xGEN capture method using the IDT Research Exome Panel plus the IDT CNV backbone. These samples were processed using Covaris fragmentation and NEBNext UltraII with the IDT xGEN capture method.

Table 1: Sample processing flow chart.

Predicted vs. observed % per lane

Samples may be run across multiple lanes to achieve higher coverage. The predicted % per lane of a sample is calculated by adding up the % each sample occupies within each lane. The observed % per lane of a sample is dependent on actual total number of reads. Below illustrates loading a set of samples on the Illumina HiSeq4000 and the predicted and observed % per lane calculation.

Observed % per lane vs. Averages Coverage

We observe a strong correlation with observed % loading and average coverage as seen in Figure 6. If you break down the samples by sample quality we identify three different categories (Figure 7). Although this is a small sample size, this is a good start into understanding how the samples need to be loaded onto the sequencer.

Data

Sample quality, genomic DNA input into the library preparation, library input into the capture, and % loading per lane on the sequencer appear to perform the same with respect to observed % loading and average coverage as seen in Figure 6. If you break down the samples by sample quality we identify three different categories (Figure 7). Although this is a small sample size, this is a good start into understanding how the samples need to be loaded onto the sequencer.

Results

Sample quality, genomic DNA input into the library preparation, library input into the capture, and % loading per lane on the sequencer appear to perform the same with respect to observed % loading and average coverage as seen in Figure 6. If you break down the samples by sample quality we identify three different categories (Figure 7). Although this is a small sample size, this is a good start into understanding how the samples need to be loaded onto the sequencer.

Acknowledgements

Thank you Dr. Katherine Miller for aiding in the statistical analysis of the samples. Thanks to the Genomic Services group for processing the samples and generating the data.