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

background

To date, IGM has processed 65 cancer cases using enhanced whole exome sequencing under a research protocol. Here, we report on the transition of the research protocol into a clinical assay to be performed in the CAP/CLIA laboratory. 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. Three samples were processed using Covaris fragmentation and NEBNext UltraII with the IDT xGEN capture method.

Sample Type	No. of Samples	Range of Genomic DNA Peak Size
Fresh Frozen Tissue	21	10,777bp – 60,000bp
Blood Normal	26	13,029bp – 60,000bp
FFPE Tumor Tissue	12	1,271bp – 31,079bp

various sample processing scenarios and pooling

The processing profile highlighted below aid our understanding of standardizing the protocols for more consistent sequencing results.

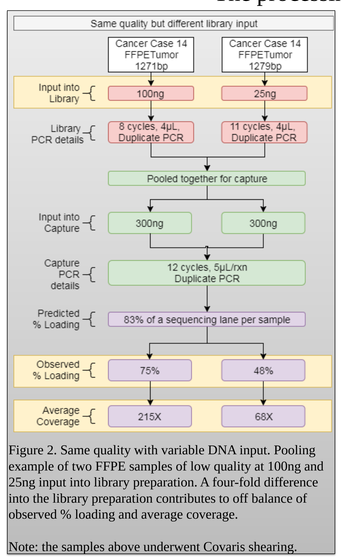


Figure 2. Same quality with variable DNA input. Pooling example of two FFPE samples of low quality at 100ng and 25ng input into library preparation. A four-fold difference into the library preparation contributes to off balance of observed % loading and average coverage.

Note: the samples above underwent Covaris shearing.

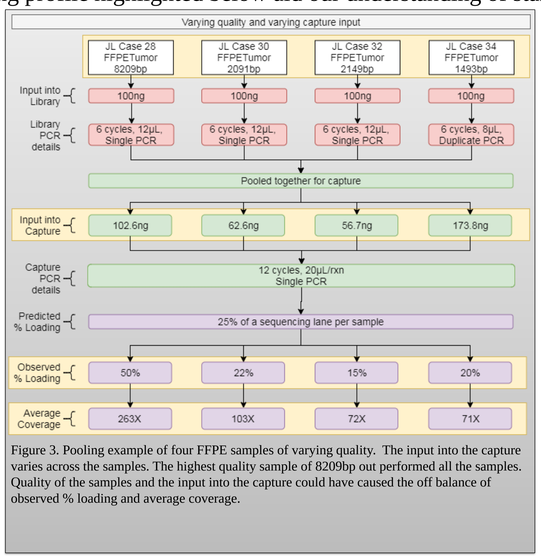


Figure 3. Pooling example of four FFPE samples of varying quality. The input into the capture varies across the samples. The highest quality sample of 8209bp out performed all the samples. Quality of the samples and the input into the capture could have caused the off balance of observed % loading and average coverage.

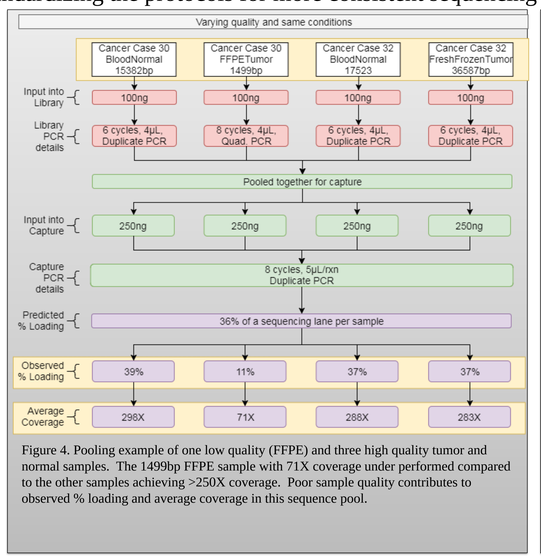


Figure 4. Pooling example of one low quality (FFPE) and three high quality tumor and normal samples. The 1499bp FFPE sample with 71X coverage under performed compared to the other samples achieving >250X coverage. Poor sample quality contributes to observed % loading and average coverage in this sequence pool.

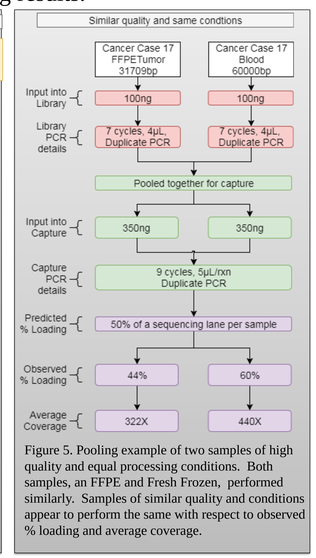
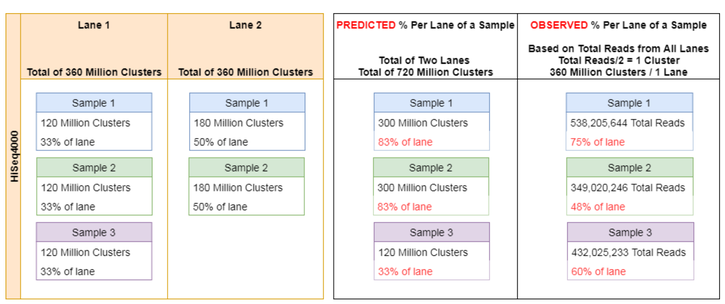


Figure 5. Pooling example of two samples of high quality and equal processing conditions. Both samples, an FFPE and Fresh Frozen, performed similarly. Samples of similar quality and conditions appear to perform the same with respect to observed % loading and average coverage.

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.



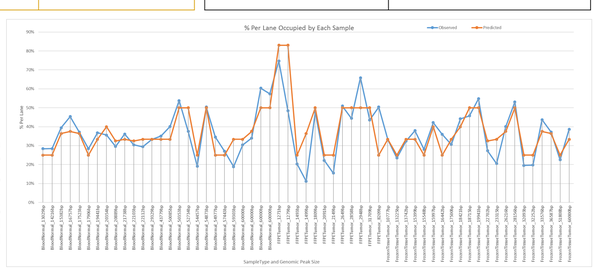


Figure 1. Blood normal, FFPE tumors, and fresh frozen tumors are arranged in ascending order of sample quality. There is a strong correlation of predicted % loading with observed % loading.

data

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.

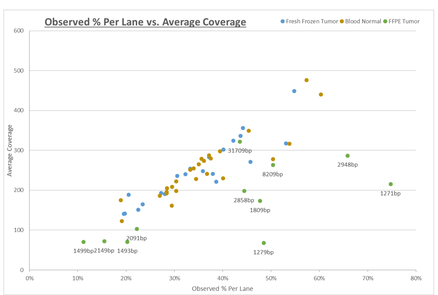


Figure 6. Observed % per lane vs. average coverage for each sample

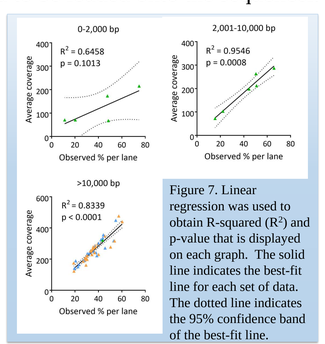


Figure 7. Linear regression was used to obtain R-squared (R^2) and p-value that is displayed on each graph. The solid line indicates the best-fit line for each set of data. The dotted line indicates the 95% confidence band of the best-fit line.

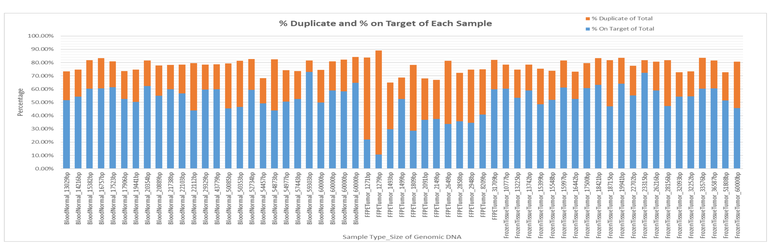


Figure 8. Overall the % on Target on Total Reads and % Duplicate of Total Reads is similar across all the samples with a few outliers on the lower quality samples which happen to be FFPE samples.

results

Sample quality, genomic DNA input into the library preparation, library input into the capture, and % loading per lane on the sequencer all are important factors. Figure 9 is a recommended sample processing flow chart. Cycle conditions are based on averages of what we have done in the past. We will adjust the processing parameters as our sample types (FFPE/frozen tissue/blood), tissue sources, and numbers increase. Optimizing the protocol is an ongoing task.

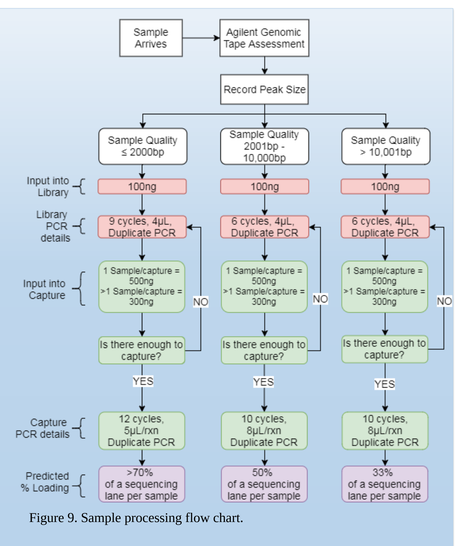


Figure 9. Sample processing flow chart.

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.