Assessment of the Utility of a Novel Massively Parallel Sequencing Platform for Multiplexed Somatic Cancer Assays.

SOPHIE LOW[1], MICHELLE CIPICCHIO[1], DANIEL STOVER[2], NICOLE FRANCIS[1], NING LI[1], MARK FLEHARTY[1], MICAH RICKLES-YOUNG[1], CARRIE CIBULSKIS[1], STACEY GABRIEL[1], NIALL LENNON[1]

[1]Broad Institute of MIT and Harvard, 320 Charles St, Cambridge, MA 02141 [2]Ohio State University Cancer Center, Columbus, OH 43210

Introduction

Tumor profiling using next generation sequencing can provide powerful information to guide patient diagnosis and care, however enabling flexibility in scale can be a challenge for core labs. Cancer research requires flexible scale, with throughput that may be too low to leverage the most cost effective sequencing options available.

Our common cancer use cases vary significantly in scope, with variable sample numbers, target territory and sequencing depth required. The Element sequencer provides an attractive cost profile for moderate yield sequencing.

	Technology	% Genome Sequenced	SNV detection	Tumor Content	Computational Analysis	Cost
Evaluated with Element	ULP-WGS	100%	NA	TFx to ~3%	Easy	
	Genotyping/ ddPCR	<0.001%	~1/1000	VAF	Easy	
Evaluated with Element	Targeted Panel	<0.1% (100s genes)	~1/1000	VAF	Moderate	
	WES	1%	~1/50	TFx to ~0.1%	Moderate	
	WGS	100%	~1/20 (~60X)	TFx to ~1%	Complex	
	Personal/ 'Bespoke' Assay	Pre: 1%-100% Assay:<.001%	~1/100k	TFx to <<0.1%	Complex	

Figure 1. Shallow depth genomes (ULP-WGS - 0.1x) are used to estimate % Tumor Fraction (%TFx) in cfDNA.

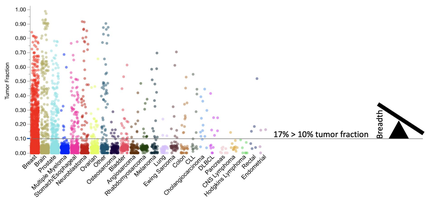
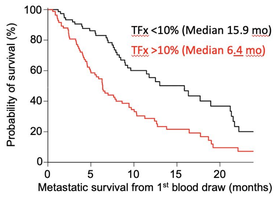


Figure 2. %TFx estimated from ULP-WGS at diagnosis or across treatment course has been shown to be prognostic.



Technology Overview

We are exploring the use of the Element Aviti benchtop sequencer, with 300 Gb output per run. This platform could provide a solution with flexible scale that would enable projects with small batch sizes or small target regions of interest while matching the most advantageous cost.

Read Length	Single Flow Cell Output (Gb)	Dual Flow Cell Output (Gb)	Run Time (hours)	Data Quality
2 x 150	300	600	48	%Q30 > 90
2 x 100	200	400	35	%Q30 > 90
2 x 75	150	300	29	%Q30 > 90
2 x 50	100	200	23	%Q30 > 90
2 x 25	50	100	17	%Q30 > 90

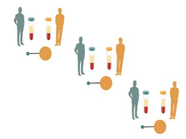
Figure 3. Performance parameters based on full read depth with 1 billion paired-end reads per flowcell (full scan). Read count based on Element control library sequencing. Actual read count might differ based on factors such as library type and preparation.

Tumor Fraction Estimation from Shallow Depth WGS

To evaluate Element performance for our shallow depth whole genome assay, two sample cohorts were analyzed.

Cohort

Dilution Series:
3 cfDNA samples (Breast Cancer) diluted into Healthy Donor cfDNA controls to assess tumor fraction accuracy and limit of detection.



Truth Comparison

Prior Illumina Shallow Depth Genome Results

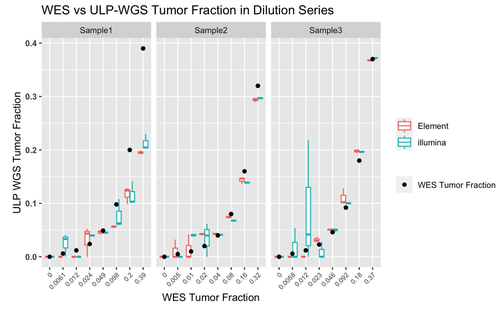
Whole Exome Sequencing and Allele Specific Purity/Ploidy Analysis

Reproducibility Series:
30 Breast Cancer cfDNAs each in triplicate (technical replicates)

Prior Illumina Shallow Depth Genome Results

Quantitative Dilution Series Results

Figure 4. ULP-WGS derived tumor fraction for each sample and platform was compared with the expected tumor fraction. Element and Illumina both show comparable performance for accuracy and reproducibility. Sample 1 indicates a known limitation of the ULP-WGS assay in which the presence of subclonal CNV events and Copy-Neutral LOH result in under-estimation of the % Tumor Fraction relative to exome which enable allele specific CNV analysis.



Reproducibility Cohort Results

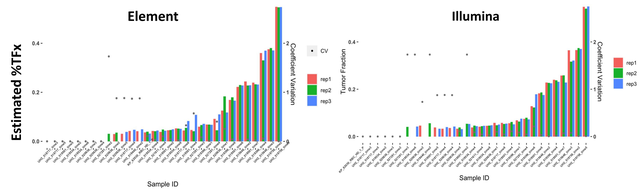


Figure 5. Reproducibility was assessed as the ability to consistently generate concordant % Tumor Fraction estimates across replicates of the same specimen on both Illumina (A) and Element (B) sequencing platforms. 30 clinical specimens and 2 healthy donor samples were sequenced in triplicate. Tumor fraction estimates as well as coefficient of variation (CV) for replicate sets are shown. Lower limit of detection is 3% TFx. Illumina and Element based sequencing results are comparable within the detection range of the assay.

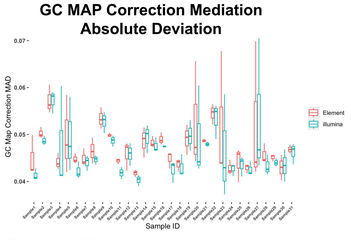


Figure 6. GC Map Correction is the median absolute deviation of the coverage of adjacent copy number bins (1 Mb Range in Genome). Passing criteria for GC Map Correction is <= 0.15. All samples, both downsampled versions and full coverage are under this threshold.

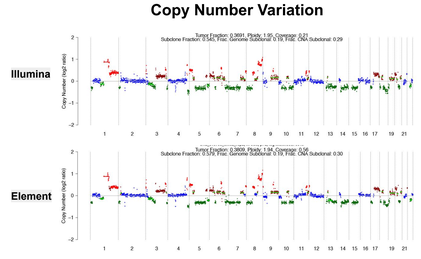


Figure 7. Copy Number Variation of representative sample sequenced on both Illumina and Element, estimated tumor fraction 36.7% and 38.0% respectively. Results show concordance of profiles across platforms.

Targeted Cancer Panel

We evaluated the sensitivity and specificity of sequencing on a custom targeted cancer panel (PanCan). Targeted regions for this panel cover 396 genes prioritized for cancer relevance over 2 Mb of territory.

Samples used to measure sensitivity were well characterized HapMap cell lines pooled together at known proportions. This enables assessment of SNV detection at a range of allele frequencies of with known variation.

- Sample 1: DNA from 20 cell lines pooled in equal proportions
- Sample 2: DNA from 10 cell lines pooled in equal proportions
- Sample 3: DNA from 5 cell lines pooled in equal proportions
- All samples were processed in triplicate

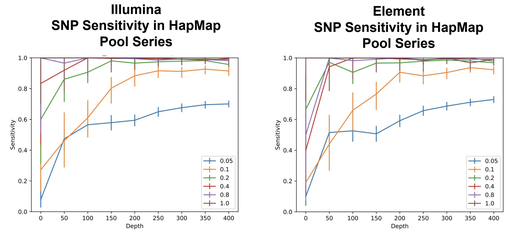


Figure 8. SNP Sensitivity analysis of HapMap Cell Line Admixtures pooled to assess sensitivity of SNV detection at a range of allele frequencies and read depth. Sensitivity across a range of read depth was comparable between sequencing technologies.

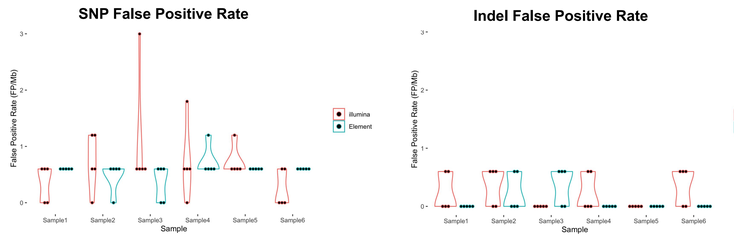


Figure 9. False positive analysis of NA12878 over 396-gene PanCancer Panel at both 1200x and 250x. Variant calling performed with mutect2. Caveats: mutect2 trained and optimized on Illumina data, as well as use of M2 Illumina Exome PON for analysis of this data.

Conclusions

We present an analysis of a Pan Cancer gene profiling panel using deep sequencing on samples with somatic variants at known allele fractions. Additionally we assess the performance of ultra low coverage whole genome sequencing for tumor fraction estimation from liquid biopsy.

The resulting variant calling sensitivity and specificity were comparable or indistinguishable between two sequencing platforms used to generate the data. The use of the Aviti, with the attractive per base cost and flexible scale, represents a solution for the complex challenges limiting the use of NGS in cancer today.

Acknowledgements

The data used in this poster was generated at the Broad Institute. For more information please visit: http://genomics.broadinstitute.org/