Comparative Analysis Of Copy Number Variants In Cell-Free DNA From Metastatic Prostate Cancer Patients Using Internal Standards and Next-Generation Sequencing

David A. Nix[1], Brian H. Dalley[1], Claire H. Hanson[1], Enos Ampaw[1], Matt Larsen[1], Bradley B. Austermiller[2], Thomas B. Morrison[2], Manish Kohli[1]
[1]Huntsman Cancer Institute and the University of Utah, [2]AccuGenomics, Wilmington, NC, USA

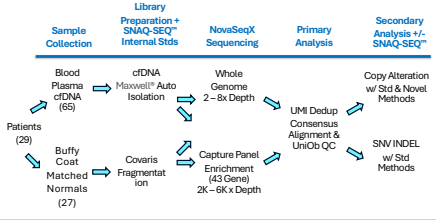
Abstract # ST071

Introduction

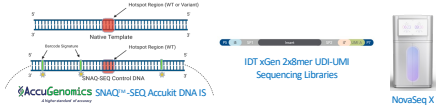
Tracking metastatic treatment response by Next-Generation Sequencing (NGS) of cfDNA provides a simple sample source but lacks methods to standardize results. An often-used analytic approach, internal standards, is being explored for its ability to standardize NGS testing results. Prospective plasma cfDNA from metastatic castrate resistant prostate cancer (PC) patients receiving standard of care treatments at a tertiary level cancer center were tested using a focused gene NGS panel. This presentation compares Copy Number Variant (CNV) detection with and without novel SNAQ™-SEQ Internal Standards (IS). All patients were followed for clinical outcomes.

Methods

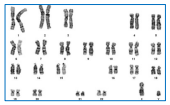
- **Small Capture Panel Design:** A PC specific, 43 gene, hot-spot capture panel was created to identify CNVs, short variants, and a gene fusion in liquid biopsy samples. Each gene contains >=20 unique 150bp capture regions covering most exons.
- **SNAQ™-SEQ IS:** AccuGenomics developed 28 IS to 10 PC driver gene mutations with clinical utility. IS contain unique base change spanning the IS target region every ~80 bases enabling bioinformatic identification. IS are randomly fragmented to 170±30 bases to simulate cfDNA fragment size. The IS input was adjusted to ~8% Variant Allele Frequency (VAF) based on each sample's plasma cfDNA or buffy coat mass (sizes 50-700 bp by TapeStation). After sequencing, IS VAFs were converted to abundance measurement from which copy number and variant plasma/ buffy coat concentration were calculated.
- **Next Generation Sequencing (NGS):** Library preparations (LPs) utilized dual index adapters with unique, inline, 8bp molecular identifiers to enable PCR duplication and error rate reduction through consensus read stack base recalling. Pre (whole genome sequencing WGS) and post capture LPs were sequenced to 2-8x and > 2K x read depth on the NovaSeq X respectively to enable sensitive detection of variants with AFs of >= 0.5%.
- **Analysis:** Alignment and UMI deduplication + consensus error correction were performed using an open-source containerized snakemake workflow. Likewise, a GATK-USeq best practice somatic copy ratio analysis workflow was run on the WGS datasets to produce a standard key of CN calls for comparing CN analysis results from the panel. Five different methods were used to generate panel CN calls, two made use of the IS (AccuGenomics, ManPck) and three without (GATK-USeq, AppPickBstSc, and AppPickGnNm).



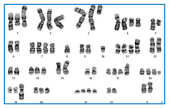
Methods



Normal Karyotype



PC Tumor



(Amparo, C, et. al., 2020 Can Gen and Prot)

PC Patient Cohort	
Age at first paired collection (years)	69 [50, 85]
Albumin (gm. percent)	4.0 [3.50, 4.60]
Serum prostate specific antigen PSA (ng/mL)	5.7 [0.1, 394.9]
Hemoglobin (g/dL)	13.4 [6.0, 15.00]
Serum alkaline phosphatase ALP (U/L)	106 [59, 590]
Gleason score at initial diagnosis:	
<8	11
>=8	13
Missing	1
De novo metastatic stage:	
Metastatic	11 (44%)
Nonmetastatic	14 (56%)
Treatment status at first collection:	
Treatment Naïve	20 (80%)
Not Treatment Naïve	5 (20%)
Patient vital status:	
Dead	22
Time to death (months)	23.2 [3.85, 43.43]
	Median [Min, Max]

Results

Sequencing QC Metrics

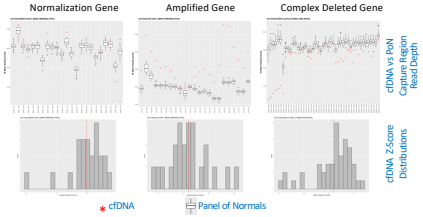
Sample Set	# Fastq Reads	# Alignments Pre Consensus	# Alignments Post Consensus	Fold Consensus Reduction	Fraction Passing QC	Fraction On Target
Panel cfDNA	182,452,880	182,181,845	22,825,228	8.4	0.97	0.54
Panel Normal	179,912,091	179,667,107	27,119,295	6.6	0.97	0.64
WGS cfDNA	231,712,674	231,621,536	175,680,395	1.8	0.99	NA
WGS Normal	248,518,172	248,338,568	206,159,011	1.2	0.99	NA

Sample Set	Mean Insert Size	Fraction Overlapping BPs	# Unique BPs	Fraction Q30 BPs	Mean on Target Coverage	Coverage @ 0.9 of Target BPs	Coverage @ 0.95 of Target BPs
Panel cfDNA	192.9	0.56	1,933,168,837	0.97	3654.9	2174.9	1754.1
Panel Normal	186.8	0.51	2,267,641,017	0.98	5040.3	3488.8	2851.1
WGS cfDNA	166.9	0.68	14,359,303,642	0.98	4.9	1.8	1.1
WGS Normal	179.4	0.56	17,441,535,019	0.98	6.1	2.9	2.1

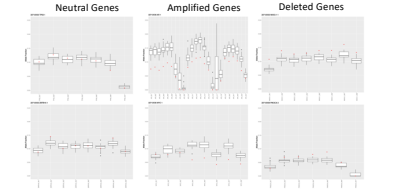
Samples with QC issues (e.g. low coverage, high read depth variability, failed concordance)	
Panel	3 of 92
WGS	1 of 92

Results

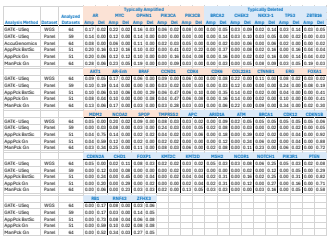
Gene Specific Read Depth Boxplots and Z-Score Histograms – Examples



Gene Specific SNAQ™-SEQ Internal Standards Allele Fraction – Examples



Copy Alteration Metrics – Genes



Results

Copy Number Method Comparison

Means	10 Genes w/ AccuGenomics IS					All 43 Genes from Panel				
	AccuGenomics	GATK	AppPckBstSc	AppPckGnNm	ManPck	GATK	AppPckBstSc	AppPckGnNm	ManPck
Fraction Datasets Analyzed (N=64)	1.00	0.92	0.80	0.80	1.00	0.92	0.80	0.80	1.00
True Positive Rate (TPR), Recall, Sensitivity	0.19	0.48	0.50	0.42	0.58	0.42	0.60	0.52	0.52
False Positive Rate (FPR)	0.02	0.04	0.19	0.08	0.08	0.09	0.23	0.17	0.16
Specificity, 1-FPR	0.98	0.96	0.81	0.92	0.92	0.91	0.77	0.83	0.84
Precision, Positive Predictive Value (PPV)	0.56	0.64	0.21	0.32	0.49	0.26	0.14	0.13	0.19
False Discovery Rate (FDR), 1-PPV	0.44	0.36	0.79	0.68	0.51	0.74	0.86	0.87	0.81
Accuracy	0.89	0.90	0.80	0.86	0.87	0.88	0.76	0.81	0.83
Copy Analysis Methods									
GATK-USeq	GATK best practice somatic variant copy ratio analysis workflow with T/N USeq filtering. Designed for exome and large capture panel datasets, not small where normalization is an issue. Used for deriving the WGS key, thus possibly bias.								
AccuGenomics	Copy alteration calling using SNAQ™-SEQ Internal Standards(IS) and AccuGenomic methods.								
AppPckBstSc (Application Picked, Best Score Normalization)	Small panel novel analysis where the best scanned scalar (>=3 genes w/ >=0.75 of abs z-scores <= 3) was used to normalize the cfDNA sample. The PoN was normalized to 1000.								
AppPckGnNm (Application Picked, Gene Regions Normalization)	Small panel novel analysis where the gene capture regions identified in AppPckBstSc as not copy altered were used to re normalize the particular cfDNA and PoN to 1000.								
ManPck (Manual IS Picked, Gene Regions Normalization)	Small panel novel analysis where the genes identified as not copy altered using the AccuGenomic IS were used to normalize the particular cfDNA and PoN to 1000.								

Conclusions

- Presented here is a comparison of several computational methods under development that make use of AccuGenomics IS for copy number analysis from a small capture panel where genome wide normalization is not possible.
- The fraction of samples showing copy alteration changes in the WGS datasets appear as expected with amplification of MYC (0.22), COL22A1(0.22), NCOA2(0.2), AR(0.17), and AR-Enhancer(0.17). Whereas NOTCH1(0.25), RB1(0.17), TP53(0.14), and NKX3-1(0.14) are deleted. The panel analysis largely replicate the WGS findings.
- Of the three methods that did not utilize IS VAFs, the standard GATK-USeq somatic copy ratio analysis performed the best with an average recall of 0.48 and average precision of 0.64. This precision is not clinically viable. Too many false positives.
- The IS copy analysis method that utilized both the IS VAFs and read depth data worked the best with a higher average recall of 0.58 and average precision of 0.49. The IS VAF alone method also performed well but utilized very stringent thresholds leading to call sets that match genes in the key but with many false negatives. Analysis is underway with relaxed thresholds.
- Copy analysis with small panel capture designs looks promising when utilizing IS.
- All analysis workflows and applications utilized here are open-source and available from https://github.com/HuntsmanCancerInstitute

Limitations

- The WGS derived truth datasets approximate the true copy number alterations in the cfDNA samples. Thus, the confusion matrix statistics are an approximation.
- The IS were designed primarily for calculating the concentration of SNV/INDEL in plasma and thus are not entirely optimal for copy analysis, which typically requires 4 to 8 IS per gene.

Presented at AMP 2024 | Vancouver, BC