Monitoring NGS plasma ctDNA sequencing using SNAQ-SEQ ONCO1LB internal controls.

Deharvengt SJ1, Green DC1, Austermiller B2, Morrison T2, Tsongalis GJ1
Laboratory for Clinical Genomics and Advanced Technology (CGAT)
1Department of Pathology and Laboratory Medicine, DHMC and NCCC, Lebanon, NH and Geisel School of Medicine, Hanover, NH
2 AccuGenomics, Wilmington, NC

Dartmouth-Hitchcock

ABSTRACT

Background: The use of Spike-in controls in Next Gene Sequencing (NGS) quality control (QC) can help monitor the testing process in individual samples which is so necessary for tumor molecular profiling tests from liquid biopsies where using VAF to monitor treatment response is sensitive up to a 2-log variation in background cfDNA. We describe the use of Standardized Nucleic Acid Quantification spike-in controls for NGS that provide an additional QC for variant calling and have the potential to standardize variant monitoring in liquid biopsies.

Methods: 30,000 SNAQ-SEQ Internal Standard (IS) (SNAQ-SEQ ONCO1LB; AccuGenomics) controls and 50 ng SeraCare (SC) 5% cfDNA (0710-0528), providing a fixed level of mutations, were added in increasing amounts (0 to 500 ng) of SC cfDNA normal (0710-0533) into 2ml of DNA negative plasma. The IS mixture is synthetic DNA, fragmented to simulate cfDNA and distinguished from genomic DNA by unique nucleotide alterations every 50bp. Targeted libraries were prepared from cfDNA extracted using the MagMAX™ Cell-Free Total Nucleic Acid Isolation Kit and the Oncomine™ Pan-Cancer Cell-free assay reagents (ThermoFisher). A modified reference genome, design and hotspot bed file directed the Ion-Torrent variant caller to report all alts for genomic and control positions. A SNAQ-SEQ R-script determined how significant each genomic variant was above the IS derived NGS background error. The NT alt count and the corresponding IS read depth at the same position was used to estimate the copies of variant per ml of plasma.

Results: There were seven SC 5% variants covered by the SNAQ-SEQ IS. The dilution series had the same input level of each variant with the resulting VAF ranging from 6% to 0.2%. ThermoFisher pipeline detected 54 of 56 possible variants and 116 false positives (FP). SNAQ-SEQ confirmed 50 of these variants and eliminated 109 FP. All four false negatives that arose from the SNAQ-SEQ analysis were a consequence of poor IS coverage and should be rescued with a redesigned IS. 5 of 7 SNAQ-SEQ FP were at one site and could be eliminated with further pipeline optimization, allowing to confirm all 54 true positives and reduce the 116 FP to 2. The %CV of each variant abundance was <25% with an average abundance of 200 copies, half of the abundance variability arose from sample stochastic effects. VAF reporting of these variants ranged >10-fold.

Conclusion: The SNAQ-SEQ internal standard provides an orthogonal QC for variant calling, pushing VAF sensitivity below the recommended 0.5% cutoff. Further, SNAQ-SEQ provides near dPCR like accuracy to the sequencing platform, standardizing LB variant reporting to abundance per ml of plasma. The IS would also eliminate the need of an extra control sample to monitor the quality from extraction to analysis.

INTRODUCTION

- Identification of actionable mutations in circulating tumor DNA (ctDNA) enables gene-targeted therapy of solid tumors based on a simple blood test.
- Because of the low and variable abundance of circulating tumor DNA (ctDNA), it is critical to ensure that the workflow of the assay performed consistently.
- Standardized Nucleic Acid Quantitation – Sequencing (SNAQ-SEQ) controls are spike-in control mixtures developed for clinical genomic applications such as Next Generation Sequencing (NGS).
- SNAQ™-SEQ is a proprietary QC approach that utilizes synthetic DNA internal standard mixtures (ISM™) spiked into every sample prior to the NGS-library preparation step. These standards thus undergo the same processes and reaction conditions as the sample to provide an ideal run control for NGS assays.

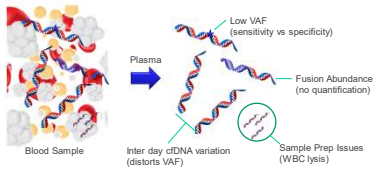

Figure 1: SNAQ-SEQ Internal Standards address issues with NGS testing of Liquid Biopsies.

Aim:

- **Background levels of normal DNA can vary over 100-fold in patient samples and leads to a concordant increase in VAF measurement variability and limits standardization. We created a proof-of-principle experiment to demonstrate SNAQ-SEQ variant QC and ability to standardized variant quantification in liquid biopsies.**

METHODS

SNAQ-SEQ control design:

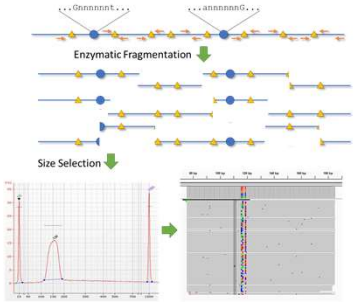

Figure 2: SNAQ-SEQ IS match reference genome sequence except in select base changes (triangles) used for bioinformatic separation. The base changes do not overlap known SNP or primer binding sites. Two locations use 6 degenerate bases (circles) as complexity capture controls, one for each primer pool (top vs bottom arrows). Every actionable mutation in targeted panel has a control sequence. For cfDNA, dsDNA controls are enzymatically fragmented and sized (Agilent Trace, 154 ± 13 bases) and diluted to 10,000 copies per target. When sequenced (inserted IGV image) the count of each unique sequence may be used for complexity capture.
SNAQ-Seq control: fragmented ONCO1LB used for Limit of Blank (LOB) for EGFR, KRAS, PIK3CA SNV, provide CNV estimates for ERBB2, and two fusions eml4-alk and tpm3-ntrkk1, and complexity capture estimates.

Workflow:

Expected VAF	5%	3.3%	2.5%	1.7%	1%	0.7%	0.6%	0.5%
Seraseq® ctDNA Complete™ Mutation Mix AF5% (ng)	50	50	50	50	50	50	50	50
Seracare cfDNA normal (ng)	0	25	50	100	200	300	400	500

Figure 3: Oncomine Pan-cancer liquid biopsy assay workflow. To simulate the cfDNA variation, 50 ng of 5% Seracare ctDNA complete (Cat#: 0710-0528) were mixed with an increase amount (0 to 500 ng) of normal Seracare (Cat#: 0710-0533) into 2 ml of DNA free plasma. Eight different dilutions were made. 30000 copies of SNAQ-SEQ IS (AccuGenomics, Onco1LB, cat# 1207) were added into each sample before extraction with the MagMax cell-free total nucleic acid isolation kit (ThermoFisher).
Libraries were prepared using 20 ng of total nucleic acid, following the manufacturer's protocol. The Chef instrument (ThermoFisher) was used for both library preparation, enrichment and chip loading. The 540 chips were sequenced on the Ion Genestudio S5 system plus.

Table 1: Dilution series.

SNAQ-SEQ Analysis Pipeline:

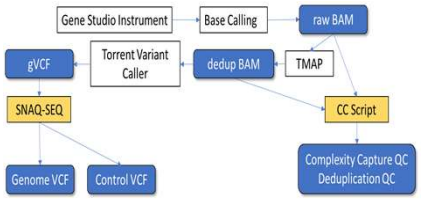

Figure 4: SNAQ-SEQ Analysis Pipeline for Ion Torrent Platform. White boxes indicate normal base calling, alignment and variant calling of Torrent Server, orange boxes SNAQ-SEQ additional steps to produce indicated data files (blue boxes). Three altered files are uploaded to the server: *reference_genome.fasta* and *designed.bed* are modified to include the control regions and the *hotspot.bed* is modified to report all genome and control positions (i.e. a gVCF). The **CC script** collects and counts the unique complexity control regions, comparing before (raw BAM) and after deduplication (.aln_not_needed.bam + .aln_needed.bam) to create a table indicating count of unique sequences added and their duplication rate before and after molecular family processing. The **SNAQ-SEQ** script examines each PASS call and outputs a genomic and control VCF changing SNV PASS calls to lowPET when not significantly different than NGS background error. Significance above background was determined by examining the same variant in the control using a Poisson Exact Test with a Bonferroni corrected 5% alpha (5% / PASS call counts). The Control VCF can be used to examine how well the pipeline is eliminating all errors as there should be SNV in the control. The variant copies per sample was calculated from the VCF alt count and depth using the formula NT_alt_count / IS_depth * IS_input_copies.

RESULTS

SNAQ-SEQ control in variant calling accuracy

SNAQ-SEQ IS should have no variants, if a variant is detected it represents a sequencing error. Any significant difference between sample and control VAF indicates a true variant, insignificant difference suggests a false positive. SNAQ-SEQ analysis used an unoptimized pipeline and is expected to be more accurate with minor tweaking to significance cutoffs and IS optimization.

- Experiment: simulated elevating background level of cfDNA while keeping ctDNA variants constant.
- Detected titrating variants:
  - 6 expected CLINVAR
  - 1 unknown SNV

Table 2: Comparison of the two pipelines

	Torrent Server Pipeline	SNAQ-SEQ IS analysis
False Positives (FP)	115#	7 (2*)
True Positives (56)	55	50 (53^)
False Negatives (FN)	1	6 (3^)
Positive predictive value	32%	88%
Sensitivity	98%	89% (95%^)

#All but 5 SNV of the 115 FP were unique and had VAF that did not respond to normal cfDNA dilution.
*Five FP were the same variant, either real or black listable.
^PIK3A IS redesign will improve coverage and eliminate at least 3 FN.

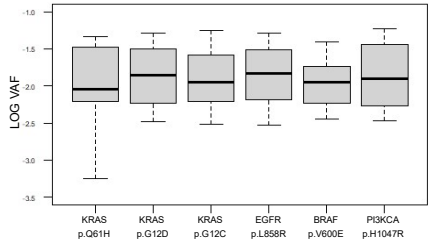

Figure 5: Boxplot representing the VAF of each variant in each dilution. Despite having identical tumor load in each sample their VAF ranged over 10-fold due to the altered background cfDNA levels.

SNAQ-SEQ control to count variants/ml in sample

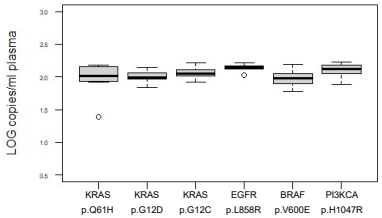

Figure 6: Boxplot representing copies/ml of plasma. SNAQ-SEQ was able to estimate the copies of each variant in the sample with near digital PCR level of accuracy (<25% CV). There was one outlier that was also detected by VAF which likely arose to stochastic sampling errors and speaks to the need properly plan for this limitation in NGS based testing. SNAQ-SEQ quantification standardizes ctDNA measurements by providing a method for measuring variants copies per ml of plasma and normalizes measurements between labs, users, reagent lots, and instruments.

CONCLUSIONS

- **SNAQ-SEQ spike-in internal standard controls for inter-sample variability, stochastic and technical errors.**
- **SNAQ-SEQ IS enable highly accurate measurement of background error rates. SNAQ-SEQ provides limit of blank measurement for EACH region in EACH specimen which dramatically improves variant caller accuracy—particularly at low variant allele frequencies that are often subject to high noise and background.**
- **SNAQ-SEQ IS is able to estimate variant concentration in patient samples, eliminating impact of background cfDNA or genomic DNA contamination on measurement. This method of variant quantification would standardize variant monitoring or minimal residual disease levels and has advantages over off target internal controls in that SNAQ-SEQ IS better monitor yield through biochemically mimicking target region and does not require altering NGS library preparation to add off target probes.**