Standardizing NGS Plasma ctDNA Measurements Using SNAQ-SEQ ONCO1LB Internal Controls.

Deharvengt SJ[1], Green DC[1], Austermiller B[2], Morrison T[2], Tsongalis GJ[1]
Laboratory for Clinical Genomics and Advanced Technology (CGAT)
1 - Department of Pathology and Laboratory Medicine, DHMC and NCCC, Lebanon, NH and Geisel School of Medicine, Hanover, NH
2 - AccuGenomics, Wilmington, NC

ABSTRACT

Background: Spike-in controls are widely used in molecular diagnostic assays for various quality purposes such as ruling out PCR inhibition. Similar controls are not routinely used in Next Gene Sequencing (NGS) methods but could help monitor the NGS process in individual samples. This is even more important for NGS liquid biopsies characterization of tumor mutation profiles from plasma, especially when VAF is sensitive to up to a 2-log variation in background cfDNA. We describe the use of Standardized Nucleic Acid Quantification spike-in controls for NGS (SNAQ-SEQ ONCO1LB; AccuGenomics) that have the potential to improve the traditional performance limits by providing direct sensitivity and specificity control for every target in every sample, and complexity capture QC.

Methods: cfDNA were extracted using the MagMAX™ Cell-Free Total Nucleic Acid Isolation Kit (Thermo Fisher). The spike-in control mixture is synthetic DNA that can be distinguished from genomic DNA by unique mono- or di-nucleotide alterations every 50bp. The controls were spiked into the plasma before extraction or before the start of the library preparation at an equimolar ratio. Seraseq™ cfDNA complete and ctDNA mutation mix at different allele frequencies (from 2% to 0.1%) were used as reference materials (SeraCare). Extraction were also made using healthy donor plasma. Targeted libraries were prepared using Oncomine™ Pan-Cancer Cell-free assay reagents (ThermoFisher). A modified reference genome, design.bed and hotspot.bed file directed the Ion-Torrent variant caller to generate a gVCF that included all alts for genomic and control positions. A SNAQ-SEQ R-script analyzed the VCF using a Poisson Exact Test to determine how significant each genomic variant was above NGS background error, generating a new VCF with the SNAQ-SEQ analysis appended to the GT field. The NT:IS ratio extracted from the VCF was used to estimate the copies of variant per ml of plasma. Further, inclusion of a complexity control (CC) provides and indication of how much of the sample template was captured in sequence and will provide a QC to monitor library preparation performance.

Results: Sequenced libraries with or without spike-in controls showed the same quality metrics. The spike-in controls did not affect variant call either when added before extraction or before library preparation(data not shown). The distribution of the complexity control sequence replicates (4064 possible sequences) followed a Poisson distribution (>0.9 Coefficient of Determination). There was no detectable bias arising by oligo synthesis as each base ranged from 24-27% across the six positions, supporting the assumption that the complexity control sequences are uniformly distributed, further. The complexity control indicated the testing procedure led to a 17% capture of plasma CC template. ctDNA variants and abundance per ml plasm demonstrated SNAQ-SEQ variant calling QC and reporting variant copies per ml of plasma.

Conclusion: These preliminary experiments demonstrated SNAQ-SEQ internal standard could be spiked-into plasma samples to monitor NGS library capture efficiency, an independent variant calling QC and abundance in plasma. These controls will provide essential missing QC for NGS tests attempting to detect low ctDNA in liquid biopsy samples.

INTRODUCTION

- Identification of actionable mutations in circulating tumor DNA (ctDNA) enables gene-targeted therapy of solid tumors based on a simple blood test.
- Due to the low and variable amount of circulating tumor DNA (ctDNA), it is critical the assay is 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 standards mixtures (ISM™) spiked into every sample prior to the NGS-library preparation step. These standards undergo the same processes and reaction conditions as the sample to provide an ideal run control for NGS assays.

Aim:

- In this proof-of-principle study, we examine the application of spike-in control to monitor the entire workflow from extraction to data analysis of an NGS-based liquid biopsy assay.

METHODS

SNAQ-Seq control design:

Figure 1: SNAQ-SEQ controls 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 (inset IGV), 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:

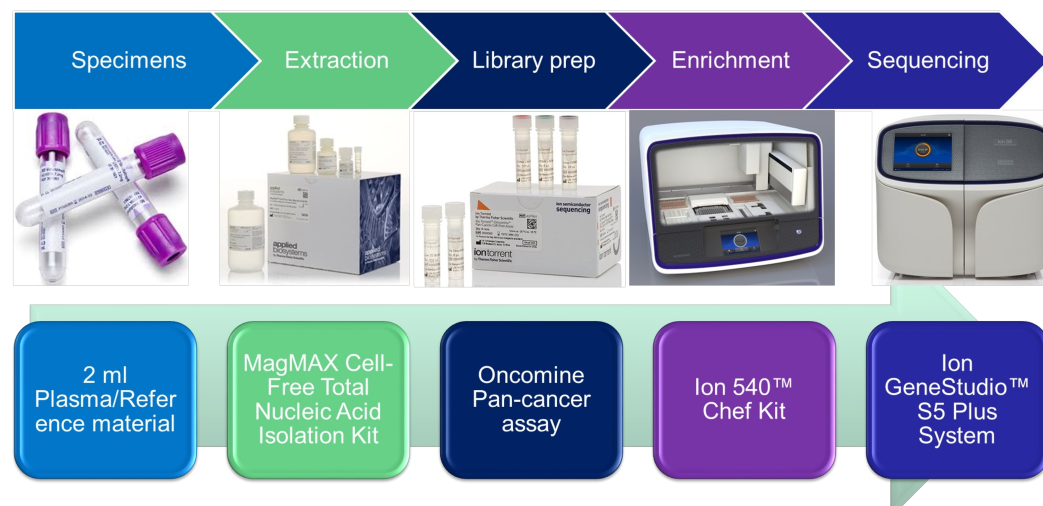

Figure 2: Oncomine Pan-cancer liquid biopsy assay workflow. Two mL of starting material were extracted using the MagMax cell-free total nucleic acid isolation kit (ThermoFisher). Seracare ctDNA Reference Material v2 and healthy human plasma were used. The libraries were prepared using 20 ng of DNA, following the manufacturer's protocol. The Chef instrument (ThermoFisher) was used for both library preparation, enrichment and chip loading. The Ion 540 chips were sequenced on the Ion Genestudio S5 system plus.

SNAQ-SEQ Analysis Pipeline:

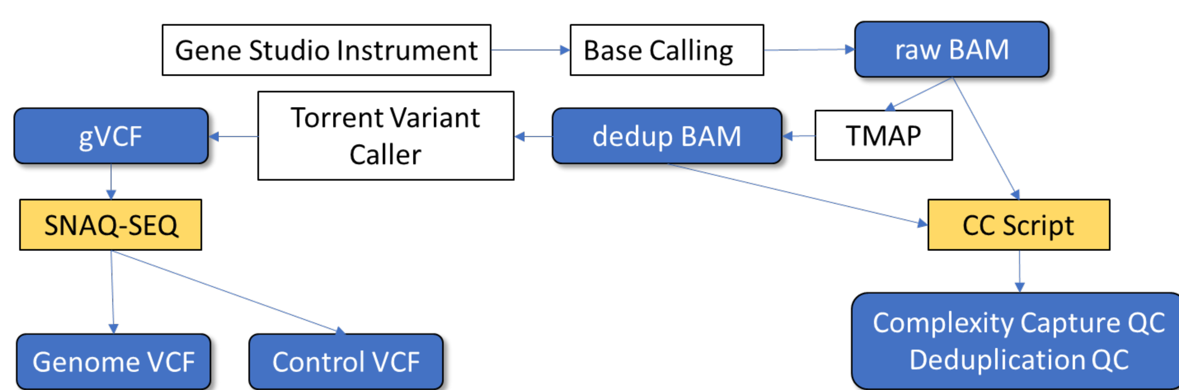

Figure 3: SNAQ-SEQ Analysis Pipeline for Ion Torrent Platform. White boxes indicate normal base calling, alignment and variant calling of Torrent Server, orange boxes SNAQ-SEQ additional steps to produce indicated data files (blue boxes). Three altered files are uploaded to the server: *reference_genome.fasta* 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 not be SNV in the control.

RESULTS

Can SNAQ-Seq control be used to monitor the entire workflow starting from extraction?

The SNAQ-SEQ control was spiked into 2 ml of Seracare ctDNA Reference Material v2 before or after sample extraction. After library preparation and sequencing, data were processed with SNAQ-Seq analysis pipeline. We compared the complexity control (CC) yields spiked before extraction or before library preparation and it was similar, showing that the control is extracted and detected. Each extraction was performed in triplicate, finding similar results each time, showing reproducibility and great consistency.

Spike in (Step)	Estimated copies added	Detected after deduplication
Extraction-1	5,051	5,804
Extraction-2	5,405	5,143
Extraction-3	5,291	5,051
Library-1	6,000	7,985
Library-2	6,000	6,263
Library-3	6,000	6,161
N/A	0	0
N/A	0	0

Table 1: Comparison of the complexity control (CC) yields spiked before extraction or before library preparation.

Does plasma affect the SNAQ-seq control?

The SNAQ-SEQ control was spiked into 2 ml aliquot of fresh mixture of different plasma samples at a different amount from 10,000 to 100 copies. After library preparation and sequencing, data were processed with SNAQ-Seq analysis pipeline. We estimated the amount of detected control after deduplication in each condition. The detected amount was proportional to the added amount of SNAqSeq, showing that the control is extracted and detected. Each extraction was performed in duplicate, finding similar results, showing reproducibility. Plasma seems to affect the consistency.

Estimated copies added in the plasma	Estimated copies present in the extract	Detected after deduplication
10,000	840	1,727
10,000	733	1,623
1,001	88.8	112
1,001	76.1	158
316.9	23.5	22
316.9	24.5	42
100.3	7.4	11
100.3	7.6	10

Table 2: SNAQ-seq Titration of in human plasma.

Is the distribution of bases random so the control can be used to make assumptions about complexity capture?

A

	Extraction step			Library Step		
	1	2	3	4	6	8
Missing count	1431	1624	1654	1128	1485	1436
% missing	35%	40%	41%	28%	37%	35%
CC input	15000	15000	15000	6000	6000	6000
Captured Input	4242	3728	3653	5209	4091	4228
%CC	28%	25%	24%	87%	68%	70%

B

Position	A	G	T	C
1	24%	27%	26%	24%
2	25%	24%	28%	22%
3	24%	26%	26%	24%
4	25%	24%	26%	24%
5	26%	23%	26%	25%
6	26%	23%	27%	24%
Mean	25%	25%	27%	24%

Table 3: A. Complexity capture **B.** Distribution of A, G, T and C among 6 degenerate CC positions.

The quantity of template captured as sequence is an important QC for overall library efficiency. Unlike Unique Molecular Indexes which indicate how many captured templates are present, the CC QC will indicate how much of the original sample was captured as sequence. The % complexity capture (**%CC**) was estimated by comparing how many CC template were added to sample (**CC INPUT**) and how many CC were detected in the sequence (**CAPTURED INPUT**). The abundance of CC controls was estimated by noting the frequency of CC replicates followed a Poisson distribution (R^2>0.90), allowing the CAPTURED INPUT to be calculated from the **FRAC MISSING** using the formula 4064/ln(1/FRACTION MISSING). FRAC MISSING is the count of CC sequences missing (out of a possible 4064). The PLASMA samples %CC averaged 22% +/- 2% whereas the DNA samples averaged 75% +/-8%. Presumably, the lower PLASMA %CC resulted from purification and volumetric losses. As a QC, significant deviations from %CC derived during test development would indicate method drift.

Can SNAQ-Seq controls be used to QC variants and measure the abundance of a ctDNA variant per ml of plasma?

				VAF lowPET			VAF PASS				ALT Counts				IS Coverage				Genome copies/ml plasma				%CV
CHROM	POS	REF	ALT	10K	1K	1K	10K	10K	1K	1K	10K	10K	1K	1K	10K	10K	1K	1K	10K	10K	1K	1K	
chr12	25398303	G	A		0.1%																		
chr17	37879562	C	T			0.1%																	
chr17	37879588	A	G				29.3%	27.8%	31.1%	30.1%	860	971	1059	1047	1992	2182	154	160	2159	2225	3438	3272	28
chr3	178952020	C	T				0.9%	0.8%	1.1%	0.7%	28	30	24	36	1369	1397	120	101	102	107	100	178	17
chr7	55242462	C	T	0.1%																			
chr7	55249063	G	A				31.8%	31.5%	31.5%	32.7%	587	748	763	720	1494	1425	108	145	1965	2625	3532	2483	29
chr7	55259450	C	T				2.7%	2.3%	3.4%	3.3%	95	92	133	126	1728	1951	112	162	275	236	594	389	36

Table 4: SNAQ-SEQ QC of variants and estimation of variant copies per ml plasma

CHROMOSOME of variant
POSITION of variant
REFERENCE BASE of variant
ALTERNATE BASE of variant
lowPET: VAF for variants not significantly above background
PASS: VAF for variants significantly above background
10K, 1K: Samples with 10,000 or 1,000 copies of SNAQ-SEQ ONCO1B
Alt Counts number of variant reads
IS Coverage number of internal standard reads
Genome copies/ml plasma is the estimated copies of variant DNA in the plasma sample
%CV is a normalized variant estimate STD x MEAN x 100%

SNAQ-SEQ enables accurate abundance measurements in samples because once control is added to sample, the ratio of genome:variant:IS is maintained through sequencing. The ability to measure variant abundance was demonstrated using 2 ml of residual patient plasma mixture spiked with 10K or 1K Onco1B SNAQ-SEQ control prior to cfDNA purification, sequenced and variants detected using method described in figure 3. Each table row indicates the Ion Torrent pipeline variants, the VAF grouped by SNAQ-SEQ QC analysis indicating the variant was not or was significantly above NGS background error: lowPET, PASS, respectively. For the PASS variants, the copies of variant per ml of plasma was calculated using the formula: ALT_counts / IS_coverage * IS_INPUT_copies / ml plasma. The level of variants ranged from 100 to 2159, consistent with a plasma sample created from a mixture of patient retains. The %CV for the copies/ml of each variant ranged from 17-36% between all the samples.
► These results demonstrate how SNAQ-SEQ measures ctDNA abundance per ml plasma.

CONCLUSIONS

- ➢ **SNAQ-SEQ spike-in reference standards covary with sample and enable ctDNA / ml plasma abundance measurement and independent QC of variants.**
- ➢ **Complexity capture should provide an independent biochemical QC to ensure every sample meets the required sensitivity by flagging samples whose complexity capture is outside a nominal range established during testing validation**
- ➢ **Controls IN your samples mean no wasted lanes for reference samples**