Replacing External Reference Materials with Internal Standards for Next Generation Sequencing

Sophie J. Deharvengt[1], Donald C. Green[1], Brad Austermiller[2], Parth S. Shah[1], Tom Morrison[2], Gregory J. Tsongalis[1].

1. Clinical Genomics and Advanced Technology, Department of Pathology and Laboratory Medicine, Geisel School of Medicine at Dartmouth, Dartmouth Hitchcock Medical Center, Lebanon, NH, USA
2. AccuGenomics, Wilmington, NC

INTRODUCTION

- Identification of actionable mutations in circulating tumor DNA (ctDNA) enables gene-targeted therapy for solid tumors based on a simple blood test.
- Liquid biopsy comes with technical challenges:
  - ➢ The use of VAF to describe variants is not standardized due to the >100-fold variability of normal cell-free DNA (cfDNA) in liquid biopsy samples.
  - ➢ The deep sequencing required to detect low variant levels can reduce sample throughput.
- SNAQ™-SEQ spike-in standards are highly multiplexed mixtures of synthetic internal standard (IS) constructs to targeted regions of highest clinical importance. They are customizable and formulated to biochemically mimic the form and function of cfDNA, acting as a per sample limit control and to support variant reporting standardized to plasma concentration.

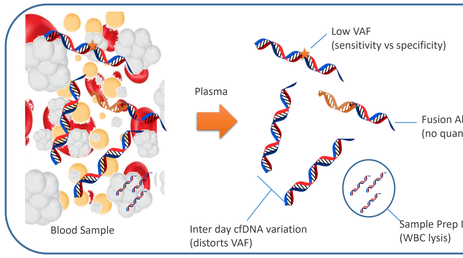
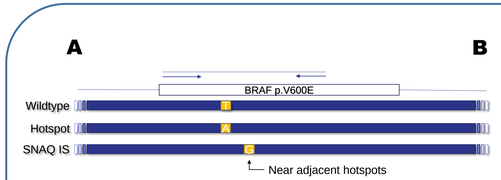


Figure 1: SNAQ™-SEQ Internal Standard (IS). SNAQ™-SEQ IS address issues with NGS testing of liquid biopsies: Variable normal background and low VAF targets.



KRAS	p.G12D/C
EGFR	p.L858R
EGFR	p.T790M
BRAF	p.V600E
AKT1	p.E17K
ERBB2	p.A775_G776insYVMA
KIT	p.D816V
NRAS	p.Q61R
GNAS	p.R201H
BRCA2	p.G3076E/V
ESR1	p.D538G; p.Y537S

Figure 2: SNAQ™-SEQ Internal Standard (IS).
A. SNAQ™ IS are synthetic constructs, each with a unique variants positioned near known hotspot positions. The IS variants can be non-synonymous like those listed in 2B, or synonymous, requiring their addition to the pipeline's whitelist.
B. The Accukit IS was pre-fragmented dsDNA to a mean length of 170 nucleotides to the 11 exonic regions containing the listed variants.

Aim: Testing the IS control as a quality metrics for variants in 11 exonic regions.

AMP2023 ANNUAL MEETING & EXPO
Poster # TT063

METHODS

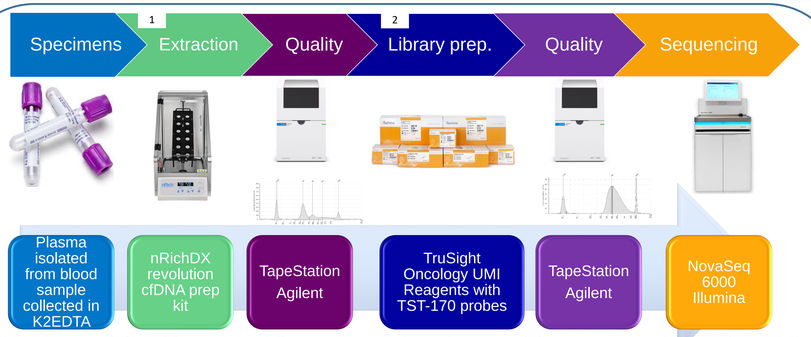


Figure 3: TruSight with UMI Workflow. Two mL of starting material were extracted using the nRichDX cell-free revolution cfDNA prep kit. The Accukit was either spiked in before extraction at 250 copies for 2 mL (1) or just before the library preparation (2). Seraseq® ctDNA complete material and human plasma were used. The libraries were prepared using 10 to 30 ng of cfDNA, following the manufacturer's protocol. The libraries were qualified and quantified using the TapeStation, then sequenced on the NovaSeq6000.

Sample	C1	C2	C3	P1	P2a	P2b	H
Description	Seracare control			Plasma Sample			Healthy Donor
Starting DNA	30			12	10	30	21
Spiked in at	Extraction			Library Preparation			

Table 1: Seracare Controls and Plasma Samples. The test step of IS addition was either at the extraction or at the library preparation step. Different amounts of cfDNA were used for library preparation.

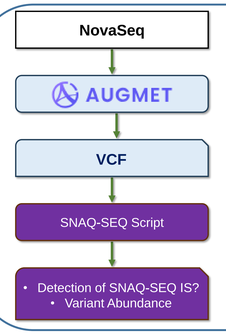


Figure 4: SNAQ™-SEQ Analysis Pipeline. Blue boxes indicate normal base calling, alignment and variant calling of Augmet, purple boxes SNAQ™-SEQ additional steps to produce indicated data files. The AUGMET bioinformatics platform was used for demultiplexing, variant calling, and sequencing quality metrics evaluation. VCF files were imported into an R markdown script to extract the IS and adjacent native template (NT) variants if present. The presence of IS variants acts as an internal positive control. The SNAQ™-SEQ Script will report the success or failure of each IS variant detection. If nearby NT variants were detected, the script will correct its VAF by removing the small IS contribution (VAFnew = NT_alt_counts / (depth – IS_alt_counts) and calculate its sample abundance using the formula NT_alt_counts / IS _alt_counts * IS_input_copies.

RESULTS

IS Spiked into Plasma

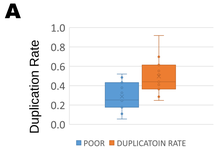
	Gene	Mutation	VAF	Allele Depth	Depth	Abundance
Mutated	*NRAS*	• p.Q61R	0.90%	26	2878	71.3
	KIT	• p.D816V	0.53%	24	4565	58.6
	EGFR	• p.T790M	0.48%	28	5863	40.1
		• p.L858R	0.51%	29	5742	31.6
	BRAF	• p.V600E	0.62%	27	4369	94.4
	KRAS	• p.G12D	0.43%	19	4455	36.9
		• p.G12C	0.38%	17	4435	33
	AKT1	• p.E17K	0.68%	27	3996	65.5
	ERBB2	• p.Y772_A775dup	0.57%	37	6506	84.1
WT	*ESR1*	• p.D538G	0%	0	4630	0
		• p.Y537S	0%	0	4630	0
	BRCA2	• p.G3076E/V	0%	0	3940	0
	GNAS	• p.R201H	0%	0	4425	0

Table 2: Detection of Native Template (NT). 250 copies of SNAQ™-Seq controls were spiked in at the first step of the nRichDX extraction of the Seraseq® ctDNA complete control at 0.5%. 100% of the variant covered by the IS mixture were detected. Abundance indicates SNAQ-SEQ measurement of variant copies added to the library preparation.

- ➢ The IS controls were successfully extracted along with the samples and sequenced.
- ➢ The IS variants were reliably identified by their unique base-substitution.
- ➢ The NT VAF were near their expected values.

SNAQ™-SEQ as Quality Control

A



B

Variant Detection Rate		
Run Quality	**IS**	**NT**
Poor	100%	62%
Good	100%	100%

C

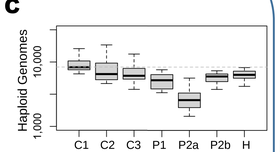


Figure 5: **Estimated Limit of Detection (LoD) Near IS Positions**. **A.** Provides an indication of library prep performance by how many of the IS input templates were captured as reads (alt_count_UMI_dedup_reads / IS_input_copies). The GOOD prep had more IS templates captured (y-axis), about a 2-fold higher than the POOR prep. **B.** Sensitivity, as indicated by positive detection rate, correlates with duplication rate. **C**. The box plot of the 7 depicted samples (Table 1) were the genomic abundance calculated from each IS position, with the grey line indicating the expected genomic copies from 30ng. Genomic abundance provides an independent measurement of sample's library preparation genomic input.

IS Spiked into Purified cfDNA

As cfDNA library preparation input was variable in each sample, we tested adjusting the IS input to maintain a single VAF between samples:

A

	Gene	Mutation	VAF	Allele Depth	Depth	Abundance
Mutated	*NRAS*	• p.Q61R	0.53%	8	1498	28.3
	KIT	• p.D816V	0.30%	4	1640	13.8
	EGFR	• p.T790M	0.71%	25	3526	24.2
		• p.L858R	0.35%	14	3911	69.5
	BRAF	• p.V600E	0.76%	13	1685	56.4
	KRAS	• p.G12D	0.93%	10	1462	29.1
		• p.G12C	0.75%	11	1457	32
	AKT1	• p.E17K	0.59%	13	2642	29.5
	ERBB2	• p.Y772_A775dup	0.25%	12	4743	28.3
WT	*ESR1*	• p.D538G	0%	0	3860	0
		• p.Y537S	0%	0	3860	0
	BRCA2	• p.G3076E/V	0%	0	2913	0
	GNAS	• p.R201H	0%	0	3966	0

B

Variant Detection Rate		
Sample	**IS**	**NT**
C2	11	9 out of 9
C3	11	12 out of 12
P1	11	3 out of 3
P2 a	11	0 out of 0
P2 b	11	0 out of 0
H	11	0 out of 0

Table 3: Detection of Native Template (NT). 45 copies of **SNAQ™-Seq** controls were spiked in at the first step of the library preparation. **A.** 100% of the variant covered by the IS mixture were detected for the Seraseq® ctDNA complete control near the expected 0.5%. **B.** Expected detection rate, variants expected for IS were based on spike-in content, and NT based on previous no IS sequencing results.

- ➢ The IS controls were successfully sequenced when added to the extracted cfDNA.
- ➢ Addition of IS did not alter variant detection.
- ➢ Once added at LoD levels, the IS control could provide an internal control for sensitivity, allowing for the removal of a per run external positive control.

CONCLUSIONS

- ➢ SNAQ™-SEQ IS variants were detectable in sample VCF without modification to pipeline software.
- ➢ Once validated, the IS added to each sample would replace an external positive control with equivalent internal standard thereby
  - Moving the sensitivity QC where it matters the most -- the patient's sample
  - Improving testing throughput and reducing testing costs