Spike-in NGS Controls for Copy Number Assessment and Improved LOD and VAF Confidence

Francine B. deAbreu[1,2], Sophie J. Deharvengt[1,2], Bradley Austermiller[3], Thomas Morrison[3], Gregory J. Tsongalis[1,2], Joel A. Lefferts[1,2]

[1]Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH; [2]The Geisel School of Medicine at Dartmouth, Hanover, NH; [3]AccuGenomics, Inc., Wilmington, NC

ABSTRACT

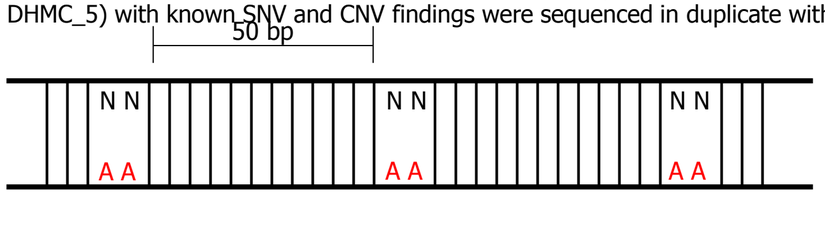
Background: Spike-in controls are a crucial component of many real-time PCR assays and rule out PCR inhibition. Similarly, next-generation sequencing (NGS) spike-in controls could help monitor the NGS process in individual samples. AccuGenomics is developing Standardized Nucleic Acid Quantification (SNAQ) spike-in controls for molecular testing that are designed to co-vary with the measured target and have the potential to improve the traditional performance limits by providing direct sensitivity and specificity control for every target in every sample. For targeted NGS, the controls mimic the NGS library sequencing errors and yield variation of the sample and allow for LOD and VAF confidence for every base in every sample. In this proof-of-principle study, we examine the application of SNAQ-SEQ in the estimation of CNV and LOD for actionable mutations in tumor FFPE samples.
Methods: The spike-in control mixture is synthetic DNA that can be distinguished from genomic DNA (gDNA) by unique dinucleotide alterations every 50bp. The controls are spiked into the gDNA being sequenced at a 1:1 molar ratio and the library preparation and DNA sequencing proceed as normal. In this study, sequencing was performed using the AmpliSeq Cancer Hotspot Panel (CHP) v2 (ThermoFisher). The control and genomic sequences are separated at the FASTQ file stage using a Perl script that identifies the dinucleotide sites. For copy number analysis the sequence counts of gDNA are normalized to the counts of the control at individual loci. To estimate limits of detection, the false positive rate of actionable variants in the control was measured. The confidence of VAF was estimated based on the previously published formula of Blomquist, et al (2015). Two normal tissue samples were sequenced in duplicate along with three tumor samples with known SNV and CNV findings.
Results: CNV findings using the SNAQ-SEQ controls with the tumor sample correlate well with known CNVs, specifically EGFR amplifications of different levels in two different tumor samples. False-positive SNV rates were low in the control and gDNA sequences but slightly higher in the gDNA, presumably due to formalin fixation. The false positive rate was significantly higher in both the control and gDNA sequence for one NRAS variant, suggesting an isolated sequencing error. Confidence interval estimates for these VAFs ranged from 2.5 to 18-fold.
Conclusion: These initial data suggest that SNAQ-SEQ controls are a useful spike-in control for NGS that not only help to more accurately define the accuracy of individual samples but can also improve the ability to identify CNVs.

INTRODUCTION

- Spike-in controls are a crucial component of many real-time PCR assays and rule out PCR inhibition.
- NGS spike-in controls are not routinely used but could help monitor the NGS process in individual samples.
- Standardized Nucleic Acid Quantification (SNAQ) spike-in controls (AccuGenomics) for NGS are being developed to co-vary with the DNA being sequenced to to improve the traditional performance limits by providing direct sensitivity and specificity control for every target in every sample for both hybrid capture and amplicon-based NGS.
- For targeted NGS, the controls mimic the NGS library sequencing errors and yield variation of the sample and allow for LOD and VAF confidence for every base in every sample.
- In this proof-of-principle study, we examine the application of SNAQ-SEQ in the estimation of CNV and LOD for actionable mutations in tumor FFPE samples.

METHODS

- The spike-in control mixture is synthetic DNA that mimics select targeted regions but can be distinguished from genomic DNA (gDNA) by unique dinucleotide alterations every 50bp.
- The controls are spiked into the gDNA being sequenced at a 1:1 molar ratio and the library preparation and DNA sequencing proceed as normal.
- Two normal FFPE tissue samples (DHMC_1 and DHMC_2) and three tumor samples (DHMC_3, DHMC_4 and DHMC_5) with known SNV and CNV findings were sequenced in duplicate with SNAQ-SEQ controls.
- NGS was performed using the AmpliSeq Cancer Hotspot Panel v2 (ThermoFisher).
- The control and genomic sequences are separated at the FASTQ file stage using a Perl script that identifies the dinucleotide sites.
- For copy number analysis the sequence counts of gDNA are normalized to the counts of the control at individual loci.
- To estimate limits of detection, the false positive rate of actionable variants in the control was measured.
- The confidence of VAF was estimated based on the previously published formula of Blomquist, et al (2015).
- Two normal FFPE tissue samples (DHMC_1 and DHMC_2) and three tumor samples (DHMC_3, DHMC_4 and DHMC_5) with known SNV and CNV findings were sequenced in duplicate with SNAQ-SEQ controls.



RESULTS

- CNV findings using the SNAQ-SEQ controls with the tumor sample correlate well with known CNVs: EGFR amplifications of different levels in two different tumor samples.
- False-positive SNV rates were low in the control and gDNA sequences but slightly higher in the gDNA, presumably due to formalin fixation.
- The false positive rate was significantly higher in both the control and gDNA sequence for one NRAS variant, suggesting an isolated sequencing error.
- Confidence interval estimates for these VAFs ranged from 2.5 to 18-fold.

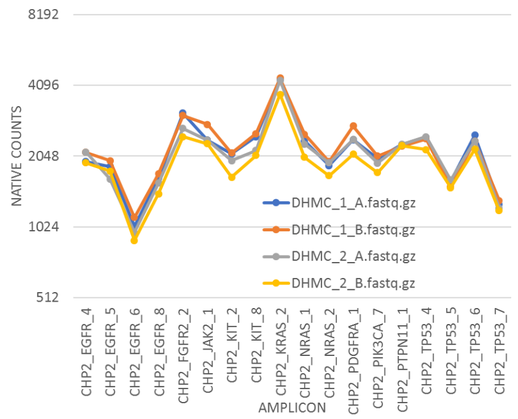


Figure 1. Native NGS counts/reads in select amplicons from DNA samples from "normal" diploid FFPE tissue.

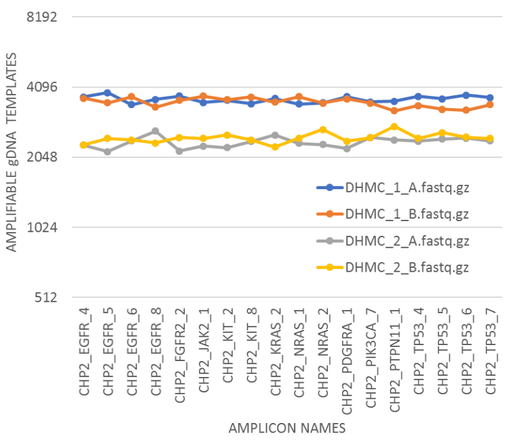


Figure 2. NGS counts (reads) in select amplicons from same "normal" DNA samples after SNAQ-SEQ normalization.

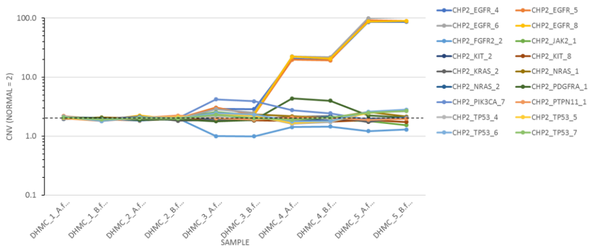


Figure 3. SNAQ-SEQ corrects for NGS variation, enabling digital PCR accuracy with NGS. The CNV estimates have a ±14% confidence interval using a single amplicon per gene. More accuracy possible with additional targets per gene.

Table 1. Estimate NGS false positive rate from normal samples

POSITION		SNP	CONTROL	NATIVE	TARGET
chr1	114713908	T>A	0.000%	0.040%	CHP2_NRAS_2 rs11554290
chr1	114716127	C>T	**0.066%**	**0.076%**	CHP2_NRAS_1 rs121913250
chr12	25227343	G>C	0.000%	0.012%	CHP2_KRAS_2 rs121913238
chr17	7673802	C>A	0.000%	0.000%	CHP2_TP53_7 rs28934576
chr3	179218303	G>C	0.000%	0.013%	CHP2_PIK3A_7 rs104886003
chr4	54733155	A>T	0.000%	0.022%	CHP2_KIT_8 rs121913507

CONCLUSIONS

- These initial data suggest that SNAQ-SEQ controls are a useful NGS spike-in control with FFPE DNA for:
  - More accurately defining the accuracy of individual samples
  - Improved CNV detection
- Additional studies evaluating the utility of these controls in other NGS assays is warranted

REFERENCES

Blomquist T, Crawford EL, Yeo J, Zhang X, Willey JC. Control for stochastic sampling variation and qualitative sequencing error in next generation sequencing. Biomol Detect Quantif. 2015 Sep;5:30-37. PubMed PMID: 26693143;