To UMI or not to UMI: Obtaining robust low-frequency variant identification for liquid biopsy with a SLIMamp-based NGS chemistry without molecular barcoding

Pillar Biosciences, Inc. 9 Strathmore Rd Natick, MA 01760
(800) 514-9307 | www.pillarbiosci.com

Andrew B. Conley, Lavanya Rishishwar, Nachen Yang, Michael Liu, Sean M. Polvino, Yue Ke, Kirsty McFarland, Zhaohui Wang, Alex V. Kotlar

ABSTRACT

Introduction. Ultra-deep targeted sequencing of cell-free DNA (cfDNA) has emerged as a powerful tool for non-invasive detection of low-frequency somatic variants, specifically in applications such as liquid biopsy. However, PCR and sequencing noise pose a challenge to detecting variants at frequencies below 1%. Unique molecular identifier (UMI) protocols can help mitigate this noise by attaching a unique barcode to each DNA molecule and calculating a consensus call per barcode. While useful, UMI protocols are complex and add significant costs compared to non-UMI methods. In this study, we present an alternative approach to improving the signal-to-noise ratio for LBx testing using SLIMamp® chemistry that does not rely on UMIs, leveraging noise rates from negative control samples.

Methods. To comprehensively evaluate the two protocols, we generated libraries from a dilution series of contrived cfDNA samples, ranging from 0.0625% to 0.5% Variant Allele Frequency (VAF), with DNA inputs varying from 10 to 30ng. Positive cell-free (cf) reference samples (n=29) were obtained from Seracare™ and Horizon Discovery, while normal reference cf gDNA was acquired from Anchor Molecular (AM; n=10) and from plasma from healthy donors. Samples were characterized using Pillar Biosciences' oncoReveal™ Core LBx panel, a research-use-only (RUO) liquid biopsy-based NGS assay kit covering 11,400+ positions across 104 genes. UMI libraries were prepared for all samples, and non-UMI libraries were prepared for Seracare™ and normal samples. All samples were sequenced on Illumina's NextSeq 550 platform with an average of 27M paired end reads, and data were analyzed with Pillar's secondary bioinformatics software PiVAT® (Pillar Biosciences' variant analysis toolkit) with an embedded denoising algorithm that utilizes noise estimates from normal samples. We assessed the impact of batch effects on overall performance using in- and out-of-batch normal controls from Anchor Molecular.

Results. We performed two comparisons (1) Direct comparison of Seracare™ samples sequenced with UMI and non-UMI protocols, and (2) in silico comparison by analyzing all positive UMI samples with and without utilizing the UMI sequence. Both comparisons revealed a significant increase in sensitivity for non-UMI protocols compared to UMI protocols, especially at lower VAF. Importantly, high specificity was consistently observed in both approaches. Potential batch effects had negligible impact on sensitivity and specificity.

Conclusion. Our findings indicate that using our SLIMamp™ chemistry, achieving high sensitivity and specificity in detecting low frequency variants in cfDNA samples by NGS testing does not necessitate the use of UMIs. Instead, implementing a robust post-sequencing computational strategy in non-UMI protocols using in- or out-of-batch negative control can yield equivalent or improved performance.

EXPERIMENT DESIGN

- The objective of this study was to evaluate the impact of UMIs on detecting low-frequency variants in liquid biopsy samples
- We developed UMI and non-UMI versions of our oncoReveal™ Core LBx panel, with almost identical amplicons
- The UMI data was analyzed
  1. Variant calling performing UMI consensus calling using the UMI sequences
  2. Variant calling without considering the UMI sequences
- Non-UMI data was using identical pipeline as #2 above
- All analysis were performed on PiVAT
- **Positive standard reference samples:**
  - Seraseq® ctDNA Mutation Mix v2 (number of mutations = 24)
  - *UMI-only:* Horizon EGFR Multiplex cfDNA Reference Standard Set (number of mutations = 8)
  - *Non-UMI only:* Seraseq® ctDNA Complete™ Mutation Mix v1 (number of mutations = 13),
- **Normal samples:**
  - Human cfDNA normal sample (n=4-5), Anchor Molecular Reference (n=10-11)

DNA input (ng)	Barcoding Method	TC/VAF (%) 0.125	0.25	0.5
10	UMI	2	2	2
	Non-UMI	2	5	8
30	UMI	3	6	3
	Non-UMI	2	6	4

RESULTS

A Statistical Modeling Approach For Noise Suppression

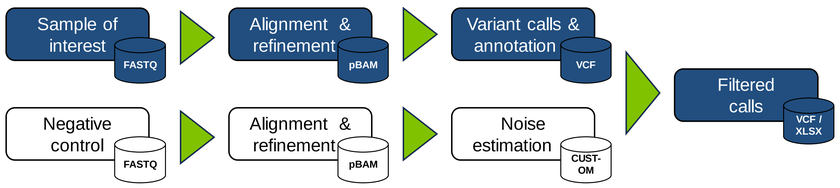

Figure 1. Overview of noise suppression approach for improving specificity. A negative control is used to dynamically model the background noise originating from various sources: (1) PCR error rate, (2) sequencing error rate, and (3) other artifactual error rates such as DNA damage. The negative control is processed identically to sample of interests, except for variant calling and annotation step. At that stage, the negative control is used to estimate the background noise model. This noise model is utilized to filter out false positive calls.

Improving Performance By Noise Suppression

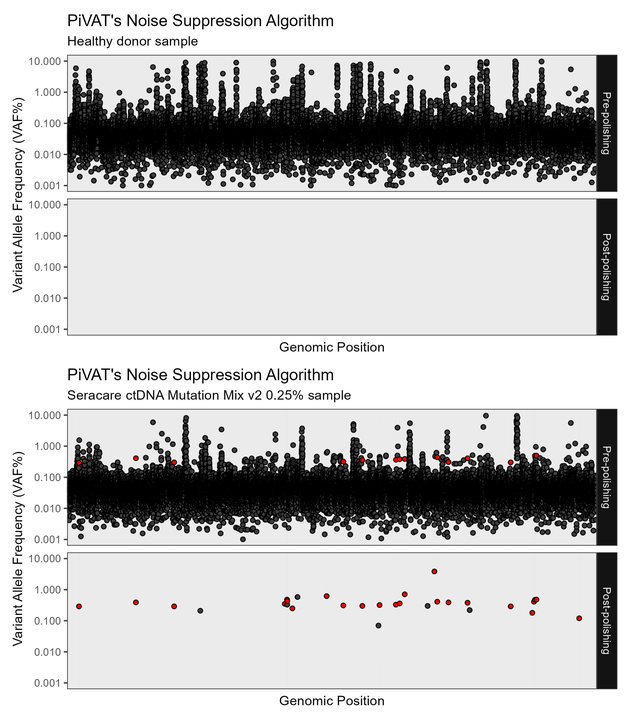

Figure 2. Noise suppression in action. The noise suppression algorithm's behavior is demonstrated on two samples: a healthy donor sample and a positive standard reference sample (Seracare ctDNA Mutation Mix v2 AF0.25%). In the healthy donor sample, several variants are detected at low variant allele frequency (VAF%), however, none of these are retained once they are passed through the noise suppression algorithm. On the contrary, in the positive standard reference sample, the expected variants (shown in red) are retained after the noise suppression algorithm. Small group of additional variants are also retained (shown in black), A small group of unknown signals remained likely introduced by the engineered materials which are expected to be filtered out during the tertiary analysis phase.

RESULTS

Performance Summary for UMI and Non-UMI Methods

Performance Metric			UMI	UMI as non-UMI	Non-UMI
LoB			2.3	2.6	1.4
NPA			99.996%	99.996%	99.998%
PPA	30ng	0.125%	52%	59%	81%
		0.25%	86%	85%	96%
		0.50%	89%	90%	100%
	10ng	0.125%	34%	41%	65%
		0.25%	48%	72%	88%
		0.50%	79%	91%	100%

LoB = FP / Sample; NPA = 100 - 100 * FP / (FP + TN); PPA = 100 * TP / (TP+FN)

Table 1. Summary of performance metrics from secondary analysis pipeline. The table summarizes the limit of blank (LoB), negative percent agreement (NPA), and positive percent agreement (PPA) for the normal and positive standard reference samples tested in this study. The higher performance of UMI data when run as non-UMI compared to UMI analysis is due to the per-UMI family size threshold filtering. *i.e.*, if a UMI family doesn't have sufficient representation, then that UMI family gets rejected, reducing our power to call variants. **The remaining false positives are expected to be filtered at tertiary analysis stage.**

Variant Allele Frequency (VAF%) Comparison Between Methods

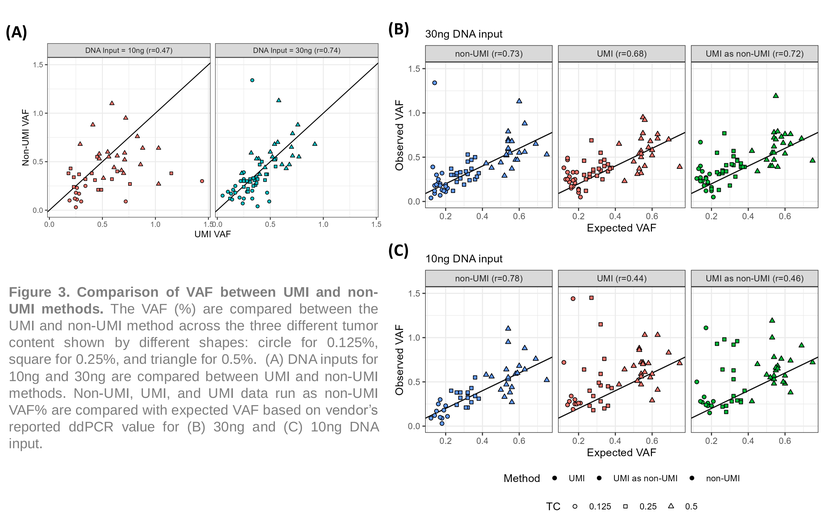

Figure 3. Comparison of VAF between UMI and non-UMI methods. The VAF (%) are compared between the UMI and non-UMI method across the three different tumor content shown by different shapes: circle for 0.125%, square for 0.25%, and triangle for 0.5%. (A) DNA inputs for 10ng and 30ng are compared between UMI and non-UMI methods. Non-UMI, UMI, and UMI data run as non-UMI VAF% are compared with expected VAF based on vendor's reported ddPCR value for (B) 30ng and (C) 10ng DNA input.

CONCLUSIONS

- We demonstrate in this study that, with the combination of our SLIMamp® chemistry and denoising algorithm, we can detect low-frequency variants with high confidence without the use of unique molecular identifiers (UMIs).
- A non-UMI approach can significantly reduce false positive calls by using a clinical normal samples for background noise modeling
- We estimated that the non-UMI protocol on average reduces turnaround time by ~1.5-hour and saves 30% in reagent cost (less PCR and purification steps) compared to protocols using UMI, enabling local NGS testing of LBx samples.