DNA Library Amplification to Produce Archived Reference Samples for R&D and Clinical Assay Development, Optimization, and Validation

Harbinger Health

Miguel Williams, Caitlin Gilley, Kyle Gowen, Zhenyu Zhang, Liudmila Elagina, Jocelyn Charlton, Rosemary Witkowski, Tony Shuber

AACC July 26, 2023 | Poster # 614

BACKGROUND

Patient derived clinical samples are finite, constraining the number of potential diagnostic analyses. Most biospecimens are distinctly allocated for discovery, technical or clinical studies. Scant nucleic acid templates are frequently increased through PCR, however, amplification is used sparingly in NGS assays. The AReS (**A**rchived **Re**ference **S**ample) platform was developed to increase the mass of index derived DNA libraries to serve as a sample reservoir from which assay development and clinical studies can be performed. The Harbinger Hx assay (HHx) is used to define cancer status using cfDNA. Comparing HHx data from both original and AReS libraries demonstrate that the two library types are highly concordant, and equivalent to sequencing replicates**.** Harbinger has already utilized AReS libraries for assay optimization, reference standards and technology development studies.

METHODS

Sequencing metrics, methylation metrics and binary classification score were compared between the original and AReS libraries utilizing the HHx assay. Briefly, cfDNA was bisulfite treated and converted into indexed libraries prior to further amplification by the AReS process. Both the original libraries and AReS samples were enriched for cancer specific regions of interest (ROIs) and sequenced to a unique median target coverage ≥100X **(Figure 1)**. Aggregate methylation events were summarized via methylation metrics and the data is then defined as either cancer or non-cancer via binary classifier. The AReS process was optimized to produce maximal yield with minimal input **(Figures 2, 3 & 4),** validated across 536 AReS libraries derived from 321 patient cfDNA samples, 121 cancer, 197 non-cancer, **(Figures 5, 6, 7)** and characterized at a molecular level through unique molecular indices tagged libraries from 4 cell line controls and 4 patient derived cfDNA samples **(Figures 8, 9 & 10)**

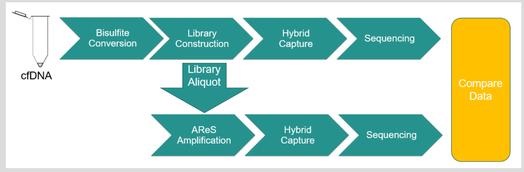

Figure 1 Schematic of the AReS Process. The Harbinger Hx assay defines cancer status by analyzing cfDNA extracted from patient plasma (top track). The cfDNA is bisulfite converted and then converted into indexed libraries. Libraries are enriched for regions informative of cancer status and sequenced. Aliquots of the indexed library can also be taken and further amplified to generate an AReS Library (bottom track). The HHx assay was used to generate a binary cancer classifier using a multi-layered logistic regression-based machine learning algorithm which was trained separately and locked prior to use in this study.

RESULTS

AReS Optimization

Increasing primers up to 3X of standard conditions improved yield by over a third, but further increases did not increase yield **(Figure 2).** Increasing input mass improved the AReS library unique rate ~50%, however, there are diminishing returns with greater than 10X input **(Figure 3).** The final AReS conditions were defined as 3X Primers with 10X input, these conditions were used for all future AReS samples. Though unique reads were less frequent within AReS libraries, additional sequencing of AReS libraries could produce an equivalent or greater number of unique reads than original libraries **(Figure 4).**

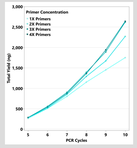

Figure 2 AReS Primer Optimization. Increasing primer concentration 3X increased PCR product yield.

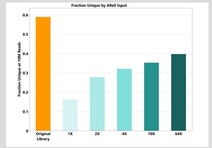

Figure 3 AReS Input Optimization. Increasing AReS reaction input up to 64X increase the fraction of unique sequencing reads, however there were diminishing returns above 10x input.

Figure 4 AReS Unique Molecule Content. Though AReS libraries have a lower fraction of unique molecules (left), allocating additional sequencing to AReS libraries (center), can generate an equivalent or greater number of unique reads (right).

AReS Concordance

The methylation metrics of each library were compared for each CGIs within the ROIs, visualized by an exemplary sample **(Figure 5).** When all 536 AReS libraries were compared to their original library pair across seven methylation metrics across all CGIs within the ROIs the Pearson correlation was ≥0.98 **(Figure 6).** Cancer classification was also concordant between original and AReS libraries, over 97% libraries had identical calls **(Figure 7).** Importantly, difficult to classify samples were specifically selected to more rigorously validate the AReS concordance. The binary classification score Pearson concordance of AReS libraries was 0.98, which is similar to what was observed across original library replicates from studies performed in-house (data not shown)

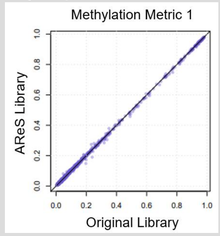

Figure 5* Individual sample *Methylation Metric Concordance. The metrics used to compare original and AReS libraries ranged from ~3-80% of events representing both rare and common occurrences.

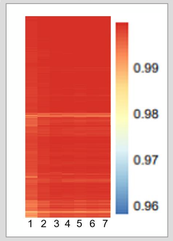

Figure 6 All sample concordance. Methylation metrics were further consolidated by Pearson correlation, which was ≥ 0.98 in all AReS Libraries.

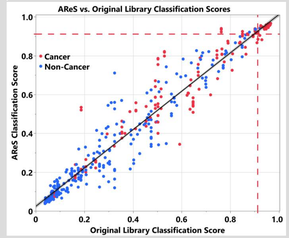

Figure 7 Classifier score concordance. The Pearson correlation between AReS and original library classifier scores was 0.98. These samples were specifically selected to include challenging classification samples.

RESULTS

Molecular Characterization

To better understand the effects of AReS amplification on a molecular level, unique molecular indices (UMIs) were incorporated into eight bisulfite-treated cfDNA libraries. UMIs enabled the quantification of individual cfDNA molecules between original and AReS libraries. Approximately 77% of reads were shared between the original and AReS libraries across all 8 paired samples. When the original library was re-sequenced, the proportion of shared reads between the first and second replicate was also ~77% **(Figure 8).** The frequency of common reads between AReS and original libraries **(Figure 9)** was similar to the frequency of common reads between replicates **(Figure 10).** Across all 8 samples the difference in Pearson correlation between these two pairs was <0.04. Together, these data indicate that there was no significant amplification bias as a result of AReS amplification, and that AReS libraries are the equivalent of sequencing replicates.

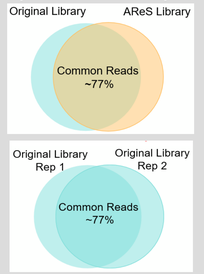

Figure 8 Fraction of common reads. The AReS and original libraries (top) shared the same fraction of common reads as sequencing replicates (bottom) ~77%.

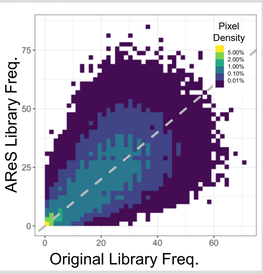

Figure 9 AReS Original library read frequency. The frequency of common reads between AReS and Original libraries was evenly distributed the Pearson correlation is 0.82. The absence of a skew indicates that there was no significant amplification bias.

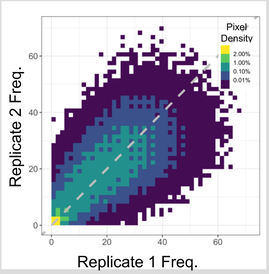

Figure 10 Sequencing replicate read frequency. Sequencing replicates had a similar distribution of reads when compared to original library and AReS pairs, Pearson correlation was 0.83.

CONCLUSIONS

In conclusion, the AReS process creates a library reservoir that is highly reproducible and is both functionally and analytically identical to original libraries. While AReS libraries do have a lower unique rate, which could affect identification of rare variants, this can be rectified through additional sequencing. There is no evidence of detectable amplification bias either through skews in read composition, methylation metrics or CYN classification. Taken together, the AReS platform has been developed, optimized, and validated for both technical research and clinical applications.