Identifying Sources of Library Preparation Artifacts in AAV Vectors with SMRT Sequencing

[1]Sarah Tong, [1]Widler Casy, [1]Jie Tan, [1]Anthony Mauri, [1]Nyasha Makoni, [1]Patrick Starremans

[1]Solid Biosciences Inc., Charlestown, MA, USA

INTRODUCTION

Comprehensive and reliable genomic characterization of adeno-associated virus (AAV) vectors is a fundamental component for the development and quality assurance of AAV-based gene therapies. Genomic characterization of AAV products is often achieved through long-read sequencing technology, which can sequence continuous stretches of DNA spanning thousands of bases.

Sequencing packaged vector genomes can be used to confirm vector identity and structural integrity, as well as to ensure consistency between manufacturing lots. Furthermore, sequencing can be used for the detection of unwanted impurities that may have been unintentionally co-packaged during production. Accurate and thorough DNA sequencing is a critical analytical tool for determining vector quality of an AAV product.

Single-molecule real-time (SMRT) sequencing incorporates long-read sequencing capabilities with exceptionally high accuracy and offers useful analysis of AAV product genome characterization. However, it is vital to differentiate between results that properly reflect the true composition of the sequenced AAV vector and results that contain library preparation-derived artifacts. An understanding of the components of library preparation that have the largest impact on sequencing results helps to achieve a more reliable analytical method that accurately depicts a comprehensive genomic characterization of AAV vectors.

METHODS AND MATERIALS

Library preparation of self-complementary AAV (scAAV) and single-stranded AAV (ssAAV) drug substance samples followed Invitrogen's *PureLink™ Viral RNA/DNA Mini Kit User Guide* for DNA extraction, and Pacific Biosciences of California, Inc.'s (PacBio) *Preparing multiplexed AAV SMRTbell® libraries using SMRTbell prep kit 3.0* protocol for the remainder of the workflow. Samples were sequenced in-house with a PacBio Vega instrument.

The amount of Carrier RNA used in DNA extraction was optimized for scAAV libraries at different concentrations in ug/sample: 5.6 (standard), 2.5, 1.0, and 0. The Carrier RNA conditions were multiplexed together for sequencing.

The manufacturer recommends targeting a loading concentration of 200-300 pM. scAAV libraries were loaded at 300 pM and 400 pM. ssAAV libraries were loaded at 400 pM, 500 pM, and 900 pM.

Thermal annealing of ssAAV was tested at the standard condition (95 C for 5 minutes and ramped down to 25 C at -1 C/min), and a modified condition (80 C for 5 minutes and ramped down to 25 C at -1 C/min). A control experiment was run without a thermal annealing step. The thermal annealing conditions were multiplexed together for sequencing.

Figure 1. AAV SMRT Sequencing Library Preparation Workflow

Astrix (***) indicate modified sections of workflow. Blue indicates ssAAV workflow step only. Orange indicates both scAAV and ssAAV workflows.

OPTIMIZING LOADING CONCENTRATION OF AAV LIBRARIES

Optimal loading concentration was outside the recommended range

Loading level is the percentage of zero-mode waveguides (ZMWs) that contain one or more polymerase-bound SMRTbell DNA complexes. The manufacturer recommends targeting a loading concentration of 200-300 pM on the Vega to achieve the defined optimal loading level of ~50-75%.

scAAV libraries achieved optimal loading level when loaded at 400 pM. ssAAV libraries did not achieve an optimal loading level at the tested loading concentrations, indicating there may have been an upstream library preparation step that inhibited loading. Furthermore, the scAAV library that achieved optimal loading level had a higher % full genome than the scAAV library that was below optimal loading level, showing that genome integrity results may be impacted by loading concentration.

Figure 2. Loading Level vs. Loading Concentration

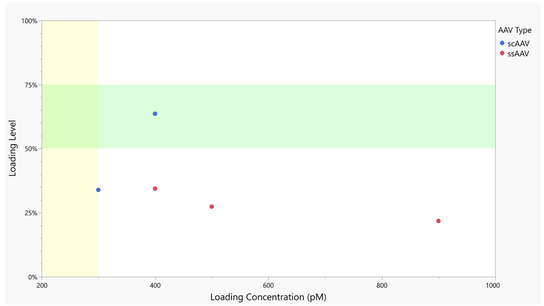

Figure 2 shows the impact of loading concentration on loading level for ssAAV and scAAV libraries. Green band range represents manufacturer defined optimal loading level (50-75%). Yellow band range represents manufacturer recommended loading concentration (200-300 pM).

Figure 3. % Full Genome vs. Loading Level

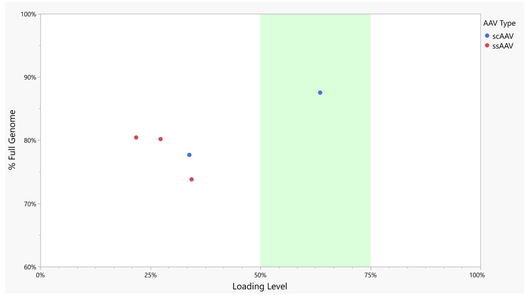

Figure 3 shows the impact of loading level on genome integrity. % full genome of scAAV was calculated by dividing the number of full scAAV reads by the total number of reads. % full genome of ssAAV was calculated by dividing the sum of full ssAAV reads and unresolved dimer reads by the total number of reads.

OPTIMAL THERMAL ANNEALING CONDITIONS FOR ssAAV LIBRARIES

Library prepared without thermal annealing had the highest % full genome

The manufacturer recommends a thermal annealing step during library preparation of ssAAV to convert the single-stranded DNA (ssDNA) into double-stranded DNA (dsDNA) for ligation of SMRTbell adapters.

The standard and modified thermal annealing protocol libraries had less full ssAAV reads and more unresolved dimer reads than the library prepped without thermal annealing. The standard thermal annealing protocol library also had the highest number of total reads, while the modified thermal annealing library and the library prepped without thermal annealing both had significantly less reads. This indicates that thermal annealing can allow for greater SMRT sequencing depth but also may create library artifacts that impact genome integrity results.

Figure 4. % Full Genome vs. Thermal Annealing Protocol

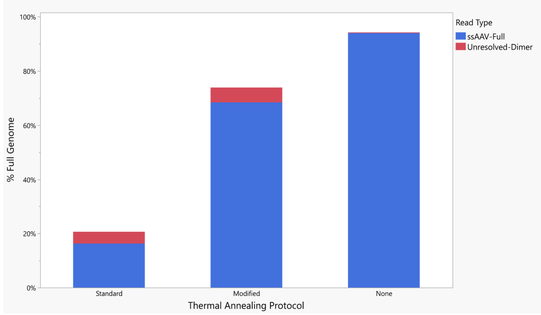

Figure 4 shows the impact of thermal annealing on ssAAV genome integrity.

Figure 5. Total Read Count vs. Thermal Annealing Protocol

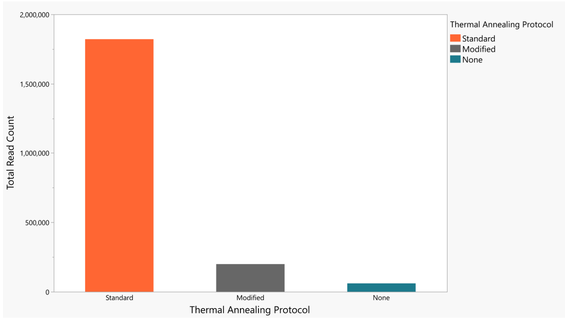

Figure 5 shows the impact of thermal annealing on ssAAV sequencing depth.

OPTIMAL CARRIER RNA CONCENTRATION FOR DNA EXTRACTION OF scAAV LIBRARIES

Carrier RNA concentration had no impact on scAAV library prep

Carrier RNA helps to increase viral DNA capture during DNA extraction, but it may also interfere with downstream library preparation. An optimal amount of Carrier maximizes the quantity of extracted DNA, while also minimizing any effects on scAAV sequencing.

The % full genome was similar across each Carrier RNA condition with an average % full genome of 90.79% and a CV of 2.40%. This shows that differences in % full genome results between libraries were likely due to assay variability, and that the concentration of Carrier RNA in DNA extraction does not impact scAAV sequencing.

Figure 6. Carrier RNA Optimization on % Full Genome

Condition	% Full Genome
5.6 ug/sample	91.79%
2.5 ug/sample	92.00%
1.0 ug/sample	91.84%
0 ug/sample	87.53%
Mean	**90.79%**
CV	**2.40%**

Figure 6 shows the impact of Carrier RNA concentration on scAAV genome integrity. CV, coefficient of variation.

CONCLUSIONS

- Our in-house scAAV libraries should be sequenced at a loading concentration of 400 pM to achieve optimal loading level. Our in-house ssAAV libraries do not seem to show a correlation between loading concentration and loading level.
- Thermal annealing of our in-house ssAAV libraries appears to increase sequencing depth, but the reads are generally of lower quality. Unresolved dimers in our in-house ssAAV libraries appear to primarily be an artifact of thermal annealing.
- Carrier RNA concentrations of up to 5.6 ug/sample do not have any impact on genome integrity results of our in-house scAAV libraries.
- Further investigation into what steps of SMRT sequencing library preparation impacts ssAAV sequencing results is necessary.

Presented at American Society of Gene & Cell Therapy, May 11-15, 2026, Boston, Ma