Full-length rAAV sequencing for mixture population characterization using highly accurate long reads

Abstract #1657 | Elizabeth Tseng, Harsharan Dhillon, Roger Volden
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Abstract

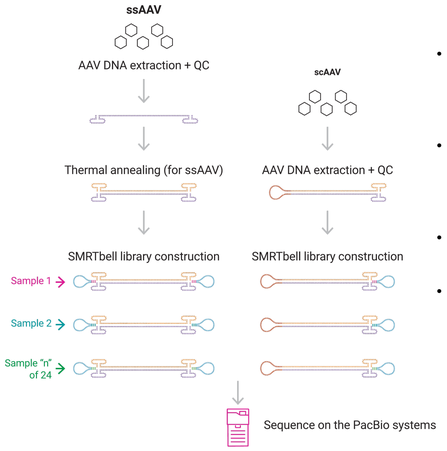
Understanding the quality of rAAV vectors by assessing the presence of cellular impurities such as truncations, chimeras, and host genome integrations are all essential for ensuring the purity and efficacy of the vector. While next-generation sequencing has been applied for assessing the quality of rAAV vectors, short read length technologies cannot sequence the full-length AAV molecules or accurately distinguish or quantify mixed populations that might contain both scAAV and ssAAV.

PacBio SMRT sequencing is a long-read sequencing technology that can produce reads of >Q20 (>99%) for insert sizes of 10–20 kb and longer, which easily cover the entirety of typical rAAV constructs.

To demonstrate that PacBio long-read sequencing can accurately characterize AAV constructs, we purchased scAAV and ssAAV constructs from commercial vendors. We created sequencing libraries from both pure (scAAV or ssAAV-only) vectors, as well as an artificially mixed (scAAV with ssAAV) population. The long-read data is able to characterize the proportion of vector genomes for truncation (full-length or partial), purity, and structure (self-complementary or single-strand).

To summarize, we show that long-read sequencing has the ability to accurately and more comprehensively assess the potency and purity of rAAV vectors, making it a valuable tool for the quality assessment of rAAV genome populations.

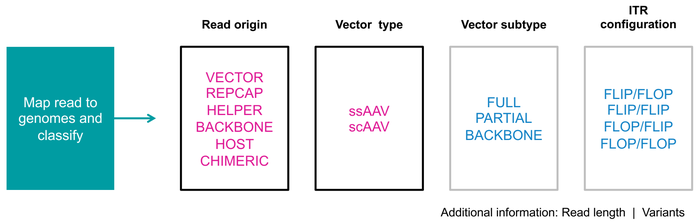
AAV sequencing using PacBio long-read sequencing



- Self-complementary AAV (scAAV) mutated ITRs form a hairpin on one end; ligate SMRTbell adapters on the other to form a SMRTbell library
- Single-strand AAV (ssAAV) requires additional thermal annealing step to produce double-stranded template before SMRTbell ligation
- Barcoded SMRTbell adapters support up to 24-plex per library
- Sequence on Sequel II/IIe platforms with on-instrument AAV mode to generate HiFi reads

Figure 1. PacBio AAV library workflow.

Analyzing AAV reads from HiFi sequencing



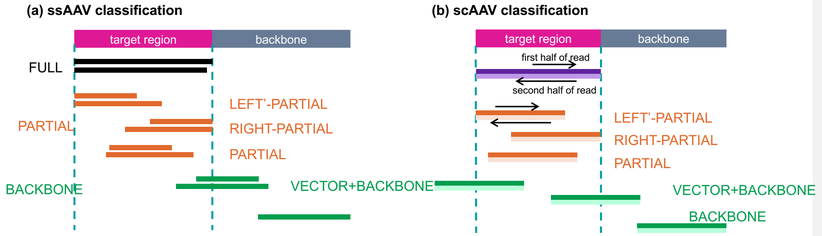
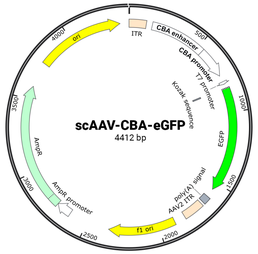


Figure 2. Classification of AAV reads from HiFi sequencing for (a) ssAAV (b) scAAV

HiFi sequencing of scAAV shows high proportion of full vector genomes

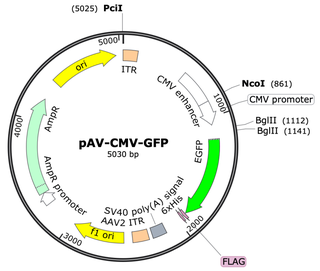


ITR-to-ITR region: 662-2739 (2.1kb)

Table 1. Read characterization of scAAV. HiFi sequencing shows the majority (95.1%) of the sequenced reads are full vectors with expected size of 2.1 kb.

		scAAV CBA-eGFP
ssAAV	Full	0.4%
	Partial	0.5%
	Backbone	0.0%
scAAV	Full	95.1%
	Partial	3.6%
	Backbone	0.1%

HiFi sequencing of ssAAV finds unexpected self-complementary structures



ITR-to-ITR region: 1-2434 (2.4kb)

Table 2. Read characterization of ssAAV. HiFi sequencing finds unexpected large (~5 kb) fragments that were self-complementary structures comprising 29% of the reads. Only 35.8% of the reads are full ssAAV genomes.

		ssAAV pAV-CMV-GFP
ssAAV	Full	35.8%
	Partial	10.4%
	Backbone	0.0%
scAAV	Full	29.0%
	Partial	19.4%
	Backbone	0.8%

HiFi sequencing of scAAV:ssAAV mixture recovers the expected vector proportions

The scAAV and ssAAV libraries (middle panel) were mixed with a 10:1 ratio and sequenced on PacBio Sequel II system with Sequel II binding kit 3.1 and Sequel II sequencing kit 2.0 with one SMRT Cell.

	Number of reads
Mapped to scAAV CBA-eGFP	259,872 (86.3%)
Mapped to ssAAV pAV-CMV-GFP	41,161 (13.7%)
Total	**301,033**

HiFi reads were first mapped to the scAAV-CBA-eGFP and ssAAV-pAV-CMV-GFP genome to bin reads back to the origin vector genome.

Conclusion

- HiFi sequencing can characterize product heterogeneity
- Commercially-purchased scAAV vector showed high proportion of full self-complementary vector genomes
- Commercially-purchased ssAAV vector showed unexpected population of self-complementary genomes with double the vector size
- HiFi sequencing able to identify vector in artificially mixed population of scAAV:ssAAV population

References

1. Procedure & checklist – Preparing multiplexed AAV SMRTbell® libraries using SMRTbell prep kit 3.0
2. https://github.com/Magdoll/AAV/wiki/Tutorial:-Analyzing-AAV-Data

Explore datasets, webinars, workflows, and inquiries at pacb.com/gene-therapy or scan QR code