AAV Sequence Analysis

Sample name: bc2182-AAV8-Adh-SOPmod-Revio5kb

Date: 2025-12-11

Contents

Analysis context

	Value
Sample unique ID	bc2182-AAV8-Adh-SOPmod-Revio5kb
Sequencing run ID	m84039_250929_210341_s4
Construct vector type	ssAAV

Read-based AAV vector type classification

Definitions

Table 2: Reference label definitions.

Reference Label	Definition
vector	Read originates from the vector plasmid.
repcap	Read originates from the RepCap plasmid. The Rep gene encodes four proteins (Rep78, Rep68, Rep52, and Rep40), which are required for viral genome replication and packaging, while Cap expression gives rise to the viral capsid proteins (VP; VP1/VP2/VP3), which form the outer capsid shell that protects the viral genome, as well as being actively involved in cell binding and internalization.
helper	Read originates from the helper plasmid. In addition to Rep and Cap, AAV requires a helper plasmid containing genes from adenovirus. These genes (E4, E2a and VA) mediate AAV replication.
host	Read originates from the host genome that is given (e.g. hg38, CHM13).
chimeric-vector	One part of the read aligns to the vector plasmid, while another part of the same read aligns to a different reference sequence.
chimeric-nonvector	The read consists of fragments that align to two or more different reference sequences, neither of which is the vector plasmid.

Table 3: Assigned type definitions.

Assigned Type	Definition
ssAAV	Single-stranded AAV vector genome where the resulting DNA template is expected to be single-stranded, as opposed to self-complementary. A sequencing read is inferred as ssAAV if it has a single alignment to the vector plasmid's ITR-to-ITR payload region and no complementary supplemental alignment.
scAAV	Self-complementary AAV vector genome where one half of the payload region is a reverse complement of the other, resulting in an intramolecular double-stranded DNA template. A sequencing read is inferred as scAAV if it aligns to the payload region in both forward (+) and reverse (-) read directions.
backbone	Read aligns to the vector plasmid sequence but fully outside of the annotated ITR-to-ITR region, indicating that the sequence fragment originated solely from the plasmid backbone.
other-vector	Read consists of a fragment mapping to the vector but with characteristics other than those listed above.

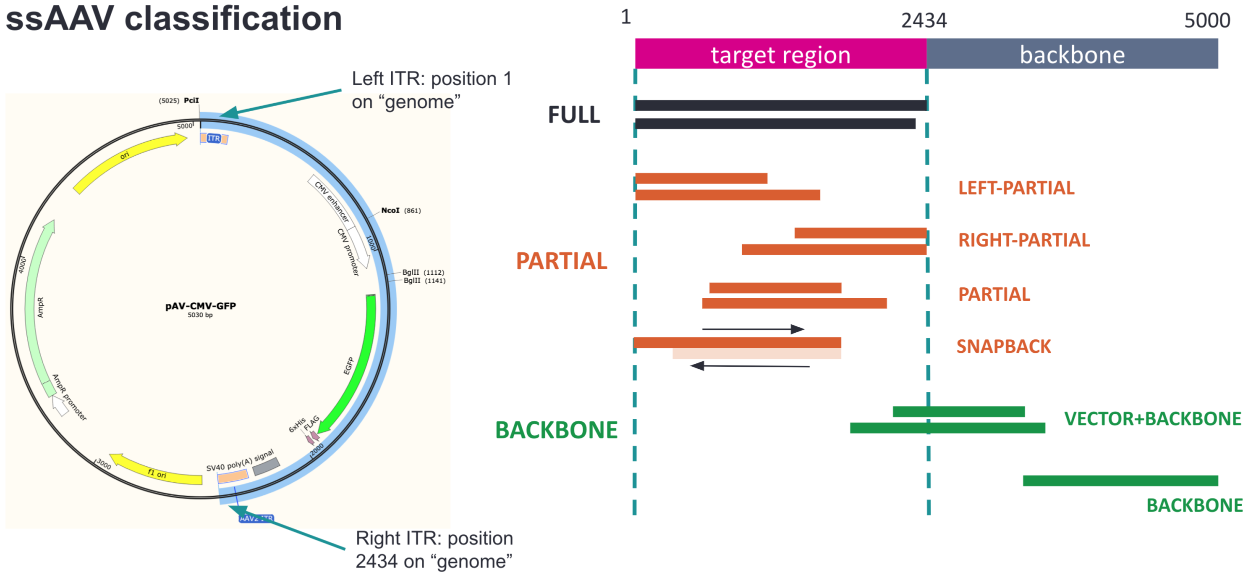


Figure 1: Single-stranded AAV types (ssAAV). ssAAV reads are expected to have a single (forward) alignment to the vector genome. "ssAAV full" reads cover the target region from left ITR to right ITR. "ssAAV left-partial" contain the left ITR but are missing the right ITR. "scAAV vector+backbone" reads overlap both the ITR-to-ITR region and the plasmid backbone, beyond the ITR.

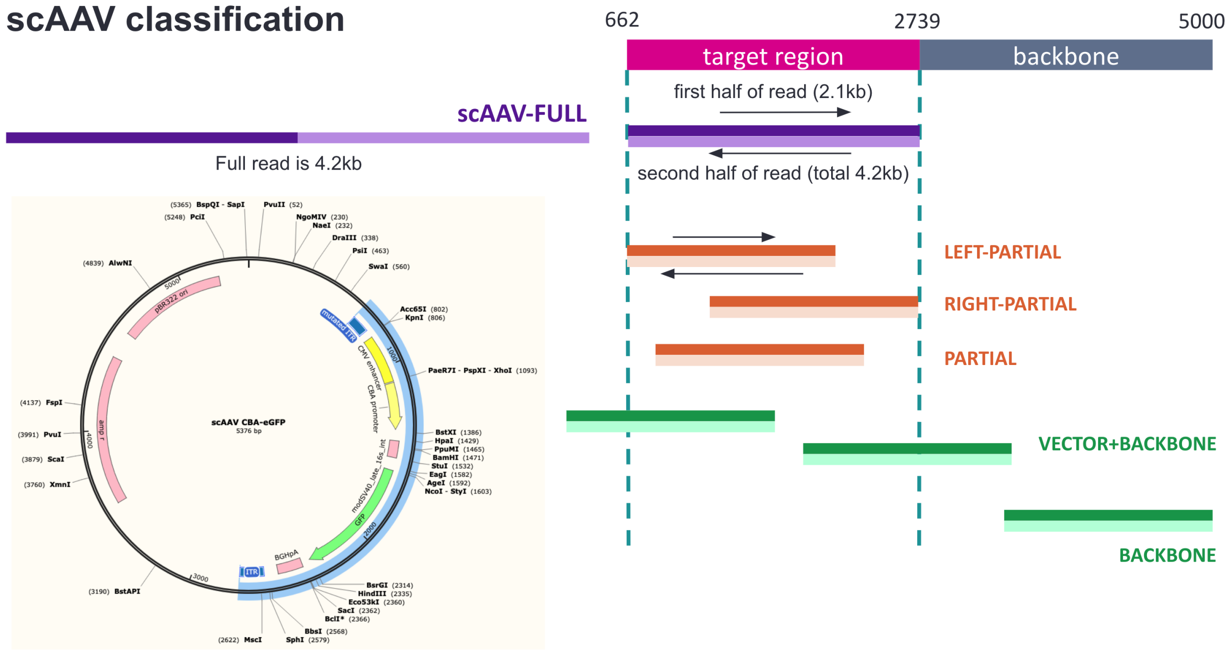
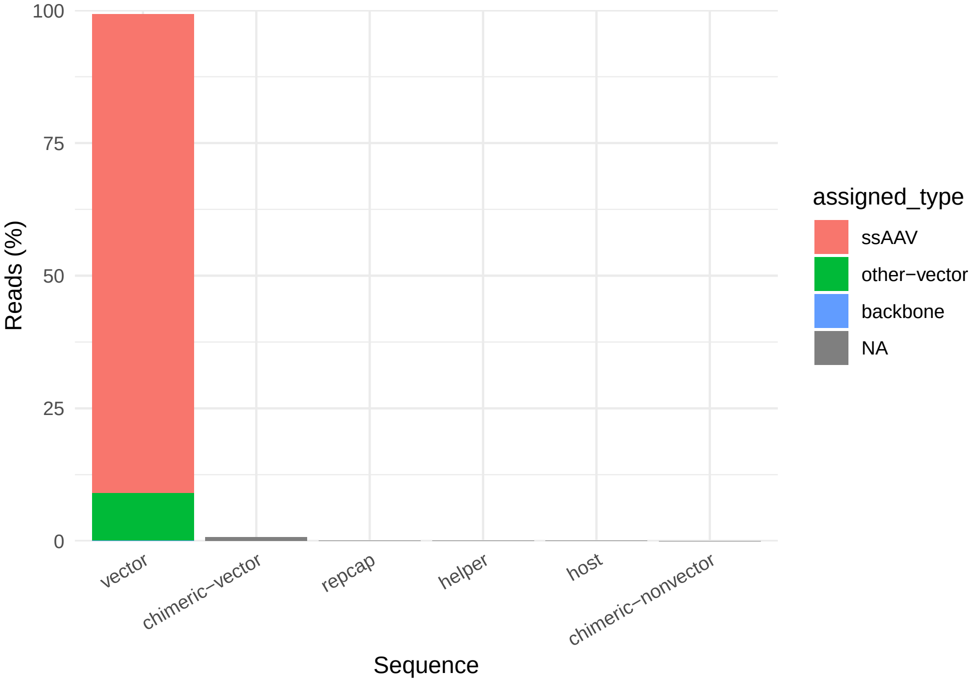
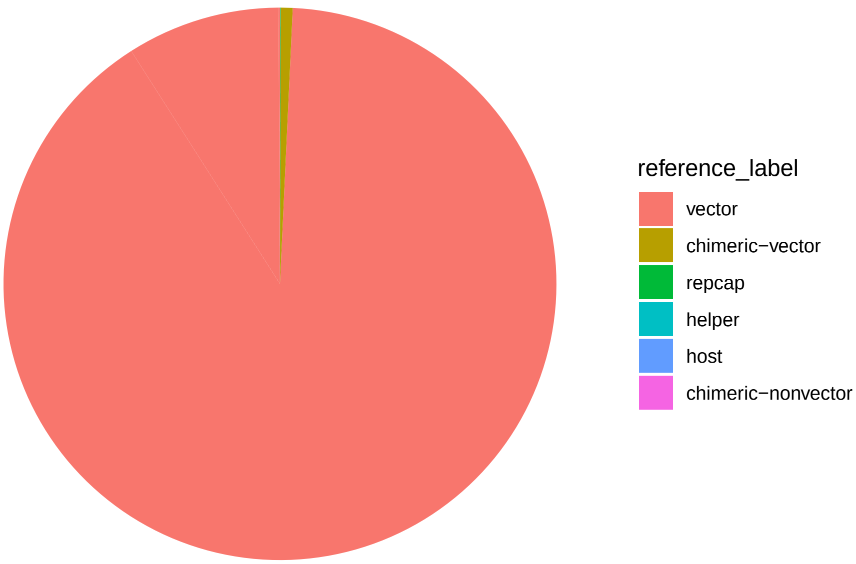


Figure 2: Self-complementary AAV types (scAAV). scAAV reads are expected to have both forward and reverse-complementary alignments to the vector genome. The definitions of full, partial, and backbone reads are the same as in ssAAV.

Assigned types by read alignment characteristics

Mapped Reference	Assigned Type	Count	Frequency (%)
vector	ssAAV	223,254	90.22
vector	other-vector	22,177	8.96
vector	backbone	195	0.08
chimeric-vector		1,723	0.70
repcap		74	0.03
helper		13	0.01
host		13	0.01
chimeric-nonvector		1	0.00

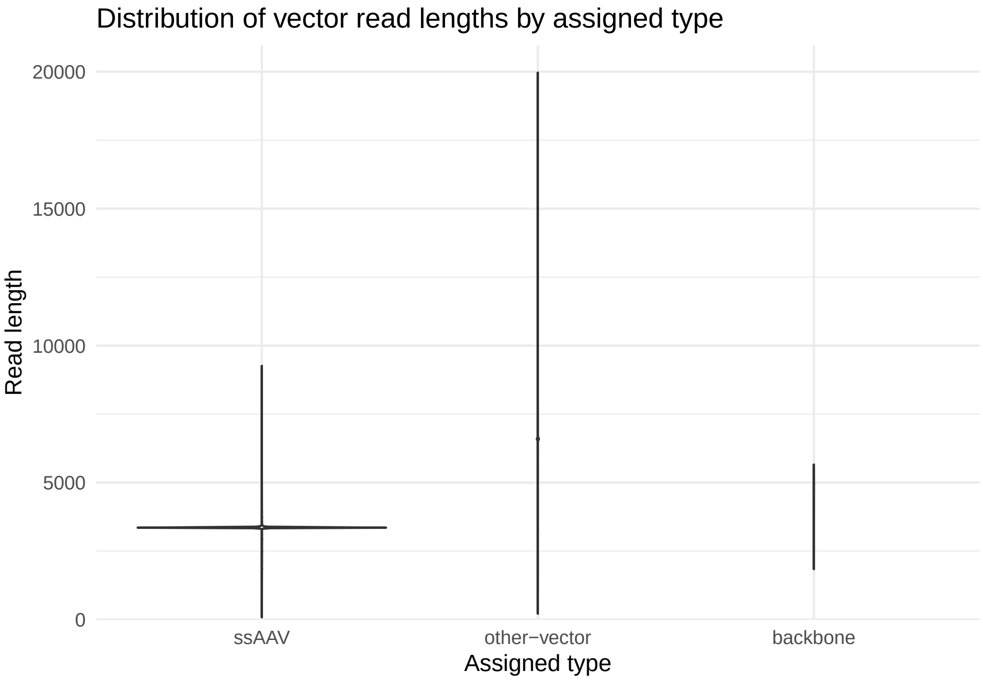




Frequency of vector assigned types

Assigned Type	Count	Frequency in AAV (%)	Total Frequency (%)
ssAAV	223,254	90.89	90.22
other-vector	22,177	9.03	8.96
backbone	195	0.08	0.08

Distribution of vector read lengths by assigned types



Assigned subtypes detailed analysis

Definitions

Table 6: Assigned subtype definitions for single-stranded input vector.

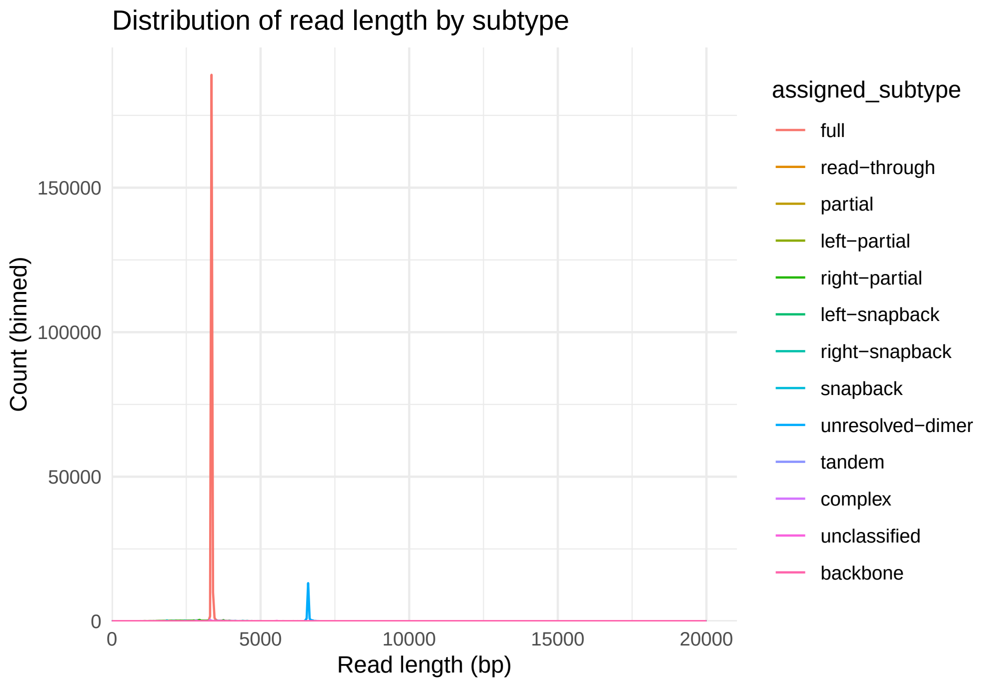
Assigned Type	Assigned Subtype	Definition
ssAAV	full	Read aligns to a fragment of the vector originating from the left (upstream) ITR and ending at the right (downstream) ITR of the vector.
ssAAV	full-gap	Read aligns to the vector ITR-to-ITR region, as with “full”, but with a significant number of gaps in the alignment between the ITRs.
ssAAV	read-through	Read aligns to a fragment including the vector as well as plasmid backbone sequence. May imply read-through beyond the right ITR, or reverse packaging if the alignment is to only the left ITR and backbone.
ssAAV	partial	Read aligns to a fragment of the vector originating from within the ITR sequences.
ssAAV	left-partial	Read aligns to a fragment of the vector originating from the left (upstream) ITR of the vector while not covering the right ITR.
ssAAV	right-partial	Read aligns to a fragment of the vector originating from the right (downstream) ITR of the vector while not covering the left ITR.
ssAAV	left-snapback	Read consists of a double-stranded, sub-genomic fragment including only the left ITR and aligned symmetrically to the (+) and (-) strands.

Assigned Type	Assigned Subtype	Definition
ssAAV	right-snapback	Read consists of a double-stranded, sub-genomic fragment including only the right ITR and aligned symmetrically to the (+) and (-) strands.
other-vector	snapback	Read aligns to a double-stranded fragment in both (+) and (-) strands, but does not include either ITR.
other-vector	unresolved-dimer	Read aligns to a double-stranded fragment covering the full ITR-to-ITR region in both (+) and (-) strands. A dimer in ssAAV context, twice the size of a ssAAVV-full vector genome.
other-vector	tandem	Read has two or more overlapping alignments on the same strand, but none on the reverse strand, indicating tandem duplication of the same region.
other-vector	complex	Read aligns to a double-stranded fragment with asymmetrical and/or multiple alignments on the (+) and (-) strands.
other-vector	unclassified	Read alignment does not match any of the above orientations.

Assigned types and subtypes

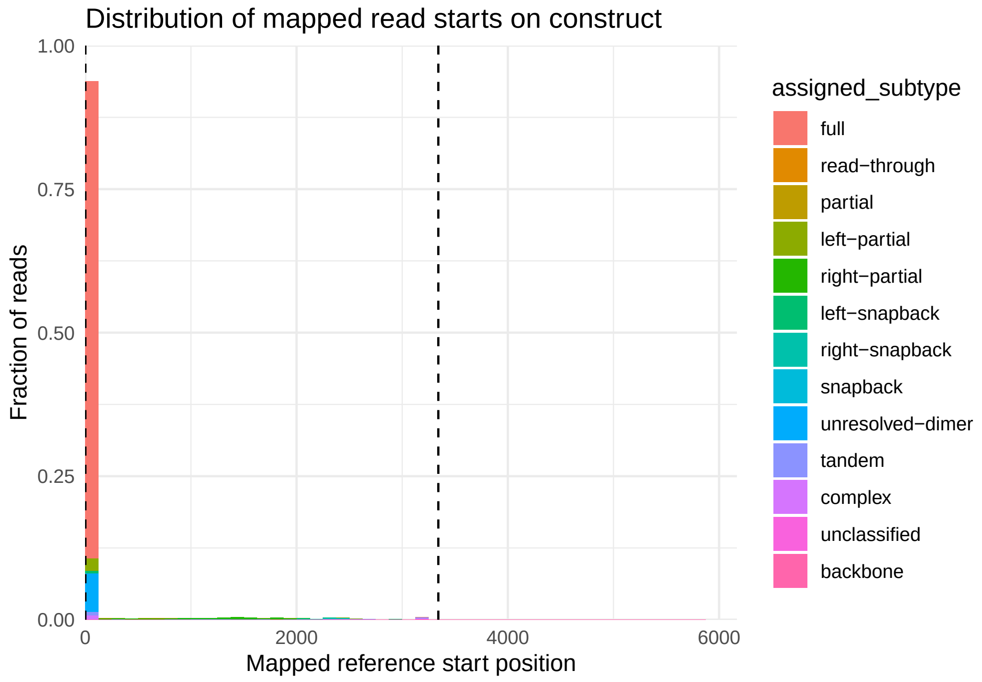
Assigned Type	**Assigned Subtype**	**Count**	**Freq. in AAV (%)**	**Total Freq. (%)**
ssAAV	full	204,042	83.07	82.46
ssAAV	right-partial	8,283	3.37	3.35
ssAAV	left-partial	5,192	2.11	2.10
ssAAV	right-snapback	3,955	1.61	1.60
ssAAV	left-snapback	1,044	0.43	0.42
ssAAV	partial	488	0.20	0.20
ssAAV	read-through	250	0.10	0.10
other-vector	unresolved-dimer	16,480	6.71	6.66
other-vector	complex	2,969	1.21	1.20
other-vector	tandem	2,071	0.84	0.84
other-vector	unclassified	402	0.16	0.16
other-vector	snapback	255	0.10	0.10
backbone	backbone	195	0.08	0.08

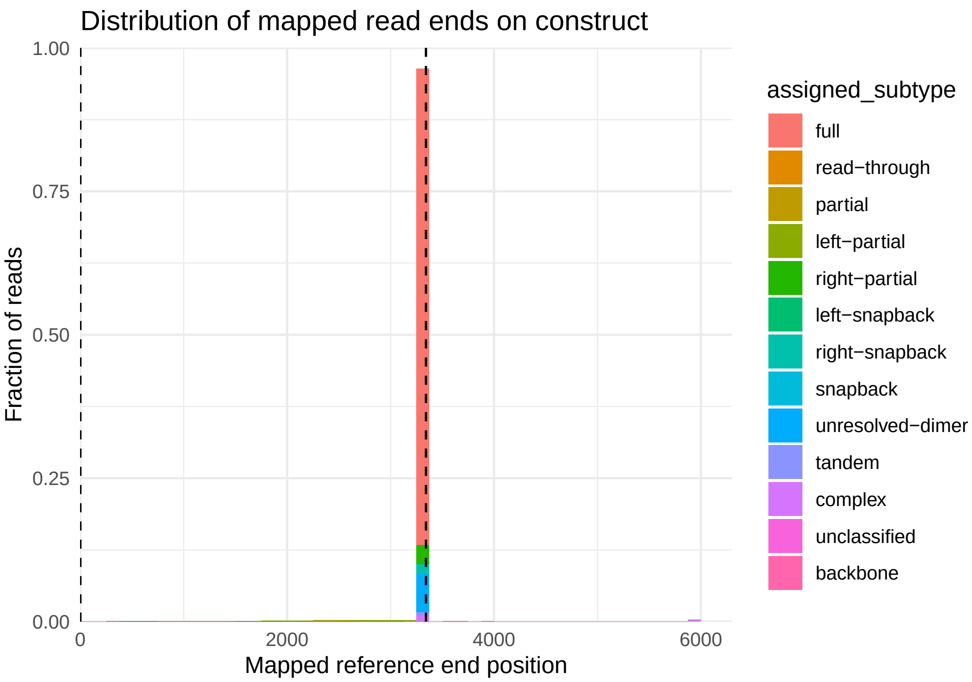
Distribution of read length by subtype

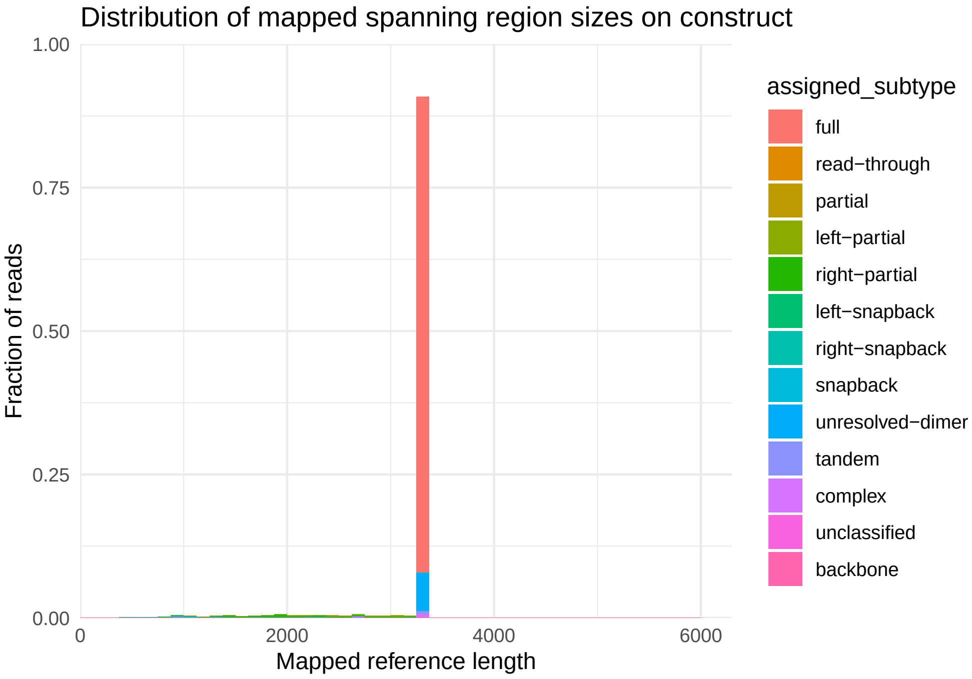


AAV mapping to reference sequence

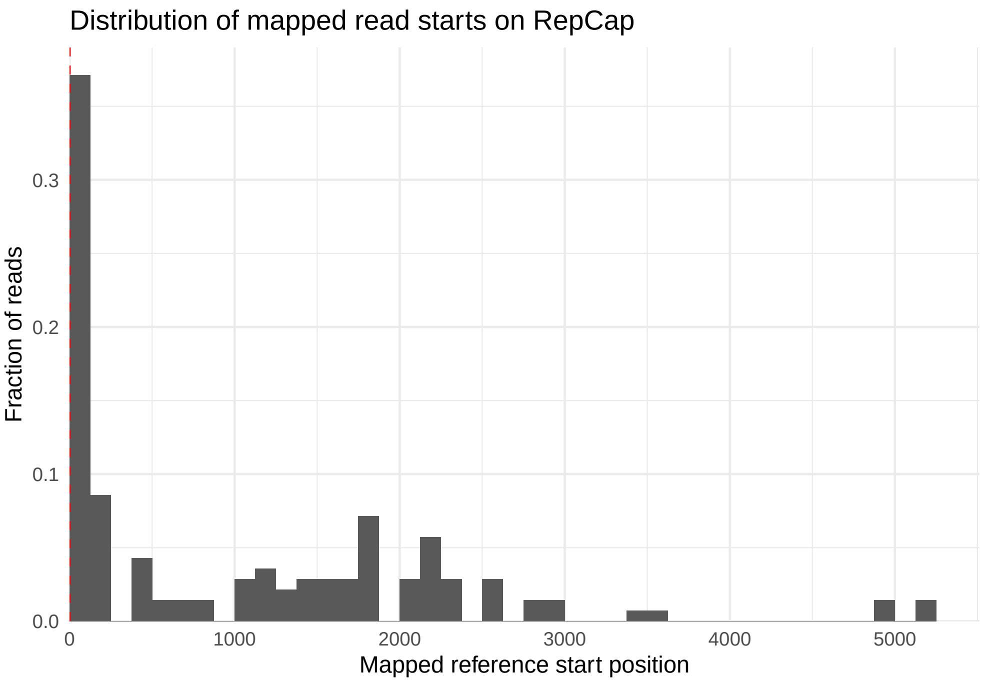
Gene therapy construct

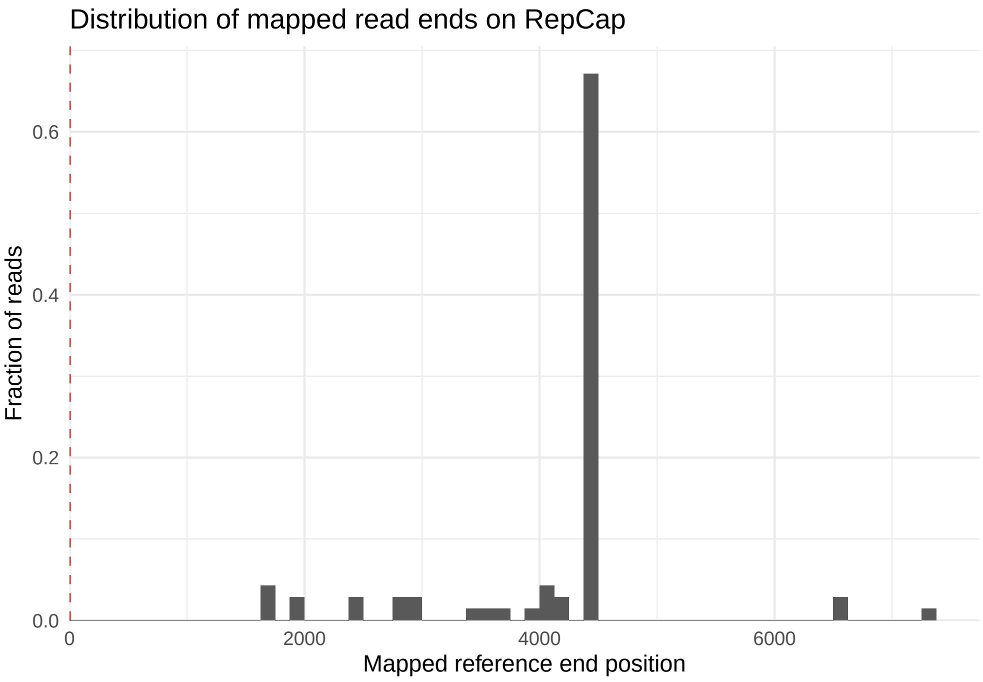


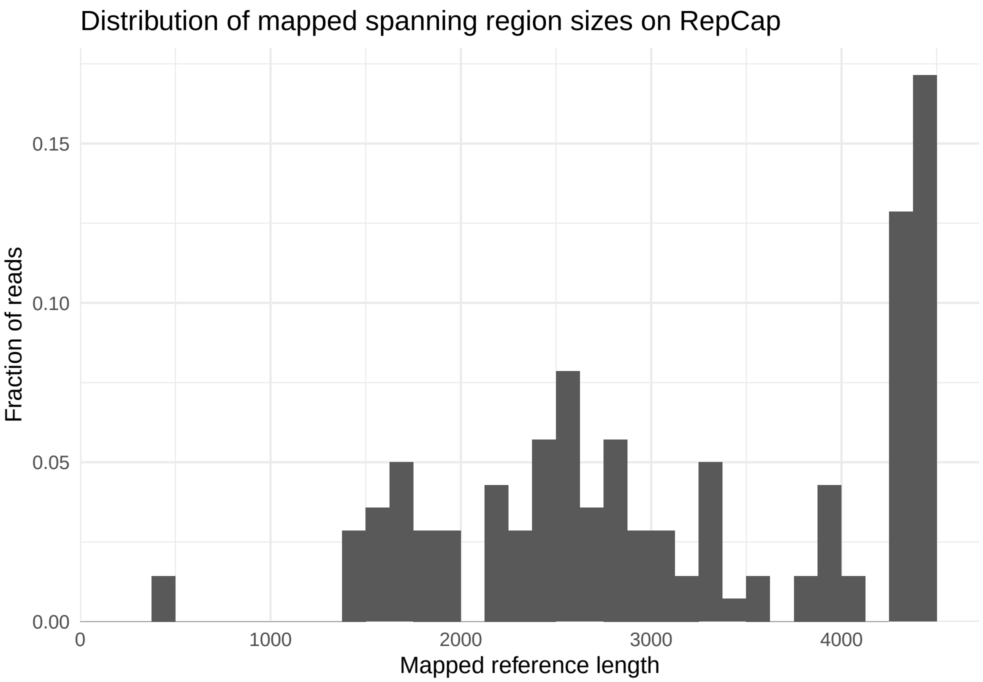

Distribution of mapped read ends on construct
Fraction of reads
1.00
0.75
0.50
0.25
0.00
0
2000
4000
6000
Mapped reference end position
assigned_subtype
full
read−through
partial
left−partial
right−partial
left−snapback
right−snapback
snapback
unresolved−dimer
tandem
complex
unclassified
backbone


Distribution of mapped spanning region sizes on construct
Fraction of reads
1.00
0.75
0.50
0.25
0.00
0
2000
4000
6000
Mapped reference length
assigned_subtype
full
read−through
partial
left−partial
right−partial
left−snapback
right−snapback
snapback
unresolved−dimer
tandem
complex
unclassified
backbone

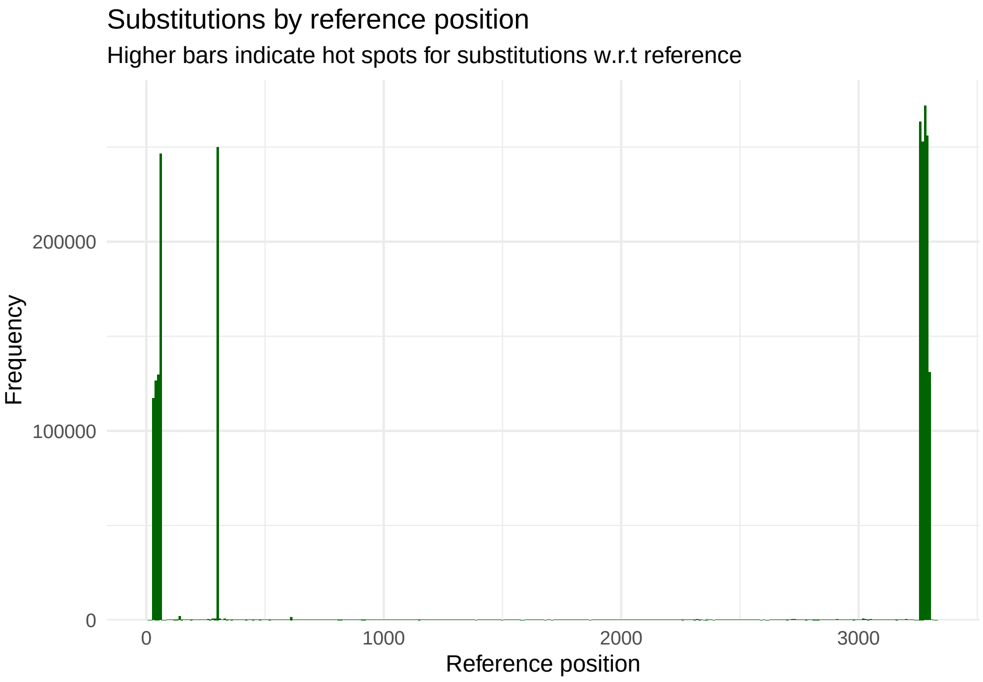
RepCap

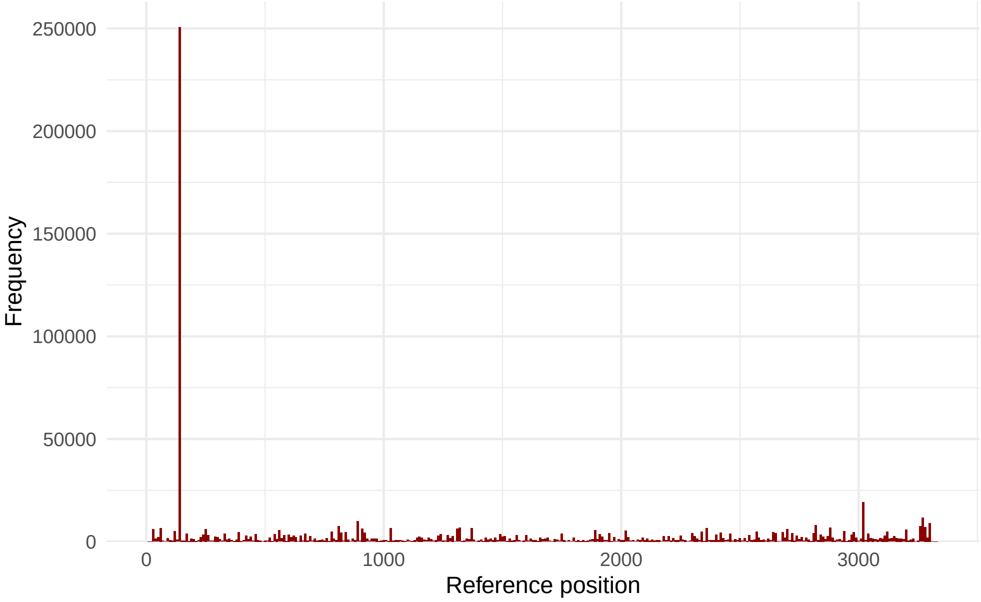

Distribution of mapped read starts on RepCap
Fraction of reads
0.0
0.1
0.2
0.3
0
1000
2000
3000
4000
5000
Mapped reference start position


Distribution of mapped read ends on RepCap
Fraction of reads
0.0
0.2
0.4
0.6
0
2000
4000
6000
Mapped reference end position



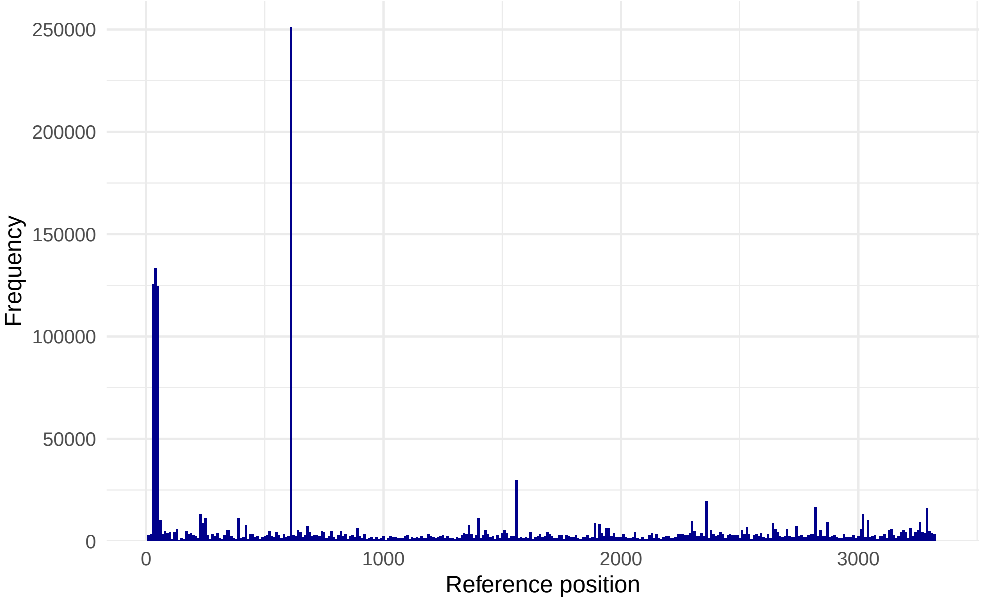
Distribution of non-matches by reference position





Deletions by reference position

Higher bars indicate hot spots for deletion w.r.t reference

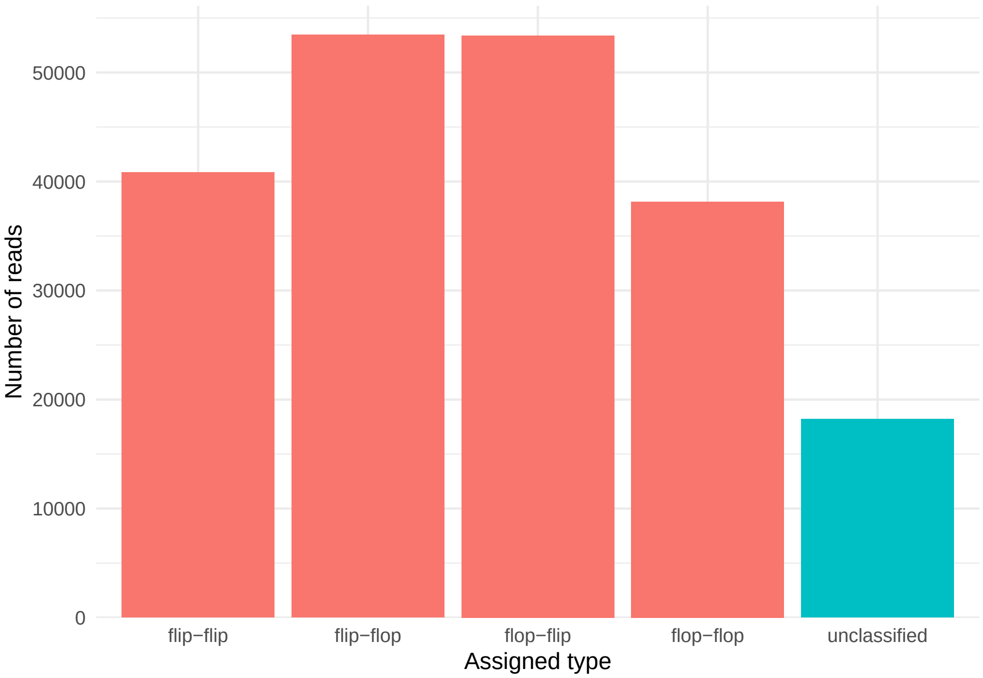


Insertions by reference position

Higher bars indicate hot spots for insertion w.r.t reference

Flip/flop configurations

Term	Definition
Flip/Flop	One ITR is formed by two palindromic arms, called B–B’ and C–C’, embedded in a larger one, A–A’. The order of these palindromic sequences defines the flip or flop orientation of the ITR. (Read more)



type	subtype	leftITR	rightITR	count
ssAAV	full	flip	flip	40,833
ssAAV	full	flip	flop	53,454
ssAAV	full	flip	unclassified	3,635
ssAAV	full	flop	flip	53,395
ssAAV	full	flop	flop	38,153
ssAAV	full	flop	unclassified	3,522
ssAAV	full	unclassified	flip	4,783
ssAAV	full	unclassified	flop	4,235
ssAAV	full	unclassified	unclassified	2,032
ssAAV	left-partial	flip	unclassified	3,276
ssAAV	left-partial	flop	unclassified	3,399
ssAAV	left-partial	unclassified	unclassified	605
ssAAV	right-partial	unclassified	flip	7,437
ssAAV	right-partial	unclassified	flop	7,521
ssAAV	right-partial	unclassified	unclassified	1,235

Methods

This report was generated by an automated analysis of long-read sequencing data from adeno-associated virus (AAV) products. The sequencing data should be from the PacBio sequencer run in AAV mode, or equivalent circular consensus sequencing (CCS) reads (Travers et al., 2010).

In this analysis, reads are aligned to the given AAV, packaging, and host reference sequences using Minimap2 (Li, 2018). The reference sequences for each primary alignment and its orientation are counted and summarized to assign read type classifications, including vector, non-vector, and chimeric reads. For reads assigned to the AAV vector, the primary alignment coordinates are compared to the annotated vector region in the reference sequence, which comprises the left and right ITRs and the genomic region between them, to assign each read to a subtype classification. Sequence variants relative to the vector reference sequence are determined directly from each read's alignment, specifically the CIGAR string indicating insertions, deletions, mismatches, and gaps.

Citations

1. Travers, K. J., Chin, C.-S., Rank, D. R., Eid, J. S. & Turner, S. W. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Research 38, e159–e159 (2010).
2. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100 (2018).