

AAV Sequence Analysis

Sample name: bc2170-AAV1-Adh-SOP-Vega

Date: 2026-02-11

Contents

Analysis context	1
Read-based AAV vector type classification	2
Definitions	2
Assigned types by read alignment characteristics	4
Frequency of vector assigned types	5
Distribution of vector read lengths by assigned types	6
Assigned subtypes detailed analysis	6
Definitions	6
Assigned types and subtypes	7
Distribution of read length by subtype	8
AAV mapping to reference sequence	8
Gene therapy construct	8
RepCap	10
Distribution of non-matches by reference position	11
Flip/flop configurations	12
Methods	14
Citations	14

Analysis context

	Value
Sample unique ID	bc2170-AAV1-Adh-SOP-Vega
Sequencing run ID	m21026_251001_215608
Construct vector type	ssAAV

Read-based AAV vector type classification

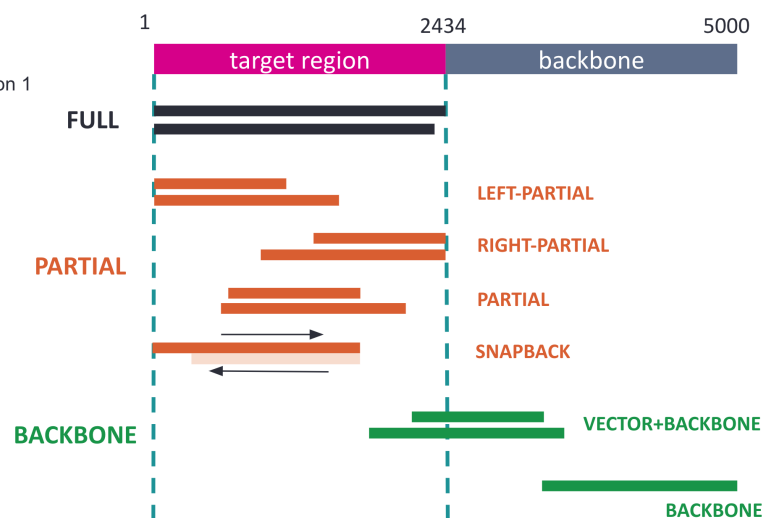
Definitions

Table 2: Reference label definitions.

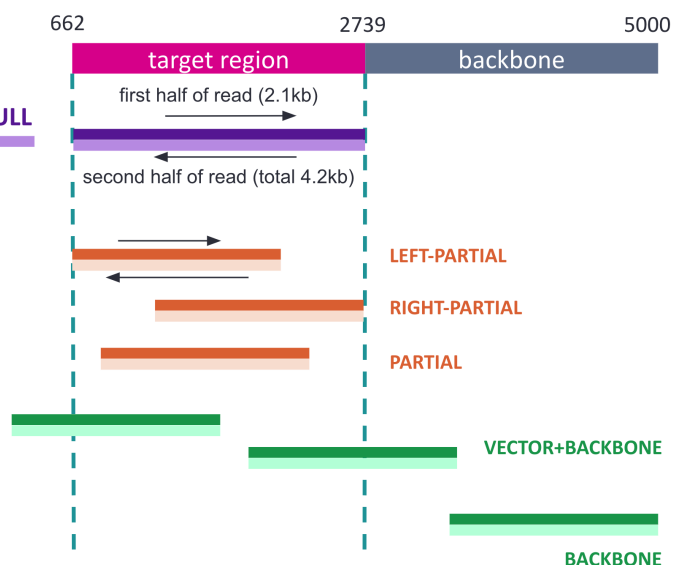
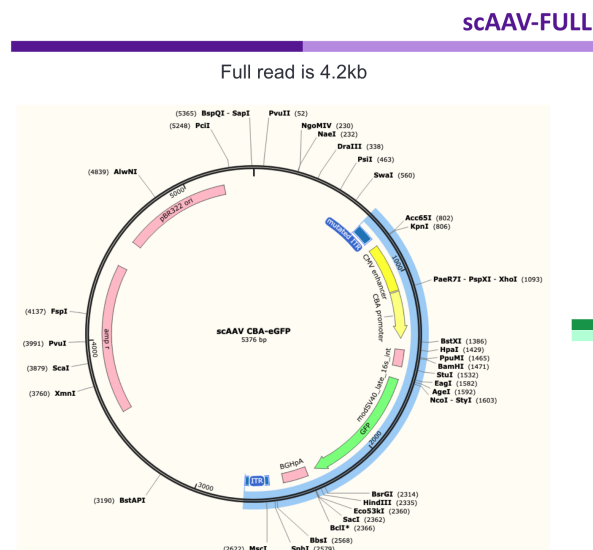
Reference Label	Definition
vector	Read originates from the vector plasmid.
repcap	Read originates from the RepCap plasmid. The Rep gene encodes four proteins (Rep78, Rep68, Rep52, and Rep40), which are required for viral genome replication and packaging, while Cap expression gives rise to the viral capsid 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.

[illegible]

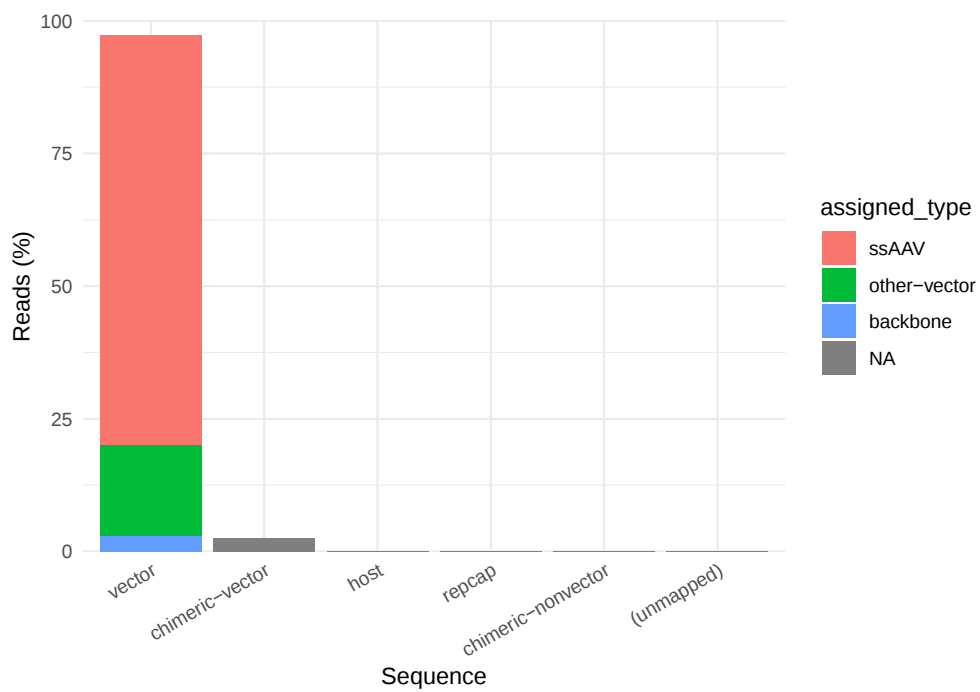
scAAV classification

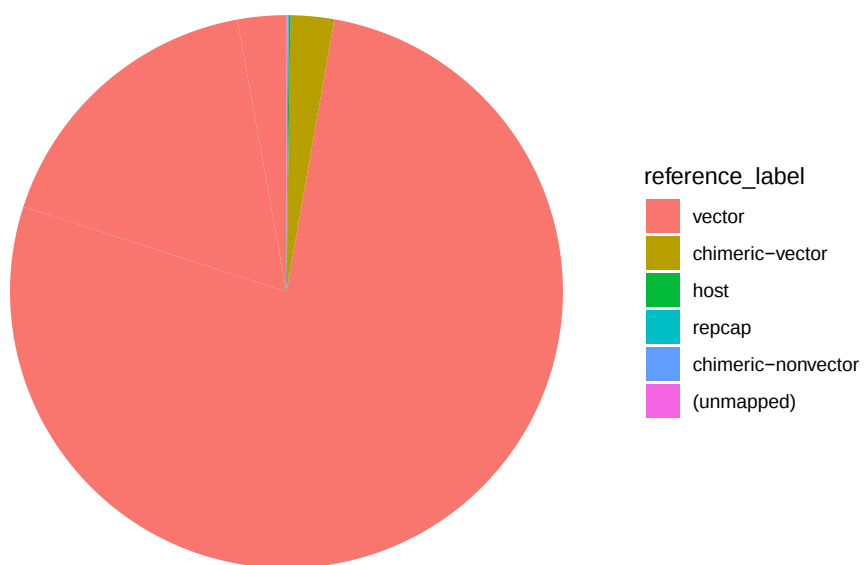


3

Assigned types by read alignment characteristics

Mapped Reference	Assigned Type	Count	Frequency (%)
vector	ssAAV	151,111	77.19
vector	other-vector	33,640	17.18
vector	backbone	5,574	2.85
chimeric-vector		4,984	2.55
host		187	0.10
repcap		148	0.08
chimeric-nonvector		1	0.00
(unmapped)		111	0.06

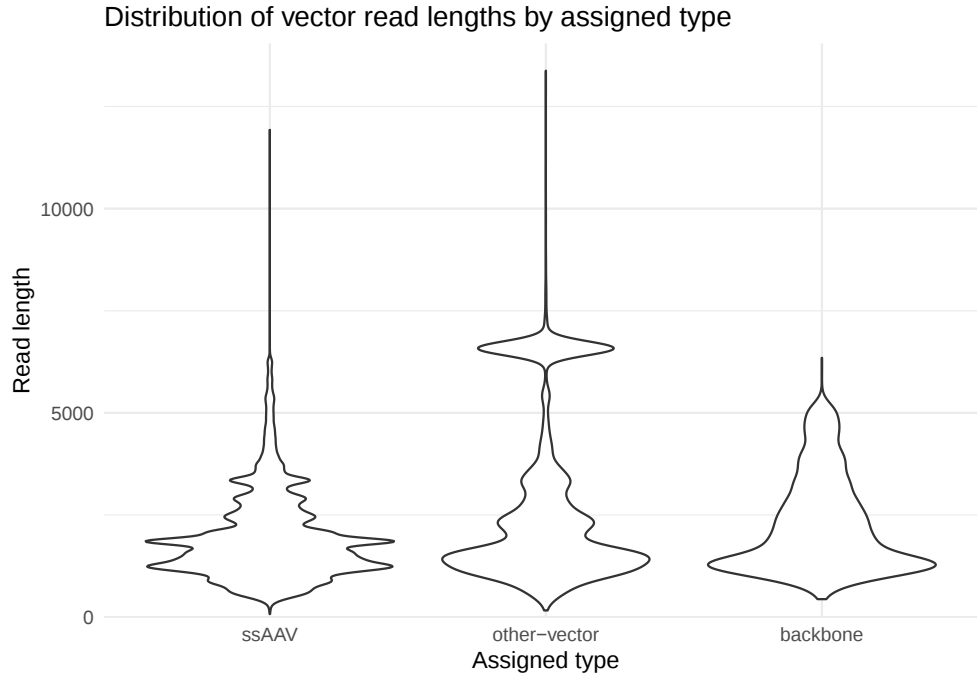




Frequency of vector assigned types

Assigned Type	Count	Frequency in AAV (%)	Total Frequency (%)
ssAAV	151,111	79.40	77.19
other-vector	33,640	17.68	17.18
backbone	5,574	2.93	2.85

Distribution of vector read lengths by assigned types



Assigned subtypes detailed analysis

Definitions

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

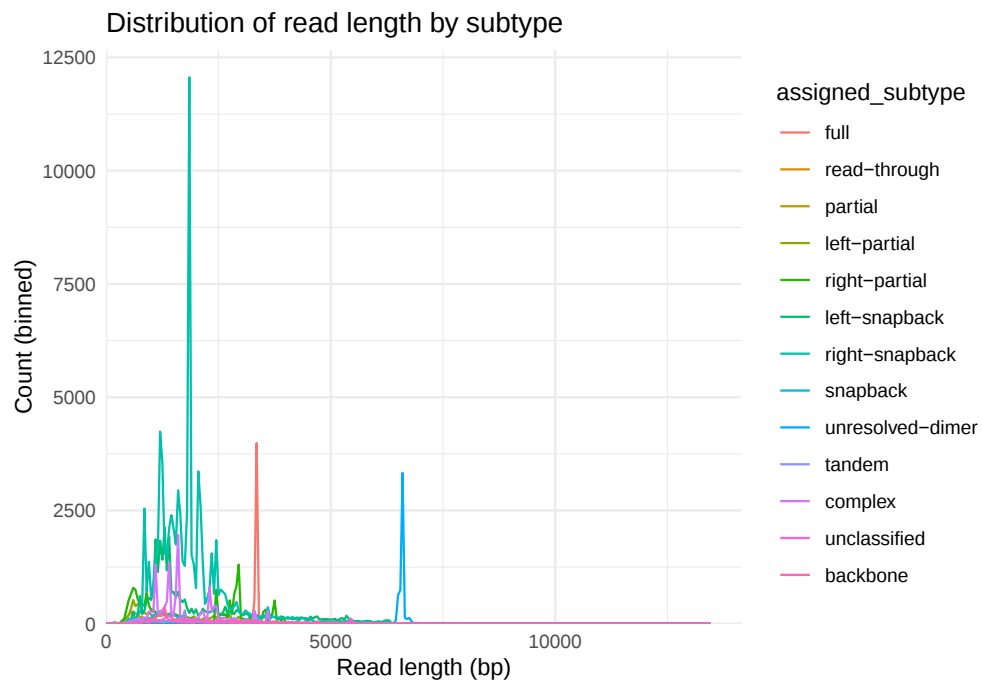
Assigned Type	Assigned Subtype	Definition
ssAAV	full	Read aligns to a fragment of the vector originating from the left (upstream) ITR and ending at the right (downstream) ITR of the vector.
ssAAV	full-gap	Read aligns to the vector ITR-to-ITR region, as with “full”, but with a significant number of gaps 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

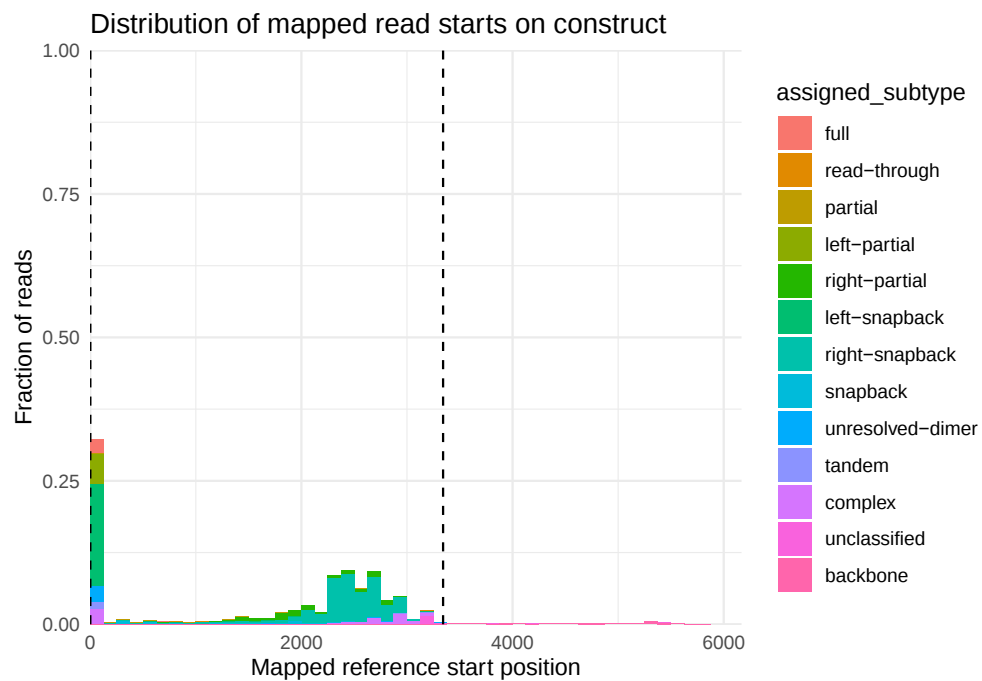
Assigned Type	Assigned Subtype	Count	Freq. in AAV (%)	Total Freq. (%)
ssAAV	right-snapback	80,400	42.24	41.07
ssAAV	left-snapback	33,786	17.75	17.26
ssAAV	right-partial	18,417	9.68	9.41
ssAAV	left-partial	10,068	5.29	5.14
ssAAV	full	4,793	2.52	2.45
ssAAV	partial	3,079	1.62	1.57
ssAAV	read-through	568	0.30	0.29
other-vector	complex	15,184	7.98	7.76
other-vector	snapback	5,660	2.97	2.89
other-vector	unresolved-dimer	5,545	2.91	2.83
other-vector	tandem	4,117	2.16	2.10
other-vector	unclassified	3,134	1.65	1.60
backbone	backbone	5,574	2.93	2.85

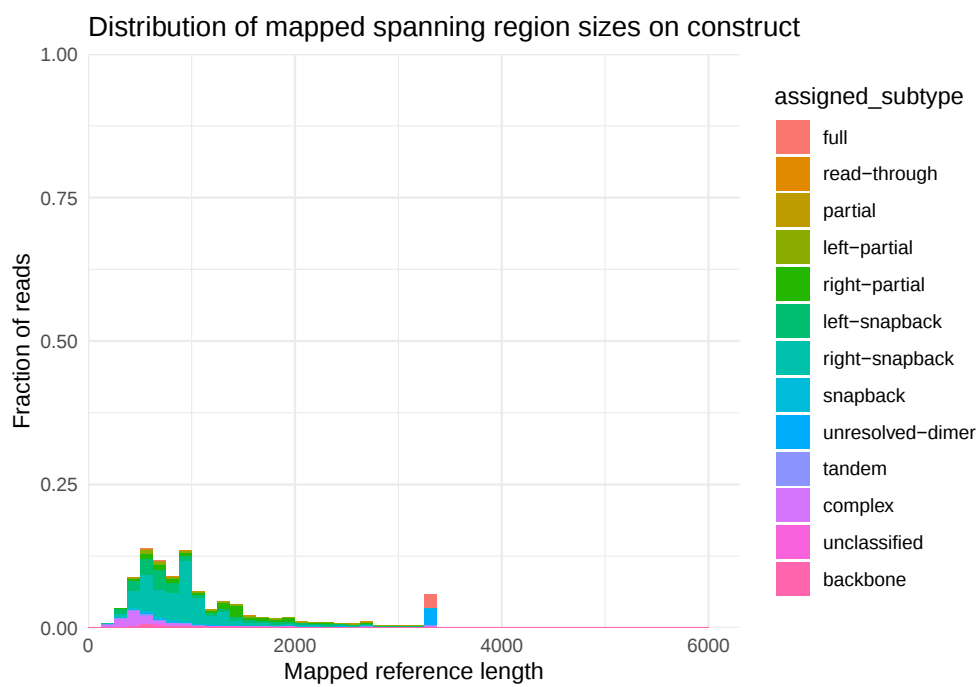
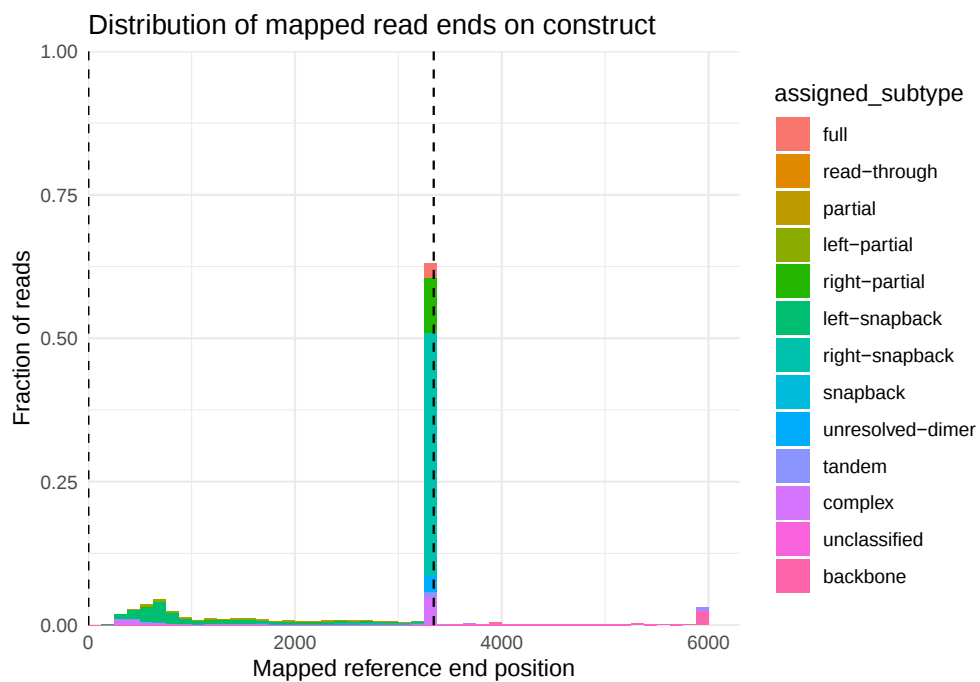
Distribution of read length by subtype



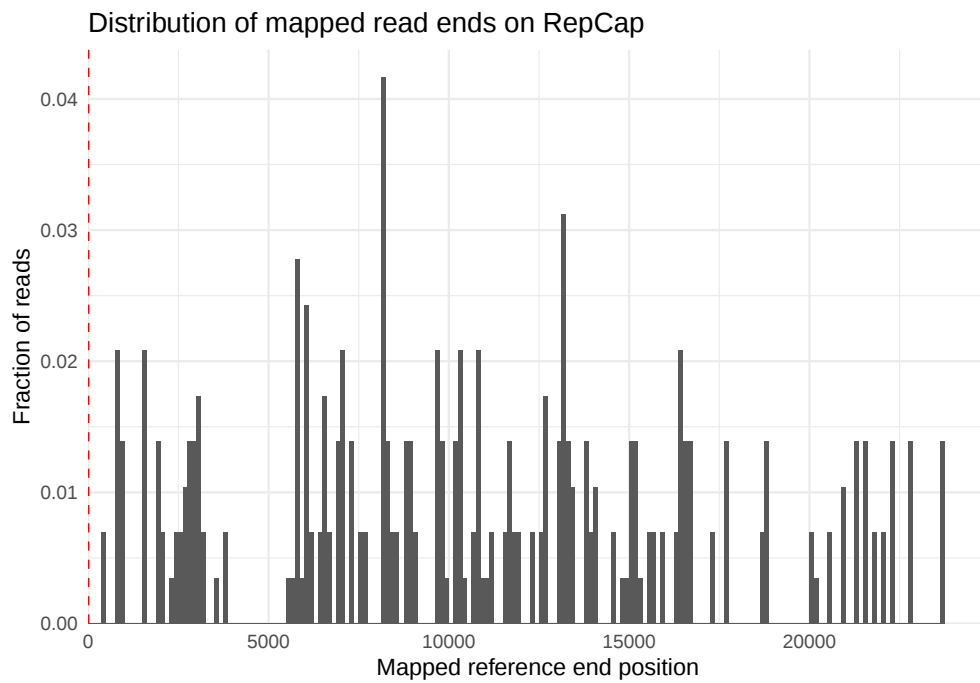
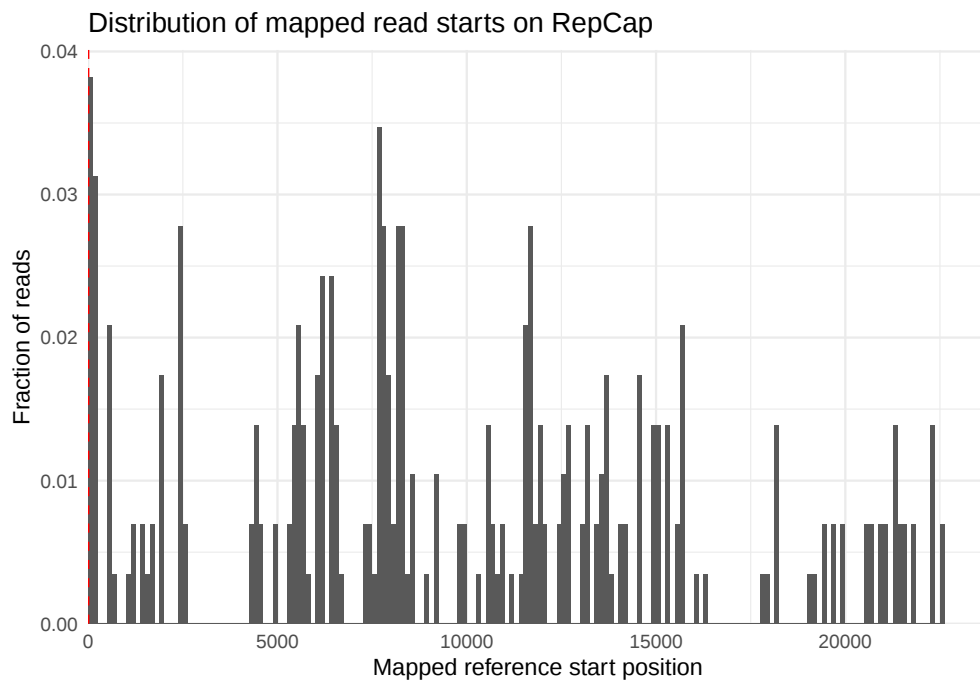
AAV mapping to reference sequence

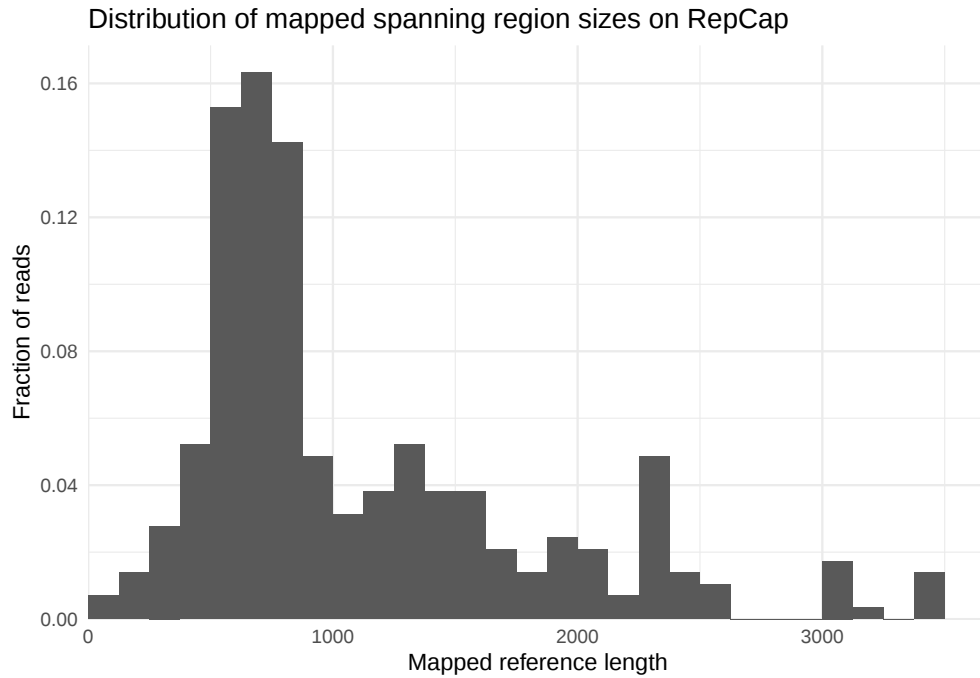
Gene therapy construct



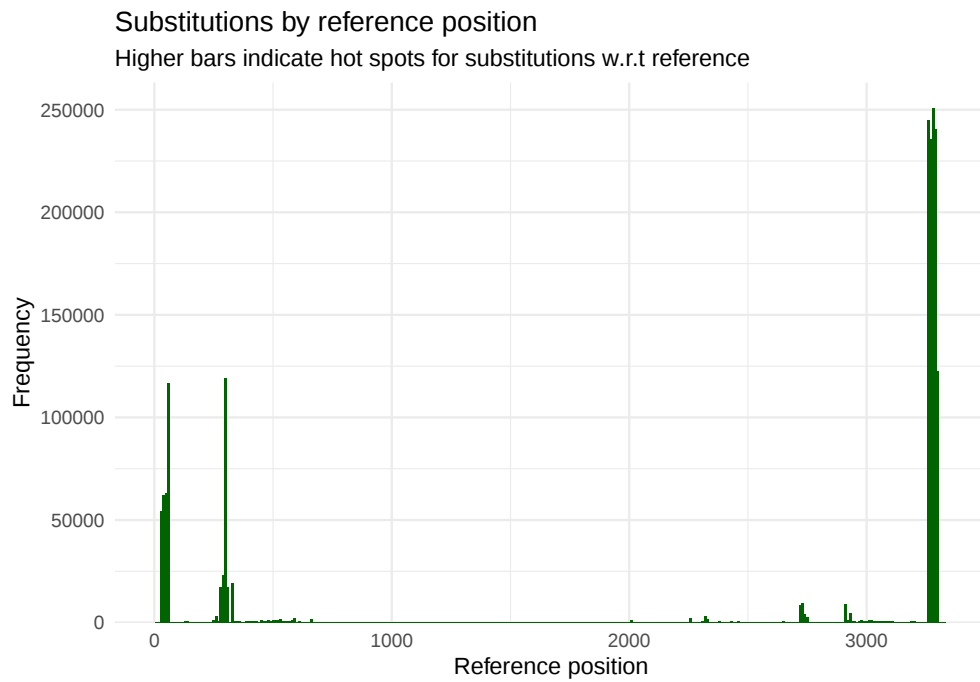


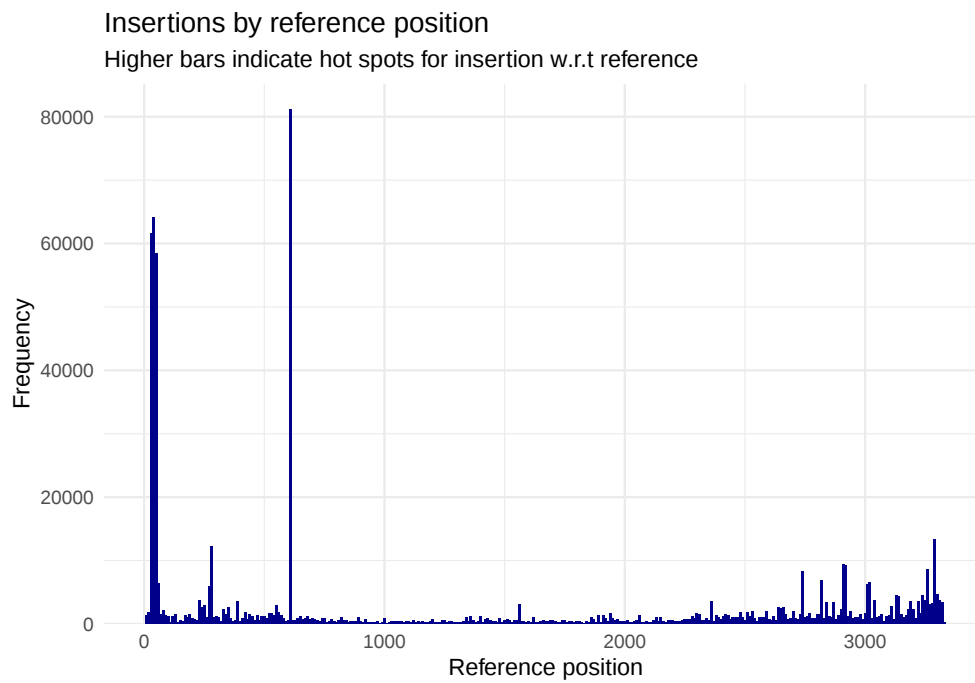
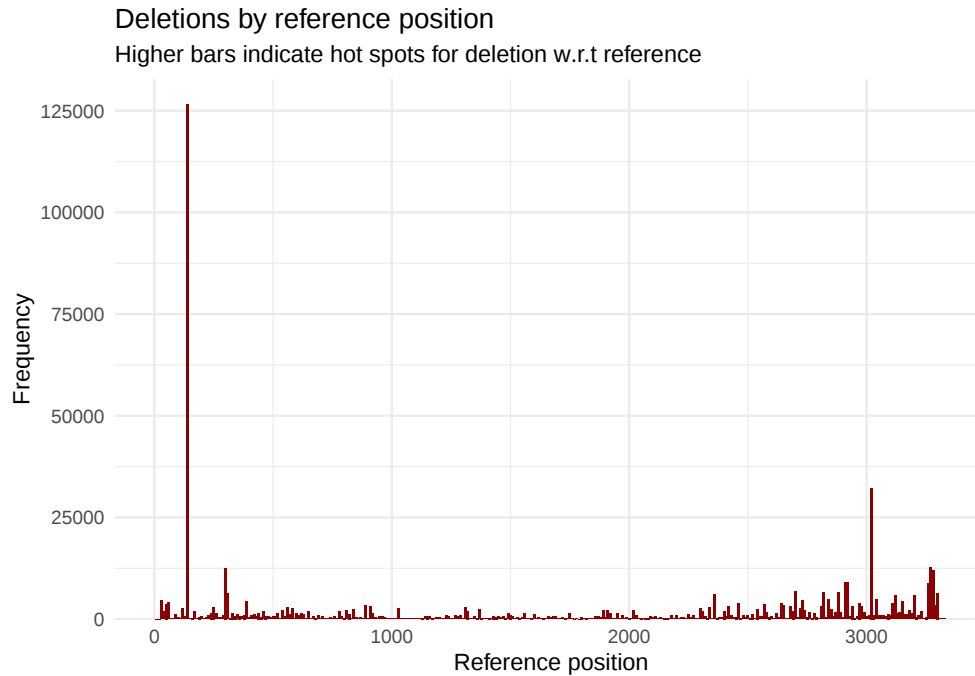
RepCap





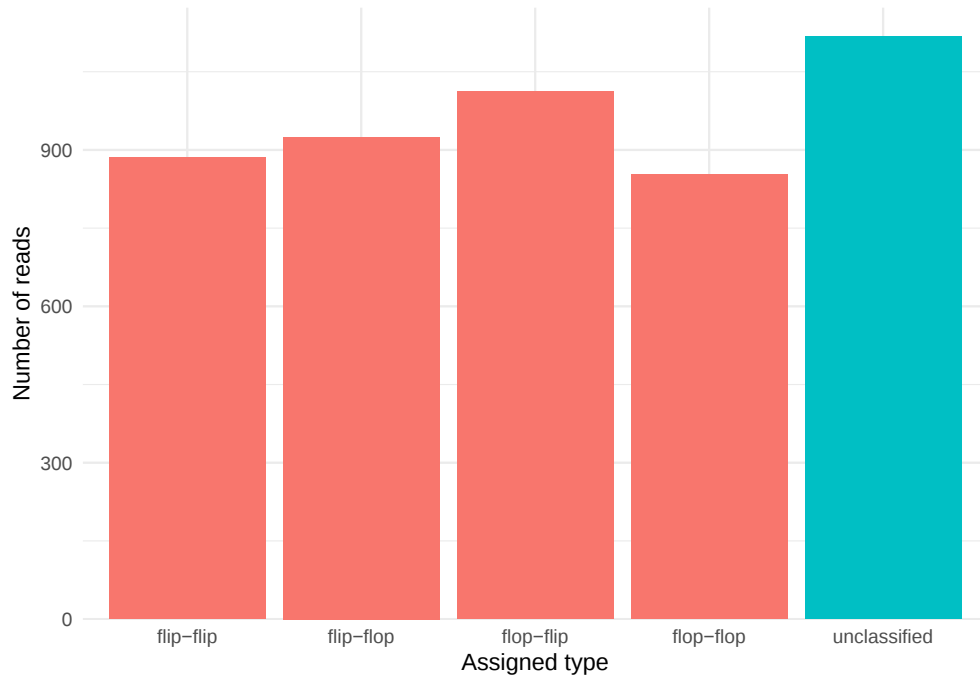
Distribution of non-matches by reference position





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	886
ssAAV	full	flip	flop	925
ssAAV	full	flip	unclassified	226
ssAAV	full	flop	flip	1,012
ssAAV	full	flop	flop	853
ssAAV	full	flop	unclassified	247
ssAAV	full	unclassified	flip	249
ssAAV	full	unclassified	flop	264
ssAAV	full	unclassified	unclassified	131
ssAAV	left-partial	flip	unclassified	33,970
ssAAV	left-partial	flop	unclassified	34,302
ssAAV	left-partial	unclassified	unclassified	9,368
ssAAV	right-partial	unclassified	flip	79,848
ssAAV	right-partial	unclassified	flop	80,261
ssAAV	right-partial	unclassified	unclassified	19,108

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).