

# AAV Sequence Analysis

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

Date: 2025-12-11

## Contents

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

## Analysis context

|                       | Value                       |
|-----------------------|-----------------------------|
| Sample unique ID      | bc2182-AAV8-Adh-SOPmod-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.                                                                                                                                                                                                   |

## ssAAV classification

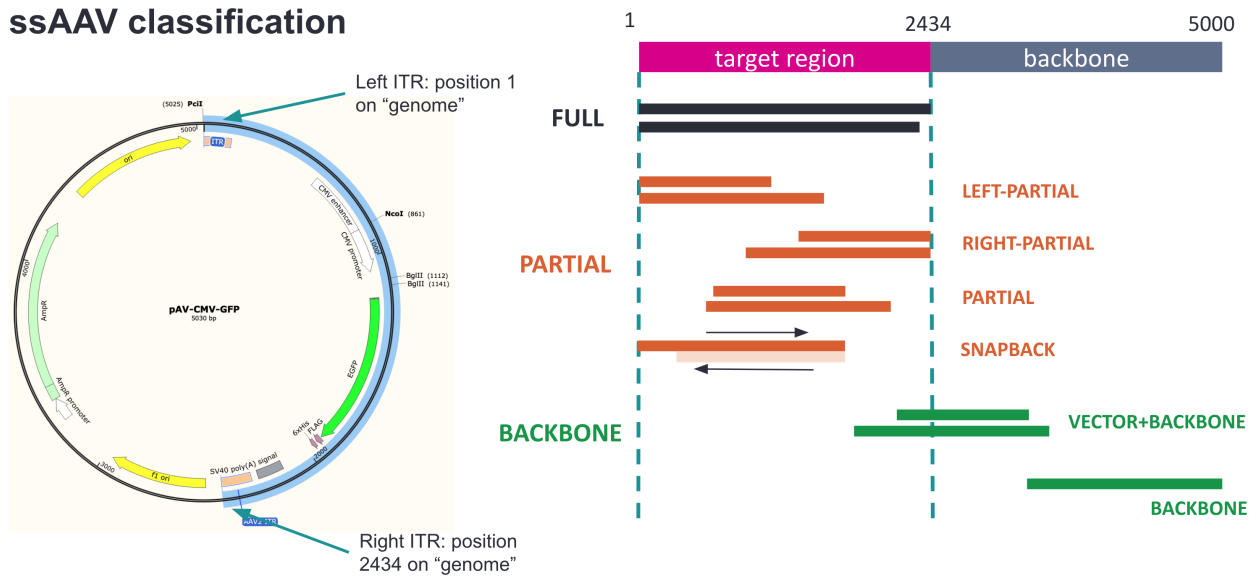


Figure 1: Single-stranded AAV types (ssAAV). ssAAV reads are expected to have a single (forward) alignment to the vector genome. “ssAAV full” reads cover the target region from left ITR to right ITR. “ssAAV left-partial” contain the left ITR but are missing the right ITR. “scAAV vector+backbone” reads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

## scAAV classification

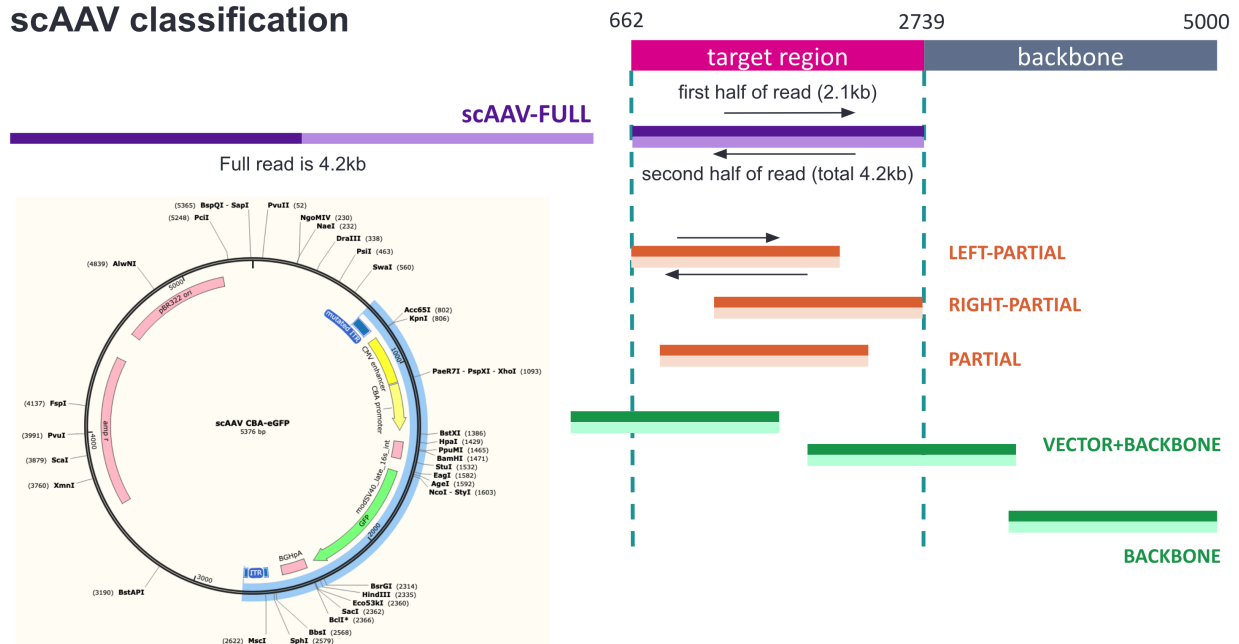
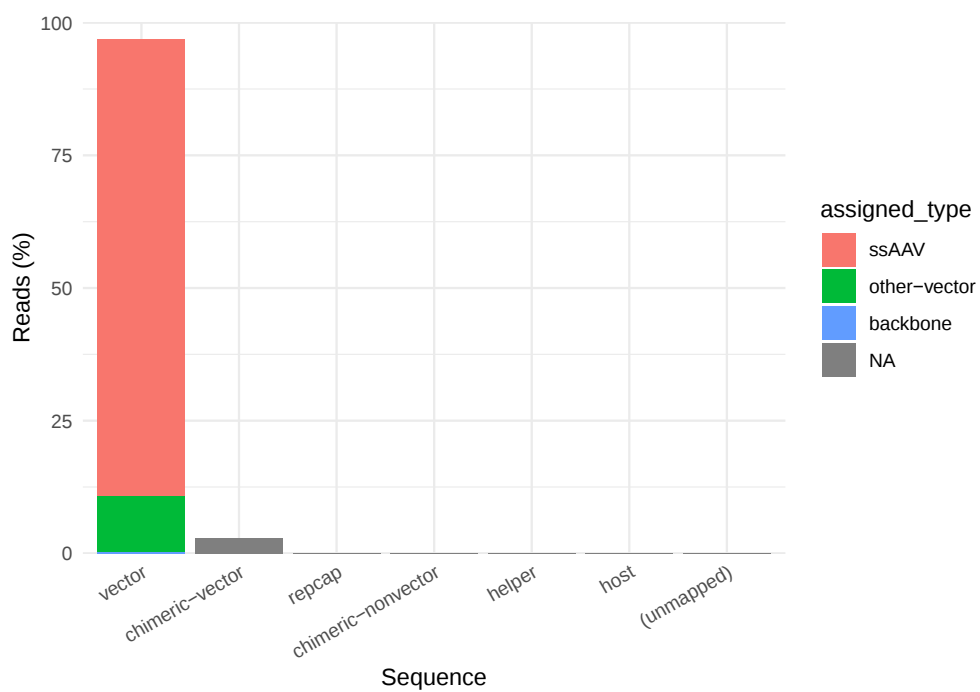
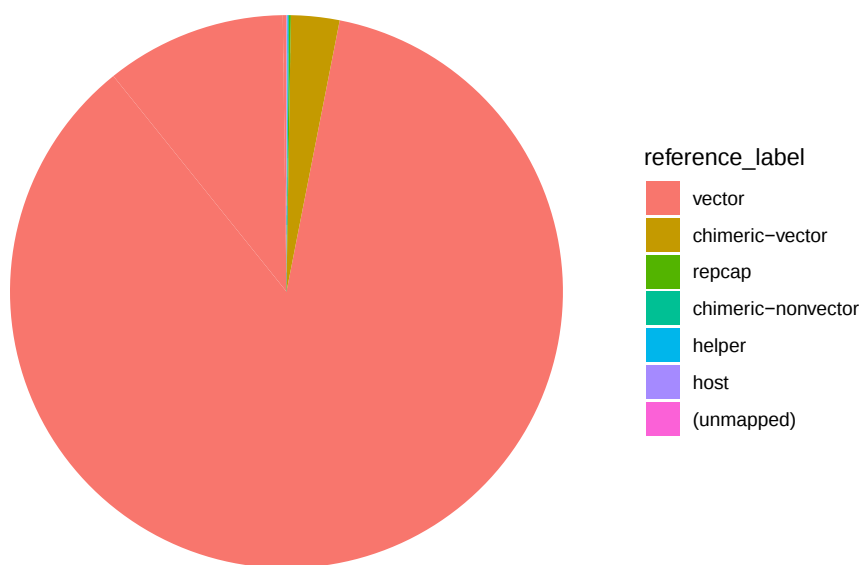


Figure 2: Self-complementary AAV types (scAAV). scAAV reads are expected to have both forward and reverse-complementary alignments to the vector genome. The definitions of full, partial, and backbone reads are the same as in ssAAV.

## Assigned types by read alignment characteristics

| Mapped Reference   | Assigned Type | Count   | Frequency (%) |
|--------------------|---------------|---------|---------------|
| vector             | ssAAV         | 112,890 | 86.15         |
| vector             | other-vector  | 13,833  | 10.56         |
| vector             | backbone      | 279     | 0.21          |
| chimeric-vector    |               | 3,710   | 2.83          |
| repcap             |               | 145     | 0.11          |
| chimeric-nonvector |               | 64      | 0.05          |
| helper             |               | 63      | 0.05          |
| host               |               | 52      | 0.04          |
| (unmapped)         |               | 10      | 0.01          |

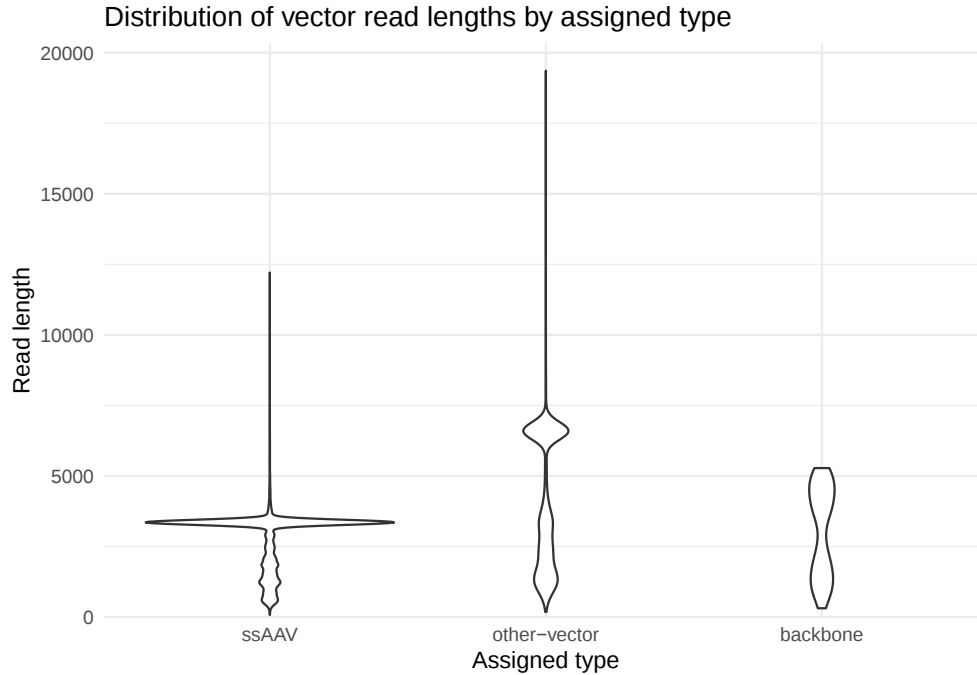




#### Frequency of vector assigned types

| Assigned Type | Count   | Frequency in AAV (%) | Total Frequency (%) |
|---------------|---------|----------------------|---------------------|
| ssAAV         | 112,890 | 88.89                | 86.15               |
| other-vector  | 13,833  | 10.89                | 10.56               |
| backbone      | 279     | 0.22                 | 0.21                |

## Distribution of vector read lengths by assigned types



## Assigned subtypes detailed analysis

### Definitions

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

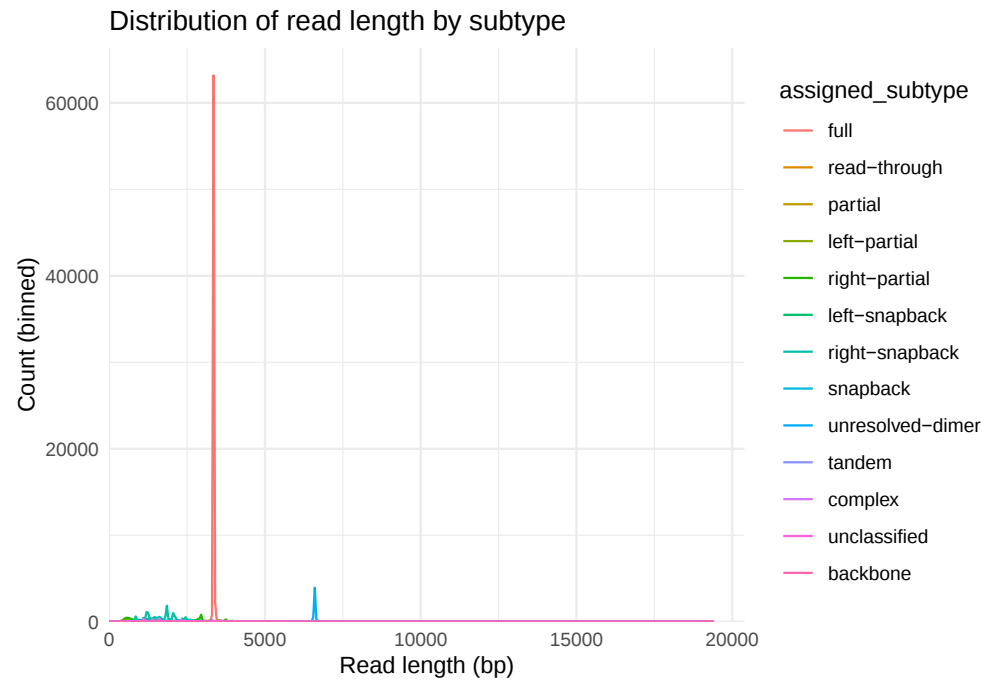
| Assigned Type | Assigned Subtype | Definition                                                                                                                                                                                                    |
|---------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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

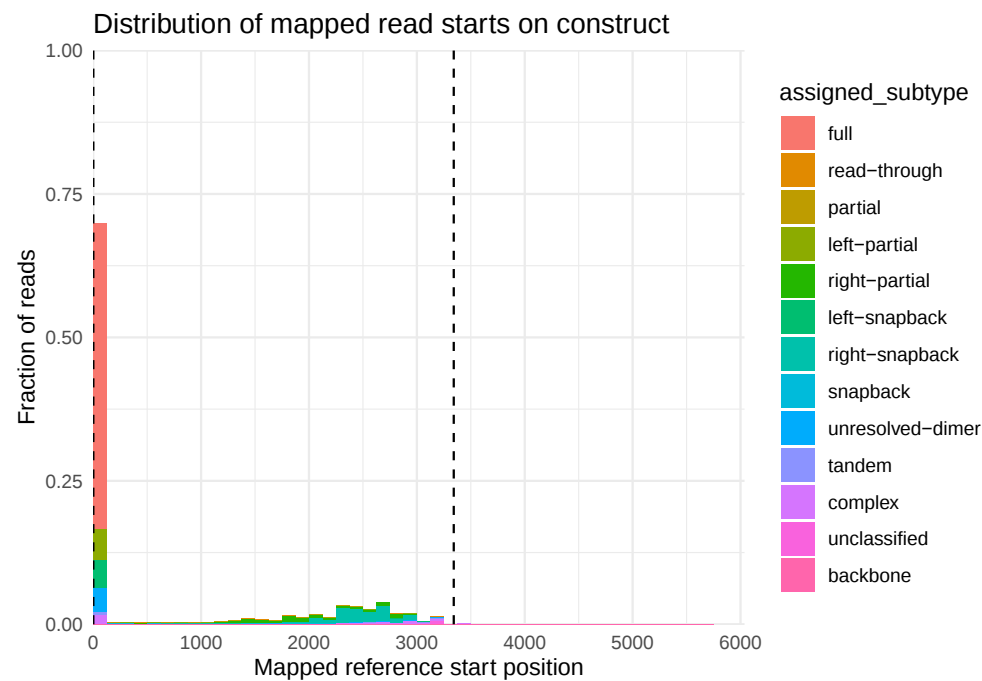
| Assigned Type | Assigned Subtype | Count  | Freq. in AAV (%) | Total Freq. (%) |
|---------------|------------------|--------|------------------|-----------------|
| ssAAV         | full             | 67,531 | 53.17            | 51.53           |
| ssAAV         | right-snapback   | 17,988 | 14.16            | 13.73           |
| ssAAV         | right-partial    | 11,056 | 8.71             | 8.44            |
| ssAAV         | left-partial     | 6,852  | 5.40             | 5.23            |
| ssAAV         | left-snapback    | 6,215  | 4.89             | 4.74            |
| ssAAV         | partial          | 3,091  | 2.43             | 2.36            |
| ssAAV         | read-through     | 157    | 0.12             | 0.12            |
| other-vector  | unresolved-dimer | 5,256  | 4.14             | 4.01            |
| other-vector  | complex          | 5,000  | 3.94             | 3.82            |
| other-vector  | tandem           | 1,568  | 1.23             | 1.20            |
| other-vector  | snapback         | 1,362  | 1.07             | 1.04            |
| other-vector  | unclassified     | 647    | 0.51             | 0.49            |
| backbone      | backbone         | 279    | 0.22             | 0.21            |

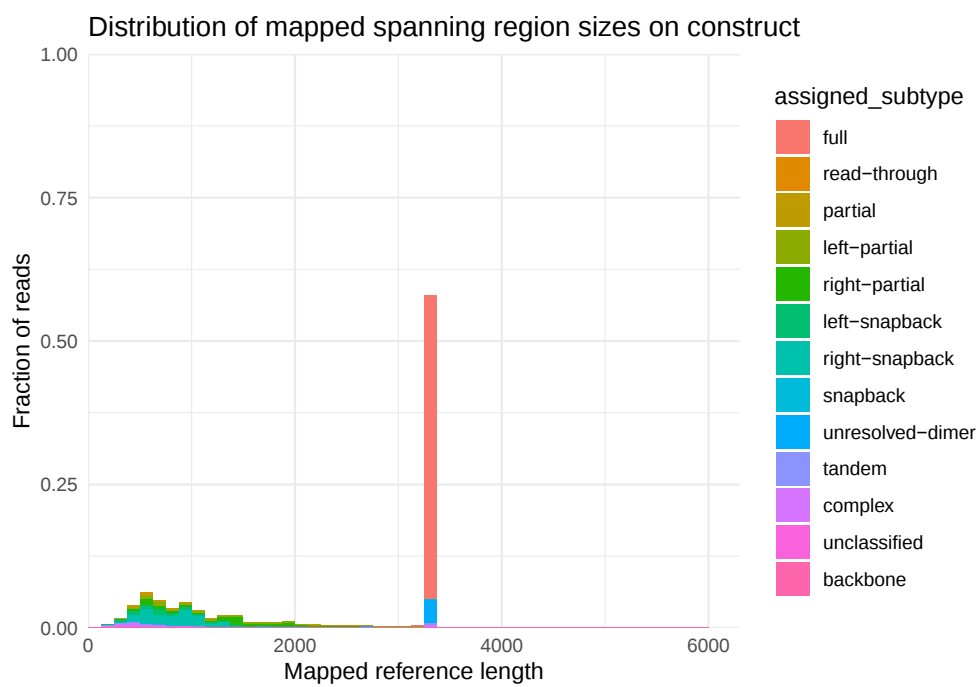
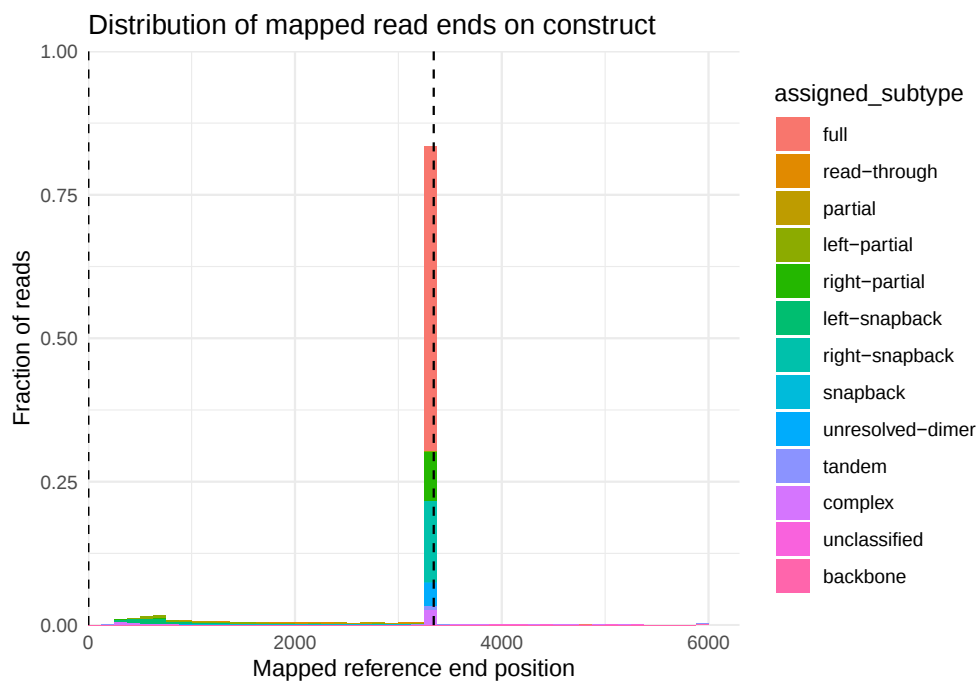
Distribution of read length by subtype



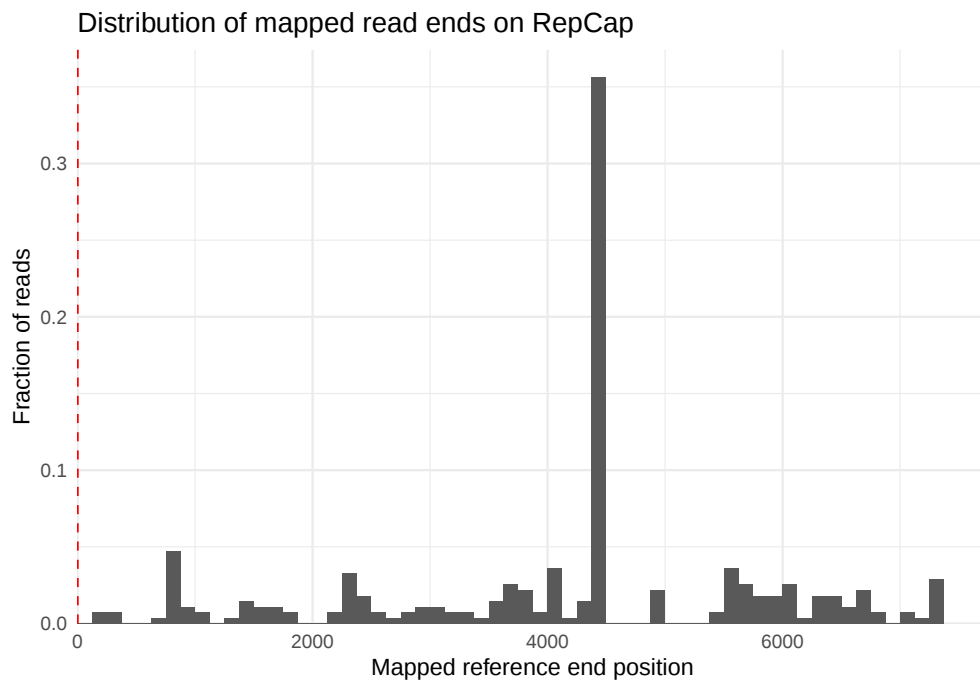
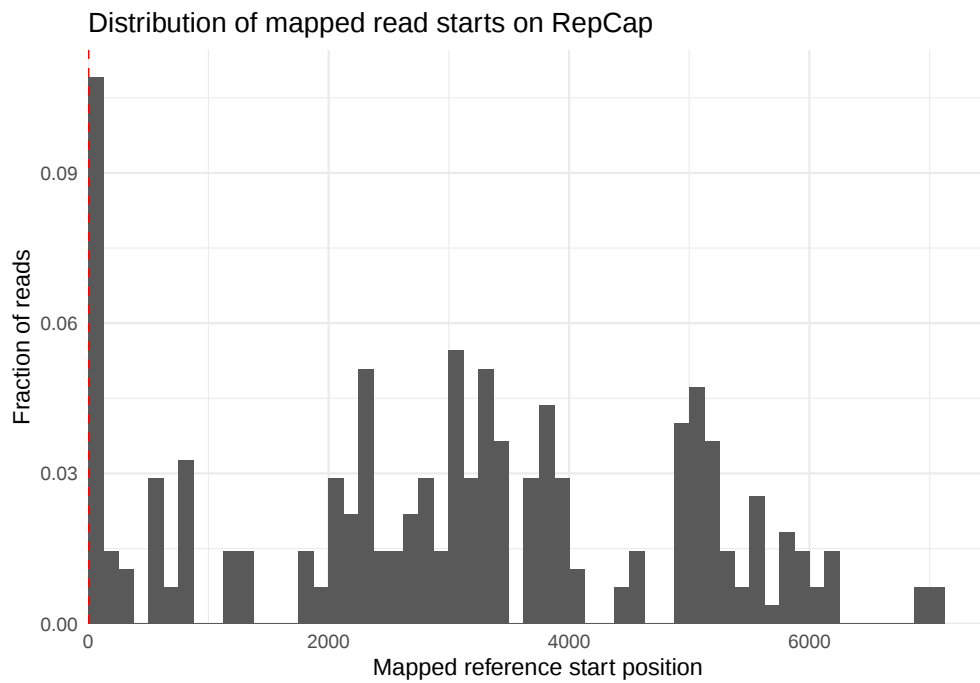
AAV mapping to reference sequence

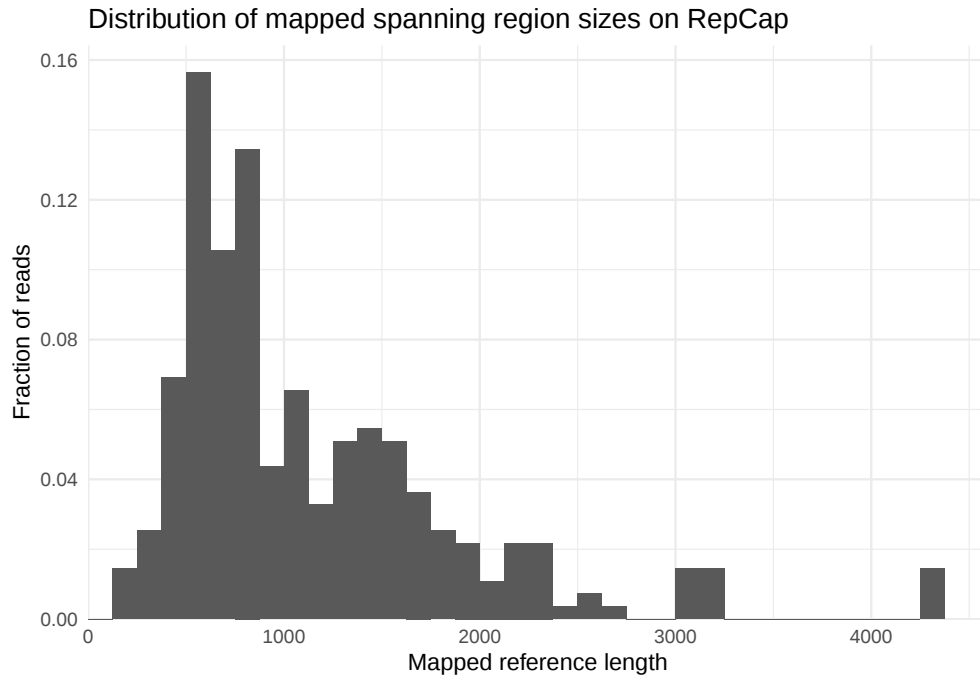
Gene therapy construct



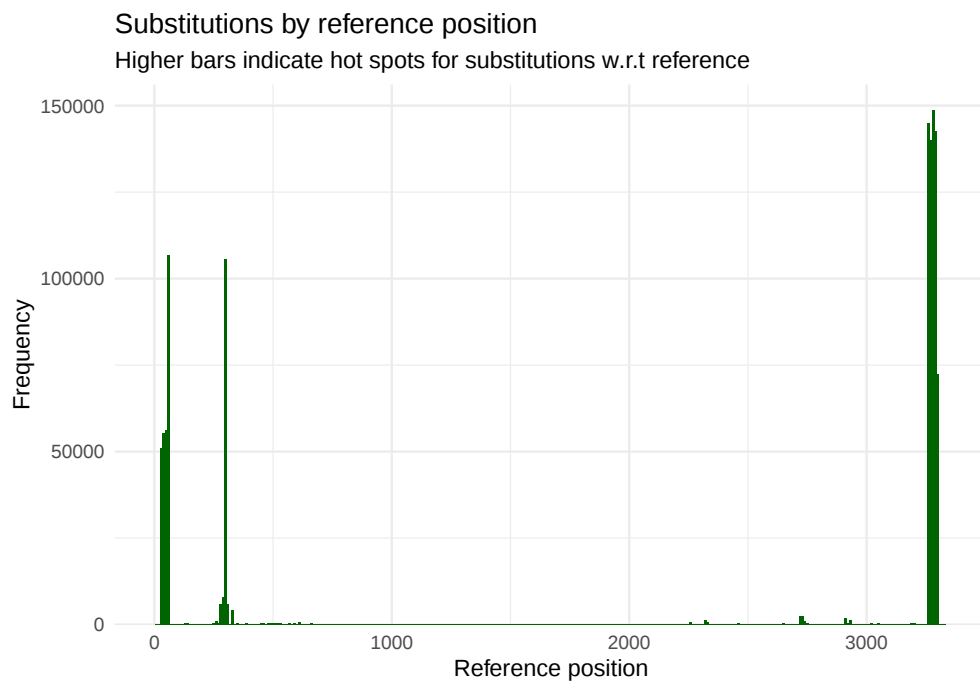


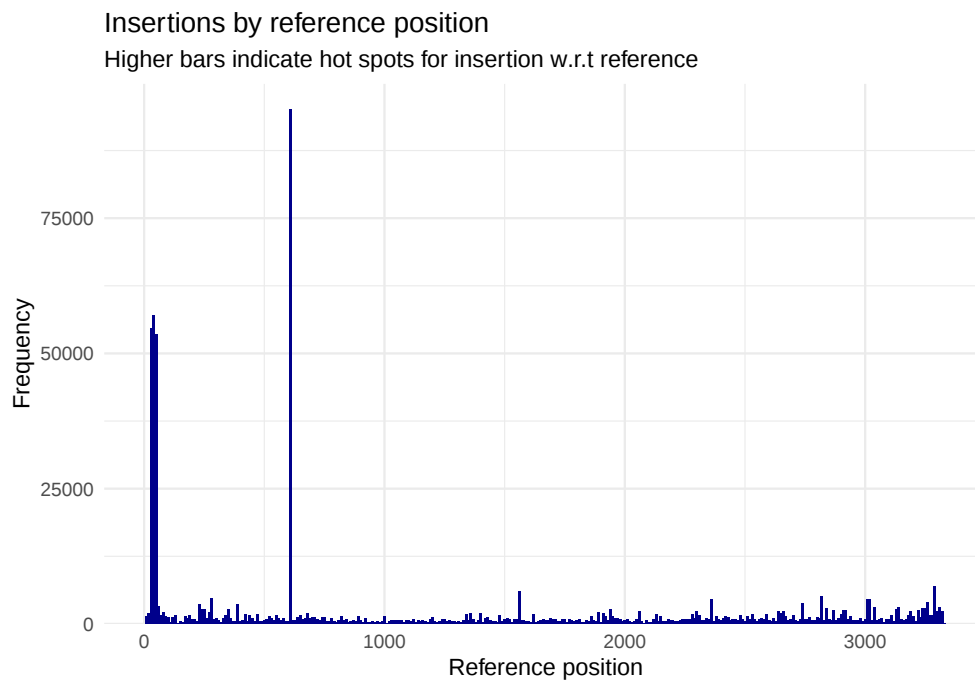
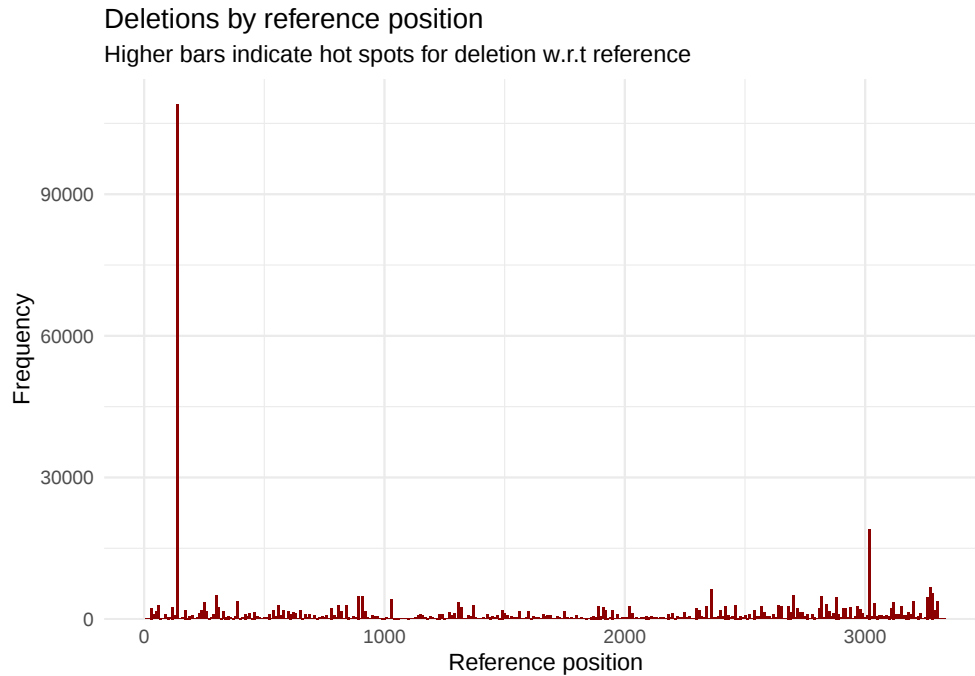
## RepCap





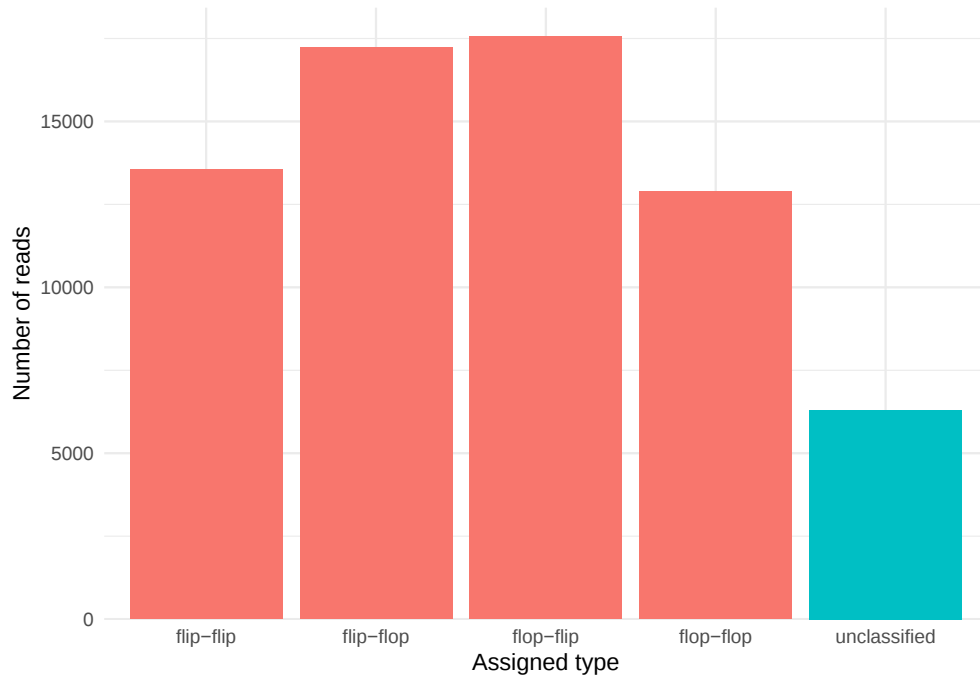
## Distribution of non-matches by reference position





## 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         | 13,562 |
| ssAAV | full          | flip         | flop         | 17,218 |
| ssAAV | full          | flip         | unclassified | 1,334  |
| ssAAV | full          | flop         | flip         | 17,551 |
| ssAAV | full          | flop         | flop         | 12,896 |
| ssAAV | full          | flop         | unclassified | 1,274  |
| ssAAV | full          | unclassified | flip         | 1,624  |
| ssAAV | full          | unclassified | flop         | 1,520  |
| ssAAV | full          | unclassified | unclassified | 552    |
| ssAAV | left-partial  | flip         | unclassified | 8,886  |
| ssAAV | left-partial  | flop         | unclassified | 8,864  |
| ssAAV | left-partial  | unclassified | unclassified | 1,532  |
| ssAAV | right-partial | unclassified | flip         | 21,346 |
| ssAAV | right-partial | unclassified | flop         | 22,357 |
| ssAAV | right-partial | unclassified | unclassified | 3,329  |

## 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).