AAV Sequence Analysis

Sample: m84026_231030_220704_s3

Date: 2024-08-16

Contents

Read type classification

Definitions

Assigned Type	Definition
scAAV	Self-complementary AAV where one half of the payload region is a reverse complement of the other, resulting in an intra-molecular double-stranded DNA template. A sequencing read is inferred as scAAV if it has both a primary and supplementary alignment to the vector genome.
ssAAV	AAV 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 AAV genome and no complementary secondary alignment.
other	Read consists of a fragment mapping to the vector but with unexpected polarities (e.g. +,-,- or +,+,+) and cannot be well-defined at the moment. This means that the algorithm was not able to distinguish the read either as ssAAV or scAAV by the definitions above. Usually this means there are multiple supplementary alignments on the vector region and/or it's a weird molecule.
host	Read originates from the host genome that is given (e.g. hg38, CHM13).
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.
chimeric	Read consists of fragments that map to one or more "genomes" (e.g. vector and host; helper and repcap).

Note: Even though ssAAV distinguishes one ITR as the wildtype (wtITR) and the other as the mutated ITR (mITR), we will still refer to them as "left ITR" and "right ITR". For example, "left-partial" would be equivalent to "mITR-partial" in the case where the mITR is the left ITR based on the given genomic coordinates.
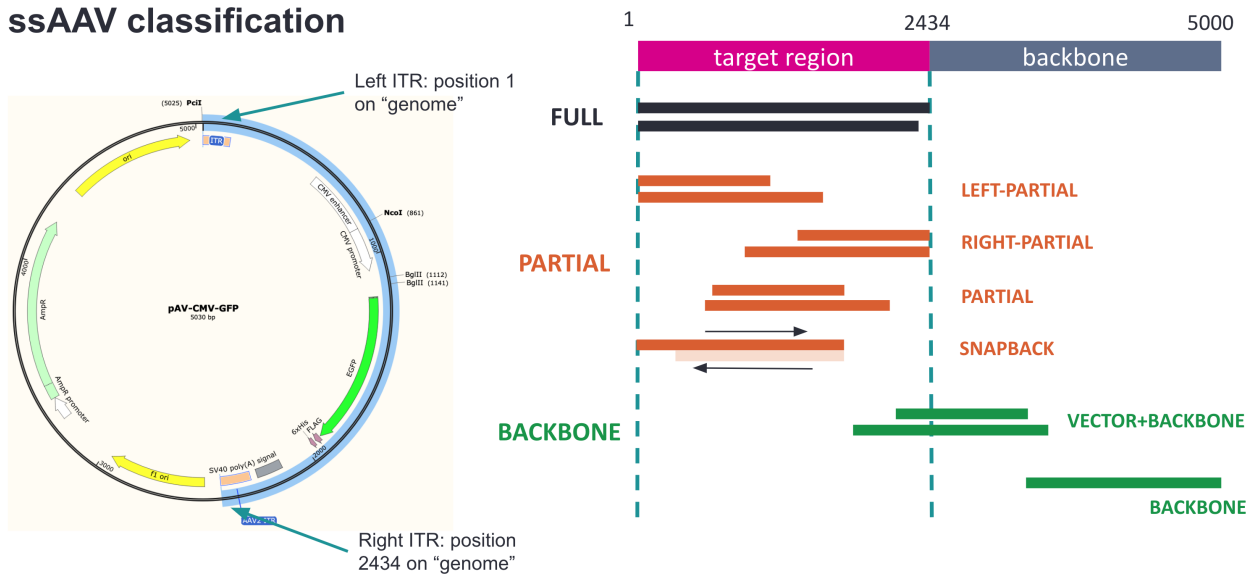
Figure 1: Single-stranded AAV types (ssAAV). ssAAV reads must fully map within the vector genome (which can contain the backbone, beyond the ITR region) consisting of a single (primary) alignment. "ssAAV-full" reads must cover from left ITR to right ITR. "ssAAV-left-partial" contain the left ITR but are missing the right ITR. "scAAV-vector+backbone" map partially within the ITR and partially to the backbone.
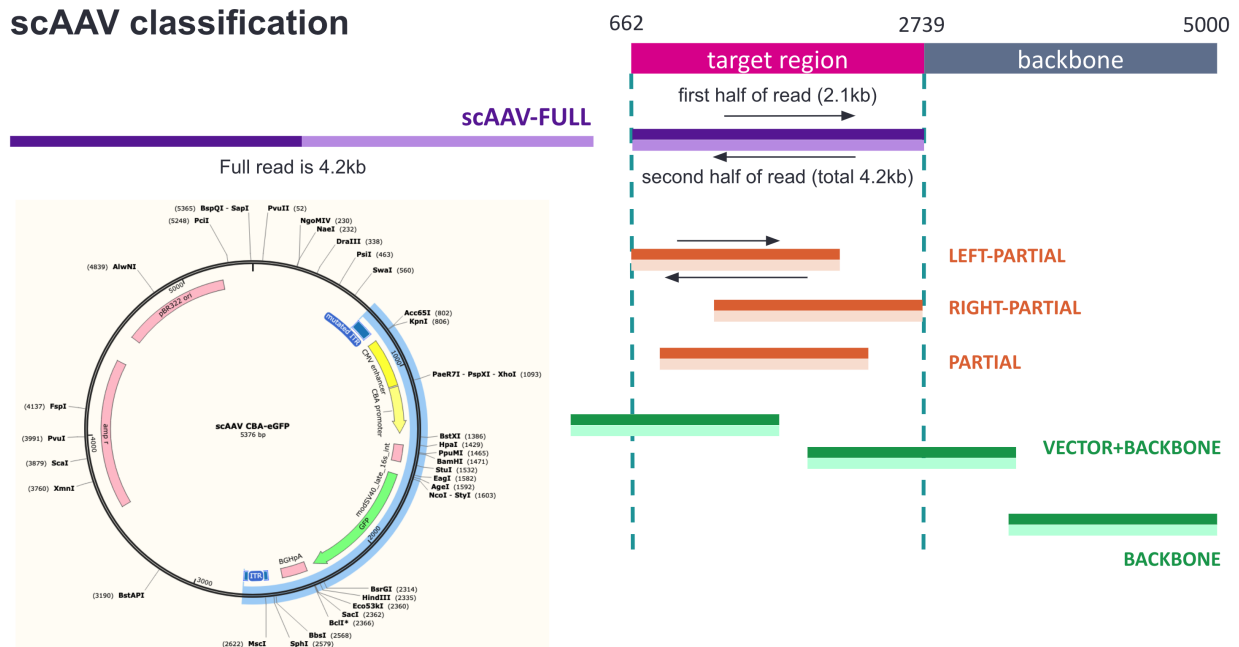

Figure 2: Self-complementary AAV types (scAAV). scAAV reads must fully map within the vector genome (which can contain the backbone, beyond the ITR region) consisting of both primary and supplementary alignments. The definitions of full, partial, and backbone reads are the same as in ssAAV.

3

Assigned types by read alignment characteristics

Assigned Type	Count	Frequency (%)
scAAV	808,813	49.44
ssAAV	703,906	43.02
other	92,562	5.66
chimeric	10,563	0.65
repcap	8,839	0.54
host	6,047	0.37
helper	4,432	0.27
unmapped	907	0.06
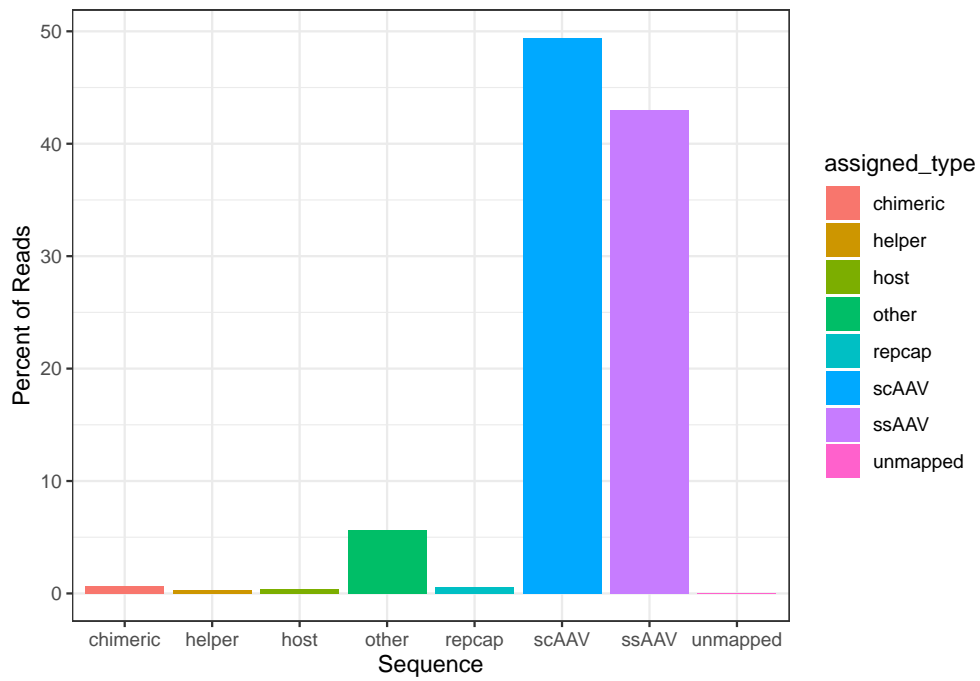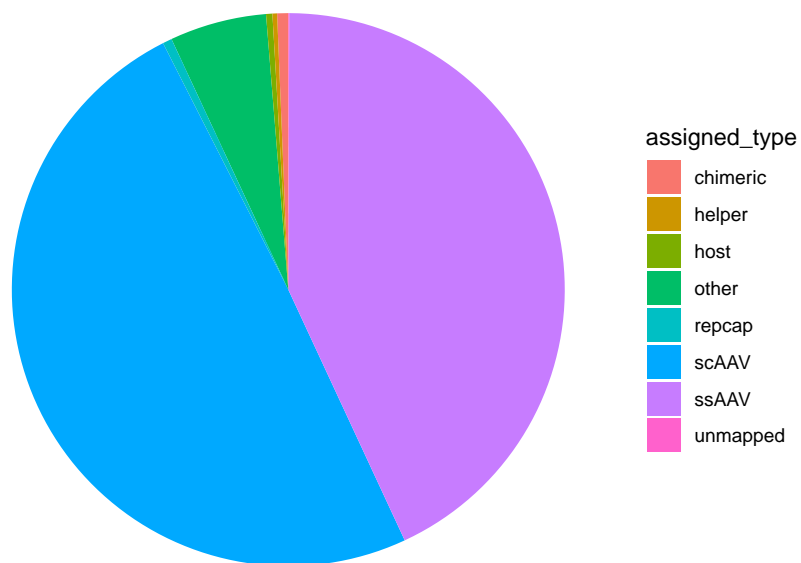
Single-stranded vs self-complementary frequency

Assigned Type	Count	Frequency in AAV (%)	Total Frequency (%)
scAAV	808,813	53.47	49.44
ssAAV	703,906	46.53	43.02

Distribution of read lengths by assigned AAV types

Distribution of read lengths by assigned AAV type
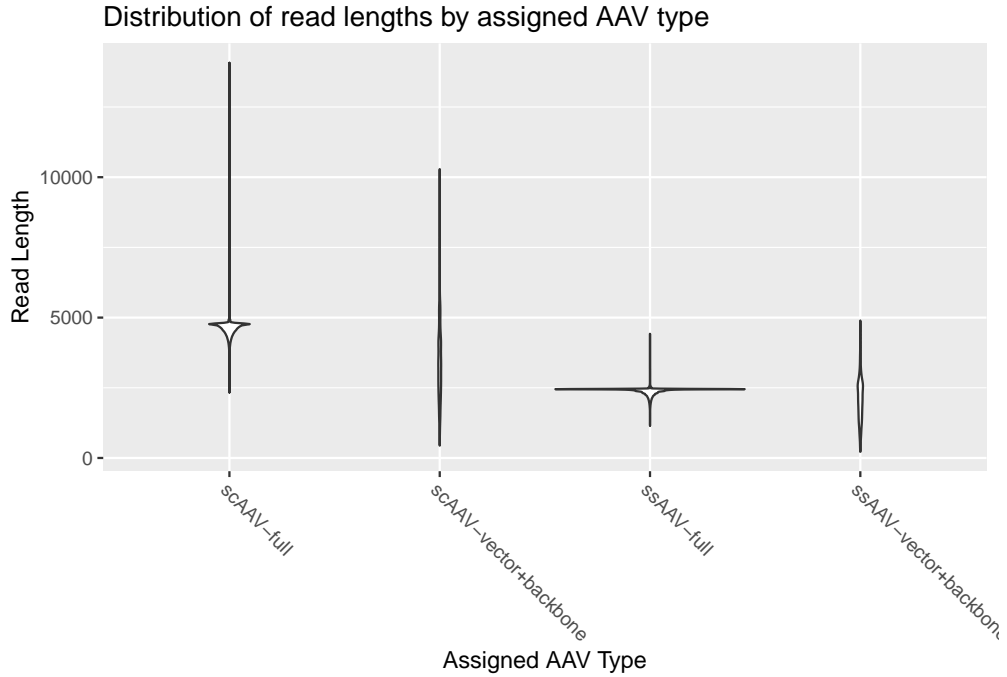
Assigned AAV read types detailed analysis

Assigned AAV types (top 20)

Assigned Type	Assigned Subtype	Count	Freq. in AAV (%)	Total Freq. (%)
ssAAV	full	383,158	25.33	23.42
ssAAV	left-partial	102,627	6.78	6.27
ssAAV	right-partial	183,353	12.12	11.21
ssAAV	partial	32,279	2.13	1.97
ssAAV	backbone	656	0.04	0.04
ssAAV	vector+backbone	1,833	0.12	0.11
scAAV	full	201,523	13.32	12.32
scAAV	left-partial	94,705	6.26	5.79
scAAV	right-partial	262,526	17.35	16.05
scAAV	partial	63,538	4.20	3.88
scAAV	backbone	11,511	0.76	0.70
scAAV	vector+backbone	8,275	0.55	0.51
scAAV	full|right-partial	41,862	2.77	2.56
scAAV	right-partial|partial	32,563	2.15	1.99
scAAV	full|left-partial	25,496	1.69	1.56

Assigned Type	Assigned Subtype	Count	Freq. in AAV (%)	Total Freq. (%)
scAAV	left-partial|partial	14,806	0.98	0.90
scAAV	partial|right-partial	10,721	0.71	0.66
scAAV	left-partial|full	9,954	0.66	0.61
scAAV	right-partial|full	8,484	0.56	0.52
scAAV	left-partial|right-partial	6,755	0.45	0.41

Definitions

Assigned Subtype	Definition
full	Read alignment includes the entire ITR-to-ITR target vector sequence.
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.
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.
partial	Read aligns to a fragment of the vector originating from within the ITR sequences.
vector+backbone	Read aligns to a fragment including the vector as well as plasmid backbone sequence.
backbone	Read aligns to a fragment originating solely from the plasmid backbone sequence.
snapback	Read consists of a double-stranded, sub-genomic fragment including only one ITR and read alignments in both (+) and (-) polarities. (ssAAV only)

Flip/flop considerations

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)

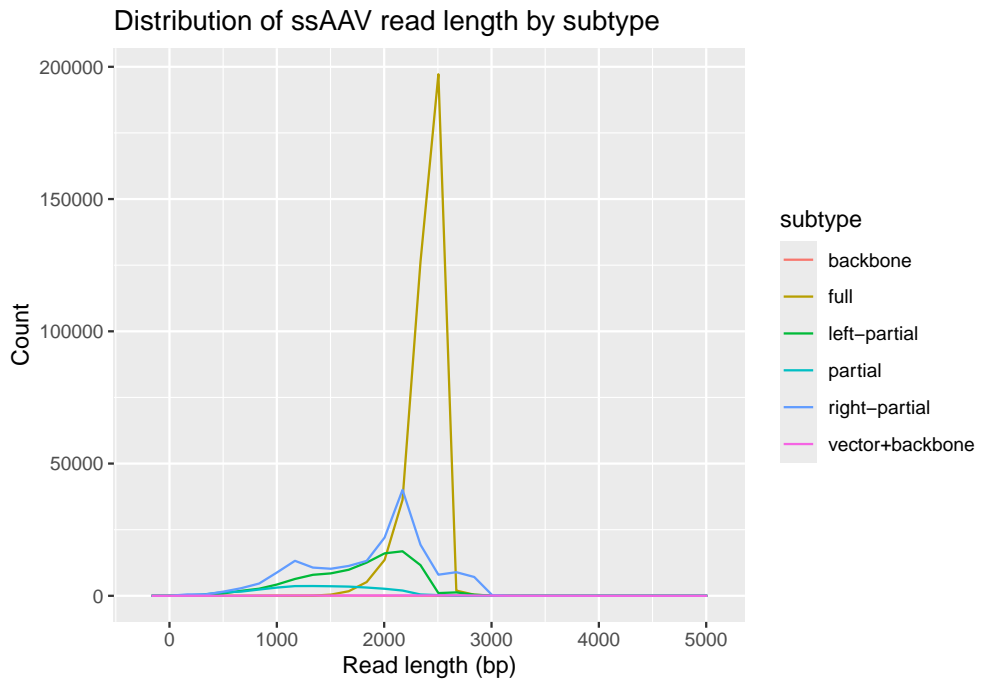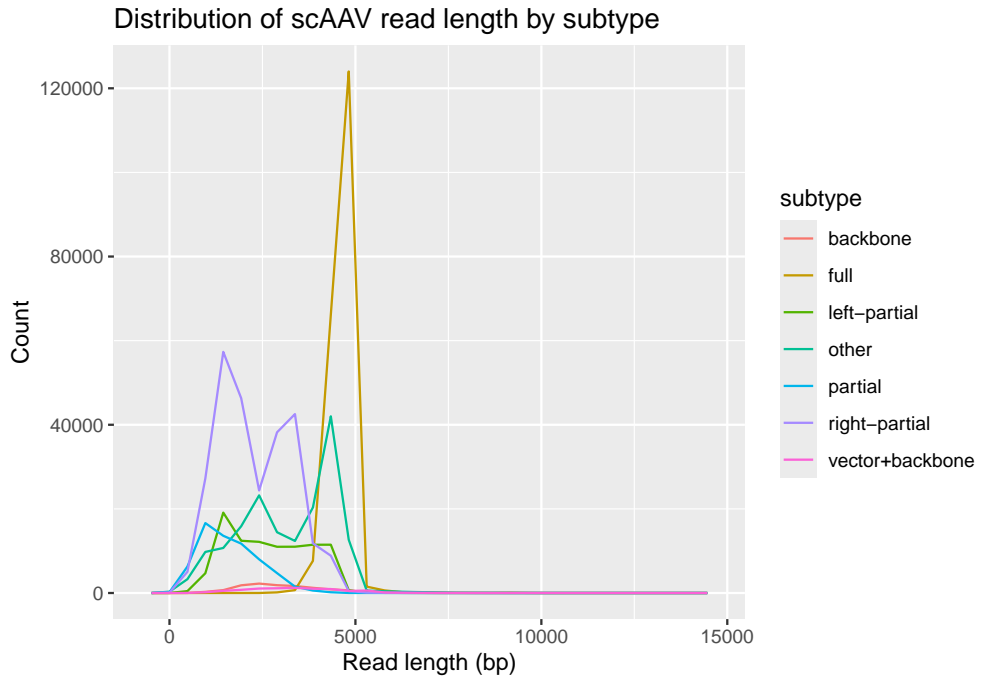No flip/flop analysis results available to display.

Flip/flop configurations, scAAV only

No scAAV flip/flop analysis results available to display.

Flip/flop configurations, ssAAV only

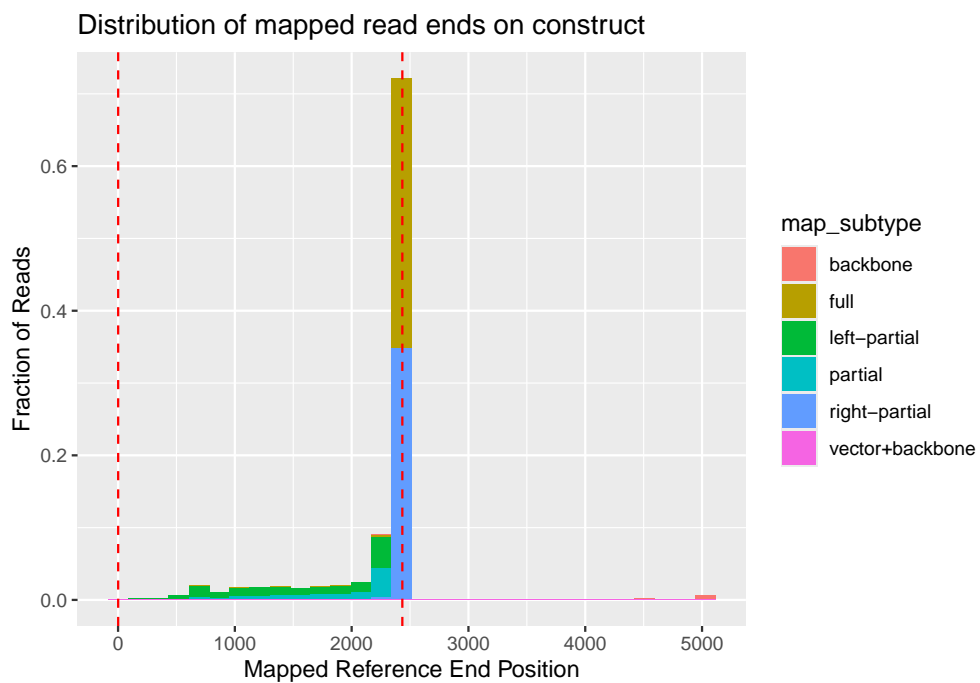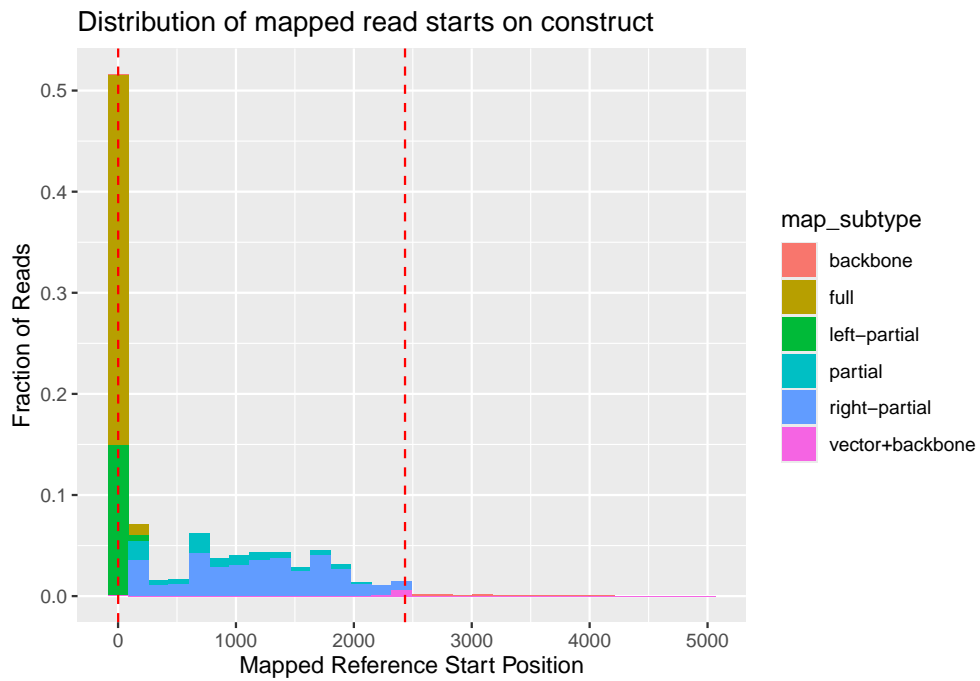No ssAAV flip/flop analysis results available to display.

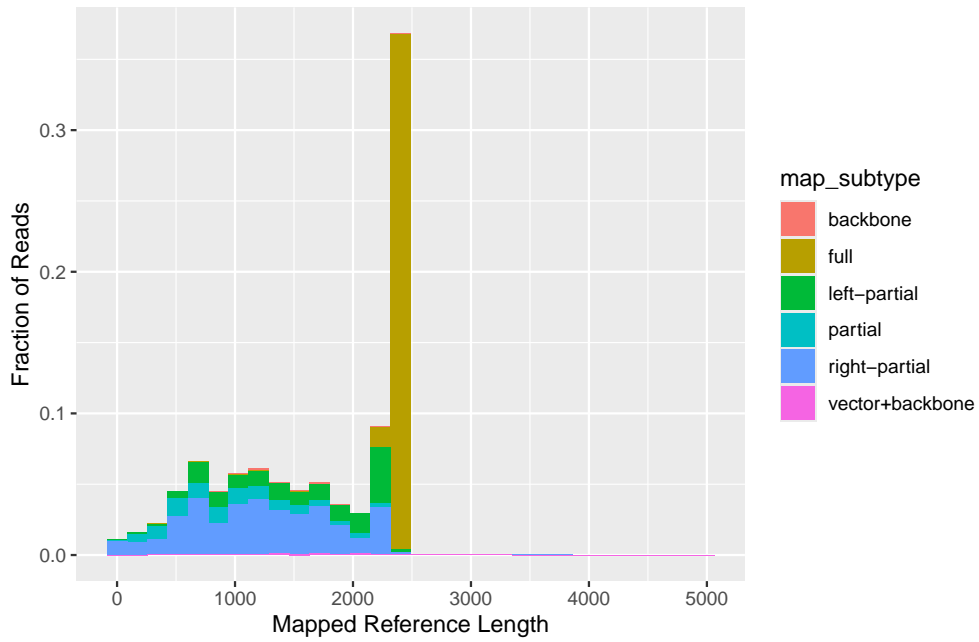# Distribution of read length by subtype

AAV mapping to reference sequence

Gene therapy construct

Distribution of mapped read starts on construct

### Distribution of mapped read ends on construct

Distribution of mapped spanning region sizes on construct
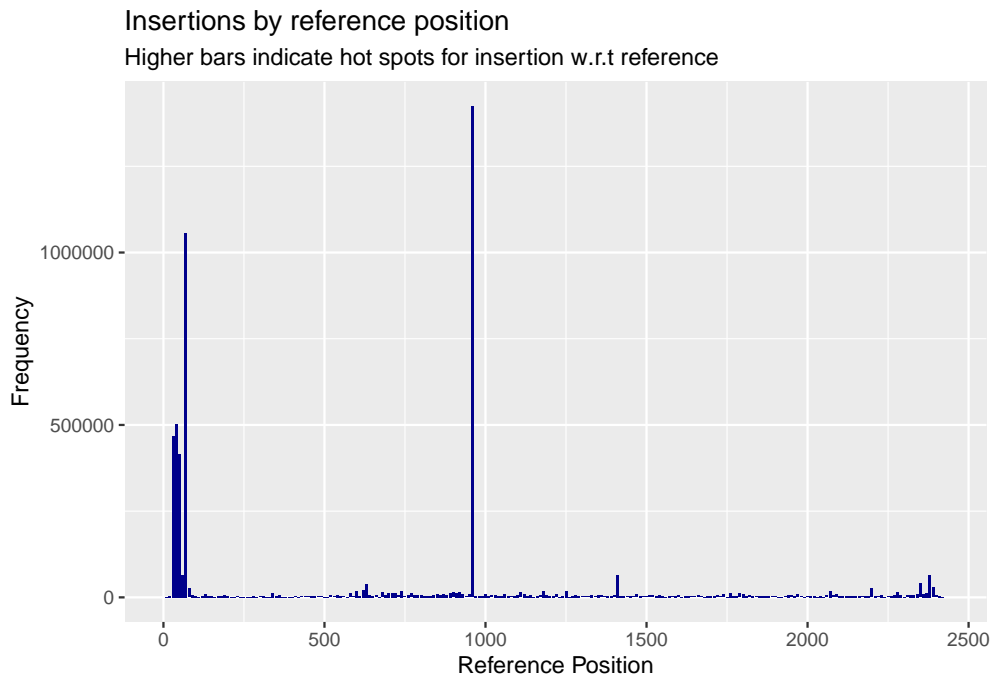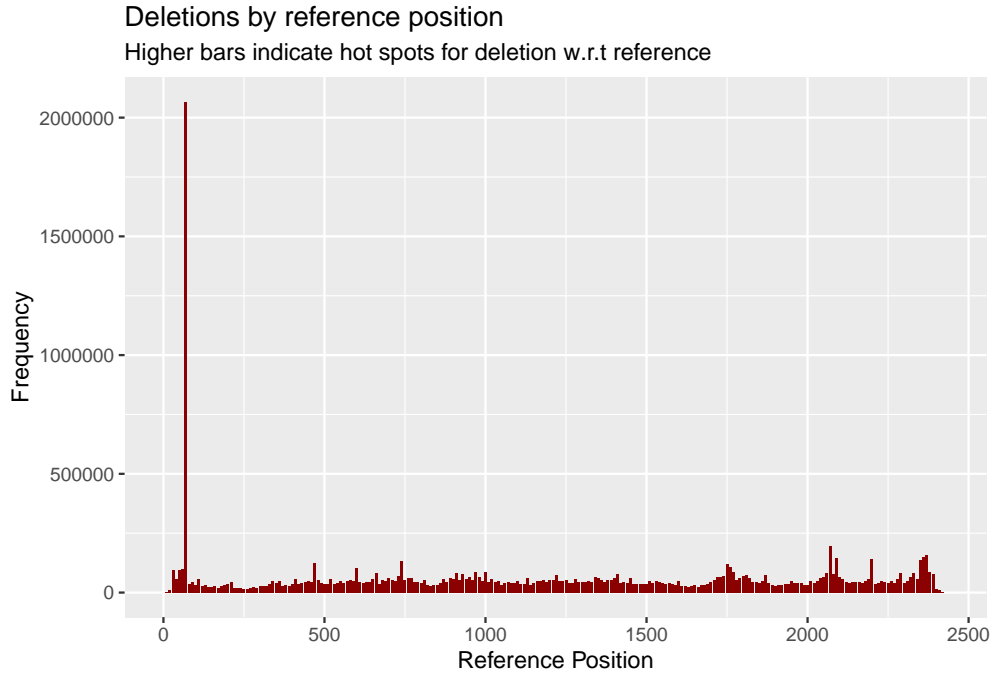

Distribution of non-matches by reference position

Substitutions by reference position

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

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
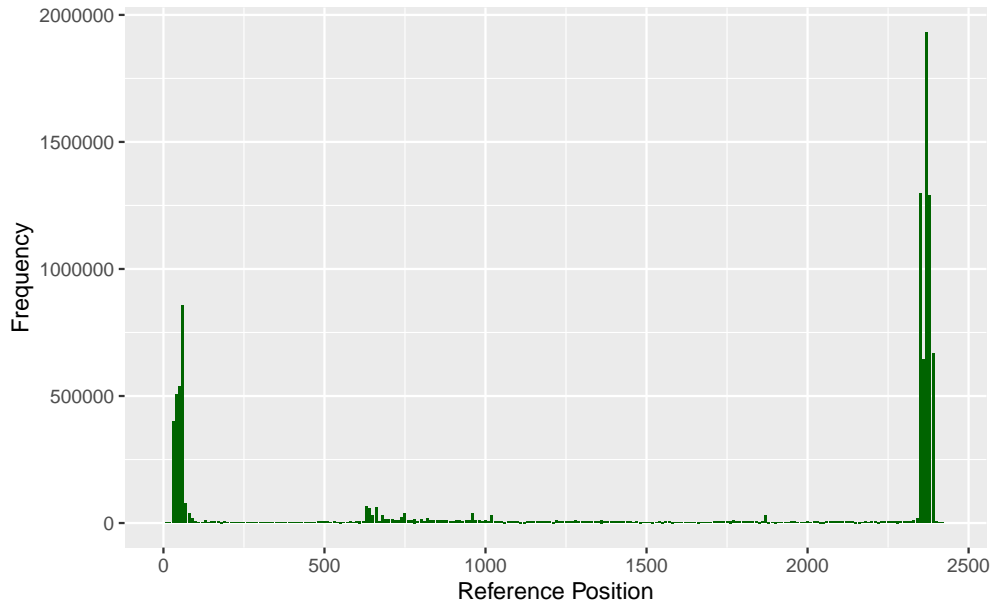
Methods

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). Reads are aligned to the AAV, packaging, and masked host reference sequences using Minimap2 (Li, 2018).

In this analysis, aligned sequencing reads were filtered for quality to include primary alignments and reads with mapping quality scores greater than 10. The alignment coordinates and orientation of reads passing

these filters were then 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 type and (for AAV reads) subtype classification according to the definitions above.

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).