# Rapid analysis of gene therapy vectors with Nanopore MinION sequencer

Suk Namkung, Ngoc Tam Tran, Alexander Brown, Dan Wang, Jun Xie, Guangping Gao[a] and Phillip W.L. Tai[a]
Horae Gene Therapy Center, University of Massachusetts Medical School, Worcester, MA
[a]Co-corresponding authors

## ABSTRACT

Adeno-associated virus (AAV) is a non-enveloped single-stranded DNA virus with a genome size of ~4.7 kb. The virus is primarily studied as a vector for gene therapy because of its non-pathogenic nature and ability to deliver engineered transgene cassettes as recombinant (r)AAV. Currently, two AAV-based therapies have been made commercially available with over 200 different rAAV-based gene therapy clinical trials ongoing. Despite their promise, quality control methods for rAAV vectors fall short in detecting contaminants and truncated genomes. These unwanted materials, which include viral and human DNA fragments related to the production process, can impact the efficacy and safety of these therapies.

In order to develop a means to analyze full-length rAAV genomes, we evaluated the capacity of long-read sequencing by Oxford Nanopore Technology's MinION platform to sequence two clinically relevant forms of rAAV genomes: the single-stranded (ss) and the self-complementary (sc)AAV vector genomes, as well as the parental plasmids used to produce these vectors. For this study, we used both MinION R9 and R10 flow cells to: 1) evaluate differences in sequencing ssAAV and scAAV genomes, 2) validate whether reads accurately reflect the vector genome population in preparations, and 3) assess the accuracy of nanopore sequencing across the inverted terminal repeats (ITRs), which are characterized by T-shaped hairpins that interfere with the processivity of standard polymerases used in sequencing.

Our results show that both R9 and R10 can identify and quantify DNA contaminants and heterogeneous vector populations; however, R9 showed higher sequencing efficiency and coverage of the ITRs. Sequencing of the parental plasmids validated that abnormal vector genomes are not an artifact of the sequencing platform. Additionally, through MinION's ability to sequence through the ITRs, we were able to reveal replication intermediates during vector production, which showed that genomes are predominantly amplified by conventional rolling-hairpin replication(RHR) and not rolling-circle replication (RCR).

## METHODS

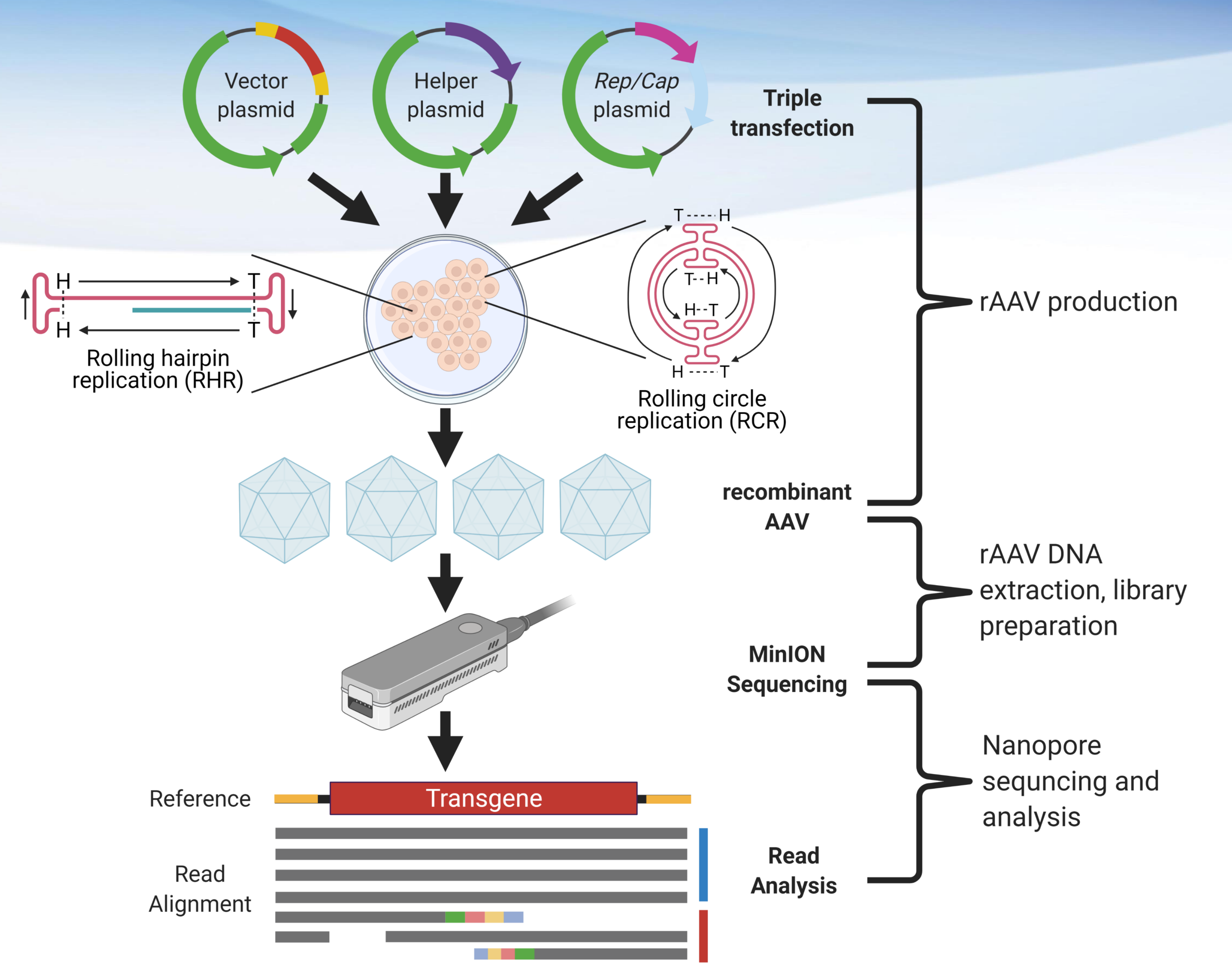
**Vector library preparation and nanopore sequencing**

Viral vector DNAs were extracted by phenol-chloroform extraction and ethanol precipitation, following methods described previously. The vector transgene plasmid control library was prepared by restriction digests. Samples were subject to library preparations for Oxford Nanopore Technologies (ONT) sequencing, following the protocol for 1D Native barcoding of genomic DNA, using the Native Barcoding Expansion 1-12 (PCR-free) (EXP-NBD103) and Ligation Sequencing Kit (SQK-LSK108) components. Libraries were then multiplexed and sequenced on an ONT MinION instrument (Flow Cell R9.4.1, FLO-MIN106D).

**Sequencing data analyses**

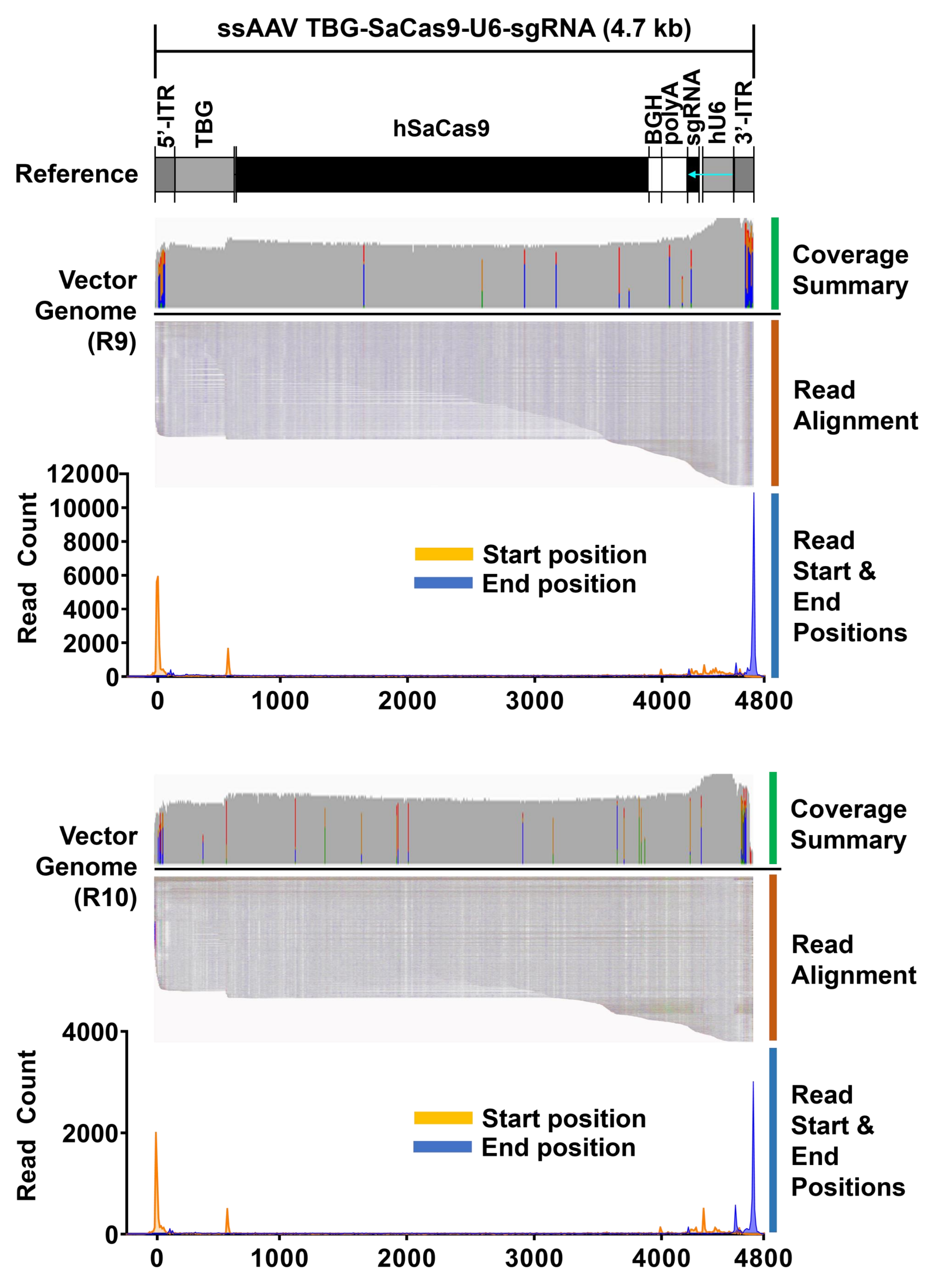
MinKNOW software was used for base calling, adapter trimming, and de-multiplexing. Sequencing reads were aligned to the appropriate reference sequences by BWA-MEM within the Galaxy web-based interface. Alignments were visualized using Integrative Genome Viewer (IGV, version 2.6.3) with soft-clipping displayed. Start and end positions of aligned reads were defined with BEDTools (version 2.27.0). The abundances of start and end positions of all mapped reads were tabulated and plotted using GraphPad Prism (version 8.2.1 for Windows).

## rAAV NANOPORE SEQUENCING WORKFLOW

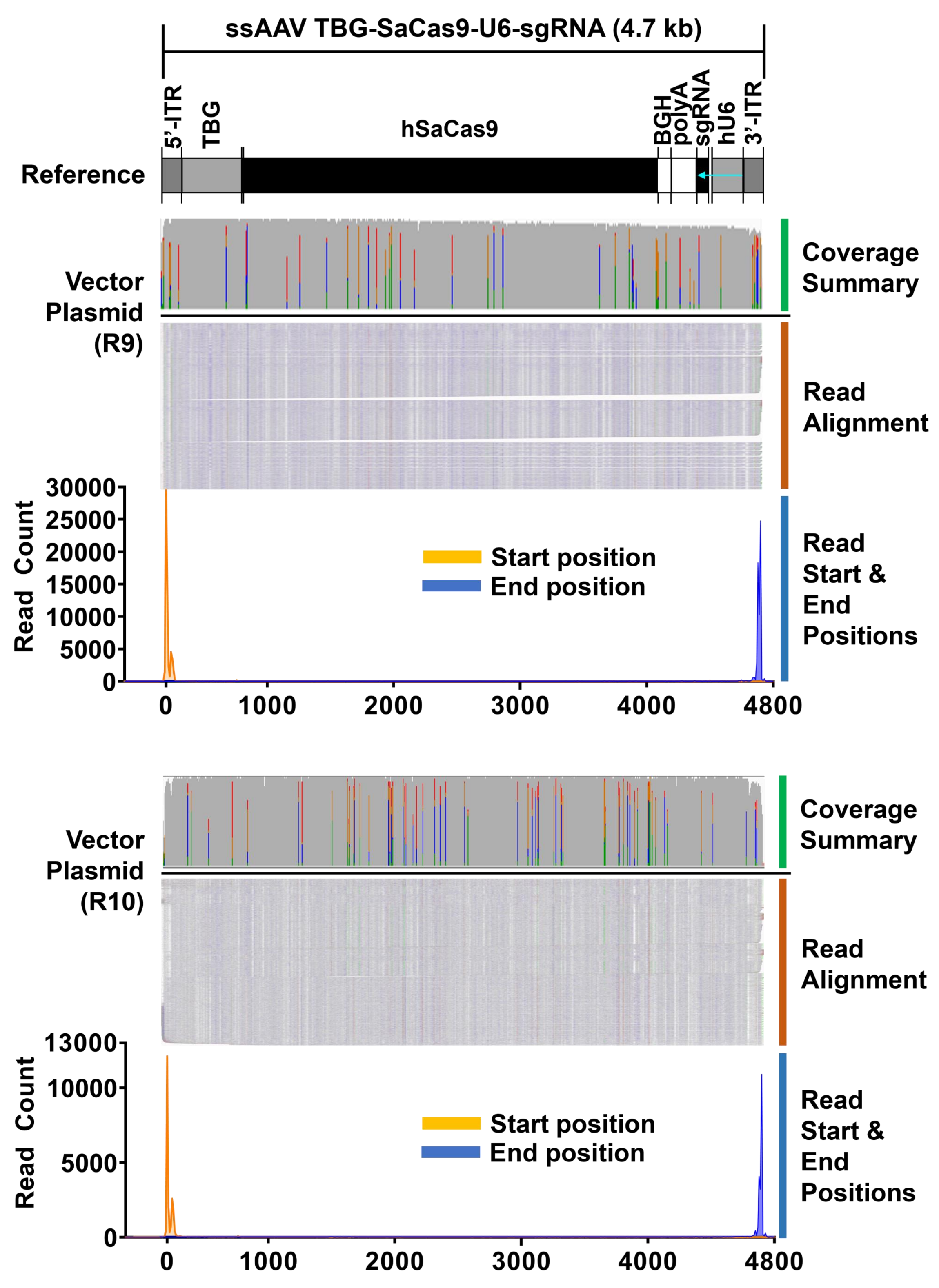


## RESULTS

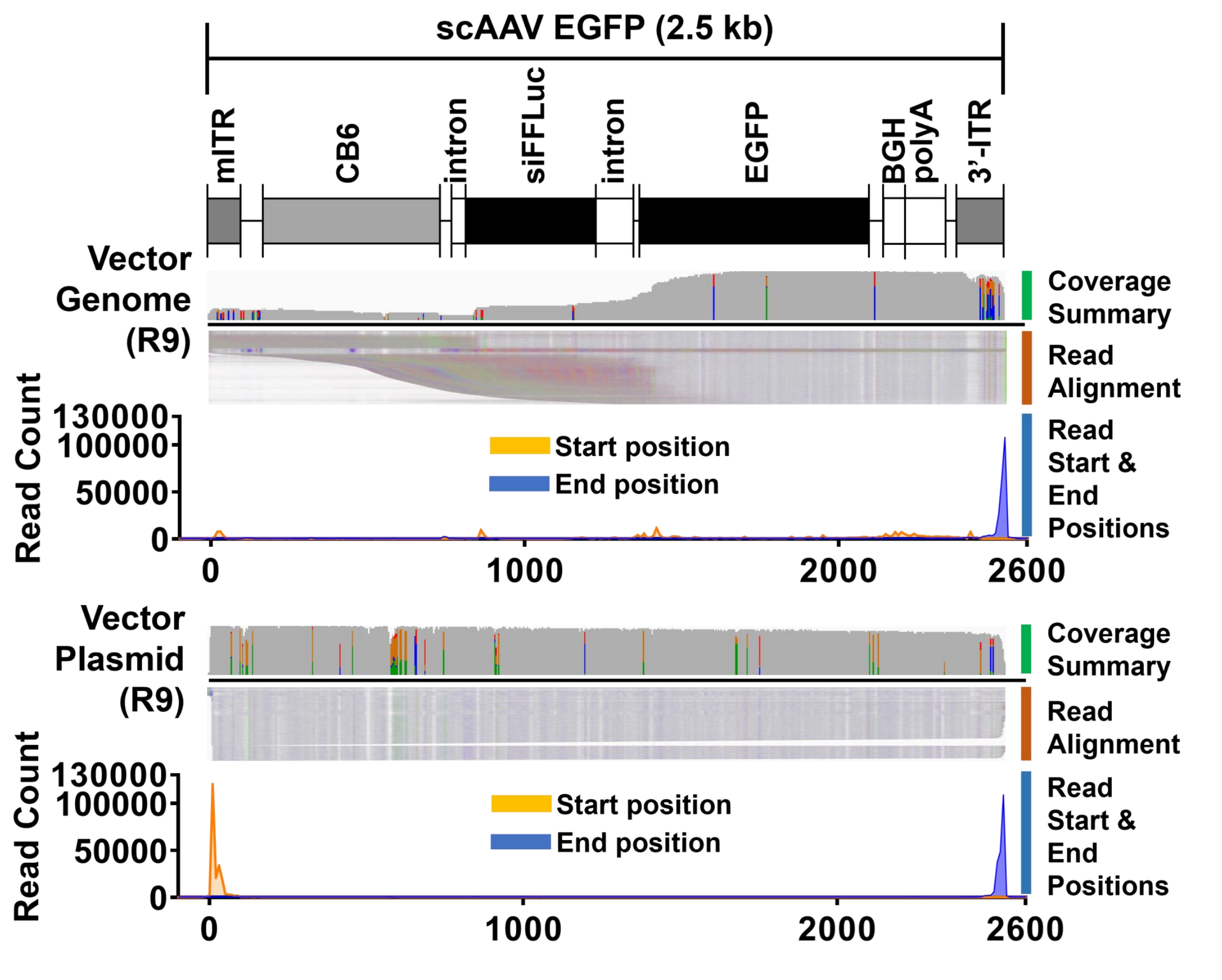
### Fig. 1a. Sequence analysis of ssAAV-SaCas9 vector genome in R9 and R10 flow cells



### Fig. 1b. Sequence analysis of ssAAV-SaCas9 vector plasmid using R9 andR10 flow cells



### Fig. 1c. Sequence analysis of scAAV-EGFP vector and vector plasmid using R9 flow cell



### Fig. 2. Vector genome with HH/TT/HT/TH conformation

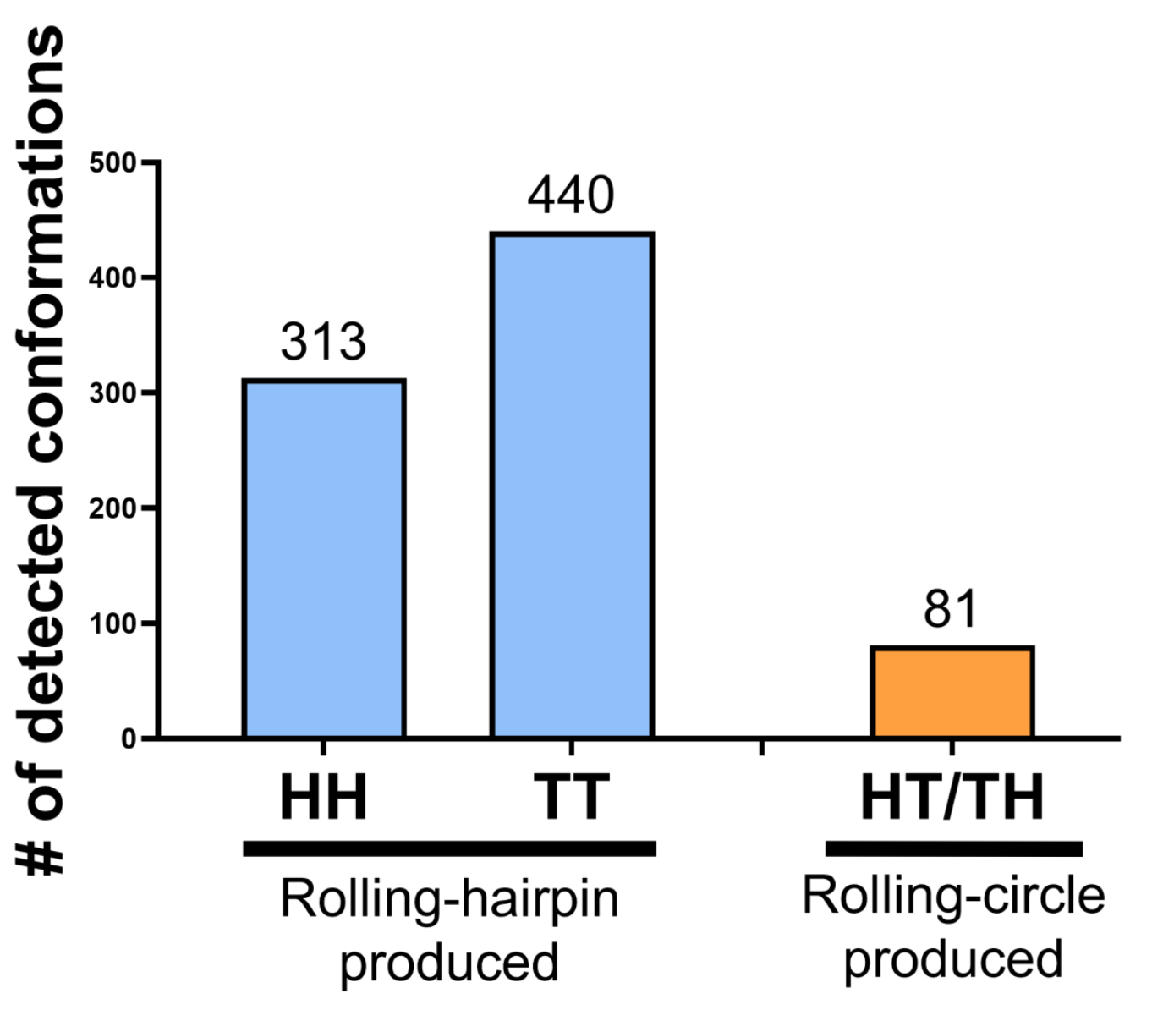


**Figure 1. Vector genomes and plasmid DNA sequenced by R9 and R10 flow cells.**
Alignments of nanopore sequencing reads representing DNAs extracted from **1a.** ssAAV-*SaCas9* vector and sequenced with R9 and R10 flow cells. **1b.** ssAAV-*SaCas9* vector plasmid sequenced with R9 and R10 flow cells. **1c**. scAAV-eGFP vector genome and vector plasmid sequenced with R9 flow cell. Diagrams of references displaying notable construct domains are shown above each alignment. Regions upstream and downstream of the vector sequences (beyond ITRs) represent the bacterial plasmid backbone. The alignment summary shows read coverage. Read alignments displayed in IGV show matching nucleotides in gray. Nucleotide mismatches are depicted in green (A), red (T), blue (C), or brown (G); gaps are in black dashes; and inserts are in purple. Alignment start and end position counts (orange and blue traces, respectively) are displayed below each alignment. The vector plasmid was subjected to restriction digestion to isolate the entire vector genome from 5' ITR to 3' ITR and sequenced. Sequencing results show very high percentage of full ITR-to-ITR reads and demonstrates the ability of ONT's platform to sequence through ITRs.

**Figure 2. RCR and RHR comparison.**
From the sequencing data, the abundance of reads with H-H/T-T (representing RHR) and H-T/T-H (representing RCR) junctions are shown. The higher abundance of H-H/T-T oriented junctions show that replication during rAAV production is predominantly through RHR.

## CONCLUSIONS

- ONT's long-read sequencing platform allows for direct sequencing and visualization of full-length rAAV genomes.
- Direct sequencing of full-length AAV genomes enables visualization of vector heterogeneity. Regions with truncation hotspots can be visualized at single-molecule resolution
- Sequencing of the rAAV genome and its parental plasmid confirms that the truncation events detected by nanopore sequencing are a true reflection of what is packaged into virions
- Sequencing of the parental plasmid confirms that nanopore sequencing is processive through the secondary structure of the ITRs
- Observation of unresolved ITR junctions represented in vector reads show that they are predominantly in head-to-head and tail-to-tail orientations. This finding implies that rAAV genomes are predominantly amplified and packaged by rolling-hairpin replication.

## AUTHOR DISCLOSURES

G.G. is a scientific co-founder of Voyager Therapeutics and Aspa Therapeutics, and holds equity in these companies. G.G. is an inventor on patents with potential royalties licensed to Voyager Therapeutics, Aspa Therapeutics, and other biopharmaceutical companies. The remaining authors declare no competing interests

## REFERENCES & ACKNOWLEDGMENT

1. Tran NT, Heiner C, Weber K, Weiand M, Wilmot D, Xie J, Wang D, Brown A, Manokaran S, Su Q, Zapp ML, Gao G, Tai PWL. AAV-Genome Population Sequencing of Vectors Packaging CRISPR Components Reveals Design-Influenced Heterogeneity. Mol Ther Methods Clin Dev. 2020;18:639-51. Epub 2020/08/11. doi: 10.1016/j.omtm.2020.07.007. PubMed PMID: 32775498; PMCID: PMC7397707.
2. Tai PWL, Xie J, Fong K, Seetin M, Heiner C, Su Q, Weiand M, Wilmot D, Zapp ML, Gao G. Adeno-associated Virus Genome Population Sequencing Achieves Full Vector Genome Resolution and Reveals Human-Vector Chimeras. Molecular therapy Methods & clinical development. 2018;9:130-41. Epub 2018/05/17. doi: 10.1016/j.omtm.2018.02.002. PubMed PMID: 29766023; PMCID: PMC5948225.
3. Xie J, Mao Q, Tai PWL, He R, Ai J, Su Q, Zhu Y, Ma H, Li J, Gong S, Wang D, Gao Z, Li M, Zhong L, Zhou H, Gao G. Short DNA Hairpins Compromise Recombinant Adeno-Associated Virus Genome Homogeneity. Molecular therapy : the journal of the American Society of Gene Therapy. 2017;25(6):1363-74. Epub 2017/05/04. doi: 10.1016/j.ymthe.2017.03.028. PubMed PMID: 28462820; PMCID: PMC5474962.
4. rAAV Nanopore Sequencing Workflow figure created with BioRender.com