Nanopore sequencing of native adeno-associated virus vectors for quality control

Adeno-associated virus (AAV) is a non-enveloped single-stranded DNA virus used in gene therapy. Accurate validation, contamination detection, and quality control (QC) of recombinant AAV (rAAV) vectors is crucial to ensure the correct rAAV genomes are packaged into cells before therapeutic use, to confirm the safety and efficacy of the therapy. However, using traditional short-read sequencing technology for QC can present limitations: features such as inverted terminal repeats (ITRs) in AAV genomes are especially hard to map due to their high GC content, palindromic nature, and complex secondary structure. Moreover, short-read sequencing has limitations in accurately characterising genome truncations and heterogeneous vector populations, which can impact functionality and efficacy.

In contrast, long nanopore sequencing reads can be used to sequence full-length, native rAAV genomes — both single-stranded and self-complementary AAV vectors — end to end, to aid their QC. This allows for ITRs to be fully characterised, enabling identification of any truncated rAAV genomes, contamination, or mutations. Transgenes and promoters of interest can also be identified to support validation of rAAV vectors.

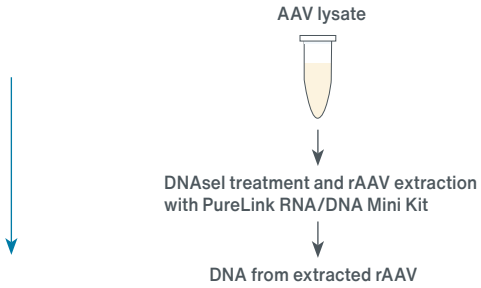
Prior to rAAV sequencing, whole-plasmid sequencing should be performed for a complete QC across the whole AAV production process.

Here we present a workflow to sequence and characterise full-length native rAAV vectors using MinION™ Flow Cells on MinION or GridION™ sequencing devices

EXTRACTION:
obtaining high molecular-weight DNA

For full-length sequencing of AAV genomes in single reads, it is important to select an extraction method that preserves the ~4.7 kb DNA fragments. We recommend using the **PureLink Viral RNA/DNA Mini Kit**, which we have seen to produce more full-length AAV genomes and ITR sequences compared to other methods, such as proteinase K and heat-based extractions. Prior to extraction, we strongly recommend treating your samples with DNAseI to remove any unencapsidated DNA contamination.

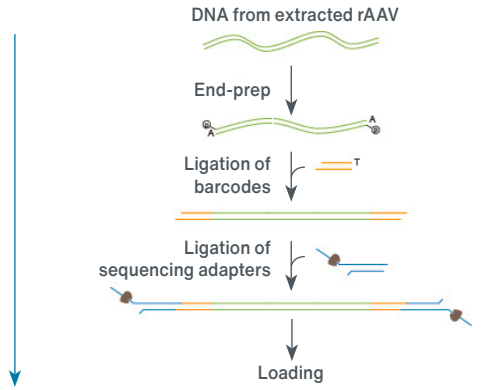
To maximise the number of full-length ITR sequences, we recommend omitting the self-annealing step after DNA extraction. We also recommend quantifying your samples using the **ssDNA** and **dsDNA HS Qubit Assay Kit** and **Qubit fluorometer** before proceeding to the library preparation.



LIBRARY PREPARATION:
sample multiplexing

To prepare your samples for sequencing in multiplex, use the Oxford Nanopore **Native Barcoding Kit 24**. This PCR-free library preparation method enables sequencing of native DNA from a low input of starting material.

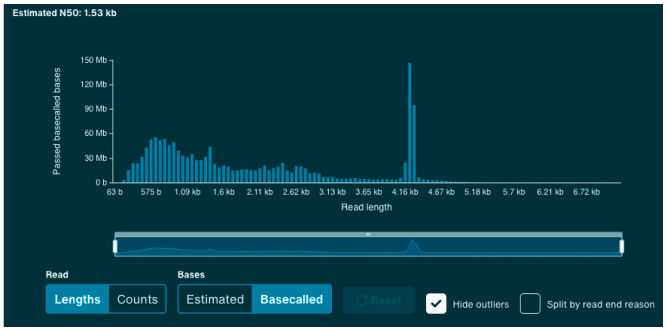
Through multiplexed sequencing of samples on a single MinION Flow Cell, the cost per sample can be reduced



SEQUENCING: running until the necessary coverage is achieved

We recommend sequencing your AAV library on a MinION Flow Cell. This can be run on a portable **MinION** device for easily accessible, routine sequencing. For higher throughput needs, the **GridION** device can be used, enabling on-demand sequencing on up to five individually addressable flow cells at a time.

Typically for six samples, sufficient data is generated after approximately seven hours of sequencing to input into data analysis workflow. To detect any shorter reads that are often indicative of contamination, such as ITR tetramers, minimum read length can be reduced to 20 bp, our sequencing software. High-accuracy (HAC) or super accurate (SUP) basecalling models can be used to produce high-quality consensus genome sequences.



ANALYSIS: using the AAV QC workflow

The analysis workflow is available to map the rAAV vector genomes for quick assessment for QC and validation. The workflow provides an intuitive user interface and incorporates a number of tools for the easy assessment of rAAV vectors. A FASTQ file is required as input into the workflow from the sequencing run, along with sequence reference files for the transgene plasmid, host cell line genome, repcap plasmid, and helper plasmid. Integrated tools used include minimap2[1] to map reads to the combined reference and Medaka[2] to generate a consensus sequence for the four different orientations of the ITR-ITR transgene cassette.

Plots produced by the workflow include a contamination graph to show read sources distribution, a truncations graph, transgene expression read coverage, and genome type frequency graph.

FASTQ input file

Read alignment and consensus generation

BAM, TSV output files and HTML report.

Add-on custom analysis is available upon request

DIVERGENE

Support@divergene.com
www.Divergene.com

References:

1. GitHub. Minimap2. Available at: https://github.com/lh3/minimap2 [Accessed: 14 November 2023]
2. GitHub. Medaka. Available at: https://github.com/nanoporetech/medaka [Accessed: 14 November 2023]