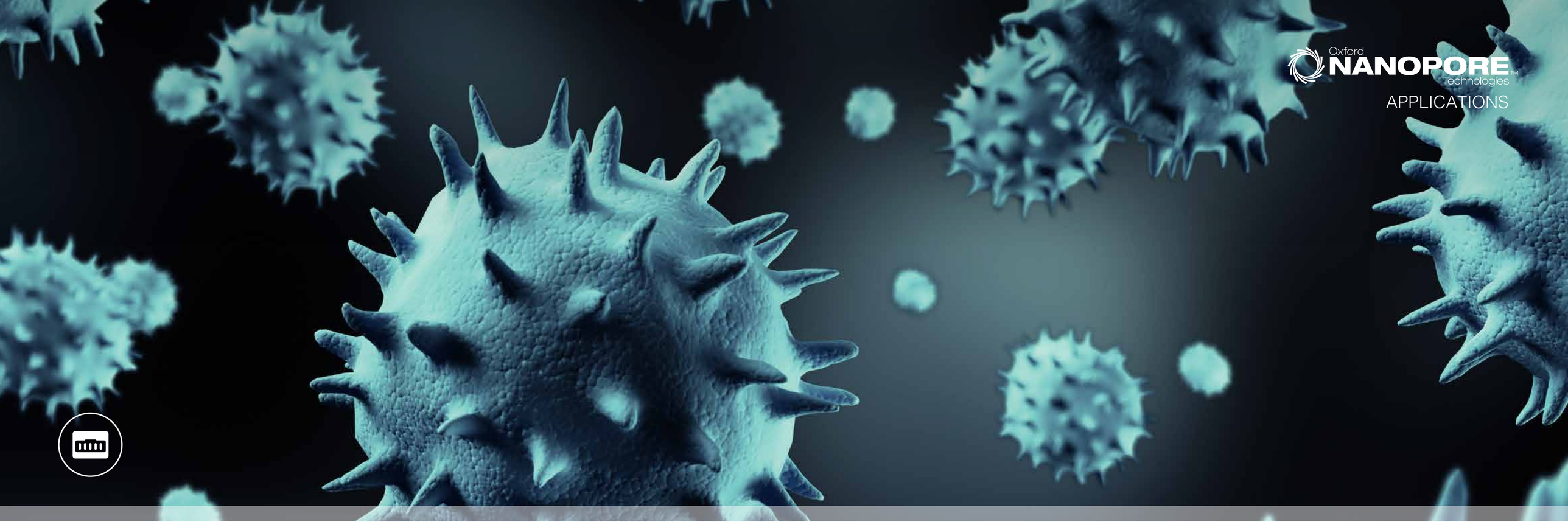

Nanopore sequencing of native adeno-associated virus genomes to aid in gene therapy vector quality control

Sequencing of full-length native AAV genomes on Nanopore devices allows identification of truncation hotspots, plasmid and cell line contamination, and mutation hotspots

Contact: **apps_posters@nanoporetech.com** More information at: **www.nanoporetech.com** and **publications.nanoporetech.com**

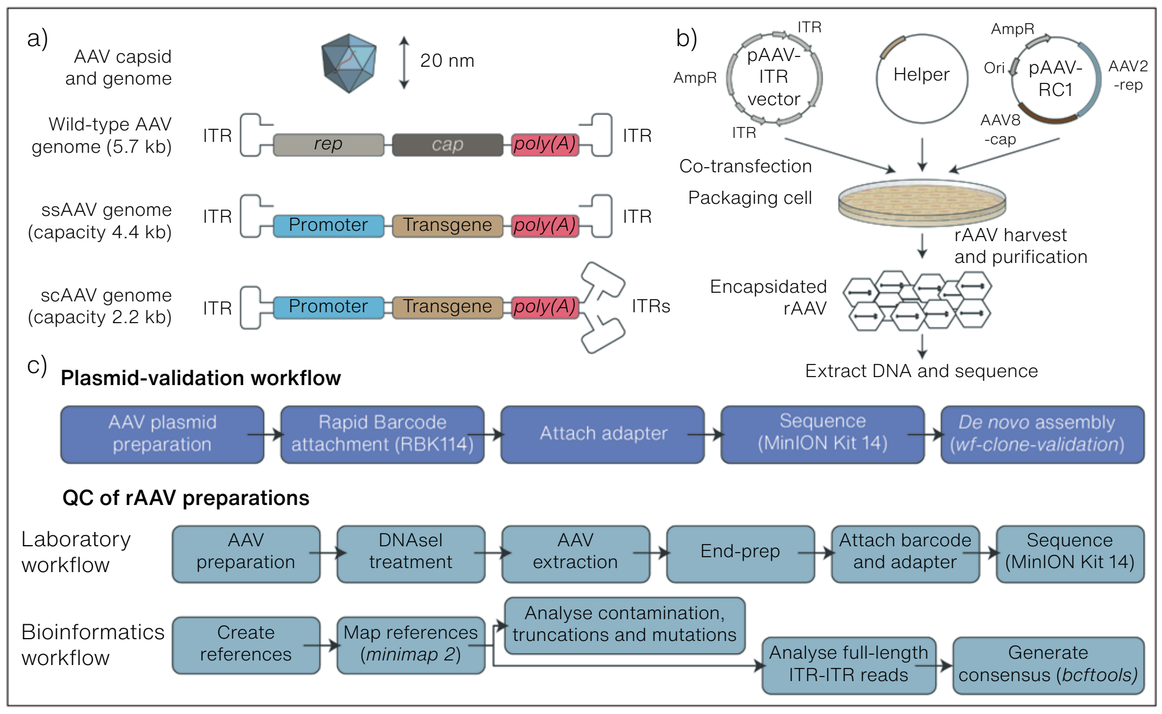

Fig. 1 AAV a) structure b) gene-therapy workflow c) library prep and bioinformatic workflows

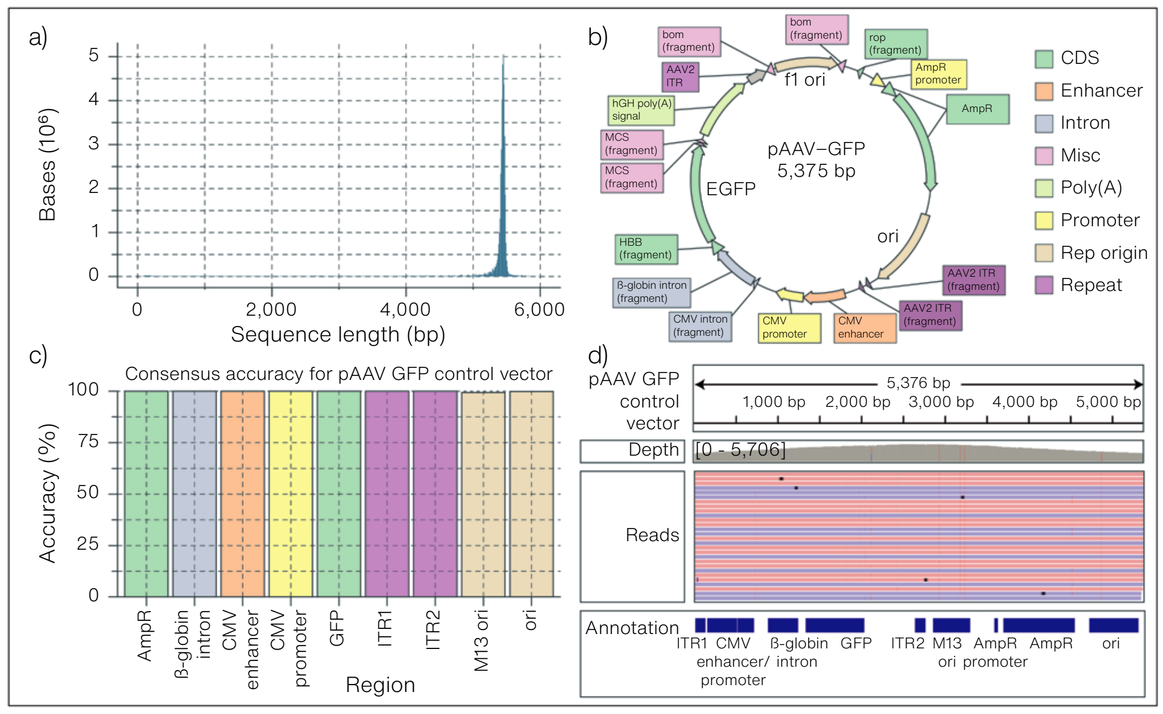

Fig. 2 Sequencing plasmid a) read length b) plasmid map c) accuracy d) coverage

Adeno-associated virus is a common vector used in gene therapy

Adeno-associated viruses (AAV) are small replication-deficient viruses containing a 4.8 kb single-stranded DNA genome. This encodes the rep and cap genes necessary for replication and capsid production, flanked by inverted terminal repeat (ITR) regions, crucial for AAV packaging and replication (Fig. 1a). AAVs are increasingly used in gene therapy for *in vivo* DNA delivery. Replacement of the cap/rep genes with a transgene of interest can produce two AAV vectors; single stranded AAV (ssAAV) and self-complementary AAV (scAAV) (Fig. 1b). Nanopore sequencing of rAAV genomes, and the plasmids used to create them, allows the generation of end-to-end reads to aid in AAV vector QC (Fig. 1c).

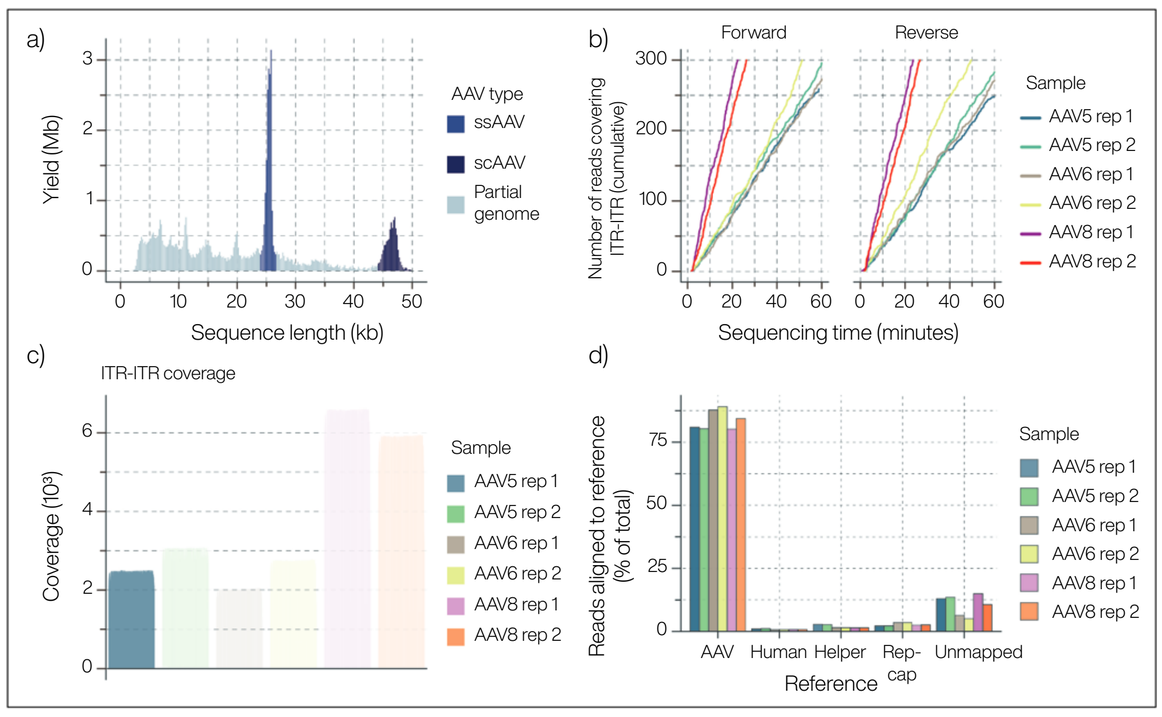

Fig. 3 a) read lengths b) coverage over time c) per-base coverage d) read alignments

Validation of AAV plasmid constructs using EPI2ME™ Labs' *wf-clone-validation*

Using the EPI2ME Labs *wf-clone-validation* and associated protocol, we put a typical control AAV transgene plasmid containing the (pAAV-GFP) through the protocol using the rapid barcoding kit. Read-length distributions show a peak at ~ 5.3 kb indicating a single-cut full-length pAAV-GFP plasmid (Fig. 2a). We used the EPI2ME Labs workflow to create a plasmid map identifying the presence of the transgene and ITR sequences (Fig. 2b) in addition to identifying all key plasmid components with high accuracy when compared with the manufacturer's reference (Fig. 2c). Alignments show even coverage indicating that circular plasmids were cut once with the rapid barcoding transposase (Fig. 2d).

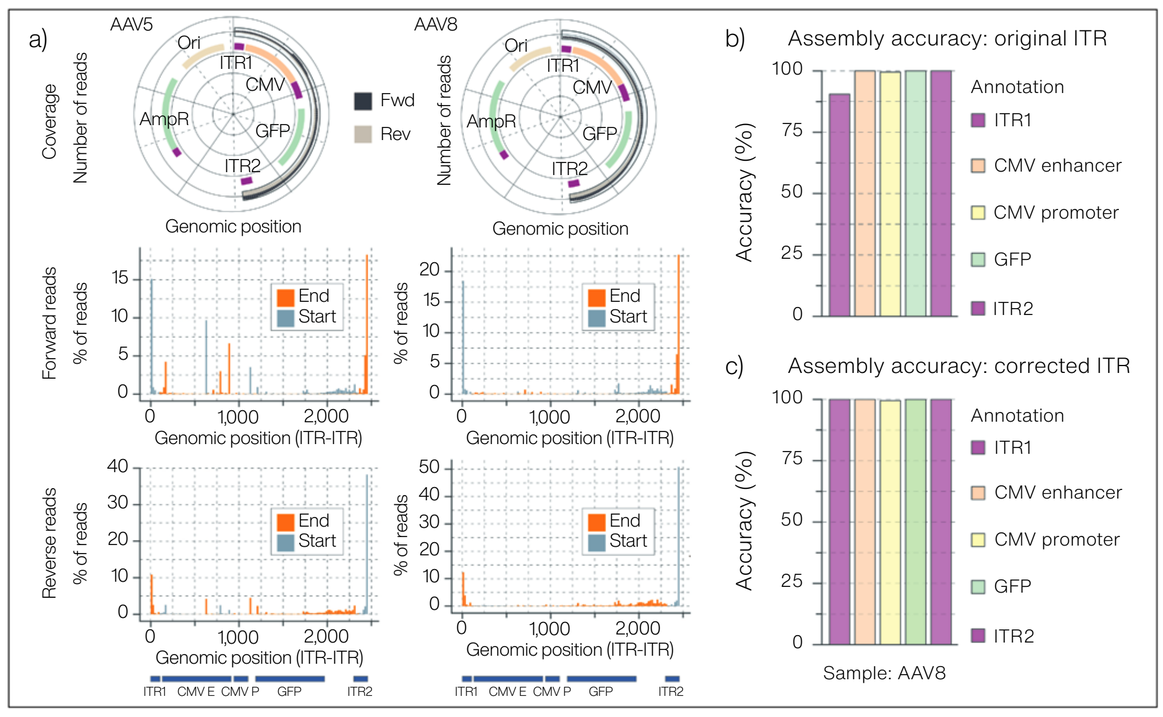

Fig. 4 a) coverage b) consensus accuracy before and c) after reference correction

Full-length native AAV genomes and vector contamination identified by sequencing

Two replicates of three serotypes, AAV5, AAV6 and AAV8, were barcoded and sequenced using NBD114 on a single GridION™ Flow Cell. Read distributions show clear peaks for ssAAV and scAAV genomes as a mixed population, plus truncated genomes and potential contamination (Fig. 3a). Within one hour of sequencing, we obtained more than 200x coverage of the ITR-ITR region for each barcoded library (> 95% read and reference coverage, Fig. 3b), and coverage was even across the ITR-ITR region (Fig. 3c). To identify potential sources of vector contamination we aligned reads to the host, helper, and rep-cap plasmids (Fig. 3d).

Generation of consensus sequence and identification of truncation hotspots

Identifying truncation hotspots and the generation of a consensus sequence of the AAV genome are common requirements for quality control of AAV preparations. Calculating the sum of the start and stop positions of primary read alignments can identify truncated genomes and their locations. Here the AAV5 sample shows truncations in the CMV promoter which are absent in the AAV8 sample (Fig. 4a). Accuracy of the *bcftools*-generated consensus sequence for the AAV8 genome showed a reduction in the first ITR sequence (Fig. 4b) when compared against the supplied reference. This was due to an incorrect ITR reference and upon correction accuracies increased (Fig. 4c). This demonstrates the importance of performing sequencing-based QC of the plasmid.