BENEFITS of HIGH THROUGHPUT SEQUENCING METHODS for CHARACTERIZATION of rAAV PRODUCTS

S. Stinus[1], S. Maestro[1], DL. Bowie[1], A. Miguel Coello[1], A. Pérez San Vicente[1], C. Trigueros Fernandez[1], A. François[1], E. Lecomte[1].
[1]Viralgen S.L., San Sebastián, Gipuzkoa, 20009, Spain.

Introduction

With the success of clinical trials, gene therapy today may represent hope for patients suffering from genetic diseases. Recombinant adeno-associated virus (rAAV) products are extensively characterised prior to release. Among the **Critical Quality Attributes (CQAs)** used for release testing, those related to the identity and purity of rAAV products are scrutinized to ensure patient safety and treatment efficacy. Although quantitative polymerase chain reaction (qPCR), PCR, and Sanger sequencing are currently among the gold standard techniques, the rapid development of high throughput sequencing technologies has paved the way for a much more exhaustive characterisation of the DNA content in rAAV products.

Studies conducted to date show that both short-read and long-read sequencing technologies are of importance for the characterization of DNA, each with its own strengths[1,2,3,4]. Next-generation sequencing offers high-sequencing depth and a low sequencing error rate, enabling exhaustive identification and quantification of contaminating DNA, and detection of mutations at unprecedented levels. On the other hand, long-read sequencing technologies offer the opportunity to sequence DNA at the single-molecule level, allowing the identification and quantification of truncation events in rAAV genomes.

We have developed and established all the laboratory protocols, from sample and library preparation to sequencing, as well as the bioinformatics pipelines for providing information on CQAs, including:

- **Payload sequence IDENTITY** (identification of single nucleotide polymorphisms, indels). **Illumina**
- **Payload sequence INTEGRITY** (truncation hotspots). **Oxford Nanopore Technologies**
- **Non-payload sequences characterization – PURITY** (identity, and relative quantification). **Illumina**

Materials & Methods

For an in-depth characterization of the DNA content in rAAV products, research-grade single-stranded AAV vectors produced by triple transfection with Viralgen's proprietary Pro10™ packaging cells were analyzed using the internally developed methods.

After sample preparation, which involves thermal annealing and DNA purification, PCR-free and ligation-based library preparation is performed. Sequencing is then conducted using either the NextSeq2000 sequencer (Illumina technology) or the MinION/GridION sequencers (Oxford Nanopore Technologies). Finally, reads were analysed by custom made bioinformatics pipelines. The main steps are:

Sample preparation → Library preparation → Sequencing → Analysis

IN-DEPTH CHARACTERIZATION of rAAV PRODUCTS

A

Reference sequence	No. of reads	Percentage
Unmapped reads	64	0.0003
Mapped reads	19 998 914	99.9997
↳ rAAV vector cassette	19 770 934	98.86

Sequence	Coverage	PID
rAAV vector cassette	100%	100%

B

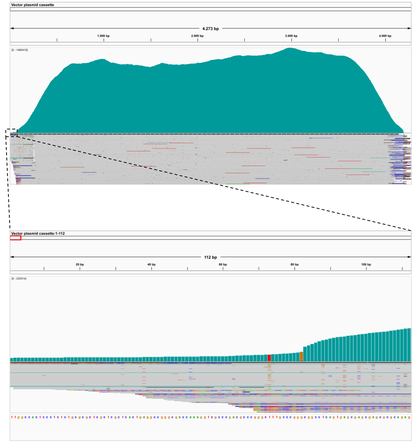


Fig.1 | PAYLOAD SEQUENCE CHARACTERIZATION | IDENTITY

(A) Number of unmapped and mapped reads, and number of reads mapping to the rAAV vector cassette. The very few unmapped reads indicate high confidence in the alignment method. Percentage of coverage and percentage of identity (PID) of the rAAV vector cassette are also shown. (B) Sequencing coverage along the rAAV vector sequence obtained with Illumina sequencing and visualized using IGV. The whole rAAV genome sequence is covered, including the ITRs. (C) IGV visualization focused on the left ITR. The two highlighted bases are located in the BB'-CC' region and can be related to flip and flop conformations of the ITRs. Therefore, these bases are not considered variants.

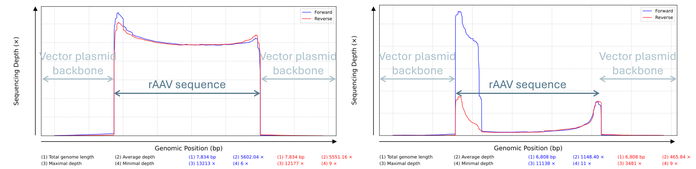
A combination of three variant callers was used to enable the detection of single nucleotide polymorphisms and insertions/deletions down to a level of 5%. No mutations were identified in the gene of interest using short-reads sequencing.

Fig. 2 | PAYLOAD SEQUENCE CHARACTERIZATION | INTEGRITY

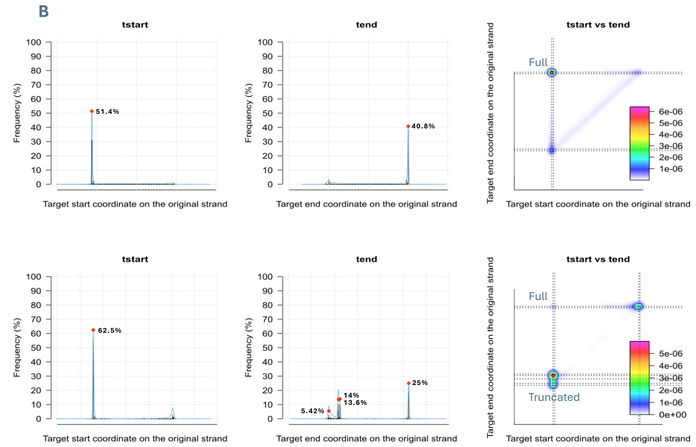
(A) Genome coverage obtained after ONT sequencing of two rAAV samples which were analyzed by alkaline agarose gel electrophoresis (data not shown), one (left panel) showing a single band corresponding to full-length genomes, and the other (right panel) showing two extra bands of lower molecular weight corresponding to truncated genomes. (B) Start and end positions of the reads obtained after ONT sequencing of non-truncated (top panel) and truncated (bottom panel) rAAV preparations. Reads were aligned on the full-length vector plasmid sequence to identify and quantify truncation hotspots and large deletions. Expected start and end positions of the rAAV genomes were found for both non-truncated and truncated rAAV. While no truncation hotspot was identified for the first sample, two main truncation hotspots were identified in the second sample.

Detection and quantification of genome truncation hotspots down to a level of 5%.

A



B



A

Reference sequence	No. of reads	Percentage
rAAV vector cassette	164 169	98.86
rAAV plasmid backbone	1 581	0.95
RepCap plasmid	286	0.17
Helper plasmid	21	0.01
Ad5 genome left end	0	0.00
Human genome	11	0.01

B

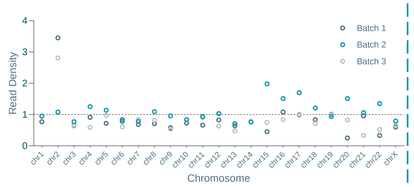


Fig. 3 | NON-PAYLOAD SEQUENCES CHARACTERIZATION | PURITY

(A) Percentages of DNA populations in the rAAV preparation. The main contamination originates from the vector plasmid backbone, as expected[5,6]. (B) Density of reads mapped per chromosome obtained after normalization to the read count of an internal control, which contained sonicated DNA extracted from Pro10™ cells. The density graph suggests an overall random distribution of host-cell DNA contaminants.

The method allows the detection and quantification of DNA down to a level of 0.01%. The host-cell DNA found in the rAAV preparation is randomly distributed across the human chromosomes, with no overrepresentation of specific regions.

Europar Batasunak kofinantzatua
Cofinanciado por la Union Europea
ELKARTEK JAURLARITZA GOBIERNO VASCO

DISCUSSION & CONCLUSIONS

We have developed methods, from sample preparation to bioinformatics analyses, to characterize both DNA starting materials (e.g., plasmids) and rAAV gene therapy products. Our methods are able (1) to identify and quantify single nucleotide polymorphisms and indels with accuracy at a level of 5%, (2) to identify and quantify truncation hotspots with accuracy at a level of 5%, and (3) to detect and quantify with accuracy the presence of low levels (0.01%) of DNA contaminants.

We have used our in-house developed methods for the in-depth characterization of DNA in rAAV products:

- Illumina sequencing confirmed the identity of the rAAV, with no mutations identified in the gene of interest. The entire rAAV genome was covered, including the extremities.
- Short-read sequencing was also used to assess the purity of the rAAV product. Most of the reads mapped to the vector plasmid cassette, emphasizing the high level of purity of the rAAV preparation. The small number of unmapped reads underscores the high confidence of the alignment. Importantly, host-cell DNA contamination was determined as random, with no overrepresentation of specific genomic regions.
- Sequencing with Oxford Nanopore Technologies enabled the detection and quantification of truncation hotspots.

High throughput sequencing is of great interest for providing information on 3 Critical Quality Attributes for rAAV products: **(1) payload sequence identity, (2) payload sequence integrity, and (3) residual DNA sequence identification.** High throughput sequencing provides information which, when combined with other analytical methods such as analytical ultra-centrifugation or mass photometry for full/partial/empty capsid characterization, ddPCR or qPCR for residual DNA quantification, and potency assays, will be beneficial to guide process development effectively toward rAAV products with improved potency and safety profiles.

Our methods can be applied for the exhaustive characterization of both DNA starting materials (e.g., plasmids) and rAAV gene therapy products. The validation of these high throughput sequencing-based methods is the next step towards safer rAAV-based gene therapies.

References

1. Lecomte E, et al. (2021). The SSV-Seq 2.0 PCR-free method improves the sequencing of adeno-associated viral vector genomes containing GC-rich regions and homopolymers. *Biotechnol. J.* 16(1):e2000016.
2. Tai PWL, et al. (2018). Adeno-associated virus genome population sequencing achieves full vector genome resolution and reveals human-vector chimeras. Mol. Ther. *Methods Clin. Dev.* 9, 130–141.
3. Namkung S, et al. (2022). Direct ITR-to-ITR nanopore sequencing of AAV vector genomes. *Hum. Gene Ther.* 33(21–22), 1187–1196.
4. Radukic MT, et al. (2020). Nanopore sequencing of native adeno-associated virus (AAV) single-stranded DNA using a transposase-based rapid protocol. NAR Genom. Bioinform. 2(4):lqaa074.
5. Chadeuf G, et al. (2005). Evidence for encapsidation of prokaryotic sequences during recombinant adenoassociated virus production and their in vivo persistence after vector delivery. *Mol Ther.* 12(4):744-53.
6. Lecomte E, et al. (2015). Advanced Characterization of DNA Molecules in rAAV Vector Preparations by Single-stranded Virus Next-generation Sequencing. *Mol Ther Nucleic Acids.* 4(10):e260.