True or artefact data : The impact of library preparation on AAV in-depth analysis by short-read NGS

K BREUNIG, F DUNKER-SEIDLER, M HAUBNER, S HENZE, F SONNTAG, M HOERER, and RC FEINER

ABL inc. ADVANCED BIOSCIENCE LABORATORIES

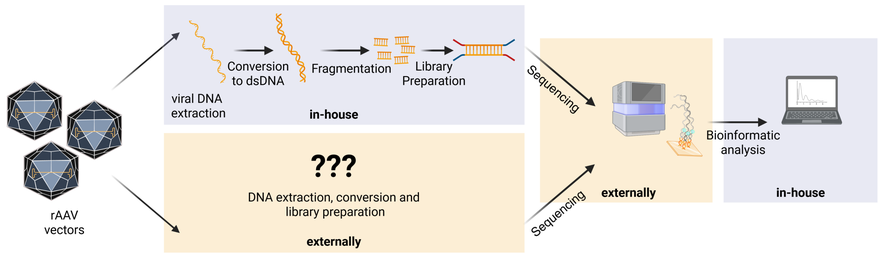
ascend

Poster: 3202

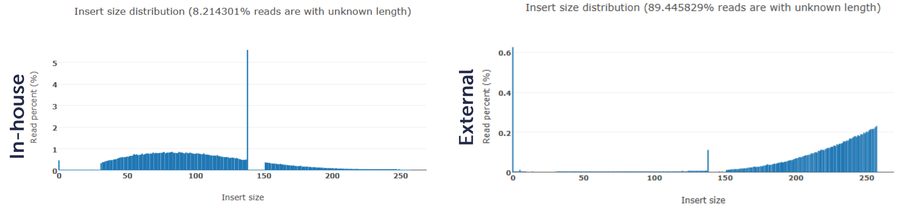
Abstract

Recombinant adeno-associated virus (rAAV) is a widely used vector in gene therapy, valued for its efficient gene delivery and low immunogenicity. Ensuring accurate genome sequence and structural integrity is essential for vector functionality and regulatory compliance. Genome integrity analysis enables detection of sequence mutations, structural rearrangements, and impurities such as truncated genomes or aberrant fusion events involving ITRs and regulatory elements. Next-generation sequencing (NGS) enables comprehensive, high-resolution analysis of rAAV genomes, allowing detection of structural variants, and low-abundance contaminants that targeted methods may miss. While our lab has successfully established long-read nanopore sequencing for this purpose, we have now additionally implemented Illumina® sequencing to complement our workflow. This poster presents the development and evaluation of an optimized Illumina sequencing workflow for rAAV genome characterization, compared to a commercially available sequencing service. Key distinguishing features include improvements in ITR coverage leading to improved alignment with long-read NGS data, as well as advantages in utilizing deep sequencing to detect rare host cell DNA impurity species

Comparative workflow for viral DNA sequencing: In-house and external approaches



Comparative workflow for viral DNA sequencing: In-house and external approaches

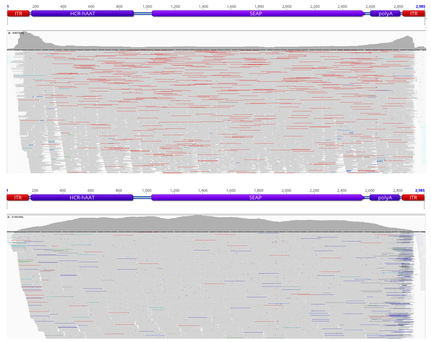


- **In-house protocol** achieves a more uniform insert size distribution with only 8.2 % unknown reads, supporting improved library quality and data usability
- **Commercial service** shows high percentage of reads with unknown insert size (89.4 %), indicating inefficient library construction or adapter trimming

Mapping distribution of aligned reads correlates with improved ITR coverage

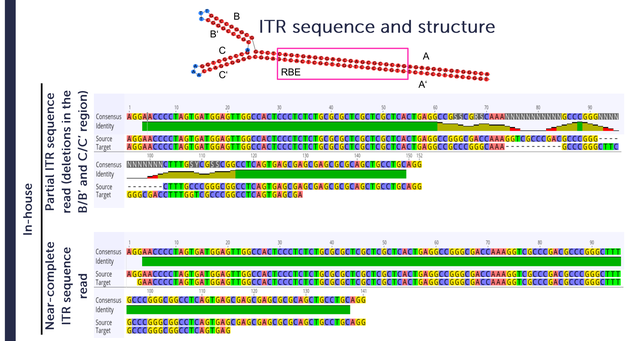
	In-house library preparation	External library preparation	Nanopore sequencing
rAAV vector genome	97.10 %	95.00 %	97.52 %
Vector plasmid backbone	2.29 %	4.00 %	1.84 %
Helper Plasmid	0.58 %	0.95 %	0.59 %
Host cell DNA	0.02 %	0.05 %	0.05 %
Sum	100 %	100 %	100 %
Total reads	141,908,349	88,915,895	1,739,776

Comparison of vector genome mapping reveals improved ITR coverage



- **In-house library** preparation shows improved read alignment in ITR regions and more consistent coverage across all genomic elements
- **External dataset** displays lower resolution in critical regions, especially ITRs and regulatory sequences, limiting detection of structural features

In-house process enables improved coverage of high-GC ITR regions

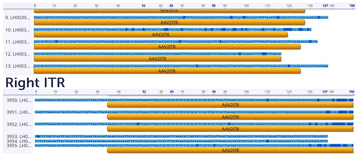


- Optimized library preparation enables sequencing into the ITR regions despite high GC content and secondary structure
- However, full ITR coverage for all sequences is limited
- Demonstrates that Illumina short-read technology can capture partial ITR sequences when protocol is properly adapted
- Target: Illumina read, Source: ITR reference

Ultradeep NGS runs allow for the detection of very rare impurities

HCD-ITR fusions

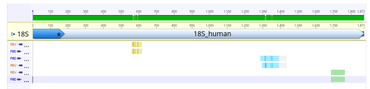
- About 10.02 % of all HCD reads contain part of an ITR



- Detection in external library not possible → No sequencing of ITRs

Detection of 18S as HCD target

- Common target in orthogonal assays e.g. ddPCR
- Detection of a very rare impurity
- High Illumina sequencing depth allows for detection of <u>three</u> 18S sequencing reads in the internal data set, <u>no</u> hit in the external data set



Summary

Short-read sequencing using Illumina technology is commonly considered incompatible with resolving structured regions like AAV ITRs. Here, we present an optimized in-house library preparation workflow that significantly improves read mapping and enables partial ITR resolution. Compared to external sequencing data, our method shows higher insert size consistency, enhanced read quality, and more accurate encapsidated DNA distribution through increased ITR (part of the rAAV vector genome) coverage. Through careful selection of fragmentation and amplification conditions, even challenging sequences like ITRs become partially accessible to short-read analysis. Although ITR quantification remains limited, our results highlight that methodological refinements can at least partially overcome previous barriers. Complementary nanopore sequencing provides long-read support for full-length genome context, enabling structural insights that remain unresolved when short-read approaches are used alone.

www.ascend-adv.com
business@ascend-adv.com

Ascend Advanced Therapies, 9800 Medical Center Dr D, Rockville, MD 20850

Scan to download: