minaris.com

NGS Bioinformatics Pipelines for Enhanced Viral Vector QC and Analytics

Mohine Alam, Brian Baier[1], Francesca Bennett, Siobhan Clerkin, Michal Fleszar, Sherin Parokkaran Johny[1], Weiheng Su[1], Vaughan Leydon[1]

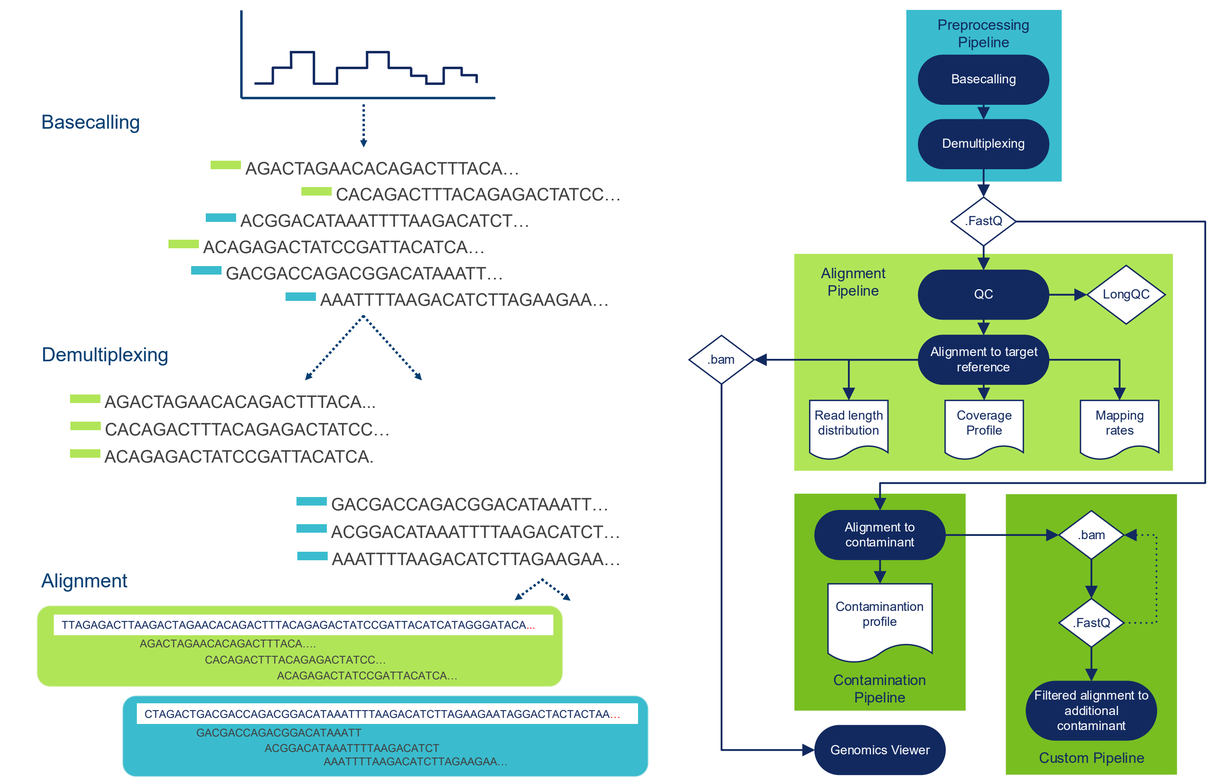
[1]Minaris Advanced Therapies, Oxford, UK.

Corresponding author: vaughan.leydon@advancedtherapies.com For information about our technologies, please contact us at www.minaris.com.

Introduction

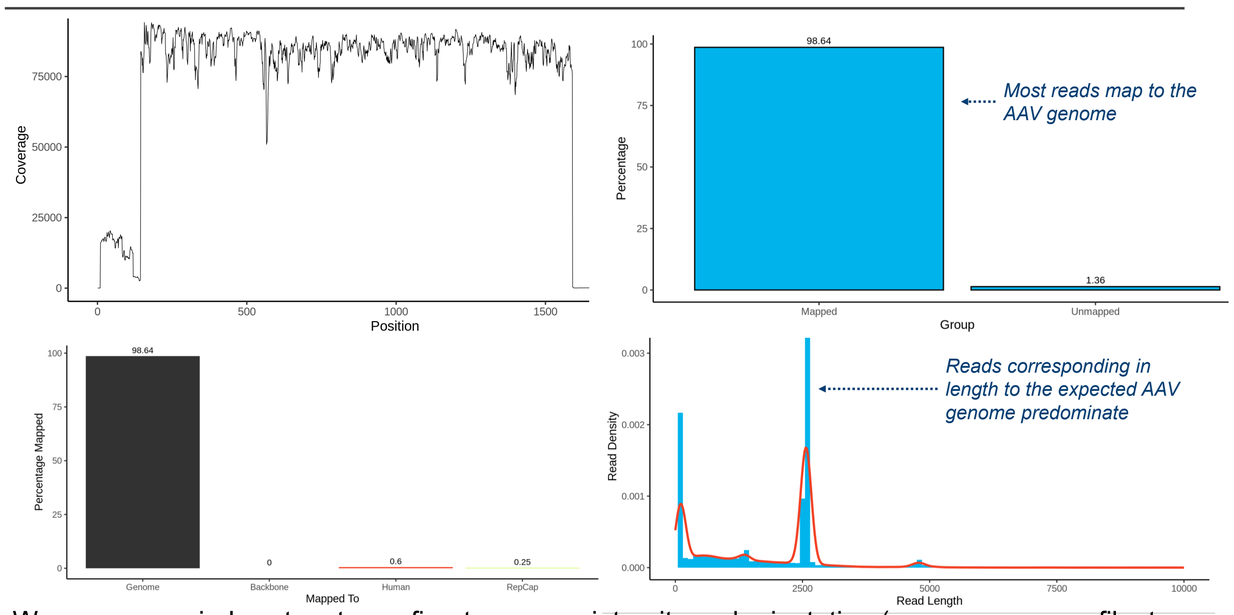
Viral vectors such as Adeno-Associated Virus (AAV) and Lentivirus are vital tools for delivering gene therapies to patients with previously untreatable monogenetic diseases. Ensuring payload homogeneity, stability, and integrity is critical for treatment safety. Next generation sequencing (NGS) plays a central role in analysing these vectors. We present bioinformatics pipelines for processing NGS data from viral vector preparations, emphasizing quality control, variant detection, transcript integrity, and splicing analysis. Key QC metrics include mapping rates, coverage, relative abundance, and read length distribution, helping identify anomalies. The pipelines also detect mispackaged plasmid or host DNA, truncated genomes, empty capsids, sequencing artifacts, and contamination, while assessing preparation homogeneity. These robust pipeline support precise evaluation of vector safety, advancing gene therapy and related research.

AAV NGS Bioinformatics Workflow



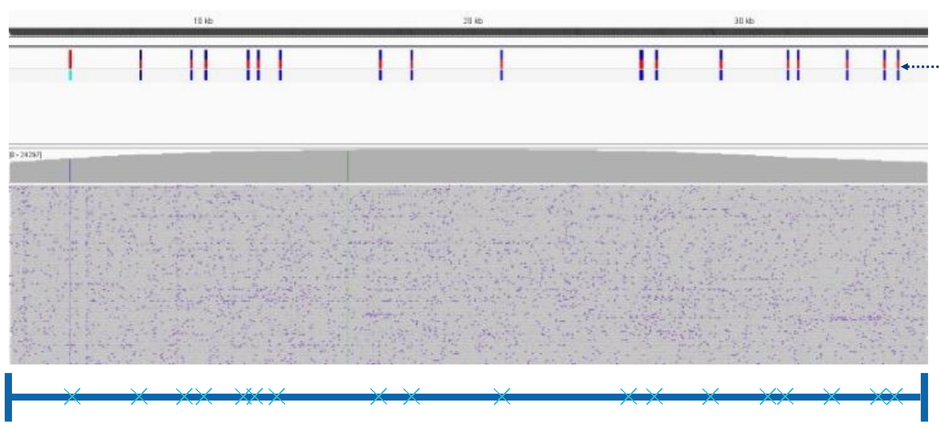
Overview of Nanopore preprocessing (basecalling and demultiplexing) and analysis (left). Modular bioinformatics pipelines can be configured to align reads to a target reference and known contaminants. Custom pipelines for novel contaminant detection filter reads from prior investigations before aligning to an additional reference (right).

Coverage and contamination profiles for Adeno-Associated Virus (AAV)



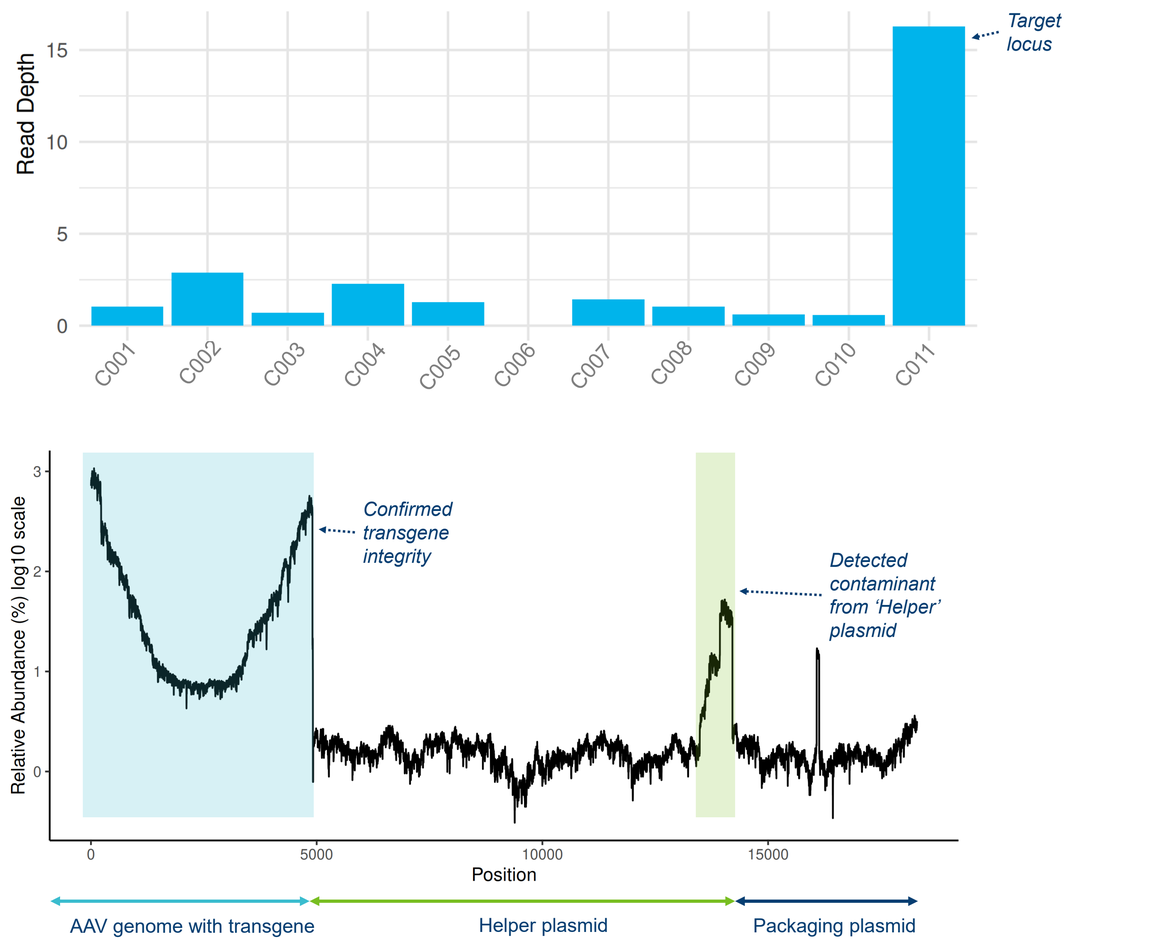
We sequence viral vectors to confirm transgene integrity and orientation (see coverage profile, top left). Rigorous library preparation reduces exogenous DNA, including reverse/co-packaged and host cell impurities. Residual contaminants are profiled in detail (bottom left). Mapping rates and read length distribution offer insights into Nanopore run performance and material homogeneity (top left, top right). These methods enable assessment of genome structure and identity, supporting evaluations of safety, efficacy, and potency.

Adenovirus (AdV) variant calling profile



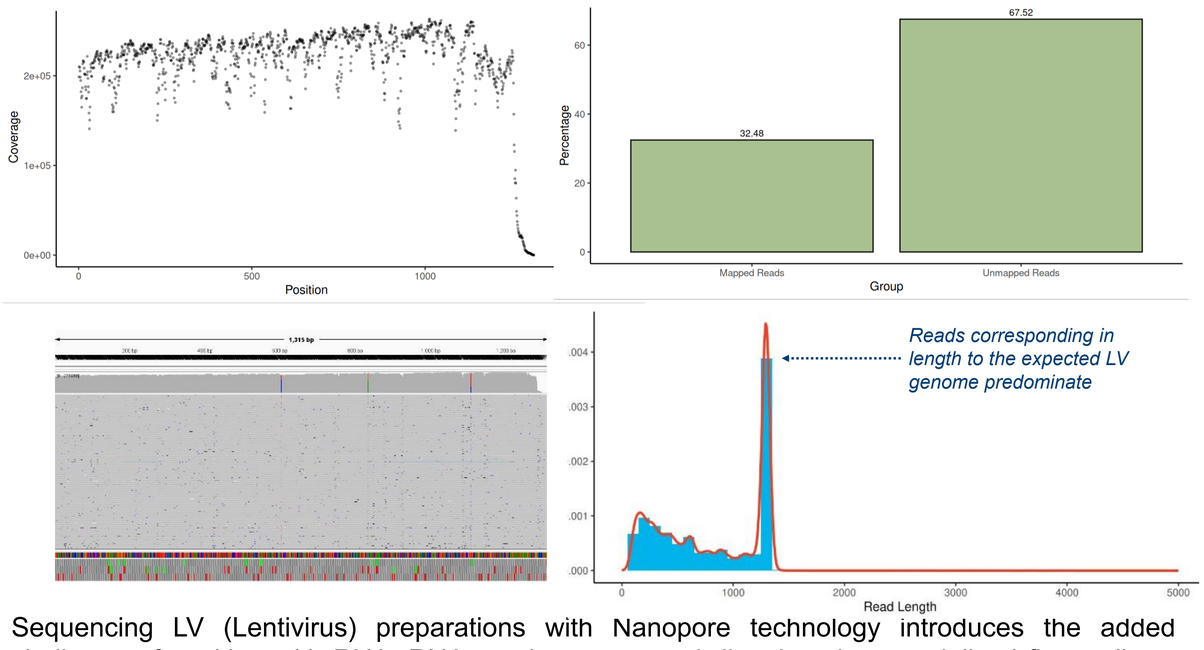
Variant analysis enables high-resolution genomic profiling, detecting small and structural variants to assess vector homogeneity. Small Nucleotide Polymorphisms (SNPs) are called when an unexpected base is found beyond a predetermined threshold as a proportion of all reads.

Custom detection for unusual AAV contaminants



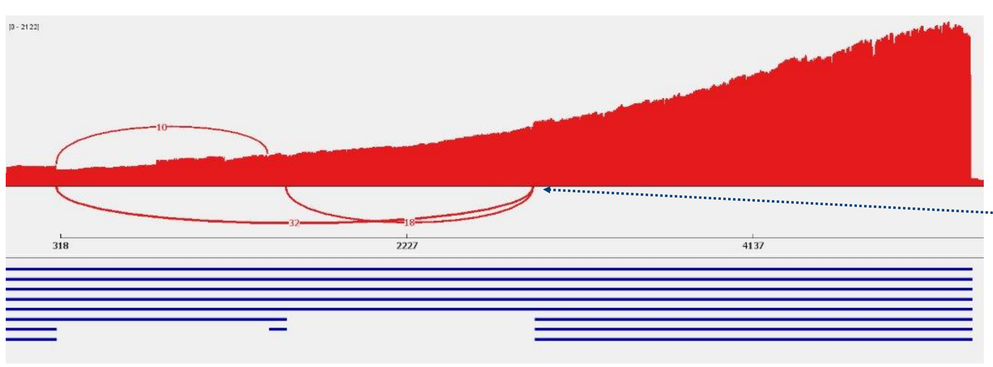
Suspected AAV contaminants deriving from AAV helper plasmids may be detected and quantified using bespoke pipeline configurations. Read depth for loci of equal size may be compared against a reference loci (top) and relative read abundance by position determined for a genome in a concatemer of references (bottom).

LTR-LTR coverage profile for Lentivirus preparations



Sequencing LV (Lentivirus) preparations with Nanopore technology introduces the added challenge of working with RNA. RNA can be sequenced directly using specialized flow cells or converted to cDNA for sequencing. Direct RNA sequencing produced good coverage (top left/bottom left), though mapping rates were generally lower than for AAV samples (top right). Read length distribution is also encouraging (bottom right).

Lentiviral RNA splicing site prediction



Gaps in coverage in which reads are consistently initiated or terminated may reflect LV RNA genome splicing, corresponding to introns. Our dedicated bioinformatics pipeline enables the prediction of splice donor and acceptor sites within LV RNA genomes, revealing unanticipated splicing patterns, as seen in the Sashimi plot (above).

Summary

Minaris Advanced Therapies offers flexible, modular bioinformatics pipelines for bespoke analysis of Nanopore sequencing data, enhancing viral vector characterisation. Beyond reference alignment and coverage profiling, the pipelines support detailed contaminant analysis and iterative mapping against multiple references. They are suitable for AAV, AdV, and LV preparations, with refined capabilities for LV that include RNA splice site prediction—potentially revealing valuable insights into transgene expression in the intended cellular background.