British Pharmacopoeia

Consultation

British Pharmacopoeia public consultation for draft guidance for Characterisation of the Capsid Particle Population in rAAV Products: Determination of Vector Genome Identity, Integrity and Encapsidated DNA Impurities.

Consultation period 18 December 2025 to 27 March 2026

Medicines & Healthcare products Regulatory Agency

British Pharmacopoeia

1. Patients, standards, and innovation

The quality of a medicine is critical to ensuring its safety and efficacy, and therefore the medicine's suitability for patients. Pharmacopoeial standards are part of an interlinked system, together with good practice guidelines and regulatory assessment, that form a foundation to ensuring medicines are of an acceptable quality. Additionally, standards have a place in supporting and enabling innovation through the availability of consistent and widely applicable quality requirements. Innovation in the field of medicines and healthcare has the potential to support patients throughout the world to live longer, healthier, and happier lives.

In recognition of the increasingly important role of biological medicines to healthcare worldwide, the Medicines and Healthcare products Regulatory Agency (MHRA) has developed and implemented a strategy for pharmacopoeial public quality standards for biological medicines.[1] This strategy, adopted following consultation with stakeholders, laid out a vision of working collaboratively to explore and develop new standard setting approaches for biological medicines. It included a commitment to investigate and take forward standard setting opportunities for innovative Advanced Therapy Medicinal Products (ATMPs).

ATMPs have the potential to be transformative to patients and healthcare globally. However, development, characterisation, and production of these innovative medicines is challenging due to their high complexity, their product specificity, and the still-emerging technologies that support them. Publications such as the Advanced Therapies Manufacturing Taskforce Action Plan,[2] the Medicines Manufacturing Industry Partnership's Manufacturing Vision for UK Pharma[3] and stakeholder feedback have emphasised the important role that standards can have in the development of these medicines. This includes a focus on the value of widely applicable standards that could support knowledge building and facilitate analytics and characterisation.

This draft guidance was written by experts in the ATMP community to support those involved in the development of analytical methods throughout the product lifecycle, and therefore contribute to the quality assurance of innovative medicines for patients.

The MHRA and British Pharmacopoeia would like to recognise and thank the numerous experts in the BP's Working Party for ATMPs that have contributed to the development of this text. The work has been supported by a joint-staff secondment scheme between the BP and the UK's Cell and Gene Therapy Catapult.[4]

2. The draft document

As part of the MHRA strategy for the creation of pharmacopoeial public quality standards for biological medicines, the British Pharmacopoeia Expert Advisory Group for ATMPs, established in March 2020, has engaged with groups across the cell and gene therapy

[1] https://www.gov.uk/government/consultations/strategy-for-pharmacopoeial-public-quality-standards-for-biological-medicines

[2] http://www.abpi.org.uk/publications/advanced-therapies-manufacturing-action-plan/

[3] https://www.abpi.org.uk/publications/manufacturing-vision-for-uk-pharma-future-proofing-the-uk-through-an-aligned-technology-and-innovation-road-map/

[4] https://ct.catapult.org.uk/

community to develop non-mandatory guidance for key analytical technologies to ensure quality throughout the product lifecycle. The working party has developed guidance to support ATMP development across a wide range of organisations, laboratory settings, and therapy types. As such, guidance is product-agnostic and does not provide a step-by-step protocol, nor constitute a prerequisite for product acceptance, but instead offers measures to ensure the production of robust, comparable, and reproducible data within and across organisations.

The ATMP industry continues to grow rapidly worldwide, with increasingly sophisticated scientific discoveries being translated into therapies. There are a variety of challenges in characterising these experimental living medicines. Any CGT product must be characterised in terms of identity, purity and potency and the choice of, and route to, validation of these assays largely lies with the developer and manufacturer. Establishing robust potency assays grows in importance throughout the development of a CGT product and becomes critical in the later clinical stages. As products move towards pivotal clinical trials and licensure, establishing the mechanism of action of the product becomes critical. This requires the potency assays to yield rich data which informs the interpretation of the outcomes in vivo, whether in models or in early human trials.

The purpose of this document is to expand guidance for the characterisation of capsid particle population in rAAV products to assess key product attributes using molecular methods: genome identity, vector integrity and contaminating DNA impurities. Efforts have been made to include the most common methodologies in use, as well as specific considerations given to the use of methods to assess different attributes. Whilst the document is intended to focus on encapsidated DNA, some DNA may be present in the final product that is outside the capsid or is associated with the capsid surface, so both encapsidated and non-encapsidated DNA are considered. Assessment of other nucleic acid contaminants such as RNA, nucleic acid determination of microbial contaminants and aspects of DNA modifications such as methylation are outside the scope of this guidance.

3. How to contribute

The draft guidance for Characterisation of the Capsid Particle Population in rAAV Products: Determination of Vector Genome Identity, Integrity and Encapsidated DNA Impurities will be posted online for public consultation for a period of two months. During this time, we are asking stakeholders to complete and return the response document, available on our website, to BiolStandards@mhra.gov.uk.

When reviewing the guidance, you may want to consider the following points:

- Do you agree with the technical recommendations made in the document?
- Are key methods for DNA characterisation and quantification covered?
- Are there any aspects which you think are missing from the document?
- Is there any terminology within the document that you think needs to be more clearly defined?
- Is the document understandable and are recommendations clear and unambiguous?
- Could the format/style of the guidance be improved?

In addition to the request for technical comments, the response form includes more general questions around the value of the guidelines and other work within the area of ATMPs where standards and standardisation could add value. This information will be used to help the BP to understand and prioritise future work related to ATMPs.

4. Confidentiality and Freedom of Information

Information we receive, including personal information, may be published, or disclosed in accordance with the access to information regimes (primarily the Freedom of Information Act 2000 (FOIA), the Data Protection Act 1998 (DPA) and the Environmental Information Regulations 2004).

Please let us know if you would like any information you provide to be treated in confidence, and please indicate any commercial sensitivities. We will maintain that confidence and resist disclosure under the access to information regimes where possible and in compliance with our legal obligations. We will also consult you and seek your views before any information you provided is disclosed.

British Pharmacopoeia

Annex 1 - Draft guidance: Characterisation of the Capsid Particle Population in rAAV Products Determination of Vector Genome Identity, Integrity and Encapsidated DNA Impurities

Medicines & Healthcare products Regulatory Agency

Characterisation of the Capsid Particle Population in rAAV Products: Determination of Vector Genome Identity, Integrity and Encapsidated DNA Impurities (Part 2)

Contents

1. Abbreviations

Acronym	Definition
AAV	Adeno-Associated Virus
AAVs	Adeno-Associated Viruses (plural)
ALARA	As Low As Reasonably Achievable (radiation safety principle)
AUC	Area Under the Curve (often in analytical chemistry or pharmacokinetics)
BWA	Burrows-Wheeler Aligner (read alignment tool)
CDMS	Clinical Data Management System
CFR	Code of Federal Regulations (U.S. regulatory framework)
CQA	Critical Quality Attribute
DNA	Deoxyribonucleic Acid
EMA	European Medicines Agency
FAM	Fluorescein Amidite (fluorescent dye used in qPCR and sequencing)
FDA	U.S. Food and Drug Administration
GATK	Genome Analysis Toolkit (variant calling and sequencing data analysis)
GMP	Good Manufacturing Practice
GOI	Gene of Interest
HEK	Human Embryonic Kidney (cell line origin)
HEK293	Human Embryonic Kidney 293 cells
HEX	Hexachloro-fluorescein (fluorescent dye)
HISAT2	Hierarchical Indexing for Spliced Alignment of Transcripts (read aligner)
ICH	International Council for Harmonisation of Technical Requirements
ICH Q2(R2)	ICH Guideline Q2 (Validation of Analytical Procedures)

ITR	Inverted Terminal Repeat (viral genome element in AAV)
NGS	Next-Generation Sequencing
PCQA	Product Critical Quality Attribute
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
SBS	Sequencing by Synthesis (NGS technology)
SDS	Sodium Dodecyl Sulfate (detergent used in protein/DNA prep)
SINEs	Short Interspersed Nuclear Elements (repetitive genomic elements)
SMRTbells	Single Molecule Real-Time bells (circular DNA used in PacBio sequencing)
SNVs	Single Nucleotide Variants
SOPs	Standard Operating Procedures
STAR	Spliced Transcripts Alignment to a Reference (RNA-seq aligner)
SYBR	SYBR Green (DNA-binding fluorescent dye used in qPCR)
TAT	Turnaround Time (can also refer to HIV TAT protein, but here likely timing)
USP	United States Pharmacopeia
VCN	Vector Copy Number (measure of transgene copies per cell)

2. Terminology

This section defines key terms and concepts used throughout this guidance document. Understanding these terms is essential for consistent interpretation and application of the guidance. Where applicable, definitions align with industry standards or organizational usage to ensure clarity and reduce ambiguity. Readers are encouraged to refer back to this section as needed to support accurate understanding of the content that follows.

Encapsidated DNA: All the DNA contained within viral capsid species classed as full, empty, partially filled / intermediate and overpackaged including particles containing contaminating DNA. This DNA can be assessed by first treating vector product with nucleases to degrade non-encapsidated DNA followed by release and analysis of the DNA within the capsid.

Vector DNA: The payload containing the gene of interest, its regulatory elements and required AAV DNA such as ITRs

Host Cell DNA (hcDNA): The production process comprises transfecting host cells with appropriate plasmids so that these host cells may assemble the desired rAAV capsids. As the process of capsid assembly tends to be imperfect, DNA derived from these host cells may be present in assembled capsids.

Plasmid DNA:

- **Gene of interest:** This plasmid provides the GOI including its regulatory elements.
- **Helper plasmid:** This plasmid contains external replicative elements which AAV otherwise might obtain from adenoviruses.
- **Rep-Cap plasmid**: This plasmid contains AAV elements required for replication and capsid assembly during production.

Antibiotic Resistance Genes: To permit plasmid production in bacterial hosts, the above plasmids contain antibiotic resistance genes.

Genome Identity Test: This test provides assurance that the final vector product contains the intended payload / therapeutic transgene.

Integrity test: This assessment refers to the integrity of the intended payload / therapeutic transgene that is encapsidated in the final vector product. It provides assurance that the packaged transgene, ITR sequences and any associated regulatory elements are present in both the correct size and full sequence, i.e. that there are no mutations, insertions, deletions or other genetic abnormalities present that could compromise efficacy/potency and safety.

Impurities test: This test can be applied to different impurities that could be present in the final vector product. In the context of this guidance 'impurities' refers to unwanted and contaminating DNA impurities such as those originating from Host Cell DNA (hcDNA) and/or from residual DNA from any plasmids, helper virus or other genetic components used during the upstream transfection step.

3. Introduction

DNA impurities are recognized as a Critical Quality Attribute (CQA) or Potentially Critical Quality Attribute (PCQA) in rAAV gene therapy products. Generally, DNA impurities can be categorized into two types: process-related impurities and product-related impurities. Process-related impurities refer to residual DNA outside the capsid that can be accessed by endonucleases, while product-related impurities pertain to residual DNA encapsidated alongside the viral vector DNA. This encapsidated residual DNA, including host cell DNA (hcDNA) or residual plasmid DNA, falls within the scope of this guidance.

Given the potential risks to patient safety, the US Food and Drug Administration (FDA) and other global regulatory agencies recommend the following limits for hcDNA levels: 10 ng per patient dose and DNA fragment sizes below 200 base pairs. These regulations are broadly applicable to vaccines, monoclonal antibodies, and other biologics, as well as gene therapy products. However, due to the inherent nature of hcDNA encapsidation in viral capsids, it is not always feasible to meet these

requirements. Regulatory agencies may, under specific circumstances, allow manufacturers to have alternative limits on the quantity and size of residual DNA impurities using a risk-based approach. However, manufacturers must ensure that any deviation from standard requirements is scientifically justified and does not compromise patient safety. Manufacturers of AAV gene therapy products must strive to meet these regulatory requirements and adhere to the ALARA (As Low As Reasonably Achievable) principle to minimize DNA impurities.

rAAV gene therapy products are produced using mammalian cell lines, particularly HEK 293 and Hela cells, which contain adenoviral genes such as E1A and E1B that were incorporated during the immortalization process. These adenoviral genes pose a potential patient safety risk, and their quantification is mandated by certain regulatory agencies.

The most likely encountered encapsidated DNA impurities in rAAV gene therapy products are host cell DNA and plasmid DNA. The presence of host cell DNA is influenced by the type of producer cells utilized in the manufacturing process, while plasmid DNA refers to residual DNA from the Rep/Cap, helper, and transfer plasmids.

Gene therapy products, such as those produced using HEK-293-based adeno-associated virus (AAV) may depend on significant amounts of plasmid DNA during manufacturing. These plasmids act as templates for the therapeutic gene of interest but must be eliminated from the final product due to safety concerns. Residual plasmid DNA can cause unintended immune responses, transfer antibiotic resistance genes, or integrate into a patient's genome. Additionally, it may result in false positives for critical safety assays, such as replication-competent AAV (rcAAV).

The final released Drug Product must be tested to provide assurance over identity and integrity of the intended therapeutic component, and to determine the levels of residual DNA impurities that may come from the production/host cells and/or plasmids and other helper components and could impact patient safety and/or product efficacy. Residual DNA impurities are a concern as they could have toxic effects such as genotoxicity and/or immunogenicity. For example, the *E1* gene in HEK293 cells, and *E6* and *E7* genes in Hela cells have oncogenic potential - both

these cell systems may be used to manufacture rAAV. It is the Sponsor's responsibility to assess any risk associated with the product and manufacturing platform.

In the context of assuring patient safety there are established guidelines (see table 1) for residual contaminating DNA for biologics which impose a limit on both the concentration of DNA per patient dose and the size of the contaminating DNA fragments. , however, no single analytical method may be sufficient to conclusively demonstrate a specific parameter. Therefore, a risk-based approach is recommended to identify and justify a combination of complementary methods that collectively provide the required level of assurance. Careful method selection and validation are essential to ensure reliability and regulatory acceptability.

Table 1 Existing guidance from organisations other than British Pharmacopoeia.

Organisation	Document Title
FDA-2006-D-0223	Characterisation and Qualification of Cell Substrates and Other Biological Materials Used in the Production of Viral Vaccines for Infectious Disease Indications
ICH Q2,	Validation of analytical procedures - Scientific guideline
WHO	WHO Technical Report Series, No 878, 1998. Annex 1. Requirements for the use of animal cells as *n vitro* substrates for the production of biologicals

4. Scope

In a previous British Pharmacopoeia document 'Characterisation of the Capsid Particle population in rAAV Products' comprehensive guidance was provided to assess and quantify all of the different rAAV species that can be generated in a rAAV therapeutic product, such as full, empty, partially filled/intermediate and overpackaged particles. The purpose of this guidance is to expand on the characterisation of the capsid particle population to assess three key product attributes: genome identity, vector integrity and contaminating DNA impurities.

This guidance has been split into sections according to methodologies that can be used to assess identity, integrity and impurities. Whilst most methods described can

be used to assess all these attributes, some may be more suitable for a particular attribute than others. Table 2 provides a high-level comparison of all methods and their associated capabilities whilst the main text provides a more comprehensive description of methods for readers who may not be familiar with the technologies described. Specific considerations for each method to assess the different attributes have also been described.

Efforts have been made to include the most commonly used methodologies to determine identity, integrity and impurities although other technologies for assessing the different attributes may be available and the user should select the most appropriate method for its intended purpose. However, the technologies covered in this guideline are not exhaustive.

Although intended to focus on encapsidated DNA, some DNA may be present in the final product that is outside the capsid or is associated with the capsid surface, even if the manufacturing process contains a nuclease step. In the context of quantifying impurities that could be present in the totality of the therapeutic product administered to the patient, from a patient safety perspective, it is essential to consider both encapsidated and non-encapsidated DNA. This aspect has been addressed in the relevant sections outlining impurity determination.

Whilst testing for identity, vector integrity and impurities is typically performed on purified product, it may also be helpful to assess impurities at different stages during the manufacturing process as this can support process development, process characterisation and process comparability if any manufacturing changes are introduced. Consideration has therefore been given to the suitability of each method to assess samples in different matrices and with varying levels of purity. Where an attribute is considered to be a critical quality attribute (CQA), it should be tested as part of release specifications with appropriate acceptance criteria. These tests should be validated in accordance with ICH Q2(R2). For tests that are not conducted as part of release specifications, while they do not need to be validated as per ICH Q2(R2), they do need to be demonstrated as fit for their intended purpose.

Determination of the vector genome titre has been discussed within a separate guidance document Determination of Genomic Titre and Vector Copy Number for rAAV and LV-RV Products and is outside the scope of this guidance. Assessment of other nucleic acid contaminants such as RNA, nucleic acid determination of microbial contaminants and aspects of DNA modifications such as methylation are also outside the scope of this guidance.

5. Encapsidated DNA Characterisation Methods

Table 2 Summary of methods for encapsidated DNA characterisation

	Next generation sequencing (Short Read)	Next generation sequencing (Long Read)	qPCR	dPCR (including ddPCR)	Alkaline denaturing gel
Integrity	Y	Y	Y – dPCR preferred	Y	Y
Identity	Y	Y	Y	Y	N
Impurity	Y	Y	Y	Y	Y (not quantitative)
Sample concentration/volume required for test	Low to medium	Medium	Low	Low	medium
Extraction	Required	Required	Required	No	Required
Time to result (length of analysis)	1 to 5 days	1 to 5 days	Same day	Same day	Same day
Expertise required	High Bioinformatics support	High Bioinformatics support	Medium	Medium	Low
Suitability as In-Process sample method	Challenge, due to long TAT	Challenge, due to long TAT	Yes	Yes	Yes

Complexity of equipment qualification in a GMP regulated environment	High	High	Low	Medium	Low – workflow considerations make QC challenging
Sensitivity of technology	High	High	High	High	low

222

223

5.1. Next generation sequencing

5.1.1. Method description

Next-Generation Sequencing (NGS) is a high-throughput method that allows for the rapid sequencing of large amounts of DNA or RNA, providing a comprehensive view of the genome or transcriptome. Usually, fragment lengths between 50 to 300 base pairs or shorter than 1 kilobases (in general) can be considered as short reads. The common feature of short-read technologies is massive sequencing of short, clonally amplified DNA molecules sequenced in parallel. The standard workflow includes nucleic acid extraction, fragmentation, NGS library preparation, sequencing and data analysis and Interpretation.

Compared to short-read sequencing, long-read sequencing, also called third-generation sequencing, can produces extended read lengths ranging from several kilobases to over 100 kilobases, enabling the resolution of complex genomic regions, structural variants, and full-length transcripts. The workflow resembles that of short reads, but utilizes longer fragments, alternative library preparation methods, different sequencing technologies, and specialized data analysis pipelines.

Next-Generation Sequencing (NGS) is a powerful tool for AAV vectors, providing detailed insights into DNA integrity and purity. NGS provides a comprehensive view of the AAV genome, identifying any variants or mutations that may be present. NGS also helps determine the genetic material encapsulated within the AAV capsids, ensuring the correct therapeutic gene is present, and identifies non-therapeutic capsids and potential contaminants, which is essential for the safety and efficacy of gene therapy products at the same time.

5.1.2. Short read advantages

Short-read sequencing technologies offer significant economic advantages, making them a reliable choice for many research and clinical applications, including AAV vector analysis. The lower cost per base allows laboratories to process many samples simultaneously, maximizing throughput and efficiency. This cost-effectiveness is particularly beneficial for AAV vector quality control, where multiple samples and high coverage are often required to ensure vector purity and integrity.

Short-read sequencing technologies have been refined over decades to achieve remarkable base-calling precision, consistently delivering accuracy rates exceeding 99.90%. This high fidelity is crucial for applications requiring precise variant detection, such as, targeted gene analysis, and AAV vector genome sequencing. The error profile is well-characterized and predominantly consists of substitution errors rather than insertions or deletions, making alignment and variant calling more straightforward. This exceptional accuracy enables researchers to confidently identify single nucleotide variants (SNVs) and small indels with minimal false positives, which is essential for verifying the integrity of AAV vector genomes and detecting any unintended mutations during vector production.

Short-read sequencing platforms have established robust regulatory frameworks that satisfy the stringent requirements of Title 21 CFR Part 11 of the Code of Federal Regulations. This compliance ensures proper electronic records management, audit trails, system validations, and electronic signatures necessary for regulated environments. The mature quality management systems associated with short-read technologies provide the necessary documentation and validation capability required for regulatory submissions in the rapidly evolving field of AAV-based gene therapies. Several long-read sequencers have full compliance with 21 CFR Part 11 and EU GMP Annex 11 regulatory requirements.

5.1.3. Long read advantages

Long-read sequencing technologies produce continuous DNA sequences spanning thousands to tens of thousands of base pairs, with some platforms capable of generating reads exceeding 100 kilobases. This extended read length represents a fundamental advantage over short-read approaches, allowing sequencing to span complex genomic elements in their entirety, including full-length AAV vector genomes without requiring prior fragmentation. By capturing long-range genetic information within single molecules, these technologies can provide a more comprehensive view of AAV vector integrity and direct visualization of full-length vs. truncated genomes. Additionally, long-read technology excels at contaminant identification, detecting and characterizing full-length contaminating sequences from host cells, helper viruses, or adventitious agents that might otherwise go unnoticed with fragmented short-read data, thereby enhancing the safety profile of AAV-based gene therapies

Long-read technologies excel at identifying structural variants (SVs) such as insertions, deletions, inversions, translocations, and copy number variations that often elude short-read detection. This capability is crucial for ensuring the integrity of AAV vector genomes, where unintended structural changes could impact therapeutic efficacy or safety. By spanning these variants with single continuous reads, long-read sequencing provides direct evidence of structural rearrangements without relying on statistical inference from paired-end mapping.

Long-read sequencing uniquely resolves AAV vector's most challenging elements—ITRs, promoters, and complex transgenes—by spanning these structures in their entirety without assembly. This capability provides critical benefits for rAAV characterisation accurate determination of palindromic ITR sequences essential for vector functionality, complete verification of repetitive regulatory elements affecting expression patterns, and comprehensive analysis of complex transgene structures that might harbour rearrangements.

Direct detection of nucleic acid modifications in unamplified DNA or RNA samples. While not typically a primary concern in AAV vector analysis, the ability of long-read platforms to detect DNA modifications directly during the sequencing process could have future applications in understanding the epigenetic aspects of AAV vector behaviour in host cells. This capability could provide insights into vector silencing mechanisms or long-term expression patterns in gene therapy applications.

5.1.4. Short read limits

Short-read sequencing technologies are fundamentally constrained by read lengths typically ranging from 50 to 300 bases, depending on the platform and chemistry used. This limitation creates significant challenges when attempting to resolve complex genomic regions that exceed these length boundaries, including full-length AAV vector genomes (typically around 4.7 kb). The fragmented nature of short-read data necessitates sophisticated computational approaches to reconstruct the original sequence, introducing potential errors and ambiguities in AAV vector genome assembly and characterization.

~~Unable to solve complex genomic regions, structural variations and large.~~ Short-read technologies struggle significantly with complex genomic architectures including segmental duplications, tandem repeats, and highly polymorphic loci. These challenging regions, which might be incorporated into certain AAV vector designs for regulatory or functional purposes, often remain inaccessible or misrepresented in short-read datasets. This limitation could lead to incomplete characterization of AAV vector genomes, potentially missing critical structural elements or variations that could impact vector performance or safety.

Packaged cellular gDNA components are not fully characterised for their potential for replication and propagation in treated cells as the limit to sequence size does not provide the full integrity of such elements. RNA is also not detected that could alter host cell response to infection and/or genotoxic events responding to this nucleic acid.

5.1.5. Long reads limits

Long-read sequencing remains more expensive per base than short-read methods, however this can be partially mitigated depending on target throughput, batching, and required depth. For AAV which has a small genome size, molecule-resolved analysis requires fewer reads, but large manufacturing QC analysis can still be cost-intensive. The reduced need for extensive bioinformatics processing and the comprehensive structural insights gained can justify the investment, particularly for critical AAV vector characterization tasks.

Accuracy: Historically, long-read sequencing technologies have faced challenges with base-calling accuracy, exhibiting higher error rates compared to short-read platforms. However, recent advancements have substantially narrowed this gap, with some platforms now achieving accuracy comparable to short-read technologies. PacBio avoids errors and is highly accurate and is, ideally the platform of choice.

- (**PacBio HiFi** — *Best when you need highly accurate, full-length per-molecule reads*: confirm full genome integrity (ITR→ITR), quantify truncated species, call point mutations in the payload, and resolve sequence heterogeneity across packaged genomes. Great for final QC of vector lots and for research that requires unambiguous SNV/indel calls across the whole genome. PacBio+1

- PacBio also enables RNA to be sequenced through generation of cDNA.
- High sensitivity is reached using CCC reads

Here is a short comparison to include PacBio if you wish

Criterion	PacBio (HiFi / CCS)	Illumina (short-read NGS)	Oxford Nanopore (ONT)
Read length	Long (HiFi reads routinely cover full AAV genome; 5–25 kb typical library inserts) — often **ITR→ITR** full reads possible. PacBio	Short (e.g., 75–300 bp). Cannot span full AAV genome in one read. Good for deep coverage of fragments. PMC	Very long (tens of kb to Mb possible) — can sequence *native* ssDNA and full intact genomes, plus host–virus chimeric reads for integration mapping. Oxford Nanopore Technologies+1
Single-base accuracy	**Very high** (HiFi consensus accuracy ≈ Q30–Q40; excellent for SNVs and indels in consensus). PacBio	**Very high** per-base (low raw error rate); gold standard for small variants and quantification. PMC	Lower raw accuracy than HiFi/Illumina historically; improving rapidly and can be polished—sufficient for structure, but SNV calling may require extra polishing/consensus. Oxford Nanopore Technologies
Ability to read ITRs / hairpins	**Good** — HiFi workflows report reads covering ITRs and GC/hairpin regions better than short reads (fewer dropouts). PacBio+1	**Poor** — short reads and PCR can fail across hairpins; ITRs often undercovered or misassembled. PMC	**Capable** — protocols exist to sequence native ssDNA AAV and recover ITRs, but library prep or hairpin stability can still cause biases. ONT can produce reads spanning ITR→ITR. Oxford Academic+1
Detecting truncations / rearrangements / full-length vs	**Excellent** — full-length HiFi reads let you directly count intact vs truncated	**Limited** — needs assembly or inference from read pairs; truncated species are	**Excellent** — long reads capture structural variants, concatemers, and full payloads; especially useful

truncated genomes	genomes and detect internal rearrangements. PacBio	inferred but not directly observed end-to-end. PMC	for complex rearrangements or integration events. Oxford Nanopore Technologies+1
Detecting AAV integration / host–viral junctions	Very good if library prep preserves chimeric molecules — HiFi read accuracy helps call junctions unambiguously. PacBio	Possible via mate-pair or specialized protocols (e.g., LAM-PCR + Illumina), but short reads complicate unambiguous mapping of large chimeric fragments. Cell+1	**Best** for single-read host–virus chimeras: ONT long reads frequently span the junction and flanking host sequence, enabling integration mapping. Oxford Nanopore Technologies+1
Epigenetic / base modification detection	No direct native methylation detection from DNA with standard HiFi (requires special prep).	Not applicable (Illumina loses native modification information because of PCR and short reads).	**Unique advantage:** ONT detects native base modifications (e.g., methylation) from current signal without bisulfite. Useful in special studies. Oxford Nanopore Technologies
Throughput & cost per Gb	Medium throughput; per-Gb cost higher than Illumina but lower relative cost for full-length small genomes when you need high accuracy per molecule. PacBio	High throughput, lowest cost per base — great for deep population profiling, barcodes, and QC assays. PMC	Flexible (scalable from MinION to PromethION). Per-Gb cost in between; good for rapid runs and small labs. Oxford Nanopore Technologies
Library prep complexity for AAV	Requires careful SMRTbell prep; can be optimized to preserve full genomes (HiFi	Straightforward (fragmentation, PCR, adapters). But PCR can bias ITRs and hairpins	Relatively simple transposase or ligation-based preps for ssDNA; ONT has published AAV workflows to sequence

protocol for AAV exists). PacBio	— special protocols often needed. PMC	native genomes. Oxford Academic+1

- **PacBio HiFi** — *Best when you need highly accurate, full-length per-molecule reads*: confirm full genome integrity (ITR→ITR), quantify truncated species, call point mutations in the payload, and resolve sequence heterogeneity across packaged genomes. Great for final QC of vector lots and for research that requires unambiguous SNV/indel calls across the whole genome. PacBio+1
- **Illumina (short reads)** — *Best for high-throughput quantitation, barcode/serotype screens, and deep population profiling*: very cost-effective to quantify variant frequencies, barcode counts, capsid libraries, or to perform high-depth mutation scanning — but **not** ideal alone for detecting full-length structure, ITR coverage, or integration sites without special protocols. PMC+1
- **Oxford Nanopore** — *Best for structural characterization and integration mapping, real-time rapid runs, and native modification detection*: ONT can sequence native ssDNA, read across large chimeric host–virus junctions, and detect base modifications. Error profile is higher than HiFi for SNVs, so combine polishing/consensus or orthogonal sequencing if you need base-level certainty. Oxford Nanopore Technologies+1

Conclusion if required:

- If you want **accurate, per-molecule full-length AAV genomes and reliable SNV calls** → **PacBio HiFi** is the safest single choice. PacBio
- If you need **cheap, deep quantification** (barcodes, variant frequencies) → **Illumina**. PMC
- If you need **structural context, integration mapping, or native modifications**, or want a fast flexible run → **Oxford Nanopore**, ideally combined with another platform for polishing SNV calls. Oxford Nanopore Technologies+1

Long-read sequencing generally produces fewer reads compared to short-read technologies. Long-read sequencing platforms typically generate substantially fewer reads per run compared to short-read technologies. This lower throughput can impact

applications requiring deep coverage across large genomes or detection of rare variants present at low allelic fractions, which might be relevant for identifying low-frequency mutations or heterogeneity in AAV vector populations.

DNA Quality Requirements: Long-read sequencing needs more DNA of higher quality. Long-read sequencing technologies generally demand higher molecular weight DNA input with minimal fragmentation compared to short-read platforms. This requirement for high-quality, intact DNA can present challenges when working with certain AAV vector samples, particularly if the extraction or handling processes lead to DNA fragmentation.

Both long-read and short-read methods of sequencing have advantages and disadvantages (Table 3). Therefore, it is important to determine what platform is fit-for-purpose, balancing accuracy, structural resolution, throughput, sample requirements, and cost. In many applications, a hybrid strategy that integrates short-read accuracy with long-read structural context yields the most comprehensive result.

Table 3 Long-Read Sequencing vs. Short-Read Sequencing

	Short reads	Zero-mode waveguides sequencing	Nanopore sequencing
Library Sample Requirement	~1–100 ng	1000 ng	200-1000 ng
Read Length (bp) up to	50-300 bp	1–25 kb	50 bp–100 kb
Accuracy	>99.90%	>99.95%	90.00 up to 99.00%
Variant calling — SNVs/ indels	Yes	Yes	Yes
Variant calling — SVs, repetitive regions, and complex regions	Possible with advanced bioinformatics algorithm	Yes	Yes
21CFR11 compliance	Yes	No	Yes

Direct detection of nucleic acid modification	No	Yes	Yes
Direct RNA sequencing	No	No	Yes
Data analysis	Post-sequencing processing	Post-sequencing processing	Real-time analysis and post-sequencing processing

5.1.6. Sample preparation library pre/seq

Before performing NGS characterization of encapsidated DNA, two critical steps must be carefully considered: sample preparation and library preparation. Both processes are integral to obtaining high-quality sequencing data, but they present distinct challenges. Sample preparation encompasses the extraction and purification of DNA from the AAV capsid, while library preparation is more restrictive, often dependent on the specific system employed for analysis. Although library preparation itself poses challenges, the quality control measures implemented by the analyst are crucial to ensuring the acquisition of reliable NGS data.

5.1.7. Enzymatic treatment

Prior to DNA extraction from the AAV capsid, it is essential to reduce as much non-encapsidated DNA as possible in the sample. This is achieved through treatment with endonuclease and exonuclease enzymes, which selectively degrade the unwanted DNA. The efficiency of the enzyme treatment should be validated by spiking the sample with DNA of known concentration, followed by comparison of treated and untreated samples to confirm effective clearance of free-floating DNA. The enzyme treatment must be both robust and efficient, ensuring sufficient reduction of contaminating DNA without compromising the integrity of the encapsidated DNA. Furthermore, it is crucial to ensure that DNase activity is completely deactivated, as any residual activity could affect the quality of the sample. Care must also be taken to avoid disrupting the AAV capsid during enzymatic treatment, as this could significantly alter the characteristics of the DNA payload. This could be assessed by spiking in a known DNA concentration once the DNase has been inhibited prior to

denaturation. It is important to inhibit DNase prior to denaturation as different capsid serotypes have different melting temperature, which could cause release of the viral payload before the DNase has been completely denatured. A DNA quantification method could then be used to assess the DNase treatment.

5.1.8. Capsid disruption & DNA extraction

To release the encapsidated DNA, the AAV capsid must be disrupted. This can be accomplished through a variety of methods, including enzymatic digestion using a serine proteinase such as Proteinase K or chemical disruption with an anionic detergent like sodium dodecyl sulfate (SDS). Heat or a combination of these methods may also be employed. The chosen disruption method should be optimized to minimize any negative impact on the integrity of the encapsidated DNA. Once the capsid has been disrupted, DNA purification methods, such as bead-based purification or spin column chromatography, can be utilized. It is important to evaluate the impact of the chosen purification technique on both the quality and size of the extracted DNA. After purification, the DNA should undergo quality assessment, including measurements of purity via UV absorbance, quantification using fluorometry or qPCR, and size determination via gel electrophoresis or capillary electrophoresis

5.1.9. Second-strand formation

The DNA extracted from the AAV is typically single-stranded, yet a double-stranded structure is required for efficient library preparation. To achieve this, it is necessary to generate complementary strands of DNA. This can be accomplished by inducing self-hybridization of the sense and antisense strands through high-temperature incubation followed by gradual cooling. Alternatively, the DNA can be converted into double-stranded DNA (dsDNA) through the use of random hexamers and a polymerase-I enzyme. The method of second-strand formation should be optimized on a product-by-product basis, as the structure and sequence of the payload DNA can influence the efficiency of self-annealing or the generation of balanced sense and antisense DNA. For self-complementary AAVs, which are designed to be manufactured as dsDNA constructs, second-strand synthesis may not be necessary, as the DNA is already present in a double-stranded configuration.

Overall, each of these steps, enzymatic treatment, capsid disruption, DNA purification, and second-strand formation, requires careful consideration to ensure the highest quality DNA is obtained for downstream NGS analysis.

5.1.10. Library Preparation

The goal of library preparation is to convert the extracted DNA into a form that can be efficiently processed by the NGS platform of choice. The library process involves several key steps, including DNA fragmentation (short read), end repair and adapter ligation. The choice of methods and reagents used during library preparation can greatly impact the accuracy and completeness of the sequencing results. Suitable library preparation ensures that DNA fragments are of appropriate size, the necessary adapters are attached for sequencing, and any biases are minimized. Quality control throughout the process is essential to verify the integrity of the DNA and ensure that the prepared library is suitable for downstream analysis.

5.1.11. Short Read Workflow

For short read library preparation the extracted DNA is required to be fragmented, traditionally this was done via mechanical shearing. More commonly DNA can be sheared enzymatically using either an enzyme the digests non-specifically or Tn5 transpose tagmentation which cut the DNA and attaches adapters at the site of the cut. Although Tn5 transposase exhibits some insertion bias this would have minimal impact on characterising the encapsidated DNA due to its small size. It is important to understand the sheering efficiency of either method to ensure consistency between library preparations.

If the DNA fragmentation was carried via mechanical shearing or non-specific enzyme sheering the fragmented DNA generated from this process typically have overhangs that require end repair, phosphorylation and A tailing. Once the ends are repaired sequencing adapters are added. If Tn5 transpose fragmentation is utilised both fragmentation and adapter ligation occurs simultaneously significantly reducing library preparation time. Barcoding can also be implemented to allow sequencing of multiple samples.

5.1.12 Long Read Workflow

Long-read sequencing involves distinct library preparation processes optimized for larger DNA fragments. Unlike short read sequencing fragmentation is not required and typically sample and library preparation methods are employed to ensure the extracted DNA does not become sheered during DNA extraction and library preparation. Once the DNA is extracted, it undergoes end repair to create blunt ends, making it suitable for adapter ligation. The next key step is the ligation of sequencing adapters to the DNA fragments, which are required for the sequencing process. These adapters allow the DNA to bind to the sequencing platform and initiate the sequencing process.

5.1.13 Quality Control

Quality control of NGS libraries is essential to ensure the quality and reliability of sequencing data. To assess library size, one common method is agarose gel electrophoresis, which separates DNA fragments based on size, allowing for visual confirmation of the expected fragment distribution. Capillary electrophoresis technologies can also be used to perform high-throughput analysis of library size, providing detailed size profiles and ensuring that the library contains fragments within the optimal size range for sequencing.

To evaluate library purity, UV spectrophotometry is commonly used to measure absorbance at 260 nm and 280 nm, providing an estimate of nucleic acid concentration and purity by calculating the A260/A280 ratio. A ratio around 1.8–2.0 indicates high purity, with minimal protein contamination. For more accurate purity assessments, fluorometric quantification methods (e.g., Qubit) offer a more specific measurement, distinguishing between DNA and contaminants. To measure library concentration, both UV-based methods and dPCR/qPCR can be utilized, with dPCR/qPCR providing a more accurate estimate by amplifying specific regions of the library. These QC techniques ensure that the library is of the appropriate size, pure, and concentrated for successful sequencing.

5.1.14 Data acquisition and analysis

Rigorous quality control (QC) of raw sequencing data is essential to assess read quality, GC content, adapter contamination, and duplication levels prior to alignment. Reads are then mapped to a reference genome using aligners. The choice of mapper and the accuracy of alignment depend on genome complexity and sequencing quality.

Consensus sequence generation follows, correcting random sequencing errors through read alignment and variant calling. Applying minimum coverage thresholds and base quality filters ensures a reliable consensus. Custom scripts may be developed to assess long-read size distributions. During pipeline development, comparative evaluation of tools is recommended to maximize accuracy and ensure methodological compliance.

5.1.15 Application of next generation sequencing for assessing attributes of encapsidated DNA of rAAV products

System selection and suitability Short-read and long-sequence platforms each have their own pros and cons. The table lists common recommendations for application usage. However, the best practice is to combine measurements from different technologies to ensure the accuracy of the measurement. Those orthogonal technologies for NGS are not limited to those listed in the guidance. Orthogonal biophysical methods, such as AUC and CDMS, provide valuable information on the size of the encapsidated genome for verification of large structure recombination, by outputting the partial/full ratio of intact capsids.

Table 4 Application of NGS for assessing attributes

Attribute	Description	Short reads	Long reads
Genomic integrity	Detect and quantify variants in the vector genome that could impact function of expressed protein	QC lot release and characterisation	Identification, semi-quantitation, and vector subtype (full or partial)
Impurity / safety	Quantify residual DNAs, such as plasmid DNA, host	QC lot release and characterisation	Characterisation

	cell DNA; SV40-T, E1A		
Identity	Identity of rAAV	QC lot release	QC lot release

5.1.16 Development and validation considerations

NGS-based methods for characterizing encapsidated DNA are complex, multi-step procedures. The performance of NGS-based methods is impacted by various steps in the experimental procedures. Therefore, comprehensive and thoughtful method development is essential. For instance, the protocol used for nucleic acid extraction, sample treatment (e.g., DNase digestion), second strand synthesis, NGS library preparation, sequencing platform selection, sequencing protocol, bioinformatics pipeline used for filtering low-quality reads, assembling reads, aligning reads to the reference sequence, and variant calling algorithms. A set of known standards, such as spike-in, is recommended for use as a control for NGS lab procedure and bioinformatics pipeline development to ensure unbiased sample preparation, data processing and analysis.

To ensure the expected performance of NGS-based analytical methods during routine method execution, a predefined system suitability criteria should be established during method development for each part of the method, including nucleic acid extraction, library preparation, sequencing, and bioinformatics analysis. The objective of validating an analytical procedure is to demonstrate that the procedure is suitable for the intended purpose.

Method validation of NGS-based methods should follow international guidelines such as ICH Q2 and ICH Q14. The analytic procedure performance characteristics to be demonstrated in the method validation study depend on the intended use of the NGS method. For instance, the NGS-based genomic integrity to output percentage of partial/mutated AAV is a quantitative measurement. Hence, the method should demonstrate sufficient specificity, range, accuracy, and precision. Moreover, the application of NGS as an impurity can be a limit measurement, where specificity and lower range limit verification are required. When the method is used for identity testing, specificity is required.

To validate an NGS-based test method, several necessary elements should be considered, including trained personnel, well-established SOPs, and record forms. All equipment must be qualified, and the bioinformatics pipeline should be validated. Suitable controls should be implemented to ensure data integrity. Additionally, assay controls and materials used for method validation should be well-characterized and qualified.

5.2. qPCR

5.2.1 Method description

PCR is a quantitative method that can be used to determine vector genome identity, integrity and to assess DNA-related impurities. The Determination of Genomic Titer and Vector Copy Number for rAAV and LV-RV Products provides a comprehensive overview of considerations for developing, performing and validating both qPCR and dPCR methods for the determination of vector genome titer. Many of these assay considerations also apply to the use of PCR to assess vector genome identity, integrity and impurities so to avoid duplication the reader is encouraged to refer to the previous guidance for a more in-depth description of PCR. Only considerations of PCR specific to assessing identity, integrity and impurities where these differ to the requirements for assessing genomic titer have been discussed in the sections below.

qPCR can also be utilized for the quantification of process- and product-related DNA impurities in gene therapy products, including residual host cell DNA and plasmid / other helper DNA components. It is also employed in residual DNA fragment size analysis, which is a critical quality attribute for patient safety.

5.2.2. Advantages

qPCR delivers results quickly, often within a few hours, It provides sensitive and quantitative measurements of the target DNA which is valuable for assessing and quantitating DNA impurities. Additionally, qPCR can detect and quantify multiple targets simultaneously in a single reaction, facilitating multiplexing via multiple fluorescent dyes and comprehensive analyses which can be beneficial when assessing vector integrity as described in Section 5.3.7 below. As a well-established

method, it benefits from numerous standardised protocols and commercially available kits for some of the more common process-related impurities, making it easy to adopt in many labs. Moreover, qPCR is relatively cost-effective which makes it advantageous for assessing many samples such as in-process samples to assess impurity clearance

5.2.3. Limitations

qPCR requires a standard curve for quantification, which can introduce variability and affect accuracy. qPCR can be sensitive to inhibitors or matrix interference in the test sample which can affect the efficiency of the reaction and lead to inaccurate results. When genome titre is used to confirm vector identity, qPCR will only target a section of the full transgene so will not provide complete sequence verification.

5.2.4. Sample preparation

Sample preparation for qPCR depends upon the intended purpose of qPCR assay, when qPCR is used to confirm the vector genome identity it may be possible to use the genomic titre assay without any changes to sample preparation as long as the genome titre assay is designed using primer /probes that target sequence specific to the vector genome such as the transgene / gene of interest (GOI) or bridges across the transgene and regulatory regions.

To confirm the identity of the vector genome by qPCR, the sample may require pretreatment before analysis. rAAV gene therapy may contain host cell DNA and plasmid DNA outside the capsid, despite the use of endonucleases during the manufacturing process. This extraneous DNA must be digested as part of the sample preparation. The sample undergoes DNase treatment to digest this non-encapsidated DNA. The DNase must then be deactivated to prevent any impact on the template vector DNA. To deactivate the DNase, a proteinase K (PK) treatment or other deactivation methods follows the DNase digestion. The PK treatment not only deactivates the DNase but also helps open the capsid, making the vector DNA available for amplification by PCR. Some highly concentrated products may need to be pre-diluted before DNase treatment and subjected to serial dilution after DNase and PK treatment (or alternative DNase deactivation method) for efficient PCR amplification.

5.2.5. Method set up and execution

As discussed above, when used for vector genome identity qPCR must utilise primers /probes that target sequence specific to the vector genome.

5.2.6. Data acquisition and analysis

For data acquisition and analysis please refer to the Determination of Genomic Titre and Vector Copy Number for rAAV and LV-RV Products guidance.

5.2.7. System selection and suitability

System selection and suitability can be found in the Determination of Genomic Titre and Vector Copy Number for rAAV and LV-RV Products guidance.

5.2.8. Method development, qualification, and validation

The ICH Q2 guideline on the validation of analytical procedures outlines the requirements for validating analytical assays based on their intended use. For an assay used to confirm identity, the only critical attribute that needs to be measured is the assay's specificity. For assessing other attributes validation requirements should be assessed according to ICH requirements.

5.2.9. Impurities

Residual DNA impurities can be process-related where DNA co-purifies with vector product but is not encapsidated, or product-related where unwanted DNA is encapsidated within the vector product. As the safety concern is associated with the total residual DNA administered to patients, when assessing DNA impurities, it is important to assess both encapsidated and non-encapsidated DNA in final product which can be done by testing samples both in the presence and absence of DNase treatment (former indicates how much of total residual DNA is encapsidated). This is the only deviation from the encapsidated DNA testing as outlined in the title.

5.2.10 Residual Host cell DNA

Residual host cell DNA (HCDNA) sometimes also referred to as residual DNA is usually quantified by qPCR or dPCR targeting a high copy gene in the human genome. The target can be a house keeping gene like 18S, a repeat sequence like

Alu, or other relevant housekeeping genes. There are commercial kits available in the market that are quite commonly used for this purpose. It is important to know that the target selection for the development of a PCR (qPCR/dPCR) based assay can be crucial in the outcome of the assay. For example, Alu elements are short interspersed nuclear elements (SINEs) present in over 1 million copies in the human genome. They are dispersed throughout the genome and are ideal for detecting fragmented DNA. Compared to 18S rRNA, Alu repeats offer superior sensitivity, fragment coverage, and specificity to human DNA. Their short amplicon size (~75 bp) also enhances assay performance on degraded or fragmented DNA.

qPCR-based assays rely on the standard curve providing results in ng/mL and widely acknowledged by the health authorities. The commercially available qPCR kits are widely used in the biopharma industry including the vaccines, monoclonal antibodies and more recently AAV gene therapy products.

5.2.11 Residual Plasmid DNA

It is recommended to use a sensitive and accurate method for quantifying residual plasmid DNA to ensure consistent removal of this process-related impurity. High-sensitivity PCR-based analytical tests are commonly employed for this purpose.

The most used target for developing assays to quantify residual plasmid DNA in gene therapy products is typically the antibiotic resistance genes incorporated within the plasmid backbone. In adeno-associated virus (AAV)-based gene therapy products, this is typically a Kanamycin-resistance gene which can be encoded by different alleles. It is therefore important to ensure the method used can quantitate the correct allele and that all plasmids contain the same antibiotic resistance cassette. There are commercially available assays designed for this purpose. A standard qPCR assay development and validation strategy, in accordance with ICH Q2guidelines, is applicable to these assays.

One of the significant challenges in developing identity and impurity assays for AAV gene therapy products is the lack of available reference standards. The United States Pharma Pharmacopeia (USP) provides solutions to accurately measure residual plasmid DNA in gene therapy products, offering a reference standard designed for

precise quantification. This standard is based on *Kanamycin* and *Ampicillin* resistance genes; it is suitable for both qPCR and digital PCR applications.

The residual plasmid DNA assays can also indicate the of level of reverse packaging, but these assays can also be designed from the other targets in the plasmid back bones like the bacterial origin of replication (Ori), however, such assays are not common.

5.2.12 Host cell DNA fragment size

In AAV gene therapy products produced using HEK 293 cells that harbour integrated adenovirus serotype 5 E1 genes which are known to have oncogenic potential fragment size is a key parameter. DNA fragments longer than 200 base pairs may encode functional genes and pose a risk of being oncogenic. Quantitative PCR (qPCR) assays used to assess fragment length are typically designed to target E1 gene regions, although alternative targets such as the 18S rRNA gene can also be employed. Commercial qPCR kits often utilize a three-amplicon design, comprising a short fragment (<100 bp), a medium fragment (~200 bp), and a long fragment (~400 bp). These assays use human genomic DNA to generate standard curves, and results are reported as the percentage of each fragment type. However, due to the need for standard curves and multiple amplicons, these qPCR assays are generally low throughput.

5.2.13. Genes of specific concern

Due to the oncogenic potential of adenovirus serotype 5 (Ad5) E1 genes, health authorities expect AAV gene therapy products to be thoroughly characterized for safety, including monitoring the presence of E1 sequences. While there is currently no formal regulatory requirement mandating this assay, authorities may request supporting data on E1 gene characterization and monitoring during regulatory submissions. Note that E1 genes are specific to HEK 293 cells. Additional genes of concern will be assessed based on the producer cell line used. For instance, if an insect cell line is employed, the evaluation will focus on genes relevant to that cell line.

As noted earlier, qPCR assays used for DNA fragment size analysis are often designed to target E1 gene regions and can typically fulfil both fragment sizing and E1

quantification purposes. However, if a different assay is used for fragment size analysis, a separate E1-specific assay may be required.

5.3. dPCR (including ddPCR)

5.3.1. Impurities

The encapsidation of DNA impurities alongside vector DNA in recombinant adeno-associated virus (rAAV) gene therapy products is a significant concern due to its potential impact on the safety and efficacy of the therapy products.

During the production of rAAV, fragmented DNA produced during cell apoptosis can be encapsidated. For example, in HEK293 cells, the amount of fragmented DNA increases significantly during production, leading to higher levels of encapsidated hcDNA.

5.3.2. Host cell DNA

The methodological considerations for quantifying encapsidated residual host cell DNA (hcDNA) have been previously addressed in the context of qPCR-based assays. With the gene therapy field progressively transitioning toward digital PCR (dPCR/ddPCR) technologies, there is increasing emphasis on the development and validation of dPCR/ddPCR assays for hcDNA quantification. In parallel, commercial vendors are beginning to introduce dedicated residual DNA quantification kits tailored for digital PCR platforms, reflecting the growing demand for higher precision and sensitivity in analytical workflows. An assay development and validation strategy like qPCR can be deployed for digital PCR, however, advance multiplexing and absolute quantification can be used for a better insight into host cell DNA in AAV gene therapy products.

5.3.3. Residual Plasmid DNA/ E1 DNA and host cell DNA fragment size analysis

Digital PCR based assays are becoming common for residual plasmid DNA, residual E1A/E1B and host cell DNA fragment analysis.

5.3.4. Genome integrity

Demonstrating the genome integrity of adeno-associated virus (AAV) particles is a a key requirement for ensuring the efficacy and safety of gene therapies. AAV particles with partial or truncated genomes may be ineffective or pose risks to patients by expressing truncated or aberrant proteins. Evaluating the genome integrity of AAV samples is essential for determining the appropriate dosage to be administered to patients.

This attribute can be assessed using multiplex digital PCR (dPCR), which utilises probe assays with distinct fluorophores that emit signals in different colour channels. These probes target various regions of the recombinant AAV genome. When all probe assays yield positive signals, it suggests that the particle likely contains a complete genome. The use of multiple probe assays in multiplex dPCR enhances the accuracy of estimating the presence of a full genome. In principal qPCR can also be used for the multiplexing assays, however, in the case of genome integrity assessment the reliability of qPCR on the standard curve while dealing with three or four different amplicons makes it challenging. Throughput, data analysis and linkage analysis can become extremely challenging if qPCR is used, hence, digital PCR is the method of choice for this type of assay.

- ***Assay design and dPCR/ddPCR multiplexing***

A triplex digital PCR assay is commonly employed to simultaneously detect three critical vector elements: the promoter, the gene of interest (GOI), and the polyadenylation signal (polyA). Primer and probe sets are specifically designed for each target region. Prior to multiplexing, individual assays must be validated to ensure specificity and efficiency.

Two multiplexing strategies can be utilized:

  - Fluorophore multiplexing, which involves selecting fluorophores with minimal spectral overlap (e.g., FAM, HEX, Cy5) to enable concurrent detection of multiple targets.
  - Amplitude multiplexing, which differentiates targets based on signal intensity.

- ***Sample preparation, partitioning and PCR amplification***

Genome integrity assessment is complex and need to be approached with great care. The tips, tricks and the best practices mentioned below will help in the effective sample preparation, partitioning and PCR amplification.

Sample Preparation

1. Optimize DNA Extraction

- Use gentle lysis methods to avoid shearing the AAV genome.
- Include DNase treatment to remove unpackaged DNA before extraction.
- Validate extraction efficiency using a spike-in control.

2. Fragmentation Strategy

- For full genome integrity assessment, avoid fragmentation.
- If assessing specific regions (e.g., ITRs, transgene, promoter), consider controlled enzymatic digestion to improve accessibility.

Partitioning (Digital PCR Setup)

1. Use Proper Master Mix

- Choose a mix optimized for low-template detection and high specificity.
- Avoid mixes with high background fluorescence.

2. Avoid Cross-Contamination:

- Use dedicated workspaces and filtered tips.
- Include no-template controls (NTCs) in every run.

3. Optimize Droplet or Chip/plate Loading and Multiplexing

- Ensure uniform partitioning by following manufacturer protocols precisely.
- Avoid bubbles or overloading, which can skew quantification.
- Use different fluorophores for different genome regions (e.g., promoter, transgene, polyA).
- Validate multiplex assays for cross-reactivity and efficiency.

PCR Amplification

1. Design Robust Primers and Probes:

- Target junctions (e.g., ITR-transgene, promoter-transgene) to assess genome integrity.
- Avoid regions prone to secondary structures (e.g., ITRs).

2. Optimize Annealing Temperature:
   - Use gradient PCR to find the optimal temperature for each assay.
3. Minimize PCR Inhibition:
   - Include internal amplification controls to detect inhibition.
   - Dilute samples if inhibition is suspected.
4. Replicate and Validate:
   - Run technical replicates to ensure reproducibility.
   - Validate with orthogonal methods (e.g., NGS, Southern blot) when possible.

- ***Data acquisition and linkage analysis***

Fluorescence data are acquired and analysed using the software provided with the dPCR system. Partitions are classified based on the presence or absence of each target. The proportion of triple-positive partitions (i.e., containing promoter, GOI, and polyA) serves as a proxy for intact vector genomes.

To correct for random co-localization of fragmented DNA, a Poisson-multinomial model is applied. Additionally, a 3D linkage analysis algorithm may be used to estimate the true physical linkage among the three targets, distinguishing genuine genome integrity from coincidental co-localization. Some dPCR platforms offer built-in analytical tools to estimate genome integrity directly.

- *Validation considerations*

Validation should follow ICH guidelines and include:

- Specificity and sensitivity testing using synthetic DNA fragments and spiked controls,
- Assessment of false-positive rates using fragmented DNA standards,

- Evaluation of assay precision, including inter- and intra-assay variability across multiple independent runs.

The multiplex ddPCR genome integrity assay can provide a powerful and scalable solution for the characterisation of gene therapy vectors. Its integration into development and manufacturing workflows could support regulatory compliance, improve product consistency, and ultimately contribute to safer and more effective therapies.

- *Best practices for vector genome integrity assay development*

The tips, tricks and best practices described below can be helpful in the development of an effective genome integrity assay.

Assay Performance & Linkage Requirements

- **Assays must be interference-free**: No steric hindrance; linked constructs should yield similar concentrations as individual fragments.
- **Use restriction enzyme controls**: Compare concentrations before and after digestion to validate linkage.
- **Optimize assay performance**: Aim for low %CV; some assays may not be suitable and require redesign.
- **Consider enzyme treatment**: For example, using enzymes to open ITRs may improve detection.
- **Pre-heating impact**: Assays should perform consistently with or without pre-heating for capsid lysis. Pre-PCR treatment can influence the measured integrity, in particular heating for capsid lysis. (*Optimizing ddPCR assay for characterizing AAV vector genome integrity, Cell & Gene Therapy Insights 2023; 9(6))*

Integrity Calculation and Normalisation

- Assessing linkage integrity requires more than counting double, triple, or higher-order positive partitions. A robust linkage algorithm is needed—one that accounts for random co-occupancy and applies Poisson correction.

These models rely on enough negative partitions, which can constrain the upper dynamic range. In practice, performance depends on the number of fragments analyzed and is part of assay validation.

- Commonly used approaches include those described by Regan et al. and Gleerup et al., though alternative models are available. (Regan 2015, Gleerup 2023)
- Normalisation to get to a percentage should be thought through. For AAV, the current consensus divides the number of integer or "linked fragments" by the sum of all the fragments. (Duong 2025)

5.4. Alkaline denaturing gel

5.4.1. Method description

Alkaline gel electrophoresis is a well-established method for separating and visualising single-stranded DNA (ssDNA) species based on their size. This technique involves the migration of negatively charged DNA molecules through an agarose matrix under the influence of an electrical field. The agarose gel act as a molecular sieve retarding the movement of larger DNA fragments more than the smaller ones, enabling size-based resolution. migrate faster than larger ones. The alkaline conditions are typically achieved with a buffer containing NaOH and EDTA to denature the DNA by disrupting the hydrogen bonding between complementary bases.

The denaturing conditions are particularly important when analysing adeno-associated virus (AAV) genomes, especially self-complementary AAV (scAAV) constructs. scAAV genomes are designed with inverted terminal repeats that enable the genome to fold back on itself, forming a double-stranded, self-annealed molecule without the need for second-strand synthesis in the host. However, during electrophoresis, this self-complementarity can result in secondary structures that complicate accurate size resolution and quantification.

While similar effects may be achieved using denaturing agents such as urea and formamide, they often compromise gel integrity and resolution, making alkaline conditions the preferred choice for this application.

Once completed, segregated samples may be visualised directly in the gel using intercalating dyes (e.g. SYBR Gold or Ethidium Bromide) or transferred to a membrane for further analysis using Southern-blotting or hybridization. This versatility makes alkaline gel electrophoresis a powerful tool for assessing genome integrity, identifying truncated forms, or confirming successful packaging of full-length ssAAV or scAAV genomes.

5.4.2. Advantages

The method enables direct visualization of native DNA species without the need for amplification or enzymatic processing. This eliminates potential biases commonly introduced by polymerase chain reaction (PCR) or restriction enzyme digestion.

It is particularly effective in differentiating between full-length genomes, truncated DNA species, and secondary structures such as self-complementary adeno-associated virus (scAAV) forms. As such, it serves as a critical tool for assessing the integrity and composition of AAV vectors during quality control procedures.

The technique requires only basic laboratory equipment and standard reagents, making it accessible and practical for use in facilities with limited resources.

Its procedural simplicity contributes to high robustness and reproducibility, supporting consistent and reliable comparisons across manufacturing batches and production runs.

5.4.3. Limitations

As stated above, the method is a direct detection technique, while this preserves the original structure it also necessitates the use of relatively larger quantities of input material to ensure adequate detection sensitivity. While principally a disadvantage, it

may not matter much since material to be analysed should be available in copious quantities.

One technical consideration is the elevated ionic conductivity of alkaline agarose gels compared to gels run under neutral pH conditions. The presence of hydroxide ions from the NaOH increases the current-carrying capacity leading to excessive heating during electrophoresis. This not only risks deformation of the gel matrix but can also negatively impact resolution and reproducibility. To minimise thermal effect, it is recommended to adjust the applied voltage gradient to 3.5V/cm or less, thereby reducing heat generation and promoting uniform migration of DNA fragments.

Additionally, the strong basic conditions can cause partial alkaline hydrolysis resulting in uneven migration patterns of ssDNA, especially particularly over extended run times or at elevated temperatures. To mitigate such effects, several preventive recommendations highlighted in the literature:

- Maintain the NaOH concentration in the running buffer at or below 50 mM, sufficient to maintain denaturing conditions while minimising hydrolysis.

- When preparing the gel, ensure that the molten agarose is cooled to ≤60°C prior to adding alkaline electrophoresis buffer to avoid thermal degradation.

- Prior to electrophoresis, allow the gel to cool to room temperature to avoid introducing additional heat when current is applied.

5.4.4. Sample preparation

Before performing alkaline gel electrophoresis, AAV DNA must be carefully extracted to ensure only encapsidated viral genomes are analysed. This first step involves enzymatic digestion of free-floating DNA using a nuclease such as DNAase I. Next, the viral capsid is disrupted to release the genome, either enzymatically with Proteinase K or chemically using a non-ionic detergent like SDS. Care must be taken during this step to avoid degradation of the AAV genome. The resulting lysate can then be purified using either ethanol precipitation, or a spin-column kit optimized for viral DNA.

It is the responsibility of the validating laboratory to determine the most appropriate capsid disruption and DNA purification method for their specific vector serotype, formulation, and downstream application. Care must be taken to avoid harsh conditions (e.g., prolonged heating, strong alkali, or high detergent concentrations) that could compromise the structural integrity or length distribution of the AAV genome, especially for self-complementary vectors.

Use a single-stranded DNA (ssDNA) ladder rather than a double-stranded DNA (dsDNA) ladder. This is critical because under alkaline conditions, dsDNA denatures and loses its native conformation, leading to unpredictable migration patterns and inaccurate sizing. ssDNA ladders maintain a consistent conformation and migrate more accurately based on length, making them suitable molecular weight markers under denaturing (alkaline) conditions.

Add the required volume of 6x alkaline loading buffer to the purified DNA and ladder. It is not necessary to denature the DNA with base or heat before electrophoresis, but it should be assessed during validation of the assay.

5.4.5. Method set up and execution

Weigh the appropriate amount of agarose powder and add it to distilled water in a heatproof container. The appropriate amount will be specific to the expected size of the AAV payload. Heat the agarose slurry until completely dissolved, swirling the container occasionally to ensure even mixing and full dissolution. If any water has evaporated during heating, top it back up to the original volume.

Once the agarose solution is clear and fully melted, cool it to about 55°C. Then add 0.1 volume of 10× alkaline electrophoresis buffer (e.g., 500 mM NaOH, 10 mM EDTA) and immediately pour the gel into a casting tray. Let the gel solidify completely. Place the gel into the electrophoresis tank and cover it with freshly prepared 1× alkaline running buffer.

Load the prepared DNA samples and ssDNA ladder into the wells. Run the gel at <3.5 V/cm to reduce heat generation. Once the tracking dye (e.g., bromocresol green) has migrated 0.5–1 cm, pause the run and place either a glass plate or heat sink on top of

the gel to help dissipate heat or perform the remaining electrophoresis in a cold room or fridge to minimise heat build-up. Resume electrophoresis until the dye reaches about two-thirds of the gel length.

Once the electrophoresis run is complete, carefully remove gel from the electrophoresis tank and submerge in a gel staining tank containing 1x TAE buffer to neutralise for 45 minutes.

Following electrophoresis, DNA bands may be visualised using one of two main approaches, depending on the analytical requirements and laboratory preferences. The first involves direct staining of the gel, this is done by soaking the gel in a solution of 0.5 µg/mL ethidium bromide in 0.5× TBE, alternatively a SYBR-based fluorescent dye can be used such as SYBR Gold.

Alternatively, the DNA can be transferred from the gel to a membrane via Southern blotting. This enables hybridization with labelled probes for sequence-specific detection of target nucleic acid fragments. The choice between direct staining and membrane transfer should be guided by the sensitivity, specificity, and downstream application needs of the participating laboratory. Southern blotting is particularly advantageous when precise identification or quantification of specific AAV genome fragments is required, whereas gel staining offers a faster and less complex method for general visualization and size estimation.

5.4.6. Data acquisition and analysis

DNA bands resolved on alkaline agarose gels or transferred to membranes can be visualised and quantitatively analysed using high-resolution digital imaging systems. These instruments offer multiple illumination modalities (e.g., UV, blue light, and white light) to accommodate a variety of DNA-binding dyes, including ethidium bromide, SYBR Gold, or related fluorescent stains. During image acquisition, the system typically performs automated focusing and exposure optimisation to ensure signal linearity and avoid saturation. However, in cases of extensive DNA smearing, low-abundance AAV genomes, or unexpected fragment sizes, manual adjustment of exposure parameters by the operator may be required to enhance detection sensitivity. Captured images, can then be processed using dedicated software for

background correction, lane and band identification, and densitometric analysis. This enables accurate assessment of genome size, integrity, and relative abundance.

Full-length AAV genomes, including both ssAAV and scAAV forms, are expected to appear as a distinct band at approximately 4.7 kilobases (kb). Although scAAV genomes are double-stranded in solution due to their inverted terminal repeat-mediated folding, the alkaline conditions denature the duplex, resulting in single-stranded forms that migrate equivalently to full-length ssDNA during electrophoresis.

Fragmented or truncated genomes are typically observed as smears or multiple lower-molecular-weight bands, indicating incomplete packaging, degradation, or vector instability. These can result from nuclease activity, mechanical shearing during purification, or inefficient vector assembly.

Empty capsids, which do not contain any DNA payload, yield minimal or no detectable DNA signal under these conditions and serve as a baseline reference when assessing genome loading.

Packaging efficiency can be estimated by performing densitometric analysis of the full-length genome band relative to the total DNA signal per lane or in comparison to known standards. This semi-quantitative approach enables evaluation of vector quality and consistency which can help distinguish between productive vector genomes and aberrant packaging products.

5.4.7. System selection and suitability

In addition to technical performance, selected instruments should offer sufficient throughput and compatibility with downstream data acquisition systems, such as high-resolution imaging platforms and densitometry software. Given the manual nature of gel preparation, electrophoresis setup, and sample loading, ease of operation is critical. These factors place increased importance on both the usability of the equipment and the technical proficiency of the operators, as small inconsistencies in handling can impact the reproducibility and resolution of results.

Safety considerations are also essential in equipment selection. Many DNA stains used in gel-based workflows, such as ethidium bromide and SYBR dyes, are

mutagenic or toxic due to their DNA intercalating properties. Therefore, systems should be designed to minimise user exposure. It is strongly recommended to ensure safe handling of hazardous reagents and minimise the risk of operator contamination.

5.4.8. Method development, qualification, and validation

Resolution The minimum resolution capability must be clearly defined to ensure that critical components such as full-length vector genomes and possible truncated fragments can be clearly separated and distinguished. The agarose concentration, buffer composition, voltage gradient, and run time must be optimised to achieve this resolution, particularly in the 2–6 kb size range where most AAV-related species are expected to migrate. Inclusion of ssDNA molecular weight markers of appropriate size intervals is essential for accurate sizing and verification of separation performance.

Repeatability As alkaline gel electrophoresis involves several manual steps the repeatability of the method is inherently influenced by the operator's technique. Therefore, training and repeated execution are critical to achieving consistent outcomes. The assays should include clearly defined acceptance criteria such as distinguishable separation of full-length vs. truncated genomes and, where possible, verification of results through quantitative imaging.

Reliability/Robustness Electrophoretic conditions are comparatively harsh at high pH causing chemical stress to the gel matrix, electrodes, and chamber materials. Such environments may lead to corrosion, warping, or brittleness of plastic and metal components over time. Routine pre-run inspection of electrophoresis equipment is essential to detect signs of wear or degradation. Components showing discoloration or cracking, should be replaced promptly to prevent run failure and ensure operator safety.

Device qualification The equipment/devices used should be qualified to demonstrate they are fir for purpose. This can involve using a DNA standard of known molecular weight and composition, such as ssDNA ladders ensuring separation occurs as expected. Additionally, a relevant control samples, such as previously characterised AAV control material can be used to compare separation against the ladder and your

sample. Image acquisition systems must be qualified for minimum sensitivity, dynamic range, and linearity of response to ensure quantitative comparability across experiments.

Method validation Method validation should confirm that the electrophoresis protocol consistently delivers reliable performance across key analytical parameters. This includes the ability to achieve baseline-resolved separation of closely migrating DNA species, such as fragments differing by less than 500 base pairs, as well as clearly defined detection limits for the minimum and maximum DNA input required for accurate visualisation and quantification. The method's reproducibility must be demonstrated across multiple operators and days, to ensure robustness under routine use. In addition, validation should assess matrix effects, including potential influences from buffer composition and sample impurities on DNA migration and band resolution.

6. Concluding remarks

This guideline concerns itself with the development and manufacture of recombinant adeno-associated virus (rAAV) medicinal products of defined efficacy at acceptable risk to patient safety.

The means to accomplish this include

- providing controlled rAAV medicinal products containing the intended payload, i.e. the gene of interest, its regulatory elements and required AAV genome remnants, i.e. inverted terminal repeats (ITRs)
- minimizing the presence of extraneous DNA, e.g. DNA derived from production cell lines/host cells, plasmids used, surplus copies of the payload or surplus elements of the intended payload
- Minimizing the presence of capsids filled only partially or not at all

The tools available to verify/test the above and associated parameters are described in some detail in this guidance. It provides a snapshot of the technology at the time of writing and is subject to changes in response to new insights and technologies.

The goal of this guideline is to support developers and manufacturers of rAAV based medicinal products. It needs to be stated clearly, however, that the responsibility of the quality of their products remains with developers/manufacturers/sponsors and that any decisions made in regard to the topics discussed is to be based on sound risk-based approaches.