PathoQuest

GMP Viral Vector and Plasmid Identity Testing with Next Generation Sequencing

Application Note

Key Takeaways

- It is a regulatory requirement to provide confirmation of the identity of the drug product. For viral vectors containing genetic material, identity testing should include a determination of the nucleic acid sequence.
- Next generation sequencing (NGS) is a vastly improved approach thanks to its much higher throughput, scalability, and speed when compared to traditional Sanger sequencing. Specifically, NGS provides deeper insights into genetic variations within the product, specifically rare and ultra-rare variants.
- Raw material plasmids should be tested with the same superior NGS technology as the final product to ensure consistent quality in manufacturing.

Summary

Identity testing is critical for ensuring quality and safety of biopharmaceuticals. As such, it is a key regulatory requirement for release testing. For viral vectors containing a genetic payload, the regulatory expectation is that identity is confirmed through the determination of the nucleotide sequence of the product. A next generation sequencing (NGS) approach for identity testing holds several advantages over traditional Sanger sequencing. NGS can be more cost effective for longer sequence reads over sequential Sanger sequencing runs. NGS can also provide a much larger and more detailed dataset on the overall sequence quality of the product, including the detection and relative quantification of ultra-low abundance genome variants in the vector product. It is also highly recommended that an NGS testing strategy is employed for any manufacturing plasmids, as their sequence accuracy and quality will have an impact on the output quality of the final viral vector product.

Introduction

When designing a testing strategy for the release of any batch of pharmaceutical product, the manufacturer must provide sufficient evidence the batch is safe, as well as being of sufficient quality and purity. For biological products, and more specifically products used for cell and gene therapies, the regulatory guidance focuses variously on identity, safety, quality, purity and potency as the critical quality attributes (CQAs) [1,2]. In this application note we will focus on the requirement for identity (often contracted to ID) testing in viral vector manufacturing and input plasmid QC.

Fundamentally, identity testing ensures that the final product is as described and distinguishable from similar products manufactured in the same facility. In the FDA Code of Federal Regulations Title 21, part 610 [3], which covers biological products (or biologics), identity may be established by:

the physical or chemical characteristics of the product, inspection by macroscopic or microscopic methods, specific cultural tests, or in vitro *or* in vivo *immunological tests.*

Extract from 21 CFR 610.14 [3]

For viral vectors such as lentivirus or AAV which carry a genetic payload, external physiochemical and immunological profiles may be very similar between different products. Therefore, identity testing must also examine the genetic component of the viral vector [1,2]. This has previously been achieved through approaches such as the PCR amplification of the region of interest, or through whole or targeted Sanger sequencing. However, with the increased adoption of NGS, both viral vector manufacturers and the regulators are finding that NGS provides them with a richer dataset giving both greater insights into manufacturing, as well as higher level of confidence in the final product. For example, NGS is successfully being used to determine the presence of genetic sub-populations, including ultra-rare genetic variants within manufactured batches of viral vector.

Overview of Sequencing

Next generation sequencing, sometimes referred to as Massively Parallel Sequencing (MPS) or High Throughput Sequencing (HTS, although this is not preferred due to potential confusion with high throughput screening); consists of a number of technology platforms which can provide a rich dataset of nucleotide sequences within a given sample. Generally, NGS platforms are classified as short read or long read technologies. Short read platforms (e.g. Illumina) provide sequence information of up to 350 bases per read. Long read platforms (e.g. Oxford Nanopore) can provide sequence information of up to several kilobases

per read. Due to the large amount of data generated on any NGS run, complex bioinformatics processing must be applied to assemble and present the data ready for interpretation (Figure 1). Short read NGS tends to be more sensitive over long read technology, making it ideal for applications such as adventitious agent detection, as well as the detection of low abundance genetic variants. However, as these shorter reads must be assembled into more useful consensus data, the bioinformatics can be more complex, especially if there is only a limited reference sequence.

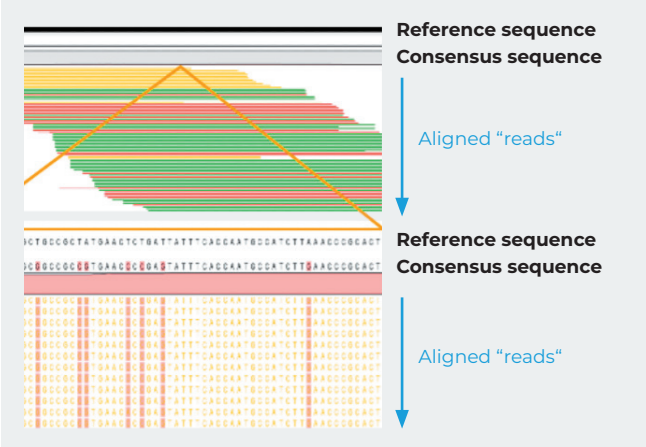

Figure 1 Representation of the alignment of a short read NGS sequencing run. Each sequencing "read" of up to 350 bases is aligned against a reference sequence to produce a consensus (or output) sequence. Upper panel A shows an expanded graphical alignment of reads. Lower panel B shows the individual base sequence alignment. Differences between the reference and consensus sequences, as well as the individual read differences, are highlighted in panel B.

Sanger sequencing has been used extensively in identity testing as well as broader characterization, and continues to be well accepted by the regulators. This approach uses a sequencing primer to generate a series of sequencing amplicons, typically of up to 1,000 bases in length. A consensus sequence is determined by capillary electrophoreses (CE) where the terminating base is detected by fluorescence (Figure 2). Samples greater than 1000 bases require additional primer design to provide coverage in separate sequencing reactions, multiplying cost and labor. In addition, sequences which are proximate to the primer and towards the end of the sequencing amplicon are often of poor quality. Therefore, to fully sequence a 1000 base fragment requires 2 or even 3 separate primer designs and Sanger sequencing reactions. As such, sequencing a full AAV or lentivirus genome (5kb or 10kb respectively) may require between 10 and 30 sequencing reactions to ensure full coverage. In addition, non canonical DNA secondary structures and complex nucleotide sequences can hinder primer extension. An example is the AAV virus inverted tandem repeat (ITR) region. The use of Sanger sequencing to decipher these motifs can be highly challenging, which is of specific concern as mutations are frequent, and these motifs play both structural and functional roles in these vectors and associated plasmids. [4].

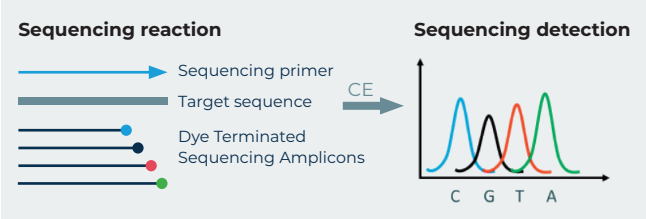

Figure 2 Illustration of Sanger sequencing. During the sequencing reaction (left panel), and oligonucleotide primer is extended against a target sequence in an environment containing fluorescent dye labelled terminator nucleotides, producing dye labelled sequencing amplicons of various lengths. These amplicons are sized using capillary electrophoresis (CE) to determine accurate base length of each amplicon (right panel). The consensus sequence is derived from the fluorescent signal for each amplicon.

PathoQuest leverages short read NGS platforms across GMP validated identity testing offering, expanding upon our GMP validated adventitious agent testing offering for viral vector testing. Figure 3 gives an overview of a short read sequencing workflow. Where input is DNA, material is fragmented into 200-500 bp size fragments. For RNA inputs, material would be reverse transcribed to DNA prior to fragmentation. The fragments are then ligated on either end with sequencing adaptors prior to sequencing on the flow cell. Sequencing occurs from either end of the fragment. Therefore, as long as the input DNA is sufficiently fragmented, a full read of each fragment will be made despite a 350 base sequencing limit.

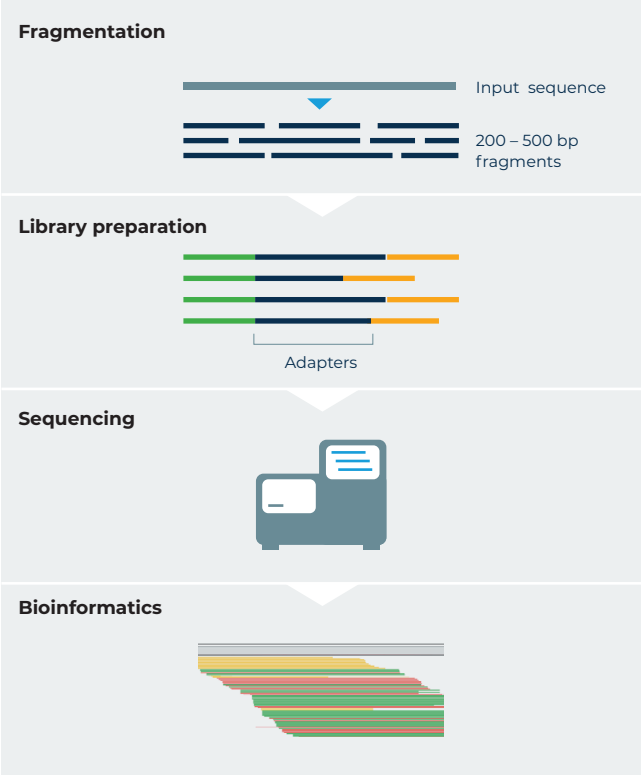

Figure 3 Illustration of a short read NGS sequencing workflow. Input sequence of DNA or reverse transcribed RNA is fragmented prior to sequencing library preparation and adapter ligation. Sequencing is undertaken on the instrument flow cell after which the output data is processed within the bioinformatics workflow.

Advantages of NGS over Sanger

Where Sanger has historically been well accepted by the regulators for identity testing, NGS is now becoming preferred due to the advantages that this technology has in providing full and comprehensive sequencing information of viral vectors as well as manufacturing plasmids:

Cost Sequencing a full AAV or lentivirus genome (5kb or 10kb respectively) may require between 10 and 30 Sanger sequencing reactions. Whereas individual runs of Sanger are often much more economical than a run of NGS, the cost of running multiple Sanger reactions in parallel can soon become more expensive. This is especially true when considering the effort required to design and validate multiple sets of primers.

Time For an individual or established sequencing run, Sanger can be much faster (typically 1-2 weeks for a GMP service). However, if the sample has not previously been sequenced, a substantial amount of time is often required in the validation of the sequencing run.

Complexity NGS can more easily sequence more complex secondary structures and highly repetitive regions, such as GC rich and poly-A motifs, and typical viral sequences like AAV ITR or lentivirus LTR, as well as some promoter and regulatory regions. Primer extension in Sanger sequencing can be significantly hindered by such areas.

Variant detection This is important when considering the quality not only of the finished product, but also of the plasmid raw materials. Sanger sequencing is by its nature a consensus method. Variations, or sub-populations in sequence can be identified, however this only becomes reliable where the variant represents >20% of the population. Even at this level, interpretation of the sequencing electropherograms can be highly subjective. In comparison, NGS can in theory detect single variants. In practice, confirmation of the variant is desirable with multiple reads. As such, variant detection is practical below 1% of the population. It is worth noting that where is it possible to estimate the relative abundance of genetic variation within a sample, NGS should be considered as a semi-quantitative rather than a fully quantitative method, since quantification depends on the depth of sequence coverage as well as the quality of the target. Release criteria should therefore reflect this when being set.

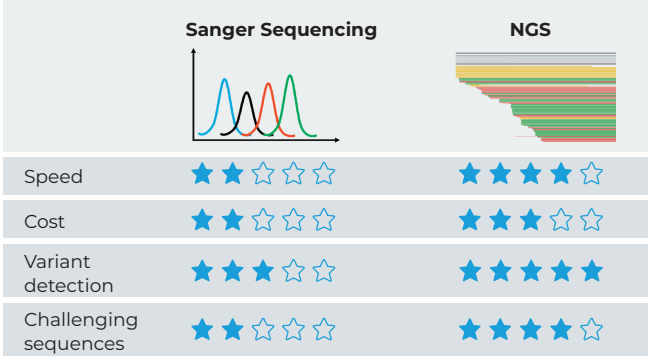

Figure 4 Summary of relative advantages of Sanger and NGS based sequencing for identity testing.

Viral Vector Identity Testing

As discussed above, a sequencing approach of the viral vector product is often expected as part of the regulatory submission to confirm identity as well as ensure batch to batch manufacturing consistency. Where Sanger sequencing can be used to meet this release criteria, an NGS approach for identity testing of viral vectors has a number of advantages.

Cost can be a compelling argument, however it is often a secondary consideration when choosing a testing strategy. That NGS is frequently more cost effective is certainly a positive aspect in its favor, given that for full coverage a Sanger sequencing run may require a relatively high number of parallel runs. It has also been discussed that Sanger sequencing can have challenges when sequencing through areas such as the AAV ITR or lentivirus LTR regions. Due to the mechanics and chemistry of NGS, such challenging sequences are much less problematic.

However, perhaps the most compelling reason for the use of NGS for viral vector identity testing is that it gives a much deeper insight into the quality of the product. Sanger sequencing is a consensus method, returning a result which is an average of the overall sequence population present. As NGS provides information on every sequence read, very low occurrence changes can be identified. This insight can be of particular importance to assess the quality of the vector batch prior to costly downstream processing. Due to the encapsulated nature of a viral vector, it is not possible to remove any changes in the viral genome in downstream processing. Indeed, even purifying out totally empty AAV particles can prove challenging [5].

Sequence variants can occur as a result of the following:

Spontaneous variants or such as mutations or recombination can arise within the viral genome transcription process, and may be more or less prevalent depending on a number of factors within the vector design and the manufacturing conditions.

Variation carried over from the **manufacturing plasmids** is likely the greatest source of any genetic variation seen in the final product. If NGS is used as a measure of quality of the final vector product, it should therefore be used to validate the quality of the input plasmid materials.

Incomplete viral genome can occur during vector manufacture if transcription is prematurely terminated, or if the genome transcript is particularly labile. Manufacturing conditions may have an impact, as conditions can put additional stress on the production cells. Moreover, sequence design may also impact the occurrence of incomplete viral genome, for example the presence of weak transcription termination signals upstream of the desired termination point. NGS can be a useful tool in the assessment of incomplete viral genome. However, as this is a measurement of the relative absence of sequences, NGS can be less sensitive than with the detection of genetic changes. It is also worth noting that for the assessment of empty:full ratios for AAV vectors, direct NGS sequencing is not particularly suitable and other assays should be considered.

For the reasons outlined above, NGS should be the first consideration when designing an identity testing strategy for viral vectors.

Plasmid Identity Testing

As part of any GMP manufacturing process, all raw materials must be adequately controlled and tested according to the risks that they present. Certain raw materials such as the growth media, the manufacturing cell line, or plasmids directly involved in manufacturing would be considered as critical raw materials. Thus, these critical raw materials are required to have appropriate supply mitigation and testing procedures.

Most current viral vector manufacturing techniques use a transient transfection approach. This is where viral components are expressed from plasmids within the manufacturing cell line prior to being assembled for subsequent downstream purification.

As the plasmid does not form part of the final product, a formal requirement for identity confirmation testing is not stated within the regulations. However, as a critical raw material, a suitable testing strategy should be applied to the plasmid which mirrors the identity testing requirement for viral vectors.

The application of NGS is particularly well suited to the QC of plasmids to:

- Ensure that errors are not present, or have spontaneously been incorporated into the plasmid during the production process. For example, AAV ITRs are known to be prone to integrate errors during amplification [4].
- Validate the clonality of the plasmid population from the start of manufacturing, and to ensure there is no cross-contamination.

It may be decided during the definition of the manufacturing process that a small sub-population of plasmid variants are acceptable, e.g. ≤5%. However, understanding the distribution of any these variants may be critical to product quality. For example, a variant may occur in a region outside of the specific coding regions of the plasmid, and be judged to have no impact on the final product. Alternatively, a variant may introduce a frame-shift within a coding sequence or sequence motif which would have significant impact on product quality. As discussed above, even at relatively high levels Sanger sequencing would be insensitive to detection of either of these cases.

Identity Testing at PathoQuest

PathoQuest is a leading expert in the provision of NGS-based GMP testing services for biopharmaceuticals. We offer a fully validated GMP service for the identity testing of viral vectors such as AAV, lentivirus, as well as non-viral vector applications such as CRISPR, used in cell and gene therapy as well as vaccine applications.

Identity testing by NGS is undertaken on short read sequencing platforms, which provides excellent detection capabilities for very low occurrence variance detection. Due to validation constraints, the GMP service has been validated at 5% variant sequence abundance. However, detection and reporting of variants below 1% is possible due to the high abundance of reads, although relative quantification can become much less accurate. Overall, the full workflow is validated to GMP, from sample preparation, through sequencing and bioinformatics. The output is a GMP Certificate of Analysis which is supported by the consensus sequence and any variant sequences >5% detected.

Sample requirements	Shipment & storage	Standard turnaround time	Fasttrack turnaround time	Sensitivity**	Output
Min sample vol of 200µl at titres of $1x10^{13}$ for AAV, $1x10^{10}$ for lentivirus, $1x10^{8}$ for adenovirus. Minimum 1ng at 0.5ng/µl for plasmids. Backup sample required BSL1 or 2*	Dry Ice / -80°C	28 days	14 days	Detection of variants validated at 5% abundance (Lower occurrence variants can be reported if required)	GMP CofA Consensus sequence Variant sequences at > 5% abundance

*Biosafety level classifications can vary between regulatory authorities – contact PathoQuest to discuss.
**NGS should be considered as semi-quantitative in this application. Abundance figures are provided at the occurrence rate of the variant as it appears within the read data. Detection of variants is possible below the validated 5% level. Please discuss with our experts if this is required.

References

1. Guideline on the quality, non-clinical and clinical aspects of gene therapy medicinal products (EMA), EMA/CAT/80183/2014, Published March 2018
2. Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs) - Guidance for Industry (FDA) CBER, Published January 2020
3. Code of Federal Regulations, Title 21 Volume 7. Part 610 – General Biological Products Standards (FDA) – accessed November 2022 at https://www.ecfr.gov/
4. Wilmott, P., Lisowski, L., Alexander, I. E., & Logan, G. J. (2019). A user's guide to the inverted terminal repeats of adeno-associated virus. Human Gene Therapy Methods, 30(6), 206-213.
5. Srivastava, Arvind, et al. "Manufacturing challenges and rational formulation development for AAV viral vectors." Journal of Pharmaceutical Sciences 110.7 (2021): 2609-2624.

Find out more: **www.pathoquest.com/index.php/ID**
or contact us via contact@pathoquest.com

BioPark – BatB 11 Rue Watt 75013, Paris

466 Devon Park Dr, Wayne, PA 19087, USA