European Federation of Pharmaceutical Industries and Associations

2.6.41. HIGH-THROUGHPUT SEQUENCING FOR THE DETECTION OF VIRAL EXTRANEOUS AGENTS

Reference: e.g. Pharmeuropa 36.2 - PA/PH/Exp. 15/T (21) 27 ANP

● Date: June 2024 ● Version: Draft

The European Federation of Pharmaceutical Industries and Associations (EFPIA) represents the research-based pharmaceutical industry operating in Europe. Through its direct membership of 36 national associations, 39 leading pharmaceutical companies and a growing number of small and medium-sized enterprises (SMEs), EFPIA's mission is to create a collaborative environment that enables our members to innovate, discover, develop and deliver new therapies and vaccines for people across Europe, as well as contribute to the European economy.

Sent to the attention of: EDQM Secretariat EPD@edqm.eu

General Comments

EFPIA welcomes the publication of a General Chapter on the detection of extraneous agents by Next Generation Sequencing. It is expected that this General Chapter will be used as a reference, and will facilitate the establishment of harmonised expectations for NGS-based method development and validation.

EFPIA appreciates the effort that has gone into creating the initial draft. The text prepared by EDQM was carefully reviewed and found informative and highly relevant. EFPIA would like to suggest some clarifications and improvements to further enhance its accuracy and clarity. This will ensure that the document meets the highest standards and effectively communicates the necessary information. Those suggestions are made taking into consideration discussions among industry representatives and the final position adopted in our Position Paper.

EFPIA looks forward to collaborating with EDQM on the final version of the General Chapter.

Background: The EFPIA Supportive Group Clonality, Characterisation and Viral Safety of Cell Lines has recently published a Position Paper entitled "Considerations for Validation and Implementation of Next Generation Sequencing for Adventitious Virus Detection for Biological Medicinal Products" https://www.efpia.eu/media/t22f5yoz/efpia-ngs-virus-detection-paper_finaljun2024.pdf

Envisaged as a practical guide, the paper addresses questions related to analytical method validation and corresponding pre-requisites originating from method development. EFPIA's position was publicly shared with the purpose to facilitate NGS use, stimulate discussion on challenging technical issues, and foster global alignment on expectations related to this innovative technology.

Terminology: we recommend using the term Next Generation Sequencing instead of High Throughput Sequencing throughout the text.

The terminology used to describe the technology varies across the field, leading to a lack of consensus in naming conventions. We recommend employing the term "Next Generation Sequencing* instead of "High Throughput Sequencing" to promote a standardized language within the international community. Notably, the term "Next Generation Sequencing" was ultimately adopted in the final version of ICH Q5A(R2), despite the initial consideration of both terms.

Additionally, we have included a summary of a survey conducted during the drafting of ICH Q5A(R2), reflecting the perspective of EFPIA members on this terminology aspect (refer to section *'HTS vs NGS' EFPIA survey : which preferred abbreviation for this technology* ?).

Diversity of technical approaches and workflows: we recommend keeping the description of technical approaches flexible and open to alternative approaches.

NGS-based methods for the detection of extraneous agents encompass a wide range of technical approaches for each step. Given the evolving nature of this technology, it is foreseeable that new technical approaches, not yet identified as potential options, may emerge in the future.

EFPIA appreciates the comprehensive efforts made in the General Chapter to distinguish various types of NGS methods: first, by including the categorisation of viromics, genomics, and transcriptomics, as well as the differentiation between non-targeted and targeted virus detection; second, by describing for each step of the workflow the main current technical approaches with specific considerations relevant for method performance.

However, there are instances where the text may appear overly prescriptive, potentially favoring specific approaches at the expense of others. Consequently, we recommend using language that allows for alternative approaches, or alternatively, incorporating a general comment at the outset of section 2 to indicate the potential use of other technically relevant approaches.

Similarly, with regard to the description of controls in section 2.10, which are essential for ensuring the proper performance of the analytical procedure, it is important to note that controls applicable to individual steps may be specific to certain technical approaches or suppliers and should therefore be presented as optional. Additionally, the inclusion of a control for the entire workflow, together with additional quality criteria for individual steps, may render controls for individual steps unnecessary.

Please see below specific (non-exhaustive) comments as examples (see section *Specific Comments on the Proposed Text*).

Reference materials: we recommend clarifying that the use of alternative reference materials to those referenced in the text (WHO panel) is acceptable.

The WHO International Reference Panel for Adventitious Virus Detection is recommended as the minimum panel of model viruses for validation. It should be made clear that alternative panels are acceptable, and possibly more suitable in some cases. Other model viruses could be selected, based on the host cell, type of sample, and the overall virus risk analysis made on the specific product. In addition, virus stocks may be obtained from other sources (including in-house sources), provided they are characterised as described in the General Chapter, and demonstrated to be of acceptable quality for the intended purpose. In addition, it may be more appropriate to specify a minimum number of model viruses than a list of viral species. Such a minimum number should be applicable to all approaches (including transcriptomics). Finally, the use of nucleic acids as spiking material should be considered for transcriptomics approaches.

Please see below specific (non-exhaustive) comments as examples (see section *Specific Comments on the Proposed Text*).

Specificity: we recommend developing further the expectations related to specificity and to distinguish it (at least partially) from breadth of detection.

The evaluation of specificity as a demonstration of the breadth of detection of the analytical procedure is not aligned with the position we have included in our position paper, and may be partially revised. Based on ICH Q2(R2) definition, specificity "describe the extent to which other substances interfere with the determination of an analyte according to a given analytical procedure. Specificity is typically used to describe the ultimate state, measuring unequivocally a desired analyte." Based on this definition, specificity was envisaged in the position paper as the method's ability to detect a viral contaminant in a complex matrix, discriminating viral nucleic acids from the background.

Two aspects should be verified:

1. The method should report positive results only in the presence of actual contamination. In this case, the specificity can be demonstrated using negative samples, represented by the matrix without viral contaminants. The analysis of negative samples is required to verify that false-positive background signals are properly identified and filtered out by the method through appropriate bioinformatic analysis.

2. In the presence of actual contamination, the method identifies exclusively the viral sequences of the contaminant virus or closely related viral species, belonging to the same taxonomic group. This assessment can be conducted by testing positive samples artificially contaminated with known viruses or cells infected with known viruses representing different virus categories. This approach allows for evaluation of the potential effect caused by the matrix in the context of method validation, and it is required to verify that the method can report the correct taxonomic group, species or strain of the viral contaminant.

The two aspects described in the position paper are globally aligned with the content of the General Chapter, which states that: "The identity of the spiked viruses should be as expected. No false positive viral signal should be detected to confirm specificity." However, it might be relevant to further highlight the need to test unspiked samples to confirm the absence of false positive viral signals.

In addition, the parallel drawn between specificity and breadth of detection is not self-explanatory and may require further clarification. Our position is that the specificity and LOD experiments contribute to the demonstration of breadth of detection, in the sense that they allow to demonstrate the method performance over a range of viruses with diverse characteristics (which may impact the performance of individual experimental steps, or of bioinformatic steps). However, demonstration of breadth of detection may also take into account the database content. In addition, it can be further demonstrated by by spiking nucleic acids, or in silico sequences, defined to mimic the variability of viral strains within viral species. Such studies are not considered to be part of the method validation, but rather part of method development.

Please see below specific (non-exhaustive) comments as examples (see section *Specific Comments on the Proposed Text*).

Specific Comments on the Proposed Text

Enter specific comments or details here

Page and Line number(s)	Corresponding text	Comments and rationale; Proposed text	Member name (for consolidation purpose, will be deleted before sending)
Diversity of technical approaches and workflows			
Page 3, Figure 2.6.41.-1	Post-extraction treatment, green part Ribosomal RNA depletion	Rationale: a differential nuclease treatment can be used to generate DNA and RNA sub-samples used to prepare RNA and DNA libraries in the library preparation step. Proposed text: "Differential nuclease treatment, highly expressed RNAs (e.g. ribosomal and others) depletion"	
Page 3, Figure 2.6.41.-1	Post-extraction treatment, blue part Ribosomal RNA depletion	Rationale: not only ribosomal RNA may be depleted. Proposal to modify as "Highly expressed RNA (e.g. ribosomal and others) depletion"	
Page 3, Figure 2.6.41.-1	The step "Library preparation" mentions "RNA-Seq"	Comment: "RNA-Seq" seems to refer to a brand name and should be replaced by a more generic and explicit description of the technique.	
Page 4, lines 42-43	An enrichment step may be applied to reduce host-cell nucleic acid content.	Comment: Please modify language to be more generic to be inclusive of other methods of enrichment such as amplicaficaiton based ones.	
Page 6, lines 27-29	The bioinformatics analysis for virus detection involves building a pipeline that generally includes initial processing of the input raw reads to obtain quality reads, which in some cases may be followed by de novo assembly to generate contigs from overlapping reads, prior to mapping or aligning the reads / contigs using a reference virus or a database of viral sequences.	Comment: The language whould be inclusive of other technologies for read classification. For example, k-mer matching is a newer technology can be used instead of read mapping. The term 'Read classification' will be inclusive of all methods. Recommendation to use the term "read classification" throughout the document.	
Page 7, Figure 2.6.41.-2	A general bioinformatics analysis workflow	Comment: the figure describes only a nucleotide-based alignment method Proposed text: "Example of a general bioinformatics analysis workflow (using nucleotide-based alignment)"	

Page and Line number(s)	Corresponding text	Comments and rationale; Proposed text	Member name (for consolidation purpose, will be deleted before sending)
Page 7, lines 41-42	For primary screening, the leftover reads are mapped against a reference virus database (e.g. the Reference Viral Database, RVDB).	Recommendation to refer to the use of a "qualified database", and to mention that the database could contain nucleotide or aminoacid sequences. In Figure 2.6.41.-2, the reference to RVDB should be deleted in order to avoid that this example is considered in the future as a requirement.	
Page 7, line 43	Section title "Counter screen of viral hits from primary screen"	Rationale: The need to do counter-screening depends on the stringency of the primary analysis Proposed text: "Counter screen of viral hits from primary screen (optional)" To add as well in Figure 2.6.41.-2	
Page 9, lines 27-42	Control of the whole workflow Control of the library preparation step Control of the sequencing step	Rationale: Controls for individual steps are not always included in the instrument manufacturer's recommendations (for example at the sequencing step). In addition, they may not be necessary if a control for the whole workflow is included and quality criteria are applied for individual steps. Proposal: to consider controls at the library preparation step and sequencing step as optional.	
Selection of spiking material for validation			
Page 10, line 15	The use of nucleic acids as spiking material is not recommended, unless otherwise justified.	Rationale: RNA nucleic acids may be relevant as spiking material for transcriptomics approaches. Proposed text: "The use of nucleic acids as spiking material is not recommended for genomics and viromics approaches"	
Page 10, line 21	WHO virus reference panel is recommended as the minimum panel of model viruses for validation.	Rationale: alternative virus panels may be used. Proposed text: "WHO virus reference panel may be used as the minimum panel of model viruses for validation."	

Page and Line number(s)	Corresponding text	Comments and rationale; Proposed text	Member name (for consolidation purpose, will be deleted before sending)
Page 10, line 47	For specificity, the validation must demonstrate the breadth of detection of different virus types, using the WHO virus reference panel.	Rationale: alternative virus panels may be used. Proposed text: For specificity, the validation must demonstrate the breadth of detection of different virus types, using a panel of model viruses representing viral diversity as described in section 3.2.	
Page 11, line 17	For the genomics or viromics approach, this validation run should at the minimum include the WHO virus reference panel.	Proposed text: For the genomics or viromics approach, this validation run should include the model viruses used during initial validation (or a subset if appropriately justified).	
Demonstration of specificity			
Page 10, line 33	Specificity shall demonstrate the breadth of detection for different types of viruses as well as the correctness of the identification.	Rationale: Specificity is typically the rate of correct identification of negative samples. The breadth of detection is typically a component of the diagnostic sensitivity in combination with analytical sensitivity. Proposed text: Specificity studies should demonstrate the unequivocal detection and identification of viral contaminants when present, and absence of positive results in the absence of contaminants.	
Page 10, line 47	The identity of the spiked viruses should be as expected.	Proposed text: For samples spiked with model viruses, the identity of the detected viruses should be as expected.	
Page 10, line 47	No false positive viral signal should be detected to confirm specificity.	Proposed text: No false positive viral signal should be detected to confirm specificity, on spiked and unspiked samples.	

‘HTS vs NGS’ EFPIA survey : which preferred abbreviation for this technology ?

From EFPIA WG Clonality, characterisation and viral safety of cell lines, December 2021

Background:

The EFPIA Working Group Clonality, Characterisation and Viral safety of Cell lines has been created in 2019 with the intent of sharing knowledge and advocating for the implementation of innovative analytical methods for the clonality demonstration, characterisation and viral safety of cell lines. The Working Group has focused its initial efforts on the use of NGS/HTS-based methods for the viral safety of cell lines and is currently working on a position paper on the topic.

The choice between NGS and HTS abbreviation was discussed early in the Working Group and has been recently rediscussed in the light of the recent ICH Q5A revision draft, where both abbreviations are used. Pros and cons for each abbreviation were debated, and a third option was proposed as a potential consensus to bridge pros of NGS and HTS: HTSeq.

Survey:

The Working Group decided to run a survey on the preferences of its members regarding various abbreviations to designate HTS/NGS technologies, and to share the outcome of this survey with EFPIA representative for ICH Q5A (Marie Murphy from Elli Lilly), for further communication with ICH Q5A EWG if deemed useful.

Participants ranked 5 designations (NGS, HTS, MPS, Deep Sequencing, HTSeq) and had the opportunity to provide a justification for their preference. Except for HTSeq, other options are listed in ICH Q5A revision draft.

Conclusion:

The survey has confirmed the absence of consensus for a preferred abbreviation to be used to designate High Throughput Screening / Next Generation Sequencing technologies, especially in the context of ICH Q5A revision.

The main concern related to NGS abbreviation is its lack of precision as a designation, and the fact that it was initially used to describe only the early technologies in the field. It is also felt that the technology is not “Next Generation” anymore, but rather routine.

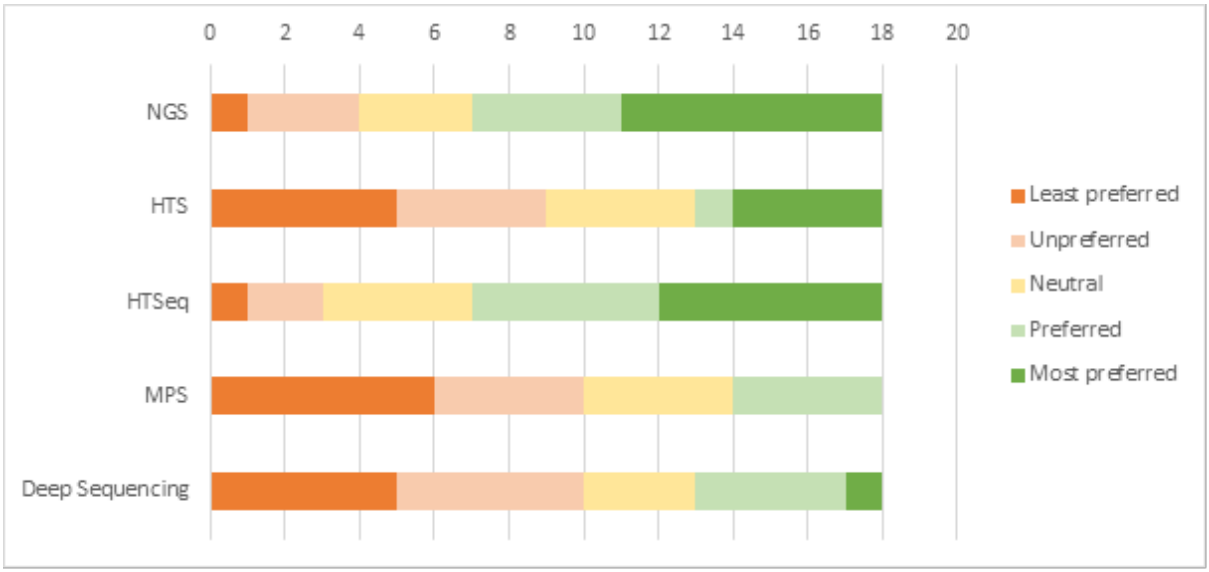
The main concern related to HTS abbreviation is its lack of precision as an abbreviation, and the possible confusion with High Througput Screening is felt strongly by some participants. Considering the Pubmed results, it seems that while the full names (High Throughput Sequencing and Next Generation Sequencing) are used at similar frequencies, HTS abbreviation is scarcely used.

In an attempt to find a consensus abbreviation that would conciliate the community, HTSeq was proposed as an abbreviation not referring to Next Generation, but informative enough by itself. The proposal received good feedback, but there are concerns that it may not impose itself in the community. It was also highlighted that the abbreviation is also used to designate a Python package and a library brand name.

Altogether, the preferred abbreviations from the participants of the survey are NGS and HTSeq, followed by HTS.

Results:

18 answers were received.



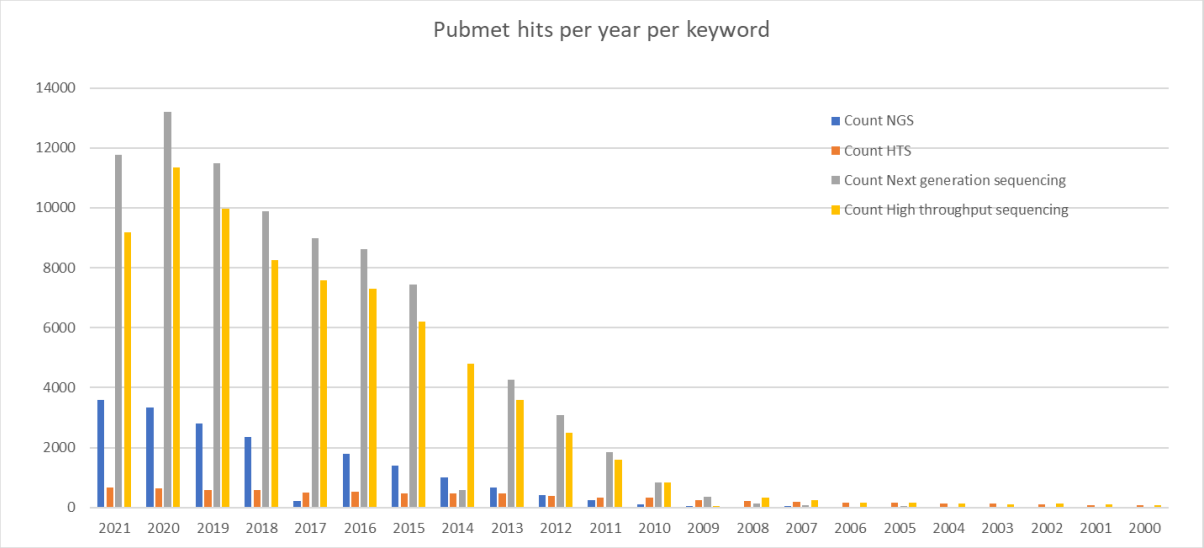
Designations ranked as most preferred are NGS, HTS and HTSeq. The high rating of NGS and HTS is consistent with them being the most established abbreviations in the field. HTSeq received high ranking as possible alternative to HTS.

Designations ranked as least preferred are Deep Sequencing, MPS, and HTS. This shows that the use of HTS is challenged by part of the group as a suitable designation for the technology.

NGS and HTSeq were also ranked more frequently than HTS as second preferred option.

Pubmed data:

In relation with a comment received about the decrease overtime of the use of NGS in papers, a quick search was done in Pubmed in complement to the survey to compare the number of publications by keyword:



From this quick analysis, there is no strong difference between the use of High Throughput Sequencing and the use of Next Generation Sequencing in the publications, with a very small advantage for Next Generation Sequencing. On the other hand, much less articles are retrieved using the abbreviations, highlighting a very low use of the HTS abbreviation compared to the NGS abbreviation (data from 2021).