Biologicals 67 (2020) 94–111

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Report of the second international conference on next generation sequencing for adventitious virus detection in biologics for humans and animals☆

Arifa S. Khan[a,*], Johannes Blümel[b], Dieter Deforce[c], Marion F. Gruber[a], Carmen Jungbäck[d], Ivana Knezevic[e], Laurent Mallet[f,1], David Mackay[g], Jelle Matthijnssens[h], Maureen O'Leary[i,2], Sebastiaan Theuns[c], Joseph Victoria[j], Pieter Neels[d]

[a] Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA
[b] Paul-Ehrlich-Institut, Langen, Germany
[c] Ghent University, Ghent, Belgium
[d] International Association for Biological Standardization for Europe, Lyon, France
[e] World Health Organization, Geneva, Switzerland
[f] Sanofi Pasteur, Lyon, France
[g] Advisor Veterinary Vaccinology, London, UK
[h] University of Leuven, Rega Institute for Medical Research, Leuven, Belgium
[i] P95 Pharmacovigilance and Epidemiology Services, Leuven, Belgium
[j] Boehringer Ingelheim, Ingelheim am Rhein, Germany

ARTICLE INFO

Keywords:
Adventitious viruses
Biotherapeutics
Bioinformatics
High-throughput sequencing
Vaccines
Biologicals
In vivo assays
In vitro assays
Massively parallel sequencing
Next-generation sequencing
Virus detection

ABSTRACT

The IABS-EU, in association with PROVAXS and Ghent University, hosted the "2nd Conference on Next Generation Sequencing (NGS) for Adventitious Virus Detection in Human and Veterinary Biologics" held on November 13th and 14th 2019, in Ghent, Belgium. The meeting brought together international experts from regulatory agencies, the biotherapeutics and biologics industries, contract research organizations, and academia, with the goal to develop a scientific consensus on the readiness of NGS for detecting adventitious viruses, and on the use of this technology to supplement or replace/substitute the currently used assays. Participants discussed the progress on the standardization and validation of the technical and bioinformatics steps in NGS for characterization and safety evaluation of biologics, including human and animal vaccines. It was concluded that NGS can be used for the detection of a broad range of viruses, including novel viruses, and therefore can complement, supplement or even replace some of the conventional adventitious virus detection assays. Furthermore, the development of reference viral standards, complete and correctly annotated viral databases, and protocols for the validation and follow-up investigations of NGS signals is necessary to enable broader use of NGS. An international collaborative effort, involving regulatory authorities, industry, academia, and other stakeholders is ongoing toward this goal.

☆ This article may be the work product of an employee or group of employees of government and other organizations. However, the statements, opinions, or conclusions contained therein do not necessarily represent statements, opinions, or conclusions of any government agency or organization. Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology or the U.S. Food and Drug Administration, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

* Corresponding author. Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD, 20993, USA.
E-mail address: arifa.khan@fda.hhs.gov (A.S. Khan).
[1] Current address: European Directorate for the Quality of Medicines, Strasbourg, France.
[2] Rapporteur.

https://doi.org/10.1016/j.biologicals.2020.06.002
Received 29 May 2020; Accepted 1 June 2020
Available online 11 July 2020

Abbreviations

AVDTIG	Advanced Virus Detection Technologies Interest Group
CBER	Center for Biologics Evaluation and Research
EMA	European Medicines Agency
Ph. Eur	European Pharmacopoeia
ECBS	Expert Committee on Biological Standardization
FDA	U.S. Food and Drug Administration
GSK	GlaxoSmithKline Vaccines
IABS-EU	International Association for Biological Standardization for Europe (IABS-EU)
ICH	International Council for the Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
NIB	National Institute of Biology, Slovenia
NIBSC	National Institute for Biological Standards and Control, U.K.
NIST	National Institute of Standards and Technology, U.S.A.
OIE	World Organization for Animal Health
OVRR	Office of Vaccines Research and Review
PDA	Parental Drug Association
WAHIS	World Animal Health Information System
WHO	World Health Organization

1. Introduction

The need for advanced virus detection technologies is increasingly recognized and has been prompted by a number of factors such as the introduction of novel cell substrates for vaccines, including mammalian, avian and insect cell lines, plants, and bacteria; and the development of cell and gene therapies, where rapid testing is needed for timely product delivery. Next-generation sequencing (NGS), also known as high-throughput sequencing (HTS) or massively parallel sequencing (MPS), has enabled the discovery of novel viruses in biological raw materials, and in clinical and environmental samples. Furthermore, NGS was also instrumental in the discovery of porcine circovirus type 1 (PCV1) in a licensed rotavirus vaccine [1] and the discovery of a novel rhabdovirus in the Sf9 insect cell line that is used for baculovirus-expressed products [2]. These cases highlighted that the currently recommended adventitious virus detection assays, even with extensive testing, can fail to detect novel and even some known viruses.

NGS has broad capabilities for the detection of known, novel, and emerging viruses, thus making it an attractive new technology for adventitious virus detection in biologics. Unlike other available nucleicacid-based technologies such as polymerase chain reaction (PCR) and virus microarrays, which are highly-sensitive, but specific for the detection of particular viruses or virus families [3], the application of NGS for *de novo* sequencing can detect all nucleic acids in a sample, regardless of whether they are associated with a known, distantly-related, or novel virus [4], or if they are cellular sequences. However, similar to other nucleic acid-based methods, rigorous handling procedures also need to be used for NGS to avoid unintended introduction of viral sequences through sample handling and reagents, and appropriate negative controls should be included. Additionally, in the case of NGS, there should be an increased awareness of potential contamination (primer hopping) if samples are pooled in a single sequencing run.

Although much progress has been made in NGS applications, some gaps and challenges still need to be addressed for using NGS in the regulatory context. For instance, appropriate viral and infected-cell standards need to be publicly available for evaluating the sensitivity and specificity of NGS for standardization and validation of the methodology. Furthermore, the development of international standards can help evaluate NGS results across different laboratories since protocols are still being developed and sequencing platforms continue to evolve. Additionally, the large volumes of data generated by NGS require extra computational and bioinformatics capacities to support the establishment of pipelines and processes for data analysis, storage and transfer, and comprehensive, annotated and publicly available reference virus databases are also needed. Finally, given that signals are commonly detected with NGS, follow-up strategies need to be developed early-on to confirm the results, and to aid in interpreting the biological significance of the signals.

The 2nd Conference on Next Generation for Adventitious Virus Detection in Human and Veterinary Biologics, held on Nov. 13–14, 2019, in Ghent, Belgium, was organized to follow-up on the scientific progress and advancements in the technical and bioinformatic aspects of NGS for virus detection made since the previous NGS for Adventitious Virus Detection in Biologics conference, which was held on October 26–27, 2017, in Rockville, Maryland, U.S.A [5]. The goal of the 2019 meeting was to assess the readiness of NGS for adventitious virus testing of human and veterinary biologics. One-hundred twenty-four scientists from 16 countries participated in the conference representing regulatory and public health authorities, industry, academia, and contract research organizations (CROs). The goal was to reach a consensus on the acceptability of NGS as a supplementary or replacement/substitution adventitious virus testing method for the *in vivo*, *in vitro* and PCR assays that are currently used for characterization and safety evaluation of human and animal vaccines.

Dr. Pieter Neels (IABS-EU) and **Dr. Sven Arnouts** (PROVAXS - Ghent University) opened the conference by welcoming the attendees and acknowledging the support received from the meeting sponsors.

2. Perspectives on NGS for adventitious virus detection

Dr. Arifa Khan (U.S. FDA) gave an overview of the past and current efforts in the last decade by regulatory health authorities and industry toward use of NGS for adventitious virus detection of biologics. The outcomes of the early meetings, which are shown in Table 1, identified challenges for using NGS for adventitious virus detection, that led to focussed discussions and scientific collaborations for addressing the technical and bioinformatics gaps for NGS applications in biologics.

Dr. Khan outlined the internal and external efforts by the FDA and the Office of Vaccines Research and Review (OVRR). These included:

- The establishment of FDA and CBER Genomic Working Groups to strengthen research and the regulatory framework for policy development and decision making relating to the use of NGS.
- The development of viral reference materials for NGS workflow

Table 1
Past public conferences on NGS and other advanced technologies for virus detection.

- July, 2009: PDA Cell Substrate Workshop. Discussed broad virus detection technologies. *Conference Proceedings in PDA J Pharm Sci and Tech, 2010; PDA Technical Report on Emerging Methods for Virus Detection (71) 2015*
- May, 2011: IABS: Adventitious Agents, *New Technologies and Risk Assessment. Conference Summary in Biologicals, 2011*
- November 2011: PDA/FDA Adventitious Agents and Novel Cell Substrates: *Emerging Technologies and New challenges*. Discussed applications and identified knowledge and technical gaps that needed to be address for further development and applications. *Conference Proceedings in PDA Journal, 2012*
- September 19, 2012: FDA VRBPAC discussions on use of human tumor cell lines and potential use of advanced virus detection technologies *Transcript pdf available*: www.fda.gov
- May, 2013: WHO Cell Substrate Study Group (Beijing). Discussed DNA based detection of adventitious agents in cell substrates
- November 2013: PDA/FDA Advanced Technologies for Virus Detection in the Evaluation of Biologicals: *Applications and Challenges*. Identified challenges and priority areas to be addressed for applications to biological products. *Conference Proceedings in PDA Journal, 2014*

evaluation and for method standardization and validation [6].
- The development of a public database of all viral sequences to facilitate detection of known and novel viruses [7,8].
- The establishment of the FDA/industry-led Advanced Virus Detection Technologies Interest Group (AVDTIG; coordinated by the PDA) to provide an open forum for scientific discussions and development of collaborative studies to address technical and bioinformatics knowledge gaps to facilitate NGS standardization [9].

Additionally, she noted community efforts that were ongoing toward advancing NGS for virus detection. These included:

- Efforts on development of virus stocks by NIBSC/WHO [10].
- Scientific discussions on standardization and validation for NGS implementation in biologics [11,12].
- The commercial availability of NGS platforms and bioinformatics processes by contract research organizations (CROs) to meet research and regulatory needs.
- Ongoing research by academia and others to expand the virus universe by discovery of viral genomes using NGS.

Dr. Khan summarized the 2017 NGS meeting [5] and noted that considerable progress has been made since then. Some of the achievements, described in other sessions, included:

- A spiking study with model viruses added in a background of different matrices representing biological materials was performed to evaluate NGS virus detection across three different laboratories [13].
- The publication of papers on upstream sample preparation [11] and downstream bioinformatics for optimisation of NGS virus detection [12].
- The publication of a reference viral database (RVDB) [7].
- The development of five live, reference virus stocks for NGS evaluation and standardization and completion of 24-month stability studies [6].
- The use of NGS for the detection of active viral RNAs by transcriptomics analysis [14,15].

Additionally, in September 2019, NIST and the FDA co-organized a Workshop on Standards for NGS Detection of Viral Adventitious Agents in Biologics and Biomanufacturing, which discussed the currently used virus reference materials, the limitations of the current standards, current industry practices, and what additional types of standards would be useful for evaluating the NGS workflow or various stages of the sequencing process [6,16]. The availability of publicly-available standards for NGS standardization could facilitate broader use of the technology. Dr. Khan emphasized that since NGS is evolving, there is a need for continued interactions with regulatory agencies, industry, and CROs to identify and address scientific challenges for using NGS in biologics.

Dr. Laurent Mallet (Sanofi Pasteur) presented on the use of NGS in the human and veterinary vaccine industry. He pointed out that NGS has a number of potential uses for adventitious viral detection, both in human and animal vaccine production, and in the overall analytical quality environment. It can be used as a specification test for release, as a substitution for compendial tests such as the *in vivo assays*, in characterization or screening studies or as an investigational tool. For both human and veterinary viral vaccines, NGS can be used at various stages in the manufacturing process, although there may some differences in applying NGS. For instance, in veterinary vaccine production, NGS may be used on the master cell bank, master seed bank or the vaccine (final product) materials, whereas for human vaccine production testing is on the cell bank (master, working and extended), seed banks (master and working) and drug substance. If the viral vaccine is difficult to neutralize, NGS could be used as a replacement/substitute for the *in vitro* test at the harvest level as it is more sensitive than the existing tests. NGS can also be used for characterization studies to reduce the risk of introducing adventitious agents through the use of contaminated raw materials, for instance by screening the pre-master cell banks and seed lots, and by testing raw materials and vaccine substrates (for instance eggs). It can also be used as an investigational tool to verify the results of other tests and to identify the contaminants detected by those other tests. If using NGS as a release test, it will be necessary to fully validate the test. If using it as a replacement or substitute test, it will be necessary to validate the test and it may be necessary to conduct a comparison head-to-head or bridging study to compare its performance to that of the existing test. Ph. Eur. Chapter 5.2.14describes the concept of substitution of *in vivo* methods by broad molecular methods, where a head-to-head comparison of the tests may not be feasible. In that context, use of existing validation data of *in vivo* tests with a particular caution regarding the interpretation of the results is recommended since the *in vivo* tests and the NGS test do not detect the same characteristics of the viral contaminant [17].

Dr. Mallet noted that the need for supplementary testing methods for adventitious viral detection, the poor performance of *in vivo* testing for viral detection, and the growing pressure, particularly in Europe, to decrease the use of animal testing, create an environment in which NGS has the potential to act as a substitute for a number of the conventional tests, including *in vivo* adventitious virus assays, antibody production tests in animals, 9CFR tests for infectious bovine and porcine viruses, and virus-specific PCR assays. However, depending on the viral vaccine that is being tested, some of the conventional tests, such as *in vitro* tests, which are broad and sensitive, are still useful and should continue to be used. In case of assays where the vaccine virus cannot be adequately neutralized, NGS may be a suitable alternative assay.

It was further pointed out that the Ph. Eur. requires that breadth of virus detection and analytical sensitivity of NGS be validated in relation to the use for which it is proposed; however, this generally requires a panel of representative, well-characterized model virus standards [17,18]. Moreover, as NGS detects the viral genome, and *in vivo* tests detect the effect of infectious viruses on experimental animals, direct comparison between the two tests is challenging [17]. There is also limited data on the validity of *in vivo* methods. One potential panel of model viruses are those used in the National Institutes of Health (NIH) study that compared sensitivities of the *in vivo* and *in vitro* assays for detection of 16 different viruses [19]. Also, of note, is the fact that the detection of a genome or genome fragments by NGS does not necessarily correlate with the presence of an infectious virus, so further work is needed to interpret the results of the NGS test.

Dr. Jean-Pol Cassart (GSK Vaccines) reported on the Advanced Virus Detection Technologies Interest Group (AVDTIG; coordinated by the Parenteral Drug Association), which was created to address the needs and challenges associated with using NGS, with the goal of preparing for the implementation of NGS for application in biologics [9]. The AVDTIG currently includes greater than 150 members from 60 organizations across government and regulatory agencies, industry, CROs, academia and technology developers.

The adoption of NGS is subject to a number of challenges, most notably related to the diversity of viral targets and biological matrices (such as cell banks, viral seeds and raw materials), along with the complexity of the NGS technologies and associated bioinformatics, which makes method validation difficult. In order to address these challenges, the AVDTIG initially identified five priority areas, and five corresponding subgroups (A to E) were formed to specifically address these priority areas through discussions and conduct of collaborative studies: A) sample selection, preparation and processing; B) the development of virus standards and reference materials; C) the development of a correctly annotated viral reference database, complete with all viral sequences, excluding bacterial viruses; D) the analysis of bioinformatics pipelines; and E) the development of follow-up strategies to confirm the identity of a positive signal. Subgroup A members have

published the results of a collaborative study [13], which compared different sample preparation approaches and technology platforms in three different laboratories in the U.S., Canada, and Belgium, in order to evaluate the sensitivity of NGS for the detection of four viruses. Comparable virus detection was achieved in the three laboratories. This pilot virus spiking study demonstrated the strength of using well-characterized virus stocks for NGS standardization by different laboratories using independent protocols and bioinformatics pipelines. Members of Subgroup A have also published a paper on sample preparation detailing best practices for extraction methods for viral nucleic acid detection, nuclease treatment for the isolation of particle-associated viral nucleic acids, whole genome amplification, quality control for upstream steps and NGS performance evaluation [11]. The results of the collaborative pilot virus spiking study and further discussions in Subgroup B led to the development of virus reference material comprising of five virus stocks, which can facilitate NGS performance evaluation, standardization, and comparison to conventional assays (presented by Dr. Arifa Khan). Discussions in Subgroup C resulted in the development of a reference virus database to reduce nonspecific cellular hits and to increase the likelihood of novel virus detection [7,8]. Subgroup D members have published a paper on considerations for optimisation of bioinformatics pipelines [12]. Subgroup E members are preparing a manuscript regarding considerations for data interpretation, including the bioinformatics criteria and thresholds for a positive hit, sources of background signals and the effectiveness of laboratory follow-up procedures (presented by Dr. Robert Charlebois). These collaborative efforts in the AVDTIG provide data and scientific knowledge for considerations of NGS as an alternative supplementary or replacement assay for adventitious virus detection in biologics.

Dr. Robert Charlebois (Sanofi Pasteur) gave an overview of bioinformatics analyses of viral NGS data and the consequences for follow-up investigation strategies. Conventional assays such as PCR are usually targeted to a particular virus, and therefore a positive PCR result has a clear follow-up strategy to confirm whether the detected virus is infectious. In contrast, NGS can detect a broad range of known and unknown viruses, and as a result, the follow-up strategy is not clear, and requires the design of an investigation protocol to test the signal. NGS is highly sensitive with a large breadth of detection, but it can also detect noise and false positives. Sources of noise include residuals within test reagents (for instance *Escherichia coli* bacteriophages), residuals within controls, human shedding during testing and environmental contamination in the testing laboratory. False positives can include sequences that are similar to the host sequences (when those segments of the host genome are not in the database), inaccurate records in public databases, records in public databases that pick-up host sequences, as well as sequencing errors or natural variation in a viral harvest dataset that lead to related viruses being misidentified. There are three scenarios that should be considered. Firstly, true positives where the viral nucleic acids are actually in the test article are detected. This can include both adventitious viral nucleic acids, as well as other background nucleic acids, and is best managed with a risk-based approach. Secondly, false negatives, where the viral nucleic acids are in the test article but are not detected; this is influenced by the design of the assay, and so in this case the limitations of the assay must be understood. Finally, false positives or noise arising from the limitations of the process; this should at least be recognized and understood, but ideally should be managed through a risk mitigation approach.

The analysis of novel cell substrates will require the characterization of unmapped sequences, which will include both “known unknowns” and “unknown unknowns”. The known unknowns can include false negatives such as viruses infecting the novel cell substrates. It is necessary to mitigate against these false negatives through the use of other assays, and in the longer-term through increasing the characterization of the virome, which represents all viral nucleic acids in a sample. It can also include false positives such as host sequences that are not in its genetic assembly but that match viruses. The unknown unknowns can include sequences not matching any reference; these can be either true or false negatives. An attempt should be made to characterize these, for instance through pattern matching or sequence assembly, and laboratory-based follow-ups may be designed using the sequence information for verification of such signals.

The decision on whether or how to investigate true positives, false positives and false negatives will depend on their potential impact. True positives usually require further investigation, unless the signal is within the range of background noise (if the background range is known). False positives usually do not require follow-up if the analyst can confidently distinguish between false positives and true positives, which may require additional learning. The decision to investigate false negatives by running additional tests to verify the NGS results depends on the potential risks and depends on whether the material being tested is a novel cell substrate or a well characterized cell line.

NGS generates a lot of information, which requires a lot of interpretation. Factors to be considered in the interpretation include factors relating to the signal itself such as the taxonomic identification, the degree of confidence of the match, the number of reads that match and where they match on the reference genome and the percentage identity profile over the reference genome. Other contextual factors also need to be considered including what is being analysed, the question being asked, where the sample came from, the biological constituents of the sample, the analytical method, comparison to previous results or other analytical methods, measures to mitigate contamination and the risks associated with the signal or sample. This can be managed through machine learning, which looks at what is expected based on the context, for instance what is the probability of a signal given the background. The learning can help make the correct decision to pass uncontaminated samples quickly and confidently, and to decide what requires further investigation, with due consideration of the potential risk.

3. Regulatory aspects

Regulatory authorities are increasingly recognizing the potential of NGS for adventitious virus agent detection. In particular, they recognize its potential to replace or substitute *in vivo* testing, which aligns with the 3Rs (replacement, reduction and refinement) initiative for animal testing [20]. They also identify its potential to replace/substitute or supplement *in vitro* testing; however further data are needed to demonstrate its suitability as an alternative method for the *in vitro* adventitious virus detection assays. Current regulations either allow for or encourage the use of NGS, or are being updated to include the use of NGS and other nucleic acid based methods as supplementary/replacement/substitution methods for broad adventitious virus detection.

The regulatory environment has evolved in recent years to allow for the streamlining of the adventitious viral agent testing package and to allow for the use of NGS as a substitute or supplement for the *in vivo* and *in vitro* adventitious tests in vaccine production and quality control. In the U.S. FDA, use of NGS for adventitious virus testing is being considered on a case-by-case basis for manufacturing of viral vaccines. A number of changes have been made to the Ph. Eur. to allow for the use of NGS for testing of human vaccines and for testing at various stages of the vaccine production process, in lieu of *in vivo* and *in vitro* tests (Table 2). For instance, NGS can now be used as a tool in virus risk assessment and testing of vaccine cell substrates and for testing extraneous agents in materials such as seed lots, and harvests of viral vaccines. For veterinary vaccines, the Ph. Eur. allows for the streamlining of the testing package and the introduction of new methods, with the focus being on an overall risk assessment [21–23] rather than performing general and specific tests. The Ph. Eur. recognizes the difficulty in conducting bridging studies to compare NGS to existing *in vivo* animal assays [17].

Dr. Robin Levis (U.S. FDA) gave an overview of the FDA's perspectives on using NGS for human vaccines. Since 2010, CBER has

Table 2
Evolvement of European Pharmacopoeia to facilitate the use of NGS.

- (Hum + Vet) Ph. Eur. Chapter 5.2.14. Substitution of *in vivo* method(s) by *in vitro* method(s) for the quality control of vaccines, version 9.3, July 2017, creation.
- (Hum) Ph. Eur. Chapter 5.2.3. Cell Substrates for the production of vaccines for human use, version 9.0 and updated version 9.3, July 2017, revision.
- (Hum) Ph. Eur. Chapter 2.6.16. Tests for extraneous agents in viral vaccines for human use", version 9.3, July 2017, revision. Corrected version in Supplement 9.4.
- (Vet) Ph. Eur. Chapter 5.2.4. Cell cultures for the production of vaccines for veterinary use,version 10.2, July 2020, revision.
- (Vet) Ph. Eur. Chapter 5.2.5. Management of extraneous agents in immunological veterinary medicinal products, version 10.2, July 2020, revision.
- (Vet) Ph. Eur. Chapter 2.6.37. Principles for the detection of extraneous viruses in immunological veterinary medicinal products using culture methods, version 10.2, July 2020, creation.

advised that characterization of novel cell substrates, in particular tumorigenic cell lines, for adventitious viruses be guided by a comprehensive testing plan and risk mitigation strategy that may include the use of new technologies, along with conventional assays [24]. Early dialogue with regulatory authorities is encouraged when considering the use of NGS for screening for viral adventitious agents.

The FDA, including CBER and many of the other FDA Centers, is widely engaged in the use of NGS, in terms of either research or regulation. Initiatives include: i) FDA/CBER Genomics Working Group to establish the research and regulatory infrastructure for development of policy and decision-making on NGS applications; ii) AVDTIG described by Dr. Jean-Pol Cassart: iii) OVRR Technical Working Group that engages in discussions with individual sponsors to facilitate development of specific products using NGS to support product characterization and product safety. CBER also conducts research, and at present there are more than twenty projects using NGS data to address a variety of scientific and regulatory questions, including for example how NGS can be used to evaluate the safety of cell substrates and the absence of adventitious agent contamination. In terms of regulation, CBER has accepted NGS data as a part of the Chemistry, Manufacturing and Controls package for the development of certain products and is encouraging discussions with manufacturers on how to best integrate NGS in product characterization. In particular, the OVRR has been involved in discussions with industry on the use of NGS for adventitious virus testing of Master and Working Virus Seeds and virus harvests; on assessments of the genetic stability of vaccine virus, and on cell substrate characterization. There are ongoing efforts to standardize and validate the use of NGS.

From the regulatory perspective, for all assays including NGS, the reliability of the assay for testing cell substrates and virus seeds, within the context of their intended use, must be demonstrated. Assays used for safety testing must be validated through the conduct of validation studies and inclusion of appropriate controls and standards. The NGS platform must also be standardized and validated through the use of appropriate reference viruses and other standards to demonstrate the efficiency of the different methodological steps, and to quantify its sensitivity and specificity.

The quality of the bioinformatics platform for data analysis is also a critical consideration for using NGS, as is the ability to access the right expertise to evaluate the data. Bioinformatics pipeline optimisation can be achieved through the use of reference datasets, by clearly defining the quality criteria for an acceptable read, and by defining parameters for short read assembly and for hybrid assembly to correct the high error rate currently seen in long-read sequencing. Current, correctly annotated, publicly available reference virus databases are also needed.

Other challenges include the development of an appropriate follow-up strategy for verifying a positive signal and assessing the biological relevance and significance of the hit.

The FDA and CBER are continuing to prepare for the use of NGS and for the review of NGS in regulatory submissions. They are actively working to identify the challenges and to address the knowledge gaps and are working to maintain and advance the necessary in-house technology and bioinformatics infrastructure and expertise. Regulatory agencies will also need to work together to harmonize their requirements for NGS. More publicly available standards are also needed.

Dr. Ivana Knezevic (WHO) presented an overview of WHO biological standardization focusing on the evaluation of cell substrates for production of vaccines and other biologicals and the WHO approach for the detection of adventitious agents. The WHO provides global written and measurement (physical) standards as a basis for the establishment or update of national regulatory standards by 194 WHO member states. These include a total of 99 written technical recommendations and guidelines, 68 of which are vaccine-specific and 9 of which are biotherapeutic product-specific, as well as more than 400 measurement standards (to inform the development, licensing and lot release of medicines, and allowing for the expression of biological activity in international units).

Dr. Knezevic noted the encouragement for developing new methods for detection of adventitious agents, published in the WHO "Recommendations for the evaluation of animal cell cultures as substrates for the manufacture of biological medicinal products and for the characterization of cell banks" [25], which was adopted by the Expert Committee on Biological Standardization (ECBS) in 2010. In addition, she explained the statements on NGS in the WHO document entitled "Scientific principles for regulatory risk evaluation on finding an adventitious agent in a marketed vaccine" [26]. The most recent consideration of NGS by the ECBS in October 2019, the endorsed proposal for replacing the mutation analysis by PCR restriction enzyme cleavage (MAPREC) for poliovirus vaccines [27], was presented as an example of the broader application of NGS for vaccine evaluation beyond adventitious agents. Dr. Knezevic noted that manufacturers from the EU and USA are aiming toward replacement of currently used assays for detection of adventitious agents; however, manufacturers and regulators from developing countries have not thus far been involved in such discussions. The role of academia and service providers is also important in the context of the increasing need for expertise in virology, molecular biology and bioinformatics. WHO leadership is expected in bringing regulators, manufacturers and academia from all WHO regions to raise awareness of the new testing methods and to facilitate regulatory preparedness for review of the data that are generated by NGS. As part of the consultation at the global level, further discussion on NGS in the context of the 3Rs is planned.

Dr. Knezevic concluded her talk by indicating that the two most important developments in terms of WHO measurement standards are panels of viruses prepared by WHO Collaborating Centers, FDA/CBER and NIBSC, as reference preparations for NGS. Development of the proposals for submission to the ECBS has been initiated and will be completed in 2020 and 2021 with the details regarding the intended use of the panels of viruses.

Dr. Johannes Blümel (PEI) described the application of NGS for adventitious virus detection during production of human biotherapeutics and advanced therapy medicinal products (ATMPs). The International Council for the Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is a consortium of regulatory authorities, industry and regional industry associations, which aims to harmonize the processes for the development and registration of safe, effective and high-quality medicines [28]. It has a number of guidelines relating to the manufacture of biotechnology products (such as proteins from cell cultures), and ATMPs, such as viral vectors and cell-based medicinal products. In light of the potential for NGS to replace or reduce animal testing, there is broad consensus among partners that some of the ICH guidelines need to be revised. Guideline ICH Q5A on the Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin is being revised, to reflect the new technologies for virus detection, qualification and validation and to

expand the scope of the guideline to include new biotechnology products such as viral-like particles and viral-vectored particles. The ICH is also establishing guideline ICH Q13 for the Continuous Manufacturing of Drug Substances and Drug Products so as to capture key technical and regulatory considerations that promote harmonisation, including Current Good Manufacturing Practices elements specific to Continuous Manufacturing (CM). This aims to allow drug manufacturers to employ flexible approaches to the development, implementation and integration of CM when manufacturing drug substances and drug products, including small molecules and therapeutic proteins for new and existing products. It also aims to clarify regulatory expectations on the development, implementation and assessment of CM technologies used in the manufacture of drug substances and drug products. Some of the key considerations for the guideline revisions include factors such as the appropriate process for sampling unprocessed bulk from continuous systems when using NGS to characterize cell banks, and the best practices for downscale models and validation. These guideline revisions are complex and lengthy processes and are expected to take up to three years.

For plasma-derived medicinal products, there is a reliance on specific validated tests for selected agents in combination with robust virus clearance, and validation requirements for specific licensed tests are in place. NGS has not yet been introduced for testing of blood donations or plasma pools. The sensitivity of NGS needs to be determined, and this can be done using established reference materials, such as those used for human immunodeficiency virus, hepatitis B virus, hepatitis C virus, human parvovirus B19, hepatitis A virus, hepatitis E virus and genotype panels. It will be challenging to use NGS for testing of blood donations, due to the large number of donations and the speed with which they will need to be tested. A conceptional advantage of NGS would be detection of multiple pathogens and their variants as well as emerging pathogens such as West Nile virus or Zika virus with a single test.

Regarding ATMPs and viral vectors, for the gene therapy products, the regulatory requirements in Europe are similar to those for vaccines. In particular, for gene transfer medicinal products for human use, the Ph. Eur. allows for the use of NGS in a similar way to which it is applied for testing viral vaccines. Specifically, NGS can be used for testing extraneous agents in viral vaccines for human use and for testing cell substrates for the production of vaccines for human use. For cell-based medicinal products, there are no specific regulations on testing, and so the use of NGS for testing these products can be decided on a case-by-case basis. Safety of cell-based products mainly relies on the safety and purity of the cell substrates and raw materials such as culture medium, as these products include living cells and therefore it is not possible to implement downstream steps for virus inactivation or removal. In many cases, living cells have to be directly used, so there is no ideal point at the end of the cell culture at which testing can be conducted. Often it is necessary to use primary cells, as there are no cell banks or there are multiple primary cell banks. Cell cultures are often supplemented with bovine or human serum or human platelet lysates. When using bovine serum or other animal derived products, the species barrier may confer some protection against infection, while this is not the case for human derived raw materials such as serum or platelet lysates. Ensuring that these raw materials are safe is critical, as is applying effective methods for virus inactivation or removal. However, testing of raw materials by nucleic acid-based detection methods such as PCR or NGS has limitations with respect to detection of low-level contaminations that could still contaminate cell cultures. In addition, these methods cannot differentiate between infectious and non-infectious virus particles. Therefore, NGS can be used to verify the safety and purity of cell banks and cell cultures, however, applying virus inactivation towards raw materials is considered an important complementary measure in order to avoid contamination of cell cultures. Inactivation methods can be overwhelmed by high loads of pathogens as observed by cases of malaria transmission with inactivated blood component [29]. Therefore, pathogen inactivation should be preferentially applied to pooled material rather than to individual components before pooling [30].

Dr. Javier Pozo Gonzalez (EMA) gave an overview of European regulatory considerations on the use of NGS for the detection of extraneous agents in biological veterinary products. In the European Union, the demonstration of absence of extraneous agents in immunological veterinary medicinal products (IVMPs) has traditionally been performed by testing, using a variety of cell culture, serological and molecular methods, in line with the requirements for testing specified in Directive 2001/82/EC, as well as general and specific monographs of the Ph. Eur. All components of animal origin, such as cell substrates, virus seeds, as well as in-process materials and the finished product must be tested. The Committee for Medicinal Products for Veterinary Use, Immunologicals Working Party guideline on the production and control of VMPs, published in 2017 [31], introduced the concept of risk assessment for the management of extraneous agents. This risk assessment-based approach consists firstly of justifying not testing for extraneous agents based on the risk assessment of the likelihood of the starting materials being contaminated, and secondly targeting testing at those agents that cannot be ruled out based on the risk assessment. There is a reference list of potential extraneous agents by species that should be consulted when deciding what extraneous agents need to be considered in each case.

The legal basis for this risk-based approach has been strengthened. A new European Union regulation governing veterinary medicinal products (2019/6) was published in 2019 and is due to come into force in January 2022. This regulation, and its associated, updated guidelines move from a prescriptive to a risk assessment approach to demonstrating freedom of materials from extraneous agents. In parallel, relevant chapters of the Ph. Eur. relating to the general and specific requirements for veterinary vaccines have been revised in a similar way to reflect this new risk assessment approach. Testing will only be required for those relevant extraneous agents that cannot be ruled out by the risk assessment. It allows more flexibility in the choice of methods that can be used for testing, which will allow manufacturers to select fit-for-purpose techniques based on the latest scientific methods. These changes are expected to reduce the overall levels and cost of testing, and also to reduce the amount of *in vivo* testing. The broad range of detection and ability to simultaneously detect multiple agents is particularly relevant for VMPs as they are often intended for use in multiple species, so testing would need to be able to detect all the different agents of importance for all the species to which the product will be administered.

In addition to using NGS for the detection of extraneous agents in VMPs, other applications include using it for the identification and characterization of vaccine strains, for the evaluation of genetic stability of vaccine strains after successive passages and for reversion to virulence studies.

Currently, regulators lack experience in using data from NGS for decision making. There is also a lack of guidance on the format and content of data on NGS testing to be provided to authorities in support of a marketing authorization dossier. Therefore, further collaboration and dialogue between researchers, manufacturers and regulators is needed, to inform the development of specific guidance on requirements for regulatory acceptance of NGS for the detection of extraneous agents, and to understand and address the current barriers to implementing NGS for the detection of extraneous agents. In order to harmonize these requirements across jurisdictions, further coordination between regulatory bodies is needed. The EMA is supporting the use of NGS for extraneous agent testing for VMPs in a number of ways. The EMA has established a multidisciplinary Innovation Task Force with scientific, regulatory and legal expertise to facilitate early dialogue on innovations in pharmaceutical development, including addressing the impact of emerging technologies such as NGS on current requirements. The Scientific Advice Working Party of the Committee for Medicinal Products for Veterinary Use of the EMA can also give product-specific advice on the types of tests and studies needed to support the

development of a VMP. The new European regulatory framework will allow greater flexibility for choosing methods for extraneous agent testing, including molecular methods with broad detection capabilities such as NGS.

4. Standardization of next generation sequencing

Activities to support the standardization of NGS for adventitious virus detection should focus on: identifying gaps and priorities for standardization, the development of readily available standards for validation of the NGS workflow (including sample preparation, sequencing, and bioinformatics), optimisation of workflows and pipelines, the conduct of multicenter standardization studies and the development of guidelines for standardization. Standards development for the most common NGS procedures should be prioritized.

Dr. Megan Cleveland (NIST) presented the outcomes of the Joint NIST/FDA Workshop on Standards for NGS Detection of Viral adventitious Agents in Biologics and Biomanufacturing [6]. The workshop aimed to identify what viral standards are publicly available and in use for NGS virus detection, and to identify the gaps in terms of availability of viral standards. It also aimed to identify opportunities for collaboration and research sharing, to help with prioritizing needs and to clarify the qualification procedures for reference materials. NGS detection of viral adventitious agents involves a series of steps which can include sample collection, reduction of host cell DNA or RNA (e.g. nuclease treatments), DNA or RNA extraction, RNA to cDNA conversion, library preparation, sequencing and bioinformatic analysis. Different standards will be needed at different steps. For instance, naked nucleic acids cannot be used at the point of reducing the host cell DNA or RNA, as they will be destroyed by this process. The DNA/RNA extraction step is a highly variable step and different viruses can be easier or harder to extract at this step. At the point of sequencing, different standards may be needed for short read sequencing technologies versus long-read sequencing technologies, although there is no current consensus on this. A number of important issues emerged during the workshop, including the fact that different standards may be required at different stages in the development of the adventitious viral agent detection process, for instance at the stages of method development, method validation and routine testing. It was generally considered that it would be best to include in the reference material all types of viral genomes (for example, single-stranded (ss)RNA or double-stranded (ds) DNA), as well as a virus with a difficult to lyse capsid. It was also agreed that digital PCR could be used to monitor the efficiency of the various steps post extraction, and that extraction is the most variable step and may require actual viruses for its validation. There was also uncertainty as to whether virus-like particles behave the same as viruses and DNA in solution. Clearly defining the types of standards needed for the various steps in the sequencing process was identified as a priority. It was also agreed that it will not be possible to make all standards for all methods and that the most common NGS procedures should be identified and that standards development should at least initially target those procedures. NIST is compiling a list of resources for standardization and reference materials.

Dr. Arifa Khan (FDA) described FDA/CBER reference materials for NGS virus detection, which included well-characterized virus stocks and a reference viral database (RVDB). The FDA is supporting NGS standardization and validation efforts for adventitious virus detection in biological products through the development of reference materials including virus stocks and cell-based materials intended to be available publicly and by conducting spiking studies to evaluate different sequencing platforms. Additionally, bioinformatics challenges are being addressed by development of a complete and correctly annotated viral database to facilitate known and novel adventitious virus detection.

The results of a multicenter study to evaluate the performance of NGS for virus detection [13], and discussions in the AVDTIG prompted the initiation of efforts by the FDA to prepare reference virus stocks for NGS platform evaluation and standardization. Large-scale preparations were made of the five viruses that were used in the previous study. These live viruses can be used for evaluating the entire NGS workflow, from sample extraction through bioinformatics. They represent diverse virus families of potential safety concerns in human vaccines and consist of different genome lengths and types (double-strand and single strand; RNA and DNA; linear and circular), particle sizes, and include enveloped and non-enveloped viruses. The virus stocks[Epstein-Barr virus (EBV), human respiratory syncytial virus (RSV), human orthoreovirus (Reo), feline leukemia virus (FeLV), and porcine circovirus type 1 (PCV1)] were well-characterized by determining infectious titer, genome copy number, particle count, full-length virus genome sequence with identification of variants, and analysis of residual host nucleic acid content. A two-year stability study was completed showing a stable infectious titer and genome copy number. Details of the reference virus stocks were presented in the NIST-FDA workshop on viral standards for NGS [6] and a manuscript is in preparation. These virus stocks were individually vialed and can be used, singly or mixed, in a number of ways to support NGS standardization including:

- To assess sensitivity of virus detection in the entire NGS workflow by performing spiking studies in different matrices relevant for the production of biologics.
- To compare NGS with current nucleic acid-based assays or infectivity assays for adventitious virus detection, such as PCR, *in vitro* or *in vivo* assays.
- To generate well-characterized NGS datasets for evaluating bioinformatics pipelines.

The virus stocks have been distributed since 2017 for collaborative spiking studies and also to others in industry with similar efforts toward NGS standardization and validation as an alternative assay to supplement or replace the conventional adventitious virus detection assays. FDA is currently participating in virus spiking studies in the AVDTIG to evaluate the limit of detection of different virus types by NGS. Additional FDA efforts are ongoing to evaluate virus detection using the Nanopore MinION and for developing well-characterized, cell-based reference materials for NGS virus detection.

In 2013, the FDA, in consultation with the AVDTIG and the National Center for Biotechnology Information (NCBI) at NIH, started work on developing a new and complete reference viral database, the RVDB [7]. The RVDB includes all viral, viral-related and viral-like sequences, including endogenous viruses and retroelements, and it has a reduced cellular content. It contains sequences of viruses from all species relevant to biologics, except bacteria. These include viral sequences that are relevant to human and animal health as well as other disciplines. The RVDB is currently available from the Center for Bioinformatics and Computational Biology at the University of Delaware (https://rvdb.dbi.udel.edu/). A corresponding protein database was generated by the Institut Pasteur [8], which is available on the RVDB site and at the Institut Pasteur (https://rvdb-prot.pasteur.fr/).

Use of the RVDB is expected to enhance NGS investigations for product safety by increasing the specificity and sensitivity of virus detection and enhancing the efficiency of large NGS data analysis, in particular for the detection of novel viruses. The reduction in non-specific cellular hits has resulted in a reduction in the volume of data for bioinformatics analysis, thus reducing the computational time. The scripts for generating the database are publicly available online (https://github.com/ArifaKhanLab/RVDB), along with an RVDB toolbox, which contains instructions and python scripts for generating the database.

Annotation of the RVDB is ongoing to identify viral and non-viral sequences to facilitate follow-up strategies for NGS hits. Additional efforts include updating the RVDB with complete representation of diverse virus families, especially for host species used in cell substrates for the production of biologics, as well as accurately annotated entries

indicating viral and non-viral regions, to provide rapid and accurate results for adventitious virus detection. The regular updates of the RVDB to include new entries in Genbank would facilitate the detection of emerging viruses. It is also planned to develop machine learning approaches for adventitious and endogenous retrovirus detection.

Dr. Edward Mee (NIBSC) gave an overview of joint NIBSC/WHO initiatives to develop standards and reference materials for use in NGS for biological products, including DNA and RNA detection assays and viral infectivity assays. These include standards for new methods that are under development and standards that can be used for NGS bridging and validation studies. The NIBSC is developing reference materials that could be used in bridging studies to compare the sensitivity and specificity of NGS for virus detection to that of *in vivo* methods. The NIBSC previously developed a highly multiplexed reference material that can be used for benchmarking NGS methods [10]. Although this reference material is still in use, it is recognized that a more refined reference material is needed for assay validation. The NIBSC is developing a definitive reference material with more precise quantification of target and non-target DNA and RNA, with well-defined infectious titers that can be used for comparative studies of different assay types. There are a number of challenges associated with selecting model viruses for inclusion in a model virus panel. The biophysical properties of the virion, such as the genome type (RNA versus DNA, single stranded versus double stranded, linear versus circular or segmented), the genome size, the particle size, the virus stability and the presence of an envelope (which can affect the extraction efficiency) can all impact the NGS detection sensitivity. The availability of model viruses of a precisely defined composition enables assay performance evaluation; however, the limits must be defined based on the virus structures. It can be challenging to decide what viruses to include in a panel. The decision on what to include can be influenced by the required biosafety levels, whether the viruses can be easily grown in cells and eggs and whether the panel viruses will be distinct from the vaccine viruses.

A wide range of tests, such as genome and particle counts, genome sequencing, infectious titres, host cell DNA and RNA quantification, consistency of fill and stability and accelerated degradation tests are being run to characterize the viral stock. This is to provide a full genome sequence and to quantify the infectious particles, as well as the background DNA and RNA from the host cells.

Each vial of the reference material will contain a known concentration of virus. The reference material can be spiked into a matrix and then run through an NGS pipeline to estimate the limit of detection. This can be done to characterize the relative limits of detection for all the viruses in the panel.

The NIBSC is producing primary or first order reference standards that could potentially be used as WHO standards. These are typically used sparingly and are intended to last over a long period of time. It is envisaged that the primary reference material will be used in collaborative studies, a limited number of bridging studies, and for definitive validation studies. In future it may also be used to make secondary reference materials which will be more widely available and can be used in method optimisation, for routine run control and for proficiency testing.

The NIBSC is also collaborating with the WHO to investigate the utility of NGS as a molecular test of virus stocks used in the manufacture of oral polio vaccine (OPV). Well-characterized mutations are known to occur in the Sabin polio vaccine strains used to manufacture OPV. These mutations are associated with reversion to neurovirulence and need to be monitored to ensure the safety of the final OPV product. This testing has traditionally been conducted using MAPREC, which is an established but technically demanding test. The collaborative study aims to compare amplicon-based NGS and MAPREC for the quantification of type 3 OPV mutations. Preliminary results indicate that there is excellent correlation between NGS and MAPREC, and that NGS could be used as an alternative to MAPREC. Testing of OPV type 1 and type 2 vaccines is ongoing.

Dr. Siemon Ng (Sanofi Pasteur) and **Dr. Cassandra Braxton** (Biogen) presented an overview of preliminary results from two ongoing collaborative spiking studies in the AVDTIG to support standardization of NGS for use in adventitious virus detection. The initial spiking study 1, was carried out by three collaborating laboratories in 2015 [13]. This study evaluated the potential sensitivity of NGS for the detection of different viruses in HeLa cells by comparing different sample preparation approaches and technologies. Five reference viruses were identified and selected based on the results of this study for use in further NGS studies.

Spiking study 2, started in 2016 with aims to further explore the impact of different sample preparation methods on the sensitivity of NGS for virus detection in common matrices. This study involves fifteen participants across academia, industry, government, and contract research organizations, and was divided into two spiking studies, which started in 2017. Spiking study 2A aims to evaluate an adventitious virus known to infect CHO cells and is of particular interest to the biopharmaceutical industry. This study goal is to evaluate detection of murine minute virus (MMV) in the presence of CHO cellular background. There are seven groups participating in the study and the study will include three main experiments. Study participants are using shared protocols for extraction and library preparation, so that the consistency of results across the different sites can be evaluated. Experiment 1 utilizes a sample at a ratio of four CHO cells to every MMV genome copy. Experiment 2 uses a CHO + MMV sample comprising a fixed CHO cell density plus serial dilutions of MMV. Experiment 3 comprises a CHO + MMV infection. Experiment 1 is complete, experiments 2 and 3 have not yet started. Experiment 1demonstrated a high variation in the number of MMV reads and in the breadth of coverage, both for the CHO + MMV sample and for the control (supernatant + MMV). All groups were able to detect MMV with a coverage of between 8 and 91%. A higher level of detection and breadth of coverage was observed in the absence of ribosomal depletion. Ribosomal depletion decreased the total number of reads and MMV specific reads. MMV was detected by different sequencing platforms. Consistent results were reported across all the participating laboratories regardless of the sequencer or the bioinformatics platform used, which were selected by each study participant.

Spiking study 2B aims to evaluate the performance and sensitivity of the NGS for adventitious virus detection in a mixed virus sample (mimicking a virus seed). The five FDA reference viruses described by Dr. Arifa Khan (FeLV, RSV, Reo, EBV and PCV1) were spiked into a background of purified, high-titre adenovirus (Ad5). At least two spiked levels were used, one common level across the participants and one variable level, along with an optional negative control. Participants used their own sample preparation protocols and bioinformatics analysis across the different sequencing platforms. Five of the eight participants have shared preliminary results in the study group. Preliminary results indicate detection sensitivity of at least 10^4 viral genome copies per mL and the detection of some additional retrovirus signals expected to be in the sample. Further examination of the unexpected and expected signals is being conducted with a comparison of the different sample preparation methods and bioinformatics pipelines used between participants. Final results will be shared at the end of the study, which is expected to be submitted for publication in 2020.

Discussions for spiking study 3 were initiated in 2019 with 15 participating groups. It aims to evaluate the sensitivity of transcriptomics for the detection of viruses in cell substrates using an infected cell line spiked into an uninfected cell line. Study details are being developed for the study to start in 2020.

Dr. Fredrik Granberg (Swedish University of Agricultural Sciences), presented on the mandate of the World Organization for Animal Health (OIE) to develop international standards, guidelines and recommendations for animal health, including standards on quality of veterinary services. These standards are prepared by working groups, specialist commissions and expert groups and are adopted by the OIE

World Assembly of Delegates. These standards are compiled in the OIE's Codes and Manuals. In 2013, the OIE updated its chapter on the Principles and Methods of Validation of Diagnostic Assays for Infectious Diseases in the Manual of Diagnostic Tests and Vaccines for Terrestrial Animals and appointed an ad hoc group on High-Throughput Sequencing, Bioinformatics and Computational Genomics (HTS-BCG). This group advised the OIE to adopt standards for sequencing production, assembly and use in the laboratory, on the farm and at any point along the value chain linking animals to consumers, and to create an OIE platform for the collection and management of pathogen genomic sequences and related metadata, integrated within the existing World Animal Health Information System (WAHIS). These recommendations were formally adopted as Resolution 33 in 2015. The resolution also included recommendations to i) address the challenges and opportunities to the OIE Member Countries' Veterinary Services and ii) that the Network of Reference Laboratories and Collaborating Centers and other partnering initiatives expand the provision of support to the OIE Member Countries.

In addition, a chapter on Standards for High-Throughput Sequencing, Bioinformatics and Computational Genomics (HTS-BCG) was included in the OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. This chapter provides information on the role of sequence data in the diagnosis and management of microbial infections and it advises that the generation, management and use of genomic sequence data should be integrated within accepted practices of veterinary investigations and within a laboratory's quality assurance system. It provides information on how to incorporate HTS-BCG in veterinary investigations, including how it can be used to diagnose, confirm and characterize infectious agents. It outlines standards for the use of HTS-BCG, in particular that the technology platform should be fit for its intended purpose. For instance, if it is intended for use in the detection and investigation of outbreaks, it should be able to quickly generate sufficient data for metagenomics-based detection. It provides guidance on sampling and reporting, including that HTS-BCG should be managed within the context of the laboratory's quality assurance system, that the sampling strategy should be tailored to the purpose of the investigation, and that epidemiological data and data on the pathogenesis of the organism in the animal species under investigation should be considered when interpreting the data. The chapter also provides guidance on specimens and sample preparation, including that precautions should be taken to ensure the integrity and quality of nucleic acids, that positive and negative controls should be included to verify each step of the procedure and that confirmation of results would require resampling of the original specimen, so the original specimen must be stored appropriately and protected from cross-contamination. It also provides guidance on the generation of sequence data, including the information on approaches to ensuring the quality control of sequencing.

WAHIS includes the World Animal Health Information Database and web-based interface. The system disseminates information on animal health and sends real-time alerts on animal disease events. A new version of this system OIE-WAHIS will be launched in 2020. OIE-WAHIS will include a pathogen genomic notification system. This web-based platform aims to collect digital whole-genome sequence data and to link these data with the corresponding epidemiological and outbreak data in OIE-WAHIS. It will provide a permanent record of the genetic sequences of pathogens that have been the subject of country reports to the OIE. It will also include quality information so that the reliability of the data can be considered when interpreting the data. This will be an open information system, so as to facilitate transparent sharing of information on the global animal health situation.

5. Validation and applications of NGS for adventitious virus testing

NGS is a promising alternative method to the traditional *in vivo* and *in vitro* adventitious virus detection assays. Validation of NGS is needed to demonstrate that the assay is fit-for-purpose and may involve a multistep strategy. Validation using panels of model viruses allows the sensitivity, specificity and breadth of detection of the assay to be confirmed. Furthermore, the strategy for follow-up of positive signals needs to be defined. The viral database used for validation studies can impact on the number of false positive and false negative results. False positive results may also be associated with the reagents and components used in the assay, in particular those that are animal-derived. In the event of the detection of a viral contaminant, the significance of this finding will depend on whether the virus is infectious A risk assessment and management approach is also advocated for preventing contamination of vaccines with extraneous viral adventitious agents.

Dr. Sebastiaan Theuns (Ghent University) gave an overview of advances in NGS sequencing technologies. It was emphasized that viral genomes are relatively small compared to bacterial and eukaryotic genomes and this is why viral genome enrichment is important when applying NGS on samples [32]. When analysing long and complex genomes, the use of long read technologies is useful; however, this is currently still evolving. A distinction between short-read platforms (Illumina, Ion torrent and MGI) and long-read platforms (Oxford Nanopore Technologies and Pacific Biosciences) was made.

In the Illumina platforms, which is currently most used, the sequences are clonally amplified and then sequenced during a process called sequencing by synthesis. This is a PCR-based method, which can introduce some bias as for instance GC rich regions can be missed. Illumina has a wide range of sequencers on the market, such as the iSeq100, MiSeq, NextSeq, HiSeq and NovaSeq, which can generate outputs ranging from approximately 1.2 GB to 6 TB per flowcell. Reads generated are highly accurate (mostly >99% accuracy) but short, ranging from 1 × 36 bp (single end) to 2 × 300 bp (paired end). The sequencing times for these devices are lengthy and range from 17 h to 3.5 days.

Pacific Biosciences produce long-read sequencers such as the Sequel II, which is a high-performance sequencer that is highly accurate (up to 99.99%) and yields long read lengths. It can process up to 8 million reads and can deliver up to 150 gigabases. Read lengths between 15 and 20 kb can be delivered, but 150–250 kb read lengths may be achieved using specific approaches.

Oxford Nanopore Technologies produces the MinION, which is a hand-held long-read sequencing device that can yield up to 30 GB per MinION flowcell. These reads can be processed in real-time once the sequencing starts. The MinION is part of the ONT family of devices which includes the Flongle, GridION X5, and PromethION 24 or 48. They also have portable sequencing hardware such as MinIT and MinION Mk1C. Flongle is a single use flowcell that can be used for small projects (1–2 GB output) that require a fast analysis and results. PromethION is useful for human or plant genome sequencing. The MinION MK1C is a mobile MinION with integrated computing capacity and a screen. There have been questions about the accuracy of the Oxford Nanopore devices, however this is improving, and they are moving towards achieving >99% single read accuracy. The most recent R10 nanopores have a dual reader head and can thus capture a larger sequence and improves accuracy (~95%). An example of the practical and promising use of nanopore sequencing was given. It was used to detect porcine viruses (porcine epidemic diarrhea virus and rotavirus A) from cell culture and in a few hours, whole genomes were assembled and allowed full identification and typing of the present strains, without random or gene specific amplification of the viral genomes [33].

In general, long-read sequences enable an easier and faster assembly of repetitive, large and complex genomes than short-read sequences. Sequencing generates a large volume of data and requires hardware that has the capacity to store and process reads that equate to terabytes of data. Sequencing analyses previously relied on central processing units (CPU) which could take days to process the data and which are designed for the execution of a single task. Recently NVIDIA developed

a graphics processing unit (GPU) architecture that facilitates processing in a few hours and allows for the simultaneous completion of a small, identical tasks in parallel. The NGS field is evolving very fast. It is challenging but important for scientists, regulators and industry to keep up with these changes.

Dr. Siemon Ng (Sanofi Pasteur) gave an overview of the Sanofi Pasteur approach to NGS assay validation for adventitious virus detection in viral seed matrices. In response to the complexity of NGS and the associated analytical and technology challenges, Sanofi Pasteur has developed a modular approach to NGS assay validation. Sanofi Pasteur's adventitious virus test is designed to test any type of biological material from cell banks, viral seeds, crude harvests, drug substances and drug products, using a whole-genome approach that takes advantage of very deep sequencing. The steps include:

1. Nucleic acid extraction from a sample, using an extraction protocol that covers the extraction of different types of DNA and RNA and that is based on a comparison of eleven different extraction methods tested using EBV, RSV, Reo and FeLV [34].
2. Reverse transcription for RNA viruses and second strand synthesis for single stranded DNA viruses.
3. Sequencing using the Illumina HiSeq1500 Sequencing System platform
4. Data processing and analysis using in-house PhyloID software, which is an automated pipeline developed by Sanofi Pasteur for adventitious agent detection from NGS data [35].
5. Follow-up investigation (if necessary).

Positive signals are further assessed to determine if it is a true positive that requires additional laboratory confirmation. Samples are considered to be free from viral contaminants and are passed if: the signal is from a single molecule (suggesting it is background noise); the nucleic acid is expected (for instance if sequencing a viral seed block); or the sequence is from a bacteriophage, a protist virus, a fungal virus, a plant virus or a retrovirus (as these are either tested using other assays or they are not considered a risk for the manufacturing process). If the signal passes these criteria, further bioinformatics analysis is required, in which case the sequence's best match against NCBI must be to a non-retroviral animal virus (as retroviruses are dealt with using another test). Furthermore, at least 80% of the hit sequence's length must align with the viral genome sequence, with at least 60% identity; this helps to remove background noise associated with a poor overall match. If the virus is expected to be present in the sample (for instance if analysing a viral seed or a bulk harvest), then the sequence identity should be at least ten percentage points greater than the match to the expected virus. Finally, the total extent of the genomic coverage for the sequences passing the above criteria must be at least half of the expected genomic coverage based on chance. Any remaining signals are then followed-up with further laboratory analysis to determine if the signals are from infectious viral particles.

Assay controls are also necessary to minimize potential environmental contaminants and to monitor sample extraction and library preparation. Quality control steps are important for checking the performance of the assay, for assessing the quantity and integrity of the starting material and for assessing the size and quantity of the sequencing library. The personnel flow is similar to that used for PCR rooms when running a PCR analysis, with personnel moving from a nucleic acid extraction room to an amplification room (for both pre and post amplifications) and finally to a sequencing room. Differential airflow is used in each room to minimize contamination, and all work is done in dedicated biosafety cabinets.

Control sample preparation involves spiking each sample with two types of pseudo-viral particles containing DNA and RNA separately prior to nucleic acid extraction. The controls are designed to be sequences that are not expected to be in any of the samples or in the environment. A system control is used to monitor the extraction efficiency, the cDNA synthesis, library generation and sequencing. The number of reads matching the expected DNA and RNA sequences are identified by PhyloID and are tracked and used as validity criteria. Each sample is dual-indexed so that each sample can be easily identified and distinguished from previous samples. All testing samples are run in individual lanes or flow cells and are not multiplexed. Nucleic acids are assessed for concentration and size using the Agilent BioAnalyzer 2100, both for quantitation and for assessing the size of the sequencing library.

A pre-made dual-indexed sequencing library is added to every sequencing run as a sequencing control. The sequencing process also requires a minimum number of reads to ensure that the test meets the necessary sensitivity, as well as a minimum average read quality score. For the PhyloID analysis, quality control steps implemented on the data include removing the adaptors and filtering out low quality reads to remove sequencing or PCR errors that might lead to misidentifications. This helps to strengthen the robustness of the assembly step and subsequent identification.

A validated panel of model viruses is used to assess the performance of the Sanofi Pasteur NGS assay for adventitious virus detection. This panel is based on viruses previously developed by an NIH study that demonstrated *in vitro* assays have greater sensitivity for virus detection than *in vivo* tests [19]. The 16 viruses were representative of potential contaminants that could be introduced during vaccine production. The viruses are: bovine viral diarrhea virus, coxsackie A16, coxsackie B3, echovirus 11, rhinovirus 12, rubella, bovine parainfluenza type 3, influenza A, measles, mumps, vesicular stomatitis virus, adenovirus 5, adenovirus 41, herpes simplex virus type 1, simian cytomegalovirus, and simian virus 40. Sanofi Pasteur's NGS adventitious virus detection assay is validated as a limit test and not as a quantitative assay. Its sensitivity has been demonstrated to be better than the corresponding *in vivo* tests and provides a justification for streamlining the testing package by replacing the *in vivo* tests with NGS for submitting to the regulator.

Dr. Blandine de Saint-Vis (Boehringer Ingelheim Animal Health) presented her team's approach to using validation data to support NGS purity testing to demonstrate absence of adventitious agents in veterinary master seeds. Purity testing for veterinary vaccine is complicated because the vaccines are developed for several target species and using different sourcing of materials, so testing must cover all the viruses that are relevant to these species. Using broad cell systems for purity testing is cumbersome due to the extraordinary diversity of viruses and species.

Dr. de Saint-Vis described the different steps involved in vaccine development. The first step typically involves working with starting materials of biological origin such as the master cell bank, working cell bank, master virus seed, working seed and other raw materials. These materials are tested with a variety of assays that screen for bacteria, fungi, mycoplasma and viruses, using mainly cell systems. The team at Boehringer Ingelheim is interested to use NGS to systematically analyze the master virus seeds, either as a replacement or as a complement to conventional testing methods. The team is developing a strategy for rapid seed purity evaluation and for risk assessment and avoidance, with a view to accelerate vaccine development. Testing using current conventional methods can take over six months, compared to approximately two months using NGS. For a master virus seed, a random sequencing of all DNA and RNA nucleic acids is performed. The study protocol is designed to detect a target of at least 130 million paired-end reads per sample. This is a complex process to validate. It has required the development of a standardized protocol for DNA and RNA extraction, as well as defining the sequencing technology and methodology, including criteria for sequencing quality. It has also been necessary to define the approach of bioinformatics analysis, including the use of internal controls, to define the criteria for a true positive, and to specify the additional tests that need to be conducted to confirm that a positive signal is actually a contaminant.

In order to validate the use of NGS for viral purity testing,

Boehringer Ingelheim, in collaboration with Viroscan3D conducted an NGS spiking study of two master virus seeds. Seven different viruses (four RNA and three DNA) of known titre were spiked. The study was designed to evaluate the sensitivity and limits of detection of the NGS method. The study utilised the Illumina NextSeq platform and a target of at least 130 million reads for DNA and RNA Seq sequencing is expected. The viruses used in the spiking study included the RNA viruses canine distemper virus (CDV), porcine reproductive and respiratory syndrome virus (PRRSV), the retrovirus RD114, and bovine viral diarrhoea virus (BVDV), as well as the DNA viruses canine parvovirus (CPV), porcine circovirus (PCV2) and turkey herpes virus (HVT). These viruses were broadly representative according to viral families and species, virion size, major veterinary concerns, and availability. The stocks were subjected to viral titration by qPCR and infectious titer.

Two master virus seeds representing two types of production were tested. Porcine circovirus type 2b (PCV2b) was produced in a baculovirus expression system, and avian influenza H9N2 in allantoic fluid (egg). The 4 RNA-viruses [canine distemper virus (CDV), porcine reproductive and respiratory syndrome (PRRS), retrovirus RD114, and bovine viral diarrhea virus (BVDV) and the 3 DNA-viruses [canine parvovirus (CPV), PCV2, and herpesvirus of turkey (HVT) were spiked simultaneously into the MSVs at four concentrations (100-10-1-0.1 CCID50/ml) and analysed by NGS, and as a control by specific digital PCR. The DNA Seq and RNA Seq read mapping for the PCV2b study resulted in a high and similar number of reads, regardless of the condition of the spiking, and the number of viral references were reduced in accordance with the dilution of the spike. As expected CPV, PCV2 and HVT were detected by DNA Seq whereas RD114, CDV, and BVDV by RNA Seq. However, no PRRSV was detected. Six out of seven viruses were observed by NGS, and the sensitivity was close to that achieved by cell culture (which is between 50 and 0.1 CCID50/ml depending on the virus type and cell culture system). To understand why PRRSV was not detected by the NGS assay, the eight spiked samples and the corresponding libraries were analysed by specific digital PCR to verify the presence of the seven viruses in the master seed after RNA and DNA extraction and during the NGS process. The investigation showed that the root-cause for the absence of PRRSV was probably the RefSeq database used for the bioinformatics analysis. The RefSeq 10 K seq database was used because it has few redundancies, but the database is not exhaustive and appears to contain insufficient data to enable the detection of highly variable RNA viruses such as PRRSV. The RefSeq database, although reliable in the event of positive results, has poor reliability in the event of negative results, leading to false negative results.

The RefSeq database was compared to the Genbank database, which contains the largest number of sequences (about 3 million). The RefSeq has few redundancies, but is not exhaustive, whereas the Genbank is exhaustive but has redundancies and as a result is associated with an increase in false positive results, and a decrease in false negative results; however, the strong redundancies resulted in a signal drop, which may impact coverage. The analysis was partly repeated using the RVDB database [7]. The use of the RVDB database resulted in better PRRSV detection, with little difference in the detection of the other viruses, demonstrating the importance of the database used, especially for highly variable RNA viruses such as the PRRSV. Further analyses using other databases are planned, as is a study comparing different bioinformatics pipelines. Boehringer Ingelheim are also considering the creation of a complementary database based on the list of viruses listed in the Ph. Eur. The overall intention is to define the best bioinformatics solution including the database pipelines to unambiguously replace the classical cell detection system assays in the future.

Dr. Dawid Walas (Merck Group, UK) presented a case associated with the investigation of a bovine polyomavirus 3 (BPV3) signal in a viral seed stock. Initial investigation included a qPCR endpoint assay. The extraction was performed using QIAGEN chemistry including carrier RNA. The extraction included plus and minus DNase steps (in separate reactions). Carrier RNA (cRNA) was included in the extractions – cRNA is routinely included in extraction of nucleic acids for many molecular assays like PCR in order to increase the yield of extracted nucleic acid, however, it is typically not used for NGS-based assays. All reactions were prepared in triplicate. The limit of detection for the BPV3 qPCR was 200 copies per reaction (500 ng starting nucleic acid quantity, including cRNA), and the target length was a 70-mer. An equine viral vector seed stock was originally tested using a qPCR panel for regulatory purposes. An unexpected signal for BPV3 was detected close to the limit of detection for the BPV3 qPCR assay in several samples. The same results were detected during a repeat qPCR. An investigation was initiated to confirm the presence of BPV3 in the test sample using NGS, and after a positive result, to further confirm if the intact genome was present, long range PCR coupled with amplicon-based sequencing by NGS was done.

The NGS analysis aimed to assess the impact of carrier RNA (cRNA) and background nucleic acids on the original result. The NGS analysis was split into two tests. The first test was designed to mimic the original qPCR conditions. For the first test, nucleic acid extraction with cRNA was conducted, no DNase was used, and the library was prepared without whole genome amplification (WGA). The assay used 100 times less starting material (5 ng). The second test was identical except that cRNA was not included in the extraction step. The samples were sequenced on both MiSeq and NextSeq and bioinformatics was performed with the aid of an in-house proprietary viral database and an adventitious agent algorithm. For the first test, MiSeq sequencing generated over 35 million reads and the NextSeq sequencing generated over 400 million reads. No BPV3 sequence (reads) were detected in either data set. For test two, in which nucleic acids were extracted without cRNA, the MiSeq sequencing generated almost 60 million reads, and three BPV3 reads, corresponding to two different GenBank accession numbers, were detected. The NextSeq results were pending at the time of the meeting. As expected, the cRNA in the sample generated significant background sequences or noise that made it more difficult to detect adventitious agents using NGS, even when generating a large number of reads on a higher throughput instrument (*i.e.* on the NextSeq). The information gained from the detection of the BPV3 reads by NGS was used for primer design for long-range PCR. That is, the PCR amplicon was designed to span a 2 kilobase region of BPV3 genome (approximately 40% of the genome length) using these anchoring sequences in an effort to confirm intactness of the genome. The long-range PCR yielded amplicons of the expected size. Surprisingly, long BPV3 amplicons were also detected in the negative no-template qPCR controls. Commercially available master mixes are often proprietary, and the vendor may not be able to confirm if the reagent contains components that could affect NGS-based adventitious agent detection. The vendor confirmed that the master mix contained bovine serum albumin. Sequencing of the amplicons demonstrated 99% identity to the BPV3 reference sequence (accession number AF406967), as opposed to 96% similarity observed for the original BPV3 fragment detected by NGS (*i.e.* derived from sequence analysis of the original viral seed stock).

This indicates that care must be taken when qualifying the reads and when deciding how to conduct further investigations of positive signals. Master mixes that contain bovine albumin may be unsuitable for the detection of bovine adventitious agents, such as BPV3, and may be problematic for the downstream detection of true hits. It is important to understand whether the reagents and components used in an assay impact the results, especially when trying to confirm a positive signal. A single round of confirmation may be insufficient and may generate a false positive result. Due diligence must be put in place to qualify the reagents for downstream applications, and not only for NGS reagents. Appropriate controls must be applied regardless of the testing.

Dr. Marc Eloit (Institut Pasteur and PathoQuest) described newly developed processes for specifically identifying cell infection by live viruses, either through the analysis of the cell transcriptome by

stranded RNA analysis or by metabolic labelling of RNAs [14,15]. Tested cells can be master cell lines or cell therapy drug substances, or cells used in *in vitro* assays to amplify very small amounts of virus that are present in for instance working or master cell lines, in raw materials, or in seeds. The technique involves detecting newly synthesized RNAs using NGS. It is non-hypothesis driven, but has a wide range of detection, including the detection of unknown viruses, and it detects only live viruses in infected cells. The technique can also be used to investigate non-inactivated viruses in vaccines or to demonstrate reversion to replication of replication-incompetent viruses.

The method relies on two pillars. Firstly, RNA is used as a biomarker of an infected cell. Virus particles can contain different types of nucleic acids, such as single stranded RNA (ssRNA), double stranded RNA (dsRNA), single stranded DNA (ssDNA) or double stranded DNA (dsDNA). However, messenger RNA (mRNA) is common to all viruses in infected cells and to all live bacteria. RNAs are expressed during the replication of all viruses, including DNA viruses and most latent viruses. RNAs are expressed by live mycoplasma. The level of viral or mycoplasma RNA expression is often high and easy to detect. Highly expressed cellular RNAs can be technically eliminated before sequencing, which increases the sensitivity of detection of viral or mycoplasma sequences.

Using RNA as a biomarker of infected cells is based on i) using the ratio of viral RNA strands (described by Dr. Audrey Brussel), and/or ii) detecting only newly synthetized RNAs. Testing newly synthetized RNAs takes advantage of the short half-life of mRNAs. The specificity of testing is based on identifying a subset of RNAs synthetized in the few hours prior to harvesting the cells, by metabolic labelling of RNAs a few hours before cells harvesting. These labelled RNAs are detected during sequencing. This can be combined with the stranded approach.

The specificity of detection for infected cells using NGS coupled to metabolic labelling, limits time-consuming follow-up studies and adds mycoplasma detection and identification to virus testing. The limit of detection using EBV-infected B95-8 cells in a background of non-infected lymphoid cells has been evaluated [15]. In terms of the range of detection of RNA-Seq, adventitious virus has been identified in a human lymphoid cell line [15], and contamination by squirrel monkey retrovirus (SMRV) in a Vero cell line has also been identified [14]. This technique has also been demonstrated to enable the differentiation between active and inert sequences in persistent viral infections [14]. Regarding analytical sensitivity of RNA-Seq, acute infection at very low doses is detected early after infection [15]. The ability to differentiate between active and inert sequences was demonstrated using a model of the tick-borne encephalitis virus (TBEV) in Vero cells [14]. The NGS transcriptome can also be used as a read-out of infectivity assays to specifically detect replicating viruses. Finally, this technique can also enable the identification of replicating viruses or live bacteria from a large background of inert sequences from the same strain as found in testing for complete inactivation of vaccines.

Dr. Audrey Brussel (LFB, Les Ulis) described efforts to improve NGS specificity for detection of replicative viruses in cells. Viral nucleic acid patterns that are specific for viral replication and that could therefore be used as a marker of viral infection were identified for each virus class. The replication-specific nucleic acid in infected cells for DNA viruses was RNA; for negative ssRNA viruses it was positive RNA; for positive ssRNA viruses it was negative RNA; for retroviruses it was DNA and positive spliced RNA; and for dsRNA viruses it was an excess of positive RNA. Using a stranded RNAseq NGS approach it was possible to identify in a single step all of these markers of infection.

The performance of this technique was then evaluated using two cell/virus systems which mimicked low levels of infection. Part 1 of the study aimed to evaluate the assay with a DNA virus (Epstein-Barr virus, EBV) establishing a latent infection, and to determine its sensitivity as compared to a PCR method. Serial dilutions (from 10^{-3} to 10^{-8}) of latently infected cells (B95-8, EBV-positive monkey lymphoblastoid cells), were inoculated into non-infected cells (Ramos, EBV-negative human lymphobastoid cells). Cell were pelleted and split into two samples, one for RNA extraction and sequencing using stranded RNA Seq NGS and the second for DNA extraction and analysis by PCR. EBV was detected by NGS in latently-infected cells with similar sensitivity to PCR, but EBV replication could only be demonstrated with the new NGS approach as it targets a viral replication-specific nucleic acid. This also identified other viral signals. In particular, a high number of murine leukemia virus (MuLV) reads were found in all samples including in unmixed Ramos cells, implicating the Ramos cells as the most likely origin of this signal. Follow-up investigation of the MuLV signal demonstrated that the reads mapped to the full MuLV genome. In addition, translated nucleic acid sequences did not indicate any unexpected stop codon and finally the presence of spliced viral RNAs, which are a marker of retrovirus replication, was detected. All these elements were consistent with Ramos contamination by an infectious xenotropic MuLV. Indeed, further follow-up investigations using an infectivity assay confirmed Ramos cell contamination by an infectious xenotropic MuLV [36]. A low number of BVDV3 reads was also detected, which may have originated from either the B95-8 cells or the serum used for its production.

Part 2 of the study evaluated whether this approach was capable of detecting an early infection of cells with a low dose of viruses with positive strand RNA genome. Madin-Darby bovine kidney (MDBK) cells were infected with a low dose (10^2 TCID50/ml) of bovine viral diarrhea virus 2 (BVDV2) (or $10^{4.7}$ TCID50/ml of gamma irradiated BVDV2 as a carry over control) and the kinetics of infection were examined over a six-day period. Total RNA was extracted from the cell pellets and both sequenced using stranded RNA Seq NGS and tested with BVDV2 RT-PCR (the reference method). For the RT-PCR, a large increase in BVDV2 signal was observed, after which it remained constant for six days. The signal for the irradiated BVDV2 was much lower and decreased over the time-period. The NGS results were largely comparable to the RT-PCR results in terms of sensitivity and early detection. BVDV2 replication was unambiguously detected by stranded RNA seq NGS which showed the presence of negative strand viral RNA [15], As a conclusion, the stranded RNAseq NGS was demonstrated to be able to specifically detect live viruses in cells, which can reduce the rate of false positive results. This NGS method was demonstrated to have a similar sensitivity to PCR or RT PCR and to also have a broader range of detection, in two virus/cell systems at a low level of infection.

Dr. Jean-Pol Cassart (GSK Vaccines) proposed a new perspective on how NGS can be used in industry to control viral contamination of vaccines and raw materials not only as a detection assay but as a mitigation tool. Viral risk assessments and their derived control actions are part of a strategy for preventing contamination of vaccines with extraneous viral adventitious agents. Risk control can include careful selection of raw materials, the conduct of viral clearance studies and virus screening with detection assays. Decisions regarding acceptance of risk must be justified by studies, which inform on how, why and when a risk could be accepted. Currently, our knowledge about virus diversity is increasing due to the performance of new, unbiased high-throughput sequencing technologies. This creates new challenges in terms of control of the risk. For instance, for newly discovered viruses there is no viral stock for performing viral clearance studies and no infectivity assays for testing. PCR could represent an immediate solution to control the risk. But PCR approaches are facing two difficulties: (1) the genetic diversity of the virus and 2) the inability to distinguish between inert and live viruses. For instance, after gamma irradiation of serum, inert virus nucleic acids are still detectable by PCR. In addition, long range PCR cannot be applied for segmented viruses. NGS allows to address virus diversity and detect unknown virus types but cannot distinguish between live and inert viruses when the method is directly applied on the sample.

The proposal is to use NGS as a read-out for infectivity assays. Using fetal bovine serum (FBS) as biological background, indicator cells can be cultured with the serum and investigated for viral RNA as a marker

of virus replication in infected cells. NGS can be used to search for a replicative signature, such as: (1) an increase in the number of viral sequence reads (for all viruses), (2) the presence of RNA transcripts (for DNA viruses), (3) a change in the ratio of RNA strands (for RNA viruses) or (4) by conducting metabolic labelling of the newly synthesized RNA with a 4-thiouridine (4sU marker) (for all viruses). These signatures are markers of virus replication in the cells.

Four different viruses were used in the study: human adenovirus 5 (Ad5) (dsDNA), human parainfluenza virus 3 (PIV3) (ssRNA), human coxsackievirus B3 (CSVB3) (ssRNA+) and non-cytopathic BVDV2 (ssRNA+). Two cell lines were selected: (1) MRC-5, a production cell substrate commonly used in vaccine manufacturing, able to detect Ad5, PIV3 and CSVB3, and (2) MDBK, a bovine cell substrate used in the present study to detect BVDV2. The design of the study was earlier described by Dr. Marc Eloit. Samples mimicking non-replicating virus consisted of cells cultured at 4 °C for 2 h with FBS spiked with a high-dose of virus (14,000 $TCID_{50}$). On the other hand, samples corresponding to replicative virus consisted of cells cultured at 37 °C for up to 21 days with FBS spiked with 1, 10 or 100 $TCID_{50}$.

The first signature presented is the evolution of the number of viral reads along the culture time. A net increase of viral transcripts is observed for all four viruses, indicating viral replication. Importantly, while conventional *in vitro* assays for adventitious virus detection were conducted simultaneously, the significant increase of Ad5 reads was detected before a positive cytopathic effect (CPE). This replication signature requires at least two timepoints to study evolution of reads number.

The second signature presented takes advantage of the stranded analysis of viral transcripts to specifically detect the appearance of antigenomic reads post-infection or a change in the RNA strand ratios between genomic and antigenomic reads. For the negative stranded RNA virus PIV3, the number of transcripts (antigenomes) clearly increases after the culture time. In addition, a net change in the RNA strand ratios is observed. For the positive stranded RNA viruses CSVB3 and BVDV2, the number of reads corresponding to the negative RNA strands (antigenome) increases, but the change in the ratios is more discrete as most of the signal comes from the positive genome and mRNA, representing almost 100% of the reads). This replicative signature also requires more than one timepoint.

The third type of signature is based on the metabolic labelling of the viral RNAs using the 4sU, a modified uridine nucleoside. Study design is the same as before, except that cell cultures are treated with 4sU at the end of the culture time. 4sU nucleosides incorporate within newly synthesized RNA. After extraction, half of the RNA is alkylated. This alkylation allows the chemical modification of the 4sU, which has consequently more affinity to G nucleoside than A. At the end, incorporation of a 4sU leads to a T > C conversion, represented by conversion indexes. In absence of 4sU treatment, the conversion index is close to one. A similar result is observed in presence of non-alkylated 4sU, On the other hand, in presence of alkylated 4sU, the conversion index massively increases indicating viral RNA synthesis. By the means of the stranded analysis of viral RNA, it is also possible to select the RNA strand labelled by 4sU and separately analyze 4sU incorporation in genomic and antigenomic reads. For PIV3, 4sU the conversion rate is higher in antigenomic reads, while for CSVB3 and BVDV2 the conversion rates are similar in both genomic and antigenomic reads.

The NGS transcriptome assay coupled with 4sU metabolic labelling represents a promising method to detect and identify viral sequences in a biological sample and to demonstrate that the identified sequences correspond to replicative virus in indicator cells. The replication/transcription activity of all virus types tested can be detected. A single timepoint and culture condition are sufficient. A consensus sequence can be derived from labelled RNA and subsequently used for the conversion analysis.

Further optimisation of the assay based on 4sU labelling is required, in particular in relation to the concentration and timing of the labelling and in relation to the selection of informative RNA strands and regions. In addition, an NGS transcriptome could be used in industry to confirm the absence of virus in raw materials used for vaccine production (such as FBS), and to demonstrate that detected viral sequences are not corresponding to replicating viruses. Finally, it is important to select relevant indicator cell substrates, such as production cells and cells from the same animal origin or broad range cells (such as Vero cells).

Dr. Katarina Bacnik (National Institute of Biology, NIB) presented on use of NGS in NIB and their approach to adventitious virus and non-product DNA detection in biologicals. The NIB uses NGS for a number of purposes. The NIB collaborates with industry and uses NGS to define the metagenome of production cell lines in the pharmaceutical industry, to define residual DNA in gene therapy products and to conduct genome sequencing and mutation analysis. The NIB uses NGS for the analysis of microbial populations in soil and water samples, for the discovery of new or unexpected pathogenic viruses, for plant virus diagnostics, for the analyses of virus population structures and their changes through evolution, for genome sequencing of new bacterial species, for metagenomic analysis of water samples, and for resolving repetitive regions in the viral genome. Finally, the NIB also conducts transcriptomics; micro RNA analyses and they screen food products for the unauthorized use of genetically modified organisms.

A series of case studies to illustrate how they use NGS were presented. The first of these was a metagenomic analysis for the detection of adventitious viruses in a selected CHO cell line with comparison of sample preparation approaches. This study aimed to test various approaches for sample preparation, to establish a pipeline for metagenomic analysis of sequencing results and to characterize the standard viral metagenome of selected production CHO cell lines. The experimental plan involved preparing two aliquots of CHO cells with growth medium. One sample was treated with liquid nitrogen to lyse the cells, the other aliquot was left untreated. In both samples the cells were separated from the growth medium. The growth medium samples were treated with ultracentrifugation or filtration, and both DNA and RNA were extracted. For the cell samples, total RNA was isolated, this was followed by ribosomal RNA depletion. All samples were shotgun sequenced using the Illumina MiSeq. Processed media, fresh media and nuclease free water were used as negative controls. The bioinformatics pipeline for the metagenomic analysis involved an initial quality control step using the CLC Genomics Workbench, followed by direct similarity search of sequencing reads on protein level (against nr database) and subsequent taxonomic classification and visualization. Viral-like reads were extracted from all the sequencing files and further analysed to possibly confirm/reject that they correspond to viral genomes. The majority of reads in the samples mapped to the host genome. The majority of viral-like reads corresponded to possible endogenous retroviruses. A large proportion of the reads were classified as human cytomegalovirus and polyomaviruses, which probably originated from expression vectors. Residual bacteriophages were also detected from some samples, which probably reflected contamination of the substrates used in the analyses. The comparison of the metagenomics composition of the samples obtained by different sample preparation approaches, showed that sequencing ribosomal RNA-depleted total RNA is a promising approach in such studies.

A second case study was presented on metagenomic analysis for the detection of adventitious viruses in selected parental CHO cell lines. The objectives of this study were to characterize three additional parental lines and to optimize procedures for sequencing ribosomal RNA-depleted total RNA from cells. This study used parental cells of diverse origins, and an increased sequencing depth (200 m reads per sample) using the Illumina HiSeq. The metagenomics investigation of the parental CHO cell lines did not indicate the presence of adventitious viral agents in the investigated samples. The study revealed expected background levels of virus-like nucleic acids in the sample, which originated from i) the remains of expression vectors, ii) the endogenized viral elements, and iii) residuals of bacteriophages nucleic acids. The

contaminants detected in the negative control included in the sequencing and samples were probably introduced during the extraction or sequencing steps and likely reflect contamination of the chemicals used during these processes.

The presentation also discussed how the NIB uses nanopore technology to define residual DNA in gene therapy products. The NIB works with samples of purified adeno-associated viruses (AAV) for gene therapy. The viral capsids of AAV have the potential to pack other nucleic acids besides the intended construct, including plasmids from transfection and host-cell DNA. The NIB wanted to identify nucleic acid impurities and their lengths and to test the applicability of the Oxford Nanopore Technologies MinION sequencer for AAV samples. The nanopore sequencing is a rapid, onsite, scalable and affordable sequencing platform that generates very long read lengths. NGS can be used to characterize therapeutic viruses and nicely complements other characterization methods.

6. Follow-up of NGS signal detection

As with any nucleic-acid based technology, there is a risk of false positives and false negatives when using NGS for adventitious agent detection. When using NGS for risk management, the use of appropriate standards and internal controls is critical for distinguishing between false positives and *bona fide* contamination. Furthermore, identification of a positive signal should be coupled with an action plan for risk mitigation, including supplementary methodologies to exclude false positives. This approach can support use of NGS to minimize risk in diagnostic or research settings, including the managing/minimizing of false positives, development of action plans to mitigate risk, and procedural changes associated with detection of true positives.

Dr. Joseph Victoria (Boehringer Ingelheim) presented three case studies on the uses of NGS in early R&D for animal health. The first case study reported on the use of NGS by their R&D team in China. This team is working on generating a Pre-Master Seed Virus (Pre-MSV), a live attenuated virus through serial passage. NGS was used to identify mutations for attenuation and to monitor the stability of the virus:the percentage of sequences obtained for the parental virus was found to be 87.5% of the total sequences. However, the pre-MSV samples exhibited only 6% of the total sequences aligning to the target genome, which raised concerns as there was no evidence of change in the viral titer. Further investigation using *de novo* assemblies was performed, and the presence of porcine parainfluenza virus was identified. Follow-up studies using parainfluenza strain-specific PCR primers were performed using retention samples at passages 0, 15, 30, 45 and 53. The retention samples were all negative until passage 53, which tested positive. Further investigation indicated that in an effort to increase the titer, samples were split into two lineages at passage 50. One of the subsamples was grown in a cell line provided by a collaborator that was contaminated with PIV. In this case study the results of the risk assessment were unambiguous and the use of NGS in early research identified the problem before the initiation of expensive animal studies. Furthermore, as a result of this finding a decision was made to include NGS screening for adventitious agents for all externally provided raw materials, as well as at multiple passages for in-process cultures.

The second case study detailed the use of NGS for the detection and discrimination of endogenous retroviral sequences (ERV). Two different cell lines that were earmarked for the growth of vaccine viral stock were tested using NGS and were found to be contaminated with ERV, the first cell line contained 1.4% of ERV and the second cell line contained 0.005% ERV. This raised the question as to whether cell line two could be considered “clean”, and to what degree these results could be used as an indicator of purity. The team set out to establish acceptable thresholds for ERVand to assess whether methodological improvements or standards could clarify this. Sample processing was modified to increase confidence in the ERV source. The DNA and RNA fraction of the cells were separated and similar results were found in terms of the presence of ERV. The supernatant was also separated into DNA and RNA and even higher levels of the ERV was found to be associated with the RNA from cell line one. These numbers increased even further for the RNA upon the application of a virus-like particle treatment. The methodological improvement increased the confidence in the ERV risk assessment. Cell line two was classified as high-risk and cell line one as low risk. As a result of this work cell line two was replaced by cell line one for the production of proof of concept vaccines. All future cells used to produce potential vaccine candidates will undergo modified NGS screening. However, an outstanding, important and unanswered question from this case study relates to the definition of thresholds for regulatory acceptance or reporting.

Case study three focused on the detection of select agent sequences. An in-house MiSeq assay was run on 60 to 70 mixed samples from ten projects. These samples included viral cultures, bacterial cultures and field and tissue samples. For one of those samples, blast analysis performed on *de novo* assembled generated sequences demonstrated the presence of *Burkholderia mallei*. This species is on the select agents list. A decision tree was developed to guide internal actions if screening detects a sequence that is similar to but slightly different to the reference genome. If all hits are 16s rRNA, the reagent is disregarded as contaminated. If the hits are not all 16s RNA, the sequence is matched to the reference genome. If it matches on a single gene or operon the next step is to investigate if the element is integrated. If not, then it can be disregarded as a reagent contamination. If it is integrated, then further examination and a report is needed. If the sample completely aligns to the reference genome, then an investigation is conducted to see if it is present in all or many samples from the run. If it is only found in a single sample, then all the samples are destroyed, and a report is made. If it is found in multiple samples then all or a subset of the samples are investigated to identify the source of the shared reagent contamination. At this point corrective action may be taken to eliminate the reagent contamination. Alternatively, all samples can be destroyed, and a report made. As a result of this work decision tree was established to address this and future identification of potential environmental and reagent contamination.

The implications of NGS for adventitious agent detection varies depending on the methods, the credibility of the data and the agent detected. For early R&D, in-house guidelines were developed based on risk management; however, there is a need for increased regulatory standardization and guidelines.

Dr. Steven Van Borm (Sciensano) presented on the OHEJP-METASTAVA project, which is a project to standardize and validate metagenomics methods for the detection of foodborne zoonoses, antimicrobial resistance and emerging threats. This project is being implemented by Onehealth EJP, a consortium of 41 public organizations covering human and animal health and food. It is co-funded by the European Joint Programme Horizon 2020 Work programme for 2017. The project is running between 2018 and 2020 and there are six partners from Belgium, France, Germany, the Netherlands and Sweden.

As part of the Metastava project a landscape analysis is being conducted to identify the existing data generation protocols, molecular diagnostic guidelines and data analysis protocols for NGS, with a view to using these to develop: i) fit for purpose standardized protocols for the model pathogens: hepatitis E, norovirus, antibiotic resistant bacteria, zoonotic pox viruses and Shigatoxigenic *Escherichia coli*; ii) quality metrics for diagnostic metagenomics, internal and external controls; and iii) analytical validation data for the model pathogens on sensitivity, specificity, repeatability, proficiency testing and interlaboratory reproducibility. These will then be integrated with existing efforts in NGS molecular diagnostics and metagenomics guidelines and standardization to develop implementing tools for the quality control of NGS in public labs for use as an accessory method to targeted diagnostic assays. In the animal health and food sector, NGS is used to test a wide variety of sample types including serum, food, feces and tissue. Given the diversity of model pathogens and samples that are covered by

Metastava, the group opted for a catch-all diagnostic approach. They are developing protocols that can potentially capture all pathogens including viruses, bacteria and eukaryotic microorganisms in a wide variety of animal-human-food matrices. The breadth of the scope may impact on the sensitivity.

Dr. Van Borm presented a veterinary diagnostic case study on how to avoid false positive results. The case relates to an undiagnosed cluster of dairy cattle presenting with high persistent fever, that were unresponsive to anti-microbial and anti-inflammatory treatment and that had redness of the nose and teat [37]. Serum and whole blood samples were taken and analysed using virus targeted NGS using the 454 GS junior. The results included bacteriophages, environmental viruses, isolated reads associated with contamination due to random amplification, 1 contig and 10 singletons indicative of the presence of Parvovirus NIH-CQV-like and BPV, a virus which is associated with neonatal diarrhea, adult respiratory and reproductive disease. The results were confirmed using qPCR targeting the contig and singletons. The extraction controls were positive. The alternative extraction controls were negative. The parvo-like sequence was traced to the nucleic acid extraction spin columns and represented background contamination [38,39].

The next case study focused on avoiding false negatives through the evaluation of quality controls for RNA virus detection. As previously mentioned, the animal health diagnostic sector deals with a large variety of sample types. The Metastava project wanted to develop follow-up strategies for individual samples. In addition to the external controls monitoring run performance, a sample level spike-in control was evaluated (mengovirus extraction control, Biomérieux). The wetlab workflow purposely avoided any enrichment strategies. The data analysis workflow used a fast mapping approach to the mengovirus genome (accession number DQ294633.1), as well as read classification against a custom database using Kraken v1.1 [40]. Two quality control points were implemented, the first being a mengovirus-specific real-time PCR after nucleic acid extraction, the second a sequencing metric consisting of normalized mengovirus read counts. In terms of reproducibility, in all sample types (feces, serum, lung tissue, swabs) the mengovirus realtime RT-PCR assay resulted in reproducible Ct values, allowing quantitative cut-offs for efficient viral RNA extraction to be set. The normalized mengovirus read counts, however, highly depended on the sample type and were more variable, resulting in the preliminary use of an arbitrary cut-off of 1 read per million raw reads, to be further evaluated in follow-up studies. Viral reads accounted for between less than 0.001 and 0.46% of all reads. The two control points can demonstrate extraction efficiency and also competition issues in the sample. They also investigated whether the addition of viral controls would decrease the sensitivity for the detection of other viruses and found that the detection of virus families and the sensitivity for the detection of astrovirus species in naturally infected porcine fecal samples and wild bird cloacal swab samples was unaffected by the addition of the viral controls.

The evaluation of a commercial exogenous internal process control for diagnostic RNA virus metagenomics demonstrated that the mengovirus extraction control kit did not interfere with diagnostic viromics, although the suitability of the virus species used may depend on the application. The evaluation of the mengovirus real-time RT-PCR confirmed the reproducible, inhibition-free nucleic acid purification in individual samples and in experimental batches. It is recommended to include sequencing platform specific QC metrics including the cDNA and library QC and sequencing metrics such as the cluster density, PhiX, read length and quality. In addition, the evaluation of normalized read counts and coverage helped to identify sensitivity issues due to competition and enabled the validation of the sequencing and analysis parts of the workflow. The sample matrix complexity for serum versus fecal versus tissue samples was an important factor in terms of reproducibility. The first results indicate that the addition of mengovirus does not affect the detected virome composition or the sensitivity.

The Metastava project demonstrates the importance of quality control schemes with external positive and negative process controls, as well as the benefits of incorporating platform specific metrics such as sequencing metrics. Exogenous internal controls can potentially be used to validate the process at the individual sample level. Sample and metadata management, good laboratory practices (GLP), follow-up confirmation and careful interpretation are also important.

Dr. Mark Kerstjens (Janssen Vaccines and Prevention B.V.) presented on strategies for managing false positive results. Janssen has adopted a risk management approach for adventitious agents which includes reviewing and assessing the risk associated with the raw materials prior to selection for use in vaccine production, having well-defined processes and procedures for process evaluation and inactivation, to minimize risk, and applying a comprehensive targeted testing package to the process. The testing package deployed by Janssen includes both *in vivo* and *in vitro* testing, as well as targeted PCR testing. Testing the crude harvest is considered to be the point with the highest chance of detecting adventitious agents. However, this is challenging as the recombinant Adenovirus platform used in Janssen vaccine production is itself a virus and consequently there are very high virus titres of $>10^{11}$ virus particles per ml at this point. The crude harvest also contains a high cell density at infection of $>10^{7}$ per ml. The crude harvest is a complex and variable matrix with different transgene proteins that may affect the ability to detect adventitious viruses.

Janssen has adopted a stepwise NGS workflow involving biological sample preparation with enrichment, nucleic acid amplification and the use of a sequencer library, followed by sequencing, a bioinformatics pipeline that uses a PathoQuest automated workflow matched to a validated database. The bioinformatics step includes filtering, identification and the generation of a comprehensive report. The report details the number of reads annotated to the contaminant and normalized to the total number of reads, the genome coverage of the contaminants and an expert conclusion.

The development of the NGS approach has included a proof-of-concept study to assess whether NGS can be used to test crude harvest samples. Crude harvest samples were spiked with RNA and DNA viruses and were analysed in two separate DNA and RNA workflows. The process included an enrichment step to increase the sensitivity.

For the DNA fraction, crude harvest Ad26 samples (2.5×10^{12} gc/ml) were spiked with Varicella zoster virus (VZV/HHV-3) at concentrations of 10^{4} gc/ml (ratio = 10^{8}), and 10^{3} gc/ml (ratio = 10^{9}). The workflow included extraction, library preparation, an enrichment step (the phenome emulsion reassociation technique (PERT)) and sequencing. The PERT step was designed to break the double strand of the DNA. For the first sample [2.5×10^{4} gc/ml (ratio = 10^{8})], the total number of reads exceeded 560 million, the VZV analysis generated over 22,000 reads and a coverage of almost 95%. For the second sample [10^{3} gc/ml (ratio = 10^{9})] the total number of reads exceeded 500 million, the VZV analysis generated over 1600 reads and a coverage of almost 59%. They repeated the experiment using a different VZV stock. For the first sample the total number of reads was over 400 million, over 141,000 for the VZV reads and the coverage was over 98%. The total number of reads for the second sample was over 485 million, almost 9000 for the VZV reads and the coverage was over 91%. This highlights the importance of characterizing the stock.

For the RNA fraction, crude harvest Ad26 samples (2.5×10^{12} gc/ml) were spiked with Rotavirus A (a double stranded RNA virus) and measles virus (a single stranded RNA virus) at concentrations of 10^{4} gc/ml (ratio = 10^{8}) (sample 1), 10^{3} gc/ml (ratio = 10^{9}) (sample 2), and 10^{2} gc/ml (ratio = 10^{10}) (sample 3). The workflow included extraction, DNase treatment, reverse transcription, library preparation and sequencing. The analysis was challenged by the high level of human reads, which accounted for approximately 90% of all reads. This was due to the presence of ribosomal RNA. A depletion step was added for ribosomal RNA, which caused a substantive reduction in the number of ribosomal RNA reads for the sample. For the rotavirus analysis, sample

one generated over 11,000 total reads and a coverage of almost 100%, sample two generated over 1000 reads and a coverage of over 57%, sample three generated over 200 reads and a coverage of almost 17%. For the measles virus analysis, sample one generated over 8,000 total reads and a coverage of over 93%, sample two generated almost 500 reads and a coverage of over 40%, sample three generated 35 reads and a coverage of just over 8%. Samples one and two for the rotavirus and sample one for the measles virus analyses gave good detection. Sample two for the measles virus analysis gave borderline levels of detection. Sample three for both analyses gave poor detection. In addition to the high level of human reads, this analysis was also challenged by the high number of Ad26 reads. Further investigation of the reads indicated that a lot of these reads came from the positive strand, from the messenger RNA from host cell production.

In summary, the DNA workflow gave a good number of VZV reads and coverage for the 10^8 and 10^9 ratios. Lower reads were observed during the repeat experiment due to the different VZV stock used for the analysis. The RNA workflow was improved by the removal of rRNA. Good recovery of rotavirus A was achieved at ratios as low as 10^9. Good recovery of measles virus was achieved at ratios as low as 10^8, and close to 10^9. Recovery could be improved by ultracentrifugation of virus particles prior to nucleic acid extraction, thus reducing the Ad26 mRNA-derived reads.

In order to manage potential false positive results, it is necessary to establish meaningful limits. These limits should be informed by expert opinion. The decision on whether to follow up positive results should be guided by a risk-based approach that should be informed by the NGS results themselves (including the number of reads and coverage), the use of assay controls, and information on potential sources of contamination.

There are a number of potential sources of false positive results including physical sources during the NGS workflow, such as kit components, controls and spikes, operator handling and cross-contamination with the environmental and other samples, as well as sources associated with the bioinformatics pipeline, such as the annotation of host-cell derived sequences, the annotation of viral sequences and the reference databases and associated annotation and scoring. False positives can be mitigated and controlled by optimizing and defining the processes, by operator training, and by careful selection of the raw materials, viral reference databases and the bioinformatics pipeline. These are also important considerations for assay validation. Assay controls are also very important to enable distinction between false and true positive results.

The need to investigate positive results depends on the NGS results for the internal controls, as well as the agreed internal procedures. It will also depend on the estimated level of the contaminant. The follow-up investigation should include considerations on the nature of the contaminant itself, such as the associated risk, potential replication in supporting cell lines and its detectability with other common assays. The infectivity should be further investigated. This depends on the permissive and susceptible cell lines or production cells. It is also important to identify and investigate potential sources of contamination such as the raw materials or starting materials and the process steps.

The results of the NGS development work are promising. Janssen is in the process of validating the assay and is considering the implementation of system suitability controls such as RNA and DNA model viruses and negative controls. There is a need for well-characterized, standardized spikes and positive control viruses. Janssen is currently drafting a strategy for assay validation and are discussing mutual expectations with the regulatory authorities. The process for the follow-up investigation on positive results needs to be validated. However, guidance on the requirements for implementation are lacking, including guidance for claiming equivalence or allowing NGS to be used as a replacement for conventional assays.

7. Panel discussion

The meeting concluded with a panel discussion convened by Arifa Khan and Laurent Mallet. The participants included: Blandine de Saint-Vis, Simone Olgiati (Merck Group), Emmanuelle Charton (EDQM), Ivana Knezevic, Johannes Blumel, Marc Eloit, Marion Gruber, Siemon Ng, Celine Lorteau (ANSES), and Frederick Granberg. Additionally, the audience comprised of representatives from regulatory health authorities, industry, contract research organizations, and academia. The discussion was focussed on the readiness of NGS for adventitious virus detection in biologics. It is recognized that the currently recommended *in vivo* animal assays and the *in vitro* cell culture assays are compendial assays without standardization and a positive result of virus detection is based on susceptibility of the animal species or cell lines used in the assays to virus infection and subsequent replication. Therefore, these conventional assays cannot detect all viruses, even including some known ones [1]. The PCR assays are highly-sensitive but are virus- or family-specific and not useful for broad detection of distantly-related or novel viruses. There was general consensus that NGS represents a powerful new tool with capabilities for known and unknown virus detection, which can be used to complement (provide additional information), supplement (provide information to fill a gap), or replace/substitute (with or without a direct comparison) some of the conventional assays. The importance of ongoing international collaborative efforts in the Advanced Virus Detection Technologies Interest Group (AVDTIG), involving regulatory authorities, industry, academia and other stakeholders, toward standardization and validation of NGS for adventitious virus detection in biologics was noted. This includes the development of reference virus standards for conducting spiking studies, and comprehensive, correctly annotated viral databases for bioinformatics analysis that can facilitate NGS validation to support applications in biologics.

It was noted that NGS is evolving quickly, and although a lot of progress has been made since the 2017 NGS meeting [5], some challenges remain regarding its implementation for adventitious virus detection. Therefore, early discussions with regulatory authorities are highly recommended and encouraged to discuss the specific use of NGS for adventitious agent detection for a particular product. The Ph. Eur. and WHO currently recommend the use of NGS as an alternative/substitution method, particularly for the *in vivo* assays, provided the method has been validated [17,18,41]. The Ph. Eur. has opened its doors to new technologies by allowing for high-throughput sequencing to be used to confirm the presence of an extraneous agent in a vaccine for human [18,41] or veterinary [22] use. By its General Notices [42], the Ph. Eur. accepts the use of alternative methods provided that they have been validated against the compendial methods. Moreover, the Ph. Eur. also recognizes the fact that innovative methods may substitute compendial *in vivo* methods without the need for a head-to-head comparison (bridging studies) with the compendial methods since the *in vivo* assays are highly variable and were established before ICH Q2, as described in Ph. Eur. Chapter 5.2.14 [17]. Regarding the management of extraneous agents in immunological veterinary medicinal products, the Ph. Eur. Chapter 5.2.5 [22]refers to the use of testing methods that fulfil specific requirements but leaves it open to any method, without need of bridging studies. The Ph. Eur. has adopted a risk assessment approach to adventitious virus detection for vaccines for human use [18,41] and for veterinary vaccines [22,23] and does not mandate the systematic conduct of bridging studies to substitute *in vivo* methods, provided that the overall testing package applied can be shown to be fit-for-purpose in terms of providing adequate assurance of freedom from extraneous agents.

Although the U.S. FDA does not have specific recommendations for using NGS at this time, CBER's Guidance for Industry [43] allows the use of alternative and novel approaches that can be demonstrated to be fit-for-purpose for adventitious virus detection. The use of NGS is not addressed in either the Japanese or Chinese Pharmacopoeia. It was

recognized that increased efforts need to be made to include additional regulatory agencies and regions that were under-represented in the meeting in order for NGS to be more generally applied as a general viral adventitious agent detection assay.

It was recognized that NGS and PCR can have similar limits of detection. In contrast to PCR, use of NGS will result in sequencing of all nucleic acids present in the test material, resulting in a large data volume. Panellists stressed that the bioinformatics pipeline and database used are critical components of the NGS workflow for obtaining accurate results for virus detection and removing background noise. False positive results as a consequence of using NGS for adventitious agent detection will need to be further investigated and verified using other assays, e.g., PCR. Furthermore, as with any nucleic-based assay, the biological relevance of a positive signal would need to be determined, e.g. by using an infectivity assay, when feasible, as part of the overall risk assessment approach.

A critical topic of discussion by the panellists and the audience was the use of NGS for replacing the *in vivo* animal assays. It was noted that, to date, there does not seem to be evidence for detection of an adventitious virus in the animal assays and that furthermore, the viruses that could be detected by *in vivo* assays could also be detected in the *in vitro* cell culture assays [19]. There was general agreement that NGS could be validated for broad virus detection and thus could substitute the currently recommended animal-based *in vivo* assays. It was noted that PCR assays are already used as an alternative method to the antibody-production animal tests (MAP, RAP, and HAP), and thus, these could also be replaced by NGS. The latter has the advantage of offering both, sensitivity and breadth of virus detection. The panellists considered substituting *in vivo* assays by NGS a priority, thus, facilitating the implementation of the 3Rs initiative of reducing animal use [20].

NGS was also recognized as a tool to supplement the *in vitro* cell culture assays particularly in those cases where the test material (e.g., virus seed or bulk harvest) causes false-positive results due to lack of sufficient neutralization of the vaccine virus or toxicity due to the media or residual host cell materials. Thus, in situations where there is no neutralizing antibody available, on a case-by-case basis, a validated NGS assay could be considered to supplement or replace the *in vitro* cell culture assays.

It was emphasized that close collaboration and communication between regulators, industry, contract research organizations, and technology developers is critical as NGS is still evolving. Additionally, the development of publicly available reference materials for NGS standardization and validation will facilitate its broad implementation and continued sharing of knowledge and experiences will advance the use of NGS for rapid and broad adventitious virus testing to enhance safety of biological products.

8. Conclusions

- NGS could be as sensitive for virus detection as PCR assays and relevant validation data could support its use as a replacement assay.
- The need to substitute *in vivo* testing with NGS was recognized to align with the 3Rs initiative to replace, reduce and refine the use of animals for medical research and development. There was support for this consideration, since NGS can broadly detect viruses and can be standardized and validated and therefore has the potential to substitute animal-based assays that are currently stipulated in various regulatory compendia and guidance documents.
- Further work is needed to replace *in vitro* cell culture-based tests: in particular, to standardize and validate the performance of NGS as compared to that of *in vitro* assays. The risk of false positives is currently limiting the ability of NGS to replace *in vitro* cell culture-based testing.
- Further work is also needed to optimize the approach to managing true and false positives and inconclusive results.
- In order to replace/substitute existing assays, regulations and applicable guidances will need to be revised, if not done so already. These already allow flexibility for the use of new technologies such as NGS in the development and production of human medicines; in Europe, regulations governing veterinary medicine are under revision to further encourage the use of NGS.
- The importance of moving from prescriptive testing to a risk management approach also needs to be considered in the overall virus mitigation strategy; this approach is already chosen by the regulators in the control of human and veterinary vaccines in Europe.
- NGS technology is evolving rapidly and it is important to keep up-to-date with the technological advances. Considerable efforts are currently being devoted to advance the use of NGS for adventitious virus detection in human biologics; such efforts should continue and be more inclusive of all regions for global acceptance.
- Concerted efforts should be made toward development of publicly available reference materials that can be used for the entire workflow or specific steps in NGS standardization and validation.
- The lack of appropriate standards and databases, particularly in the veterinary domain, make it premature at the present time for NGS to be part of compulsory regulatory requirements. However, all stakeholders were encouraged to continue the work highlighted at this meeting with the aim of NGS becoming part of the routine methodological tool kit for manufacturers and regulators.

9. Post-meeting note

The manuscript was submitted during the current COVID-19 pandemic. There is an urgent need to develop vaccines against SARS-CoV-2 to protect against COVID-19 disease. Regulators and industry recognize that development of international virus standards and use of NGS as a rapid method for adventitious virus detection assays could aid in accelerating SARS-CoV-2 vaccine development.

Declaration of competing interest

All authors declare no competing interests.

Acknowledgments

In addition to the organizing committee members who were session chairs and are coauthors on the paper, the contributions of the following committee members are appreciated: William Egan and Sven Arnouts. We are grateful to the speakers and panellists for their participation and providing valuable discussions in the meeting. All of the speakers and panellists have reviewed their talks and the panel summary for the meeting report. Ms. Abigail Charlet and Ms. Debbie Colman are acknowledged for providing excellent administrative and technical support.

References

[1] Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, et al. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. J Virol 2010;84:6033–40.
[2] Ma H, Galvin TA, Glasner DR, Shaheduzzaman S, Khan AS. Identification of a novel rhabdovirus in Spodoptera frugiperda cell lines. J Virol 2014;88:6576–85.
[3] Taliaferro LP, Galvin TA, Ma H, Shaheduzzaman S, Williams DK, Glasner DR, et al. Evaluation of the broad-range PCR-electrospray ionization mass spectrometry (PCR/ESI-MS) system and virus microarrays for virus detection. Viruses 2014;6:1876–96.
[4] Khan AS, King KE, Brack K, Cassart J-P, Chiu C, Dehghani H, Duncan P, Jaing C, Kolman J, Munroe D, Palermo A, Plavsic M, Sampath R, Slezak T, Takle G, Taliaferro LP, Toso E, Vacante D, Willkommen H. Emerging methods for virus detection. PDA Tech Rep 2015;71.
[5] Khan AS, Benetti L, Blumel J, Deforce D, Egan WM, Knezevic I, et al. Report of the international conference on next generation sequencing for adventitious virus detection in biologicals. Biologicals 2018;55:1–16.
[6] Cleveland MH, Anekella B, Brewer M, Chin PJ, Couch H, Delwart E, et al. Report of

the 2019 NIST-FDA workshop on standards for next generation sequencing detection of viral adventitious agents in biologics and biomanufacturing. Biologicals 2020;64:76–82.
[7] Goodacre N, Aljanahi A, Nandakumar S, Mikailov M, Khan AS. A reference viral database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. mSphere 2018;3.
[8] Bigot TTS, Pérot P, Eloit M. RVDB-prot, a reference viral protein database and its HMM profiles. F1000 2019;8:530.
[9] Khan AS, Vacante DA, Cassart JP, Ng SH, Lambert C, Charlebois RL, et al. Advanced virus detection technologies interest group (AVDTIG): efforts on high throughput sequencing (HTS) for virus detection. PDA J Pharm Sci Technol 2016;70:591–5.
[10] Mee ET, Preston MD, Minor PD, Schepelmann S. Development of a candidate reference material for adventitious virus detection in vaccine and biologicals manufacturing by deep sequencing. Vaccine 2016;34:2035–43.
[11] Ng SH, Braxton C, Eloit M, Feng SF, Fragnoud R, Mallet L, et al. Current perspectives on high-throughput sequencing (HTS) for adventitious virus detection: upstream sample processing and library preparation. Viruses 2018;10.
[12] Lambert C, Braxton C, Charlebois RL, Deyati A, Duncan P, La Neve F, et al. Considerations for optimization of high-throughput sequencing bioinformatics pipelines for virus detection. Viruses 2018;10.
[13] Khan AS, Ng SHS, Vandeputte O, Aljanahi A, Deyati A, Cassart JP, et al. A multicenter study to evaluate the performance of high-throughput sequencing for virus detection. mSphere 2017;2.
[14] Cheval J, Muth E, Gonzalez G, Coulpier M, Beurdeley P, Cruveiller S, et al. Adventitious virus detection in cells by high-throughput sequencing of newly synthesized RNAs: unambiguous differentiation of cell infection from carryover of viral nucleic acids. mSphere 2019;4.
[15] Brussel A, Brack K, Muth E, Zirwes R, Cheval J, Hebert C, et al. Use of a new RNA next generation sequencing approach for the specific detection of virus infection in cells. Biologicals 2019;59:29–36.
[16] National Institute of Sandards and Technology. Workshop on standards for NGS detection of viral adventitious agents in biologics and biomanufacturing. 2019https://www.nist.gov/news-events/events/2019/09/workshop-standards-ngs-detection-viral-adventitious-agents-biologics-.
[17] Substitution of in vio methods by in vitro methods for the quality control of vaccines. General chaper 5.2.14. Ph.Eur. tenth ed. Strasbourg, France: Council of Europe; 2020.
[18] Cell substrates for the production of vaccines for human use. General Chapter 5.2.3. Ph. Eur. tenth ed. Strasbuorg, France: Council of Europe; 2020.
[19] Gombold J, Karakasidis S, Niksa P, Podczasy J, Neumann K, Richardson J, et al. Systematic evaluation of in vitro and in vivo adventitious virus assays for the detection of viral contamination of cell banks and biological products. Vaccine 2014;32:2916–26.
[20] European Medicines Agency. Guideline on the principles of regulatory acceptance of 3Rs (Replacement R, Refinement) testing approaches. 2016. (EMA/CHMP/CVMP/JEG-3Rs/450091/2012). http://www.ema.europa.eu/docs/en_GB/document_library/Scientificguideline/2017/03/WC500224306.pdf.
[21] Cell cultures for the production of vaccines for veterinary use. General Chapter 5.2.4. Ph. Eur. tenth ed. Strasbourg, France: Council of Europe; 2020. supplement 10.2.
[22] Management of extraneous agents in immunological veterinary medicinal products. General Chapter 5.2.5. Ph. Eur. tenth ed. Strasbourg, France: Council of Europe; 2020. supplement 10.2.
[23] Principles for the detection of extraneous viruses in immunological veterinary medicinal products using culture methods. General Chapter 2.6.37. Ph. Eur. tenth ed. Strasbourg, France: Council of Europe; 2020. supplement 10.2.
[24] US Food and Drug Administration. FDA briefing document: vaccines and related biological products advisory committee meeting: cell lines derived from human tumors for vaccine manufacture. 2012http://wwwfdagov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/BloodVaccinesandOtherBiologics/VaccinesandRelatedBiologicalProductsAdvisoryCommittee/UCM319573.pdf.
[25] Recommendations for the evaluation of animal cell cultures as substrates for the manufacture of biological medicinal products and for the characterization of cell banks. WHO expert committee on biological standardization: sixty-first report. Geneva: World Health Organization; 2013 Annex 3. WHO Technical Report Series, No. 978 https://www.who.int/biologicals/expert_committee/TRS_978_61st_report.pdf?ua=1, Accessed date: 29 April 2020.
[26] Scientific principles for regulatory risk evaluation on finding an adventitious agent in a marketed vaccine. WHO expert committee on biological standardization: sixty-fifth report. Geneva: World Health Organization; 2015 Annex 2. WHO Technical Report Series, No. 993 https://apps.who.int/iris/bitstream/handle/10665/173739/9789240694095_eng.pdf;jsessionid=A55EA3492466DE356E42C29FDEB8E93F?sequence=1, Accessed date: 29 April 2020.
[27] Proposed second WHO international reference reagents for MAPREC analysis and NGS of poliovirus types 1, 2 and 3 (Sabin). WHO expert committee on biological standardization: seventieth report. Geneva: World Health Organization; 2020. p. 84 WHO Technical Report Series, No. 1024; section 9.2.4 https://www.who.int/biologicals/technical_report_series/en/, Accessed date: 25 May 2020.
[28] International Council for harmonisation of technical requirements for pharmaceuticals for human use. https://www.ich.org/.
[29] Allain JP, Owusu-Ofori AK, Assennato SM, Marschner S, Goodrich RP, Owusu-Ofori S. Effect of Plasmodium inactivation in whole blood on the incidence of blood transfusion-transmitted malaria in endemic regions: the African Investigation of the Mirasol System (AIMS) randomised controlled trial. Lancet 2016;387:1753–61.
[30] Blumel J, Schwantes A, Baylis SA, Stuhler A. Strategies toward virus and prion safe human platelet lysates. Transfusion 2020;60:219–20.
[31] European Medicines Agency. Guideline on requirements for the production and control of immunological veterinary medicinal products. EMA/CVMP/iWP/206555/2010-Rev.1. 2017.
[32] Conceicao-Neto N, Zeller M, Lefrere H, De Bruyn P, Beller L, Deboutte W, et al. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. Sci Rep 2015;5:16532.
[33] Theuns S, Vanmechelen B, Bernaert Q, Deboutte W, Vandenhole M, Beller L, et al. Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobuvirus as an important enteric virus. Sci Rep 2018;8:9830.
[34] Sathiamoorthy S, Malott RJ, Gisonni-Lex L, Ng SHS. Selection and evaluation of an efficient method for the recovery of viral nucleic acids from complex biologicals [corrected]. NPJ Vaccine 2018;3:31.
[35] Charlebois RL, Ng SH, Gisonni-Lex L, Mallet L. Cataloguing the taxonomic origins of sequences from a heterogeneous sample using phylogenomics: applications in adventitious agent detection. PDA J Pharm Sci Technol 2014;68:602–18.
[36] Uphoff CC, Lange S, Denkmann SA, Garritsen HS, Drexler HG. Prevalence and characterization of murine leukemia virus contamination in human cell lines. PLoS One 2015;10:e0125622.
[37] Rosseel T, Pardon B, De Clercq K, Ozhelvaci O, Van Borm S. False-positive results in metagenomic virus discovery: a strong case for follow-up diagnosis. Transboundary Emerg Dis 2014;61:293–9.
[38] Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J Virol 2013;87:11966–77.
[39] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 2014;12:87.
[40] Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 2014;15:R46.
[41] Tests for extraneous agents in viral vaccines for human use. General chaper 2.6.16. Ph. Eur. tenth ed. Strasbourg, France: Council of Europe; 2020.
[42] General Notices. General chapter 1. PH. Eur. tenth ed. Strasbourg, France: Council of Europe; 2020.
[43] US Food and Drug Administration. Guidance for Industry. Characterization and qualification of cell substrates and other biological materials used in the production of viral vaccines for infectious disease indications. 2010http://wwwfdagov/downloads/biologicsbloodvaccines/guidancecomplianceregulatoryinformation/guidances/vaccines/ucm202439.pdf.