Vaccine 34 (2016) 2035–2043

Contents lists available at ScienceDirect

Vaccine

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

Development of a candidate reference material for adventitious virus detection in vaccine and biologicals manufacturing by deep sequencing

Edward T. Mee[a,*], Mark D. Preston[b], CS533 Study Participants[1], Philip D. Minor[a], Silke Schepelmann[a]

[a] Division of Virology, National Institute for Biological Standards and Control, Medicines and Healthcare Products Regulatory Agency, South Mimms, Hertfordshire EN6 3QG, United Kingdom
[b] Division of Technology, Development and Infrastructure, National Institute for Biological Standards and Control, Medicines and Healthcare Products Regulatory Agency, South Mimms, Hertfordshire EN6 3QG, United Kingdom

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form
30 November 2015
Accepted 5 December 2015
Available online 19 December 2015

Keywords:
Adventitious virus
Vaccine
Deep sequencing
Reference material
Collaborative study

ABSTRACT

Background: Unbiased deep sequencing offers the potential for improved adventitious virus screening in vaccines and biotherapeutics. Successful implementation of such assays will require appropriate control materials to confirm assay performance and sensitivity.
Methods: A common reference material containing 25 target viruses was produced and 16 laboratories were invited to process it using their preferred adventitious virus detection assay.
Results: Fifteen laboratories returned results, obtained using a wide range of wet-lab and informatics methods. Six of 25 target viruses were detected by all laboratories, with the remaining viruses detected by 4–14 laboratories. Six non-target viruses were detected by three or more laboratories.
Conclusion: The study demonstrated that a wide range of methods are currently used for adventitious virus detection screening in biological products by deep sequencing and that they can yield significantly different results. This underscores the need for common reference materials to ensure satisfactory assay performance and enable comparisons between laboratories.
© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author. Tel.: +44 1707 641191.
E-mail address: edward.mee@nibsc.org (E.T. Mee).
[1] The members of the CS533 Study Participants are: Xuening Huang[c], Jenny Nguyen[c], David Wall[c], Stacey Hargrove[c], Thomas Fu[c], George Xu[c], Li Li[c], Colette Cote[d], Eric Delwart[e,f], Linlin Li[e,f], Indira Hewlett[g], Vahan Simonyan[g], Viswanath Ragupathy[g], Alin Voskanian-Kordi[g], Nicolas Mermod[h], Christiane Hill[i], Birgit Ottenwälder[j], Daniel C. Richter[j], Arman Tehrani[j], Jacqueline Weber-Lehmann[j], Jean-Pol Cassart[k], Carine Letellier[k], Olivier Vandeputte[k], Jean-Louis Ruelle[k], Avisek Deyati[k], Fabio La Neve[l], Chiara Modena[l], Edward Mee[m], Silke Schepelmann[m], Mark Preston[m], Philip Minor[m], Marc Eloit[n], Erika Muth[n], Arnaud Lamamy[n], Florence Jagorel[n], Justine Cheval[n], Catherine Anscombe[o], Raju Misra[o], David Wooldridge[o], Saheer Gharbia[o], Graham Rose[o], Siemon H.S. Ng[p], Robert L. Charlebois[p], Lucy Gisonni-Lex[p], Laurent Mallet[p], Fabien Dorange[q], Charles Chiu[f], Samia Naccache[f], Paul Kellam[r], Lia van der Hoek[r], Matt Cotten[r], Christine Mitchell[s], Brian S. Baier[s], Wenping Sun[s], Heather D. Malicki[s]
[c] Advance Biosciences, Houston, TX, United States
[d] Bioreliance, Rockville, MD, United States
[e] Blood Systems Research Institute, San Francisco, CA, United States
[f] University of California, San Francisco, United States
[g] Center for Biologics Evaluation and Research/Food and Drug Administration, Bethesda, MD, United States
[h] Institute of Biotechnology, University of Lausanne, Switzerland
[i] Charles River Biopharmaceutical Services, Erkrath, Germany
[j] Eurofins Genomics, Ebersberg, Germany
[k] GSK Vaccines, Rixensart, Belgium
[l] Merck-Serono, Colleretto Giacosa, Italy
[m] National Institute for Biological Standards and Control, South Mimms, Hertfordshire, United Kingdom
[n] Pathoquest, Paris, France
[o] Public Health England, Colindale, London, United Kingdom
[p] Sanofi Pasteur Ltd., Toronto, Ontario, Canada
[q] Texcell, Evry, France
[r] Wellcome Trust Sanger Institute, Cambridge, United Kingdom
[s] WuXiAppTecInc, Philadelphia, PA, United States

http://dx.doi.org/10.1016/j.vaccine.2015.12.020
0264-410X/© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Production of live viral vaccines on animal cell or egg substrates carries the risk of adventitious virus contamination of the final product [1,2]. Testing for adventitious viruses is therefore an essential quality control step in the manufacture of vaccines and other biological medicines. Non-specific screening for adventitious viruses is partly based on animal tests which have served well for decades, but there are legal and ethical imperatives to replace such tests. Cell culture tests largely solve the ethical issues and are cheaper to perform, and recent efforts to compare sensitivity and specificity with animal tests have been promising [3]. Nevertheless, cell and animal tests are limited by the restricted tropism of some viruses and may not detect non-cytopathic, non-pathogenic or non-haemadsorbing viruses. For example, porcine circovirus (PCV) DNA was detected in two rotavirus vaccines [1,4,5] despite these routine adventitious virus tests showing no evidence of contamination. PCR-based tests offer sensitive and specific detection of their target pathogens, however screening for all potential viruses by PCR is impractical, and non-target viruses would remain undetected.

Deep sequencing (DS, also referred to as massively parallel or high throughput sequencing) offers the potential for identification of extraneous nucleic acid in samples without *a priori* knowledge of the likely contaminant and without the requirement for propagation of the virus. Such methods have already been successfully applied to detection of adventitious agents in vaccines [1], cell lines [6,7], serum [8,9] and bioreactors [10] and multiple laboratory and informatics methods for viral metagenomics have been developed for clinical and other biological specimens [11–17]. There is substantial interest among vaccine manufacturers, contract research organisations, regulators and medicines control laboratories in evaluating the method for routine safety testing, and potentially replacing some or all of the existing *in vitro* and *in vivo* tests. A major challenge to the realisation of this potential is the identification of a robust, sensitive and specific assay design. A wide range of methods exist for viral metagenomics, many of which are early in their development: multiple options exist for generation of sequencing libraries; several commercial sequencing platforms exist, based on fundamentally different chemistry, with more in development; numerous bioinformatics pipelines are used for sequence classification, both academically and commercially developed; and the databases against which the reads are searched are constantly evolving. Given these parameters, it is important to have suitable reference materials to ensure that different methods generate comparable results. In addition to reagents for comparison of methods and determination of run performance, well-characterised materials of defined virus concentration will be required in order to determine limits of detection for particular viruses or virus types. We describe here a candidate material for qualitative comparison of methods and run performance and its evaluation in an international collaborative study encompassing 15 laboratories. The study highlighted that a broad range of laboratory and informatics techniques are in use, and no consensus exists on the most appropriate combination of methods to achieve maximum sensitivity. We discuss the major challenges for the incorporation of deep sequencing into adventitious agent testing workflows, highlight areas requiring particular attention and describe the requirements of future reference materials to enable validation and comparison of methods.

2. Methods

2.1. Aim and scope

The primary aim of the study was to evaluate the suitability of reagent 11/242-001 as a reference material for deep sequencing-based adventitious virus detection by comparing the results obtained from 15 independent laboratories using a variety of sample preparation, sequencing and informatics methods. Identifying the optimal processing parameters for each step of the process was not feasible given the large number of variables. This study did not aim to assess sensitivity of any particular method, nor the proficiency of the individual laboratories. An outline of this project was presented to the World Health Organisation (WHO) Expert Committee for Biological Standardisation (ECBS) at the 2013 meeting and the committee felt that the project could provide useful information on the value of the reference material and the merits of currently used methods [18].

2.2. Participants

Participants were identified through existing networks of contacts and *via* the Parenteral Drug Association (PDA)/Food and Drug Administration (FDA) Advanced Virus Detection Technologies Interest Group. Participants included vaccine manufacturers, contract research organisations, academic laboratories, regulatory agencies and medicines control laboratories with an interest in virus detection in biological medicines. A full list of participating laboratories is shown in Collaboration Group.

2.3. Material

An existing multiplex quantitative polymerase chain reaction (qPCR) run control reagent, 11/242-001, was available for the study. This reagent contains 25 viruses representing a range of common hazard group 2 human viruses (United Kingdom Advisory Committee on Dangerous Pathogens classification) with a variety of genome and envelope types (Table 1).

Individual viruses were propagated in cell culture or by egg passage, and non-cultivable viruses were isolated from clinical specimens. The origin of each virus is described in Table 1. Real-time PCR (RT-PCR) Cycle Threshold (Ct) values were determined for individual virus stocks, and the viruses were then pooled such that the predicted Ct value of each would be approximately 30. Pooled virus was formulated in 10 mM Tris, pH 7.4, supplemented with 2% foetal calf serum. 1 ml of reagent was filled into 2856 2 ml screw-cap Sarstedt vials and frozen at −70 °C. Samples of pooled material were assessed by in-house RT-PCR (see Supplemental table* 1 for PCR conditions) to determine the presence of the 25 viruses (Table 1). Not all viruses were detected following formulation – hence development of the reagent as a qPCR control was ceased and the material was deemed an ideal candidate for the current study. Infectivity of pooled viruses was not confirmed as the intended use was in nucleic-acid based detection methods. The precise concentrations of individual viruses are not known, however RT-PCR data suggest the viruses are present at a range of nucleic acid concentrations (Ct values range from ~24 to not detectable, Table 1). A previous study found that up to 22 of the 25 viruses were detectable by sequencing at modest read depth (~2,000,000 reads) [14]. The presence of additional viruses in the reagent was considered a possibility due to the isolation of several of the target viruses from human clinical specimens, the propagation of others in cell culture and the addition of foetal calf serum to the reagent. The presence of such viruses was not known in advance.

2.4. Study design

Two vials of reagent were shipped to each laboratory on dry ice. The list of target viruses was known to the laboratories to facilitate import and appropriate biocontainment. Laboratories processed the reagent according to their preferred method. Technical replicates were requested, but not mandatory due to the high costs

Table 1
Virus composition of multiplex reagent 11/242-001.

Group	Family	Envelope	Species/serotype	Genome size (kb)	PCR Ct value	Sample origin
dsDNA	Adenoviridae	No	Adenovirus 2	35.9	29.71	293 cell culture
			Adenovirus 41	34.2	ND	Clinical specimen
	Herpesviridae	Yes	Human herpesvirus 1	151.2	30.59	MRC5 cell culture
			Human herpesvirus 2	154.7	32.48	MRC5 cell culture
			Human herpesvirus 3 (VZV)	124.8	29.02	MeWo cell culture
			Human herpesvirus 4 (EBV)	171.7	31.27	B95-8 cell culture
			Human herpesvirus 5 (CMV)	233.7	28.95	MRC5 cell culture
dsRNA	Reoviridae	No	Rotavirus A	18.5	24.49	Clinical specimen
ssRNA (+)	Astroviridae	No	Astrovirus	6.8	30.53	Clinical specimen
	Caliciviridae	No	Norovirus GI	7.6	ND	Clinical specimen
			Norovirus GII	7.5	ND	Clinical specimen
			Sapovirus C12	7.5	33.37	Clinical specimen
	Coronaviridae	Yes	Coronavirus 229E	27.2	ND	MRC5 cell culture
	Picornaviridae	No	Coxsackievirus B4	7.4	30.72	Hep-2 cell culture
			Rhinovirus A39	7.1	31.16	MRC5 cell culture
			Parechovirus 3	7.2	29.35	LLC-MK2 cell culture
ssRNA (−)	Orthomyxoviridae	Yes	Influenza A virus H1N1	13.2	32.02	Egg passage
			Influenza A virus H3N2	13.6	ND	Egg passage
			Influenza B virus	14.2	ND	Egg passage
	Paramyxoviridae	Yes	Metapneumovirus A	13.3	31.86	LLC-MK2 cell culture
			Parainfluenzavirus 1	15.5	34.43	PRF5 cell culture
			Parainfluenzavirus 2	15.7	33.87	PRF5 cell culture
			Parainfluenzavirus 3	15.4	ND	PRF5 cell culture
			Parainfluenzavirus 4	17.4	31.83	PRF5 cell culture
			Respiratory syncytial virus A2	15.2	34.33	Hep-2 cell culture

ds double-stranded, *ss* single-stranded, *VZV* Varicella Zoster Virus, *EBV* Epstein Barr Virus, *CMV* Cytomegalovirus, *ND* not detectable. Ct values provide a crude estimate of viral genome abundance; quantitative PCR data are not available.

involved. Laboratories were asked to analyse the entire data set, plus a random subset of 2 million reads.

2.5. Analysis

The parameters reported in the primary analysis were (a) total number of target viruses detected; (b) rankings of target viruses; and (c) correlation between read depth and number of target viruses detected. The majority of laboratories ran two technical replicates, with the replicate detecting most target viruses selected for primary analysis. In the event that both replicates detected equal numbers of target viruses, the replicate containing fewer total sequencing reads was selected. Secondary analysis included (a) reporting of non-target viruses detected by three or more labs; (b) ranking of all target and non-target viruses detected by three or more labs; and (c) consistency of virus detection between replicates. Rankings were determined based on the absolute numbers or proportions of reads matching the indicated viruses, to account for the fact that total numbers of reads, and proportion of reads identified as viral differed greatly between laboratories.

One lab reported difficulty in distinguishing Norovirus strains due to a large number of closely related sequences available in public databases. This lab provided an explanation and reported hits to 'Human norovirus' rather than to serotypes GI and GII. For result plotting, hits to 'Human norovirus' (for this laboratory only) were considered to be to Norovirus GII that had a higher overall rank. Hits to Ad2 and AdC were merged as the reference sequences are identical. Similarly hits to Mastadenovirus F were merged with Ad41.

2.6. Reagent availability

The reagent is available *via* the NIBSC catalogue (nibsc.org/products), reference 11/242-001.

3. Results

3.1. Return of data

Data were returned from 15 of 16 laboratories. One laboratory was unable to complete analysis within the study time frame. A wide range of sample preparation and informatics methods was used by the laboratories, with no two laboratories using identical methods (Table 2, Supplemental Tables 2 and 3). Four laboratories returned data generated using two different methods. The majority of laboratories employed informatics methodology that identified viruses in a blind manner rather than specifically targeting the 25 known viruses (Supplemental table* 3).

3.2. Number of target viruses detected and effect of read number

Participating laboratories generated differing numbers of reads depending on the library preparation and sequencing platform used. Using a subset of 2,000,000 reads, a single lab detected all 25 target viruses (range 6–25). Using all reads, two labs detected all 25 viruses (range 6–25, Fig. 1). The majority of methods detected at least 20 target viruses using 2,000,000 reads (median 20.5) and at least 21 viruses using all reads (median 22) (Fig. 2). The number of target viruses detected did not correlate with total read numbers (Fig. 2), though it is expected that the underlying variation between methods masked any effect. For a given method, increasing read depth would be expected to increase the probability of detecting a given virus, though this analysis was beyond the scope of the current study.

3.2.1. Consistency of replicates

Ten laboratories (13 methods) performed technical replicates. The vast majority of viruses detected were present in both replicates (Fig. 1). The most common inconsistency was the presence of Norovirus GI, GII or Influenza B in one replicate but its absence

Table 2
Summary methods by laboratory.

Lab	Nuclease	Extraction	Primary lib	Seq library	Platform	Database(s)	Primary identification	Blind/targeted
L01-A	Yes	Column/silica	Ion Torrent	Ion Torrent	Ion Proton	In-house	Smith-Waterman/BLAST	Blind
L01-B	Yes	Column/silica	Ion Torrent	Ion Torrent	Ion Proton	In-house	Mapping to targets	Targeted
L02	Yes	Column/silica	Fragmentation-ligation	Ion Torrent	Ion Proton	Virus RefSeq/in-house	Proprietary	Blind
L03	DNAse	EZ1	cDNA amplification Nucleic acid extract	Nextera XT	MiSeq	nt (Jun 2014)	SURPI	Blind
L04	No	Column/silica	cDNA synthesis	Nextera XT	MiSeq	NCBI	Align to NCBI	Blind
L05	No	Column/silica	Adaptor ligation	TruSeq	HiSeq2500	NCBI Viral Genome Neighbor	Alignment to references	Blind
L06	No	Beads	Adaptor ligation	Custom	HiSeq2000	In-house	BWA alignment	Blind
L07	No	Phenol/chloroform	Confidential	Nextera	MiSeq	Virus RefSeq/NCBI	Mapping to all viruses	Both
L08	No	Maxwell	None	ScriptSeq	MiSeq	Virus RefSeq/nr	BLAST	Blind
L09	No	Column/silica	Proprietary	Custom	454	In-house (Ref/Seq GenBank derived)	Proprietary	Not specified
L10-A	No	Column/silica	Fragmentation-ligation	TruSeq	HiSeq2500	GenBank 2013, clustered viral partition	BLASTn	Blind
L10-B	No	Column/silica	Fragmentation-ligation	TruSeq	HiSeq2500	GenBank 2013, clustered viral partition	BLASTn	Blind
L11	No	Column/silica	cDNA	Nextera XT	HiSeq1500	In-house	BLAST	Blind
L12-A	No	Column/silica	Random RT-PCR	Nextera XT	MiSeq	Virus RefSeq/nt	BLASTn	Blind
L12-B	No	Column/silica	MDA/SPIA	Nextera XT	MiSeq	Virus RefSeq/nt	BLASTn	Blind
L13	No	Magnetic beads	TruSeq	TruSeq	MiSeq	Virus RefSeq	BLAST/ CENSUSCOPE	Blind
L14	DNAse	Column/silica	Fragmentation-ligation	Illumina PCR	MiSeq	GenBank	SLIM	Both
L15-A*	Yes	Column/silica	MDA/SPIA	Nextera XT	MiSeq	In-house	BWA then BLAST	Targeted
L15-B*	Yes	Column/silica	MDA/SPIA	Nextera XT	MiSeq	In-house	BWA then BLAST	Targeted

Individual laboratories are represented by coded identifiers unrelated to the order in Supplemental Table 2. Separate methods performed by the same laboratory have the suffix-A/-B. Blind indicates methods where viruses were identified without reference to the 25 target viruses. Targeted indicates methods where these 25 viruses were specifically targeted. *L15-A and L15-B represent similar methodology, but performed using variable amounts of starting material (Supplemental Table 2).

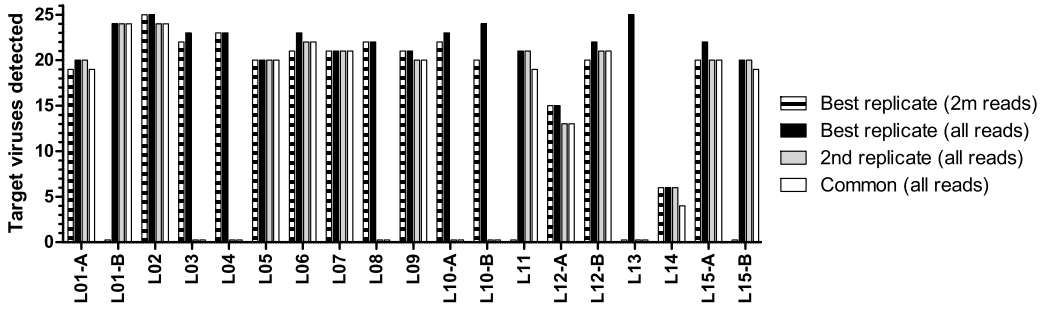


Fig. 1. Number of target viruses detected by individual laboratories and methods. *Horizontal hatched bars*, target viruses detected in best replicate using 2 million reads. *Solid black bar*, target viruses detected in best replicate using all reads. *Grey bar*, target viruses detected in second replicate using all reads. *White bar*, target viruses detected in both replicates using all reads. Laboratories L01-B, L11, L13 and L15-B performed analysis only on the total read set.

in the other (Supplemental table* 4). The nine viruses where discrepancies were observed were predominantly the bottom ranked viruses in the positive replicate, *i.e.* the replicate in which the virus was detected (Supplemental table* 4 and Fig. 4).

3.2.2. Consistency of virus detection across methods and laboratories

A detailed breakdown of viruses detected by each laboratory is shown in Supplemental table* 5. Ad2, Human Herpesvirus (HHV)-3

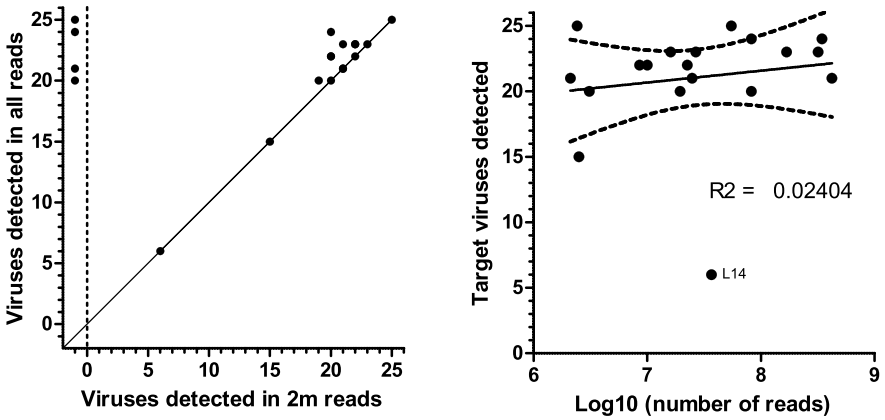


Fig. 2. Frequency of target virus detection in all methods. *Left panel*, Correlation between number of target viruses detected in 2,000,000 reads (median 20.5) and entire read set (median 22). Four laboratories performed analysis only on the total read set and these are shown left of the dotted line. *Right panel*, correlation between number of target viruses detected and read depth. Graph shows best fit and 95% confidence bands for regression line.

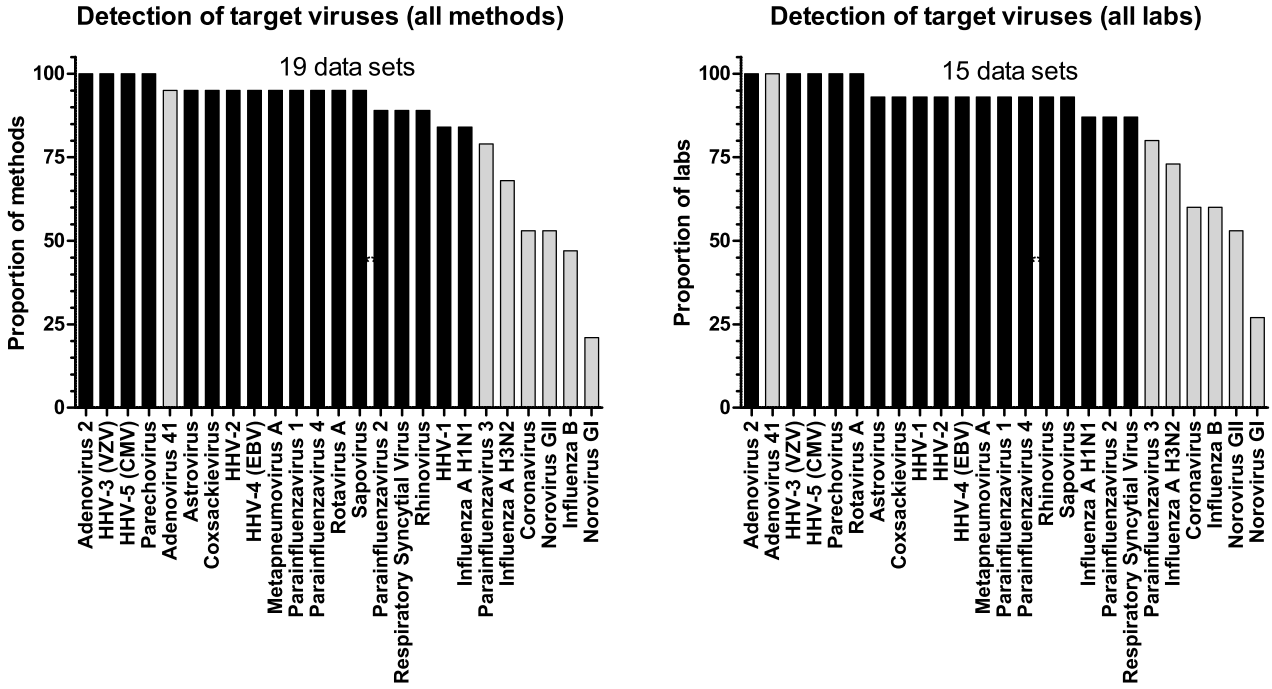


Fig. 3. Consistency of virus detection across different methods and laboratories. *Left panel*, proportion of all methods detecting target viruses. *Right panel*, proportion of all laboratories detecting target viruses using best method. *Grey shading* indicates viruses not detected by real-time PCR.

and HHV-5 were detected by all methods (Fig. 3, left panel). Ad2, Ad41, HHV-3, HHV-5, Parechovirus, and Rotavirus A were detected by all laboratories (Fig. 3, right panel). Norovirus GI and Influenza B were detected by fewer than 50% of methods while Norovirus GI was detected by fewer than 50% of laboratories. The current study did not aim to directly compare PCR and deep sequencing for detection of these viruses, however with the exception of Ad41, the viruses detected by the fewest laboratories and methods were those that were not detected by real-time PCR.

3.2.3. Additional viruses detected

Ten of 15 laboratories reported the detection of additional viruses; those detected by three or more laboratories are described in Table 3. At least 20 additional viruses were reported by single labs; these are not reported here as their presence was not corroborated by a second lab.

3.2.4. Rank order of target viruses

Of the 25 targets, Parechovirus had the highest rank based on proportion of reads returning a hit (median rank 1.5, Fig. 4), while Norovirus GI had the lowest rank (median 26, *i.e.* not detected). The ranking varied significantly between laboratories, most notably rotavirus A, which varied from a rank of 1–26 (median 10). This may suggest that different methods are differentially likely to detect a given virus or family of viruses, however the low sample numbers and wide range of methods used precluded statistical analysis of the major factors influencing virus detection. When rankings were calculated with the inclusion of the additional viruses detected by three or more laboratories, the non-target Bovine Viral Diarrhoea

Table 3
Non-target viruses detected by three or more laboratories.

Virus	Number of laboratories (methods)	Laboratory/method
Bovine viral diarrhoea virus	10 (13)	L01-A, L02, L03, L06, L07, L09, L10-A, L10-B, L11, L12-A, L12-B, L15-A, L15-B
Human bocavirus	7 (8)	L01-A, L02, L03, L06, L10-A, L10-B, L11, L12-B
Human enterovirus (multiple[a])	6 (8)	L02, L07, L09, L10-A, L10-B, L11 L12-A, L12-B
Aichi virus	4 (5)	L03, L10-A, L12-A, L15-A, L15-B
Bovine parvovirus	4 (5)	L01-A, L02, L10-A, L10-B, L11
Porcine/other circovirus	3 (3)	L03, L06, L15-B

[a] Multiple similar results are consolidated to 'Human enterovirus'. Some laboratories did not report non-target viruses.

Virus had a median rank of 8 and several other non-target viruses had rankings higher than target viruses (Fig. 4, bottom panel).

3.3. Summary results

The current study was not a proficiency test, and it is recognised that multiple experimental design considerations affect the ability to detect particular viruses. The results obtained in the current study, may be used for reference when using this material:

All laboratories detected: Ad2, Ad41, HHV-3, HHV-5, Parechovirus 3, and Rotavirus A.

Greater than 90% of laboratories detected Astrovirus, Coxsackievirus B4, HHV-1, HHV-2, HHV-4, Metapneumovirus A, Parainfluenzavirus 1, Parainfluenzavirus 4, Rhinovirus A39, and Sapovirus C12.

Greater than 50% of laboratories detected: Coronavirus 229E, Influenza A H1N1, Influenza A H3N2, Influenza B, Norovirus GII, Parainfluenza virus 2, Parainfluenzavirus 3, Respiratory Syncytial Virus A2, and the non-target Bovine Viral Diarrhoea Virus.

Fewer than 50% of laboratories detected: Norovirus GI, as well as the non-target Human Bocavirus, Human Enterovirus, Aichi Virus, Bovine Parvovirus, and Porcine/other Circoviruses.

4. Discussion

The detection of infectious PCV-1 in a human vaccine [19] demonstrated that existing safety tests may not detect some virus contaminants, and highlighted the potential for deep sequencing methods to form part of an improved testing scheme [1]. This study employed a reagent containing diverse virus families, genome and coat types, for comparison of the different strategies. The large diversity in laboratory and informatics methods used and the variability in detection of target viruses underscore the need for such materials to facilitate assay development and method comparison. The study highlighted a number of issues facing the successful implementation of deep sequencing within a manufacturing/regulatory environment and these are discussed below.

4.1.1. Issues in assay design and sample preparation

The nature of the material being tested will be determined by the product and production stage and different materials may have distinct upstream processing requirements (*e.g.* filtration, concentration, centrifugation). The discrepancies in detection of target viruses may in part be attributable to the assay design. For example, assays targeting only particle-protected nucleic acid and using nuclease may discriminate against certain signals, relative to those targeting total nucleic acid, though notably one laboratory that identified all 25 viruses (L02) reported using nuclease. Extraction methods may also affect the ability to detect different viruses [14] and variable efficiency of reverse transcriptase steps may bias against detection of RNA viruses. Sequencing library preparation typically requires nanogram to microgram amounts of DNA that may be challenging to obtain from certain starting materials. Lower amounts of DNA may yield adequate sequencing libraries in some cases, but deviating significantly from recommended inputs may be problematic in a Good Laboratory Practice (GLP) or Good Manufacturing Practice (GMP) environment. Amplification by PCR or Phi29-based systems was employed by two laboratories, however the risk of bias due to different template preferences should be considered. As with any molecular technique, the inclusion of no template controls is essential to avoid or identify false positives due to trace contamination of the sample extraction columns [7,20], molecular biology reagents [7,21] or sample-to-sample contamination during library preparation or sequencing [22].

4.2. Matrix effects

Matrix effects may be broadly defined as any change in the sensitivity and specificity of a detection assay due to substances in a sample which inhibit extraction of the target and/or co-purify and interfere with downstream processing. Adventitious virus detection is likely to be performed on diverse samples including raw materials, culture supernatants and bulk harvests, some of which may interfere with nucleic acid detection methods [23]. Competition by non-viral nucleic acids, *e.g.* from host cells, will negatively impact limits of detection, and the concentration of such may vary significantly between sample types. In such cases, nuclease treatment may increase sensitivity for particle-protected viruses, but should be used with caution for the reasons described in the previous section. A detailed investigation of matrix effects was beyond the scope of the study, however future reference materials should be compatible with a variety of matrices, and methods should be validated using a matrix similar to that of the test article.

4.3. Issues in bioinformatics and databases

It is important that bioinformatics algorithms strike an appropriate balance of speed, sensitivity and specificity. The size of sequencing datasets and reference databases continues to grow, offering obvious advantages in terms of sensitivity; however this has necessitated the development of new sequence classification algorithms to enable data analysis within a reasonable time, at a cost of potentially increased false negative rate [24–27]. A range of methods were used in the current study and thresholds for assigning hits varied, with some laboratories reporting detection of a virus on the basis of a single read but others requiring that additional criteria were met. Viral identification stringency should be determined by the context. A low stringency will maximise the chance of detecting novel viruses which may be highly divergent from databases references at the risk of increased false positives, and may be appropriate *e.g.* for screening of master cell and virus banks where follow-up testing can be performed. In a routine testing scheme, or where a defined set of contaminants is being screened for, higher stringency may be appropriate to minimise false positives, *e.g.* due to matches with host sequences similar to viral genes. While the current study focussed on the detection of known viruses, many of the pipelines employed are entirely compatible with virus discovery investigations simply by altering the processing and search parameters.

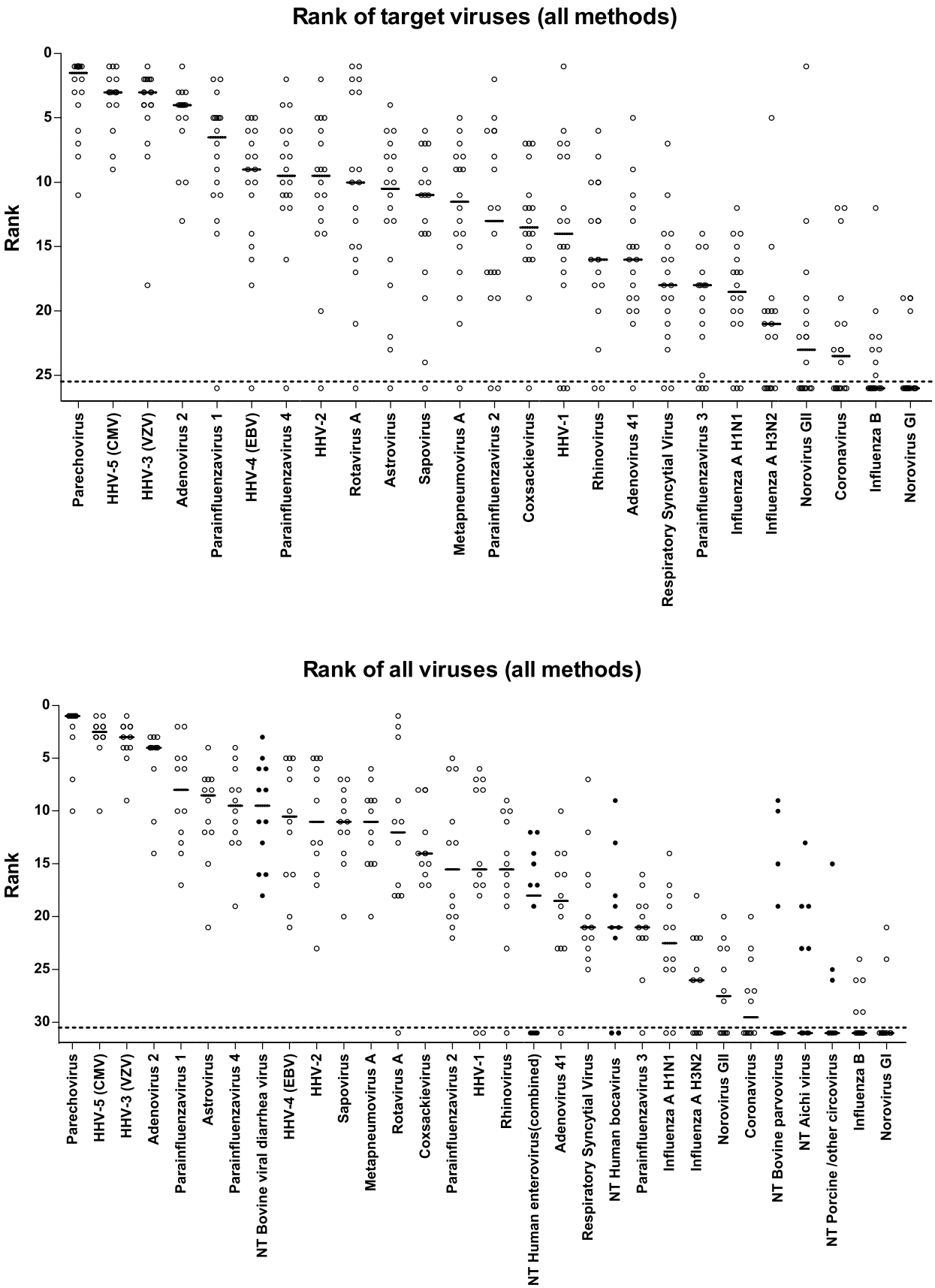


Fig. 4. Rank order of viruses using all methods. *Top panel*, ranking of target viruses. *Bottom panel*, ranking of all viruses, excluding results where no additional viruses were reported. *Horizontal bars* indicate median rank of viruses; *open circles* indicate individual data points. *Solid circles* and *prefix NT* indicate non-target viruses reported by three or more laboratories. Data points below dotted line indicate that virus was not detected – such viruses were assigned a rank order of 26 (for target viruses) or 31 (for all viruses) for plotting and calculation of median.

We also observed variation in the databases used for sequence classification. Publicly available databases such as those hosted by NCBI [28] are valuable but incomplete and not curated, and consideration is needed of what action should be taken if sequencing reads that are currently non-classified are later identified as being of viral origin. The size, breadth and complexity of databases can pose a challenge to sequence identification pipelines, especially when databases contain entries whose taxonomy is erroneously assigned. It may therefore be advisable for adventitious virus screens to be performed using curated databases, even though their breadth of coverage may be more limited.

4.4. General issues

Once hits are identified by deep sequencing, it will be important to confirm the presence of the contaminant by a second molecular method. Infectivity should be assessed if an appropriate assay exists. The presence of certain viral nucleic acid, *e.g.* that remaining after viral inactivation or reduction processes, may be considered acceptable if there is no evidence of infectious virus, though the fact that some viruses produce infectious nucleic acids [29–31] necessitates a cautious approach. Investigation procedures both within a Good Manufacturing Practice and regulatory setting will likely follow existing guidelines.

4.4.1. Intended use and limitations of the reference material

The reagent is intended to be used as a control to assess assay performance relative to the results presented herein, or to results from historical runs of the reagent. While not a proficiency testing material (since some assays have different design parameters), the reagent will enable users to perform inter-assay, intra-laboratory and inter-laboratory comparisons.

The reagent does not purposefully contain a single-stranded DNA virus, a class of virus of particular interest given its detection in two vaccine products [1]. However, circovirus and parvovirus sequences were detected in the sample by three and four laboratories respectively (Table 3), providing evidence for the presence of these ssDNA viruses in the reagent, albeit at a low level. The most commonly reported non-target virus, Bovine Viral Diarrhoea Virus, likely originates from the foetal bovine serum used in the reagent, as does the Bovine Parvovirus. The detection of Human Bocavirus and Aichi Virus most likely reflects the use of faecal samples as a source for several of the target viruses (Table 1), while the 'Human Enterovirus' category likely reflects a combination of authentic enteroviruses from the faecal samples and potentially mis-classification of reads originating from Coxsackievirus and Rhinovirus. Only short sequence contigs were obtained for the circoviruses and definitive identification was not achieved, hence the origin of these viral reads cannot be determined.

Absolute quantification of the components, in terms of infectious units, particles or genomic equivalents is not available. The reagent is therefore not suitable for determination of limits of detection or quantification. A number of reports have begun addressing limits of detection of particular viruses by deep sequencing [32–35], and the potential replacement of existing *in vitro* and *in vivo* tests [10,23]. Empirical definition of limits of detection for all possible viruses is impractical, but limits of detection for a set of viruses representing all major genome and coat types is a minimal starting point, with potential extension to include the full set of viruses for which screening is mandatory.

Future reference materials will contain purified virus particles representing the diverse size, genome structure, GC-content and particle structure (enveloped or not) of the virus Kingdom. The materials should be subject to comprehensive characterisation including precise quantification, be compatible with a variety of sample matrices commonly encountered in biologicals manufacturing and not be restricted to particular molecular biology techniques or sequencing platforms, since these are likely to change over time.

4.5. Summary

The collaborative study highlighted the diversity of methods currently employed for adventitious virus screening by deep sequencing, and the variability in target virus detection underscored the need for a suitable reference material to enable assay comparison. Reagent 11/242-001 will serve as a useful first generation reference material for evaluating and improving such methods, monitoring of intra-laboratory consistency and enabling inter-laboratory comparisons to support this promising but nascent application of deep sequencing.

Author contributions

ETM, PDM and SS conceived the study. ETM, PDM, SS and study participants designed the study. ETM co-ordinated the study and data collection. ETM and MDP analysed the data. ETM drafted the manuscript. All authors revised and approved the manuscript.

Acknowledgements

We are grateful to the participants of the study for contributing time and resources. We would like to thank Dr. Rob Anderson, Dr. Sophie Collot-Teixeira and Dr. Kathryn Doris at NIBSC for provision of reagent 11/242-001 and advice on production and characterisation. We are grateful to the NIBSC Sales and Dispatch Team, in particular Mark Harris and Paul Bolton for co-ordination of reagent shipments and thank members of the FDA/PDA Advanced Virus Detection Technologies Interest Group for helpful discussions and feedback. This report is independent research which, in part, was commissioned and funded by the Department of Health Policy Research Programme (NIBSC Regulatory Science Research Unit, 044/0069). Funding reported by participants: NIH grant number R01-HL-105770 to Eric Delwart. The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health.

Conflict of interest: The authors declare no conflicts of interest related to this study. An edited version of this report has been submitted to the World Health Organisation for consideration by the 2015 meeting of the Expert Committee on Biological Standardisation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.vaccine.2015.12.020.

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(Jun (12)):6033–40, http://dx.doi.org/10.1128/JVI. 02690-09, pii:JVI.02690-09.

[2] Cutrone R, Lednicky J, Dunn G, Rizzo P, Bocchetta M, Chumakov K, et al. Some oral poliovirus vaccines were contaminated with infectious SV40 after 1961. Cancer Res 2005;65(Nov (22)):10273–9, http://dx.doi.org/10.1158/0008-5472.CAN-05-2028, pii:65/22/10273.

[3] 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(May (24)):2916–26, http://dx.doi.org/10.1016/j.vaccine.2014.02.021, pii:S0264-410X(14)00194-7.

[4] Dubin G, Toussaint JF, Cassart JP, Howe B, Boyce D, Friedland L, et al. Investigation of a regulatory agency enquiry into potential porcine circovirus type 1 contamination of the human rotavirus vaccine. Rotarix: approach and outcome. Hum Vaccin Immunother 2013;9(Nov (11)):2398–408, pii:25973.
[5] Gilliland SM, Forrest L, Carre H, Jenkins A, Berry N, Martin J, et al. Investigation of porcine circovirus contamination in human vaccines. Biologicals 2012;40(Jul (4)):270–7, http://dx.doi.org/10.1016/j.biologicals.2012.02.002, pii:S1045-1056(12)00026-7.
[6] Onions D, Cote C, Love B, Toms B, Koduri S, Armstrong A, et al. Ensuring the safety of vaccine cell substrates by massively parallel sequencing of the transcriptome. Vaccine 2011;29(Sep (41)):7117–21, http://dx.doi.org/10.1016/j.vaccine.2011.05.071, pii:S0264-410X(11)00818-8.
[7] McClenahan SD, Uhlenhaut C, Krause PR. Evaluation of cells and biological reagents for adventitious agents using degenerate primer PCR and massively parallel sequencing. Vaccine 2014;32(Dec (52)):7115–21, http://dx.doi.org/10.1016/j.vaccine.2014.10.022, pii:S0264-410X(14)01385-1.
[8] Onions D, Kolman J. Massively parallel sequencing, a new method for detecting adventitious agents. Biologicals 2010;38(May (3)):377–80, http://dx.doi.org/10.1016/j.biologicals.2010.01.003, pii:S1045-1056(10)00004-7.
[9] Gagnieur L, Cheval J, Gratigny M, Hebert C, Muth E, Dumarest M, et al. Unbiased analysis by high throughput sequencing of the viral diversity in fetal bovine serum and trypsin used in cell culture. Biologicals 2014;42(May (3)):145–52, http://dx.doi.org/10.1016/j.biologicals.2014.02.002, pii:S1045-1056(14)00006-2.
[10] Richards B, Cao S, Plavsic M, Pomponio R, Davies C, Mattaliano R, et al. Detection of adventitious agents using next-generation sequencing. PDA J Pharm Sci Technol 2014;68(Nov (6)):651–60, http://dx.doi.org/10.5731/pdajpst.2014.01025, pii:68/6/651.
[11] Alfson KJ, Beadles MW, Griffiths A. A new approach to determining whole viral genomic sequences including termini using a single deep sequencing run. J Virol Methods 2014;208:1–5, http://dx.doi.org/10.1016/j.jviromet.2014.07.023, pii:S0166-0934(14)00295-X.
[12] Cotten M, Oude MB, Canuti M, Deijs M, Watson SJ, Kellam P, et al. Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. PLOS ONE 2014;9(4):e93269, http://dx.doi.org/10.1371/journal.pone.0093269, pii:PONE-D-13-49127.
[13] Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, Strydom H, et al. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. J Virol Methods 2014;195:194–204, http://dx.doi.org/10.1016/j.jviromet.2013.08.035, pii:S0166-0934(13)00382-0.
[14] Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S, et al. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. J Virol Methods 2015;213:139–46, http://dx.doi.org/10.1016/j.jviromet.2014.12.002, pii:S0166-0934(14)00461-3.
[15] Matranga CB, Andersen KG, Winnicki S, Busby M, Gladden AD, Tewhey R, et al. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol 2014;15(11):519, http://dx.doi.org/10.1186/PREACCEPT-1698056557139770, pii:s13059-014-0519-7.
[16] McClenahan SD, Uhlenhaut C, Krause PR. Optimization of virus detection in cells using massively parallel sequencing. Biologicals 2014;42(Jan (1)):34–41, http://dx.doi.org/10.1016/j.biologicals.2013.11.002, pii:S1045-1056(13)00134-6.
[17] Moore NE, Wang J, Hewitt J, Croucher D, Williamson DA, Paine S, et al. Metagenomic analysis of viruses in feces from unsolved outbreaks of gastroenteritis in humans. J Clin Microbiol 2015;53(Jan (1)):15–21, http://dx.doi.org/10.1128/JCM. 02029-14, pii:JCM. 02029-14.
[18] World Health Organization, www.who.int/en/ WHO Expert Committee on Biological Standardization sixty-fourth report; 2013.
[19] McClenahan SD, Krause PR, Uhlenhaut C. Molecular and infectivity studies of porcine circovirus in vaccines. Vaccine 2011;29(Jun (29–30)):4745–53, http://dx.doi.org/10.1016/j.vaccine.04.087, pii:S0264-410X(11)00642-6.
[20] 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(Nov (22)):11966–77, http://dx.doi.org/10.1128/JVI. 02323-13, pii:JVI.02323-13.
[21] 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, http://dx.doi.org/10.1186/s12915-014-0087-z, pii:s12915-014-0087-z.
[22] Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. PLOS ONE 2014;9(4):e94249, http://dx.doi.org/10.1371/journal.pone.0094249, pii:PONE-D-13-49866.
[23] Ng SH, Azizi A, Edamura K, Malott RJ, Charlebois RL, Logvinoff C, et al. Preliminary evaluation of next-generation sequencing performance relative to qPCR and in vitro cell culture tests for human cytomegalovirus. PDA J Pharm Sci Technol 2014;68(Nov (6)):563–71, http://dx.doi.org/10.5731/pdajpst.2014.01013, pii:68/6/563.
[24] Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res 2014;24(Jul (7)):1180–92, http://dx.doi.org/10.1101/gr.171934.113, pii:gr.171934.113.
[25] Shamsaddini A, Pan Y, Johnson WE, Krampis K, Shcheglovitova M, Simonyan V, et al. Census-based rapid and accurate metagenome taxonomic profiling. BMC Genomics 2014;15:918, http://dx.doi.org/10.1186/1471-2164-15-918, pii:1471-2164-15-918.
[26] Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 2014;15(3):R46, http://dx.doi.org/10.1186/gb-2014-15-3-r46, pii:gb-2014-15-3-r46.
[27] Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 2012;9(Aug (8)):811–4, http://dx.doi.org/10.1038/nmeth.2066, pii:nmeth.2066.
[28] Brister JR, Ako-Adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. Nucleic Acids Res 2015;43(Jan (Database issue)), http://dx.doi.org/10.1093/nar/gku1207, pii:D571-D577.gku1207.
[29] Maassab HF. An infectious ribonucleic acid derived from influenza-infected cells. Proc Natl Acad Sci U S A 1959;45(Jun (6)):877–81.
[30] Alexander HE, Koch G, Mountain IM, Van DO. Infectivity of ribonucleic acid from poliovirus in human cell monolayers. J Exp Med 1958;108(Oct (4)):493–506.
[31] Alexander HE, Koch G, Mountain IM, Sprunt K, Van DO. Infectivity of ribonucleic acid of poliovirus on HeLa cell mono-layers. Virology 1958;5(Feb (1)):172–3, pii:0042-6822(58)90014-X.
[32] Cheval J, Sauvage V, Frangeul L, Dacheux L, Guigon G, Dumey N, et al. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. J Clin Microbiol 2011;49(Sep (9)):3268–75, http://dx.doi.org/10.1128/JCM. 00850-11, pii:JCM. 00850-11.
[33] Wang CJ, Feng SF, Duncan P. Defining a sample preparation workflow for advanced virus detection and understanding sensitivity by next-generation sequencing. PDA J Pharm Sci Technol 2014;68(Nov (6)):579–88, http://dx.doi.org/10.5731/pdajpst.2014.01010, pii:68/6/579.
[34] Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, et al. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics 2014;15:96, http://dx.doi.org/10.1186/1471-2164-15-96, pii:1471-2164-15-96.
[35] Moore RA, Warren RL, Freeman JD, Gustavsen JA, Chenard C, Friedman JM, et al. The sensitivity of massively parallel sequencing for detecting candidate infectious agents associated with human tissue. PLoS One 2011;6(5):e19838, http://dx.doi.org/10.1371/journal.pone.0019838, pii:PONE-D-10-06050.